WOFEX 2011

proceedings of the 9th annual workshop

Faculty of Electrical Engineering and Computer Science,
VŠB – Technical University of Ostrava

ISBN 978-80-248-2449-9

WOFEX 2011

http://wofex.vsb.cz/2011/

9^{th} annual workshop,
Ostrava, 8^{th} – 9^{th} September 2011
Proceedings of papers

Organized by

VŠB – Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science

BIOM and SoftComp workshops – organized as a part of WOFEX 2011 workshop – have been realized as a part of BIOM (CZ.1.07/2.3.00/20.0073) and SoftComp (CZ.1.07/2.3.00/20.0072) projects. These projects are co-financed by ESF and Czech state budget.

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

ISBN 978-80-248-2449-9

Faculty of Electrical Engineering and Computer Science,
VŠB – Technical University of Ostrava

Page count:	693
Impression:	200
Edition:	1st
First published:	2011

This proceedings was typeset by PDFLaTeX.
Printed and bound in Tiskárna Tiskservis, gen. Sochora 1764, 708 00 Ostrava-Poruba, Czech Republic.

Published by Faculty of Electrical Engineering and Computer Science, VŠB – Technical University of Ostrava

Preface

The workshop WOFEX 2011 (PhD workshop of Faculty of Electrical Engineering and Computer Science) was held on September 8^{th} – 9^{th}, 2011 at the VŠB – Technical University of Ostrava. The workshop offers an opportunity for students to meet and share their research experiences, to discover commonalities in research and studentship, and to foster a collaborative environment for joint problem solving. PhD students are encouraged to attend in order to ensure a broad, unconfined discussion. In that view, this workshop is intended for students and researchers of this faculty offering opportunities to meet new colleagues.

This book of proceedings includes 110 papers of faculty PhD students and 4 papers of external authors. The proceedings of WOFEX 2011 are also available at WOFEX Web site http://wofex.vsb.cz/2011/. I would like to thank the authors and the Organizing Committee from Department of Computer Science, namely Jiří Dvorský and Pavel Moravec, for their arduous editing work.

September 2011

Michal Krátký
Program Committee Chair
WOFEX 2011

Organization

Program Committee

Chair:

Michal Krátký (VŠB – Technical University of Ostrava)

Members:

Pavel Brandštetter (VŠB – Technical University of Ostrava)
Zdeněk Dostál (VŠB – Technical University of Ostrava)
Petr Chlebiš (VŠB – Technical University of Ostrava)
Stanislav Rusek (VŠB – Technical University of Ostrava)
Václav Snášel (VŠB – Technical University of Ostrava)
Vilém Srovnal (VŠB – Technical University of Ostrava)
Vladimír Vašinek (VŠB – Technical University of Ostrava)

Organizing Committee

Jiří Dvorský (VŠB – Technical University of Ostrava)
Pavel Moravec (VŠB – Technical University of Ostrava)
Renáta Mostýnová (VŠB – Technical University of Ostrava)

Workshop Location:

Campus of VŠB – Technical University of Ostrava
17. listopadu 15, 708 33 Ostrava–Poruba, Czech Republic
8th – 9th September 2011

http://wofex.vsb.cz/2011/

Table of Contents

Study Programme P2649 Electrical Engineering

Electrical Power Engineering (3907V001)

Electrical Machines, Apparatus and Drives (2642V004)

Electronics (2612V015)

Technical Cybernetics (2612V045)

Study Programme P1807 Computer Science, Communication Technologies and Applied Mathematics

Computer Science (1801V001)

Computational and Applied Mathematics (1103V036)

Communication Technologies (2601V018)

External Authors

The Light Spectrum of Obtrusive Light

Daniel Diviš and Karel Sokanský

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
daniel.divis@vsb.cz, karel.sokansky@vsb.cz

Abstract. The world around us is changing faster than we can recognize. Some changes that are directly related to our lives are not even noticed, but some of them are very intensively perceived and we respond to them. If it is somewhere obtrusive light (light pollution), where is it fault? Is it caused by poorly designed lighting system, the current state of the weather or high content of dust particles and aerosols in the air? The first case is the outdoor lighting system as a source of obtrusive light and second one (night sky) only displayed the obtrusive light and it can be disturbing.

Keywords: obtrusive light, light pollution, light spektrum, high pressure sodium lamp

1 Introduction

The aim of this paper is to find out the proportion between the public lighting radiation (both direct and reflected) to the upper hemisphere scattered in the atmosphere under the specific meteorological conditions (cloudy, rainy, foggy, dusty weather), when is the reflected radiation best visible and becomes a target of criticism. These weather conditions cause the highest luminance of the sky and therefore it has highest obtrusive potential. In this particular experiment, we focused on studying the impact of high pressure sodium lamps, which are mostly used in public lighting systems, to the obtrusive light. As these sources are the most used in the public lighting, we can expect their important influence to obtrusive light itself. Structure of this paper is following:

- *used measuring equipment*
- *a description of our experimental measurements*
- *measured results*

2 Measuring equipment

Complete measuring equipment consists of a spectrophotometer Jeti 1211 together with a laptop, which is essential for its controlling. This spectrophotometer was chosen because of its high sensitivity and long integration time up to 10 minutes. High sensitivity together with long-time integration allow to evaluate very low luminances of the night sky.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 1–6.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

The spectrophotometer can measure the following values:

- *spectral radiation*
- *luminance*
- *illuminance*
- *chromatic coordinates x, y, u, v*
- *correlated color temperature*
- *dominant wavelength*
- *color rendering index*

The spectrophotometer Jeti 1211 is specified by the following parameters:

- *measured wavelengths - 350 nm - 1000 nm*
- *wavelength resolution 5 nm*
- *viewing angle 1.8 °*
- *luminance accuracy ±2 % (1000 cd/m2; 2856 K)*
- *wavelength accuracy ±0.7 nm*

We are mainly interested in measurements of spectral radiation in examining the obtrusive light. These values are used for evaluation.

3 Description of our experimental measurements

Measurement of the spectral qualities of the light reflected from the night sky took place on August 12, 2010 on the roof of the Institute of Construction and Architecture in Academy of Sciences in Bratislava. The position of the measuring point is marked with intersecting axes of the compass (see Fig. 1). The measurements were done from 5:00 p.m. to 6:30 p.m. During the measurements the whole sky was overcast and the air temperature was 5 ° C.
Red circle with a number 1 in Fig. 1 and Fig. 2 shows the position of the shopping centre. The second circle (number 2) shows the high buildings that are near the measurement location and thus they are an important source of the obtrusive light in the measured area. The upper floors of tall buildings emit light radiation easier in the upper half-space. The shopping centre and high buildings are seen from the fish-eye lens snapshot in Fig. 2.

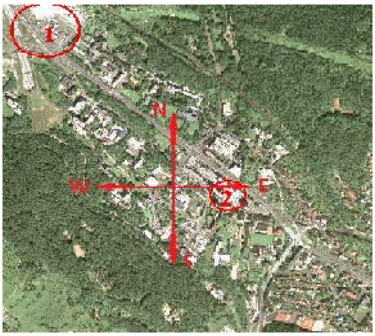


Fig. 1. Measuring point with intersecting axes of the compass

Fish-eye lens is a type of lens that allows the camera to capture a whole half-space. The measurement of the night sky parameters were done by the cut made by axe starting on the angle 42 ° on the north and go through the centre of the compass card. This direction was chosen because of two major sources of obtrusive light on both edges of the measured area above the horizon (shopping centre and the centre of Bratislava city). The measurement was done from red plus to red minus marks in the figure 2.

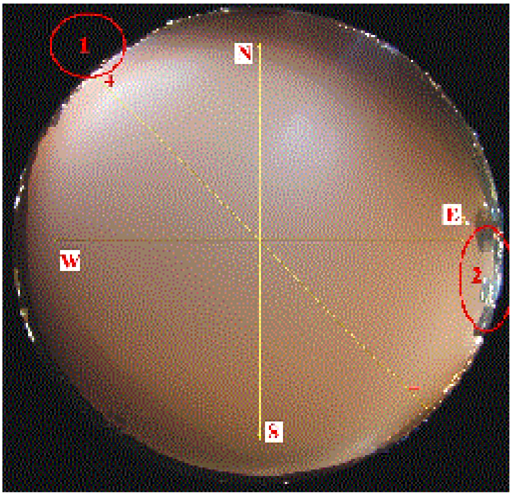


Fig. 2. Night sky photo made by fish-eye lens from the measurement point in Bratislva

Measured part of the sky from the shopping center to the zenith was marked by positive values of measured angles and the opposite side was marked by negative angles. Measurement the upper hemisphere of the night sky was done in step 15 degrees from the horizon through the zenith to the opposite horizon. The aim of the measurement was to obtain information about differences between incident radiations and differences of the spectral characteristics of the night sky in the direction from the centre of Bratislava through the zenith to the shopping centre.

3.1 Processing of the measurement

For the experiment there was chosen a region wavelength of the light radiation 550 nm - 700 nm. It is area of high pressure sodium lamps radiation because high pressure sodium lamps radiated about 589 nm, particularly in their surrounding (see Fig. 3 and Fig. 4).

The evaluation was done by comparing the total measured energy amount of reflected radiation from the night sky to our defined parts of the spectrum, where the high-pressure sodium lamps radiation is dominant. There was measured the reflected spectral power from the night sky at various angles from horizon to zenith. It was also examined if there is no major change in the measured spectral characteristics of the locality due to presence of different particles in the atmosphere and other light sources than high pressure sodium lamps.

Results show that proportion of reflected components of radiation in the visible wavelength range 550 nm - 700 nm (high pressure sodium) is about 60% of the total measured power of reflected visible radiation. These results are shown in Fig. 5. In the measured area of the upper hemisphere does not occur any extreme deviation from the 60% value. It was confirmed that the values of reflected visible radiation power decrease towards to the zenith. It is clearly visible in Fig. 6. High values measured at 15 degrees were caused by the shopping centre, which was in the direction of measurement. Measured power from the direction of the shopping centre was significantly higher than from the direction of the city centre. Lower impact of the city centre we can explain by further distance from the measuring point compared to a distance of the shopping centre.

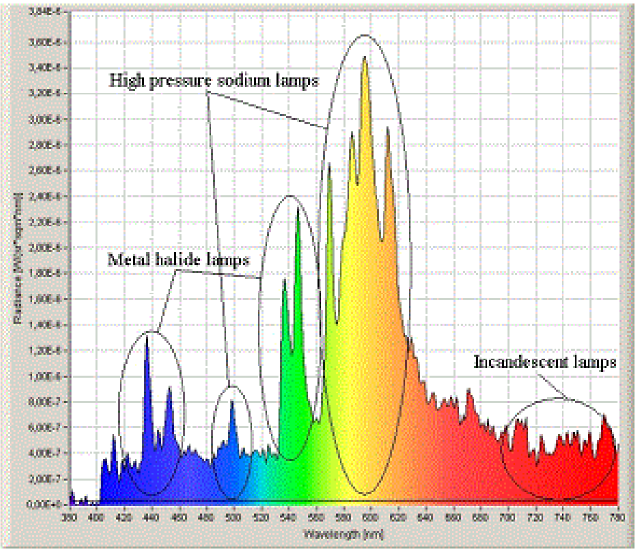


Fig. 3. The measured spectrum of reflected radiation from the night sky in Bratislava

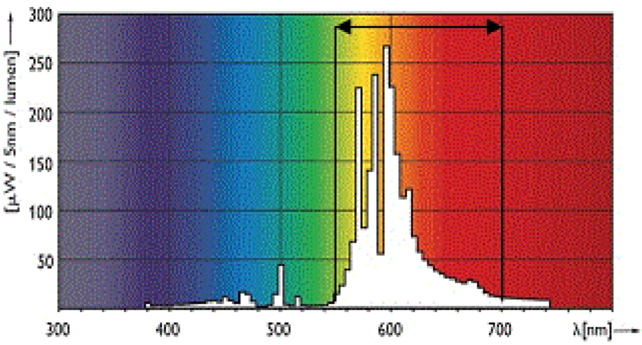


Fig. 4. The spectrum of high-pressure sodium lamp radiation [5]

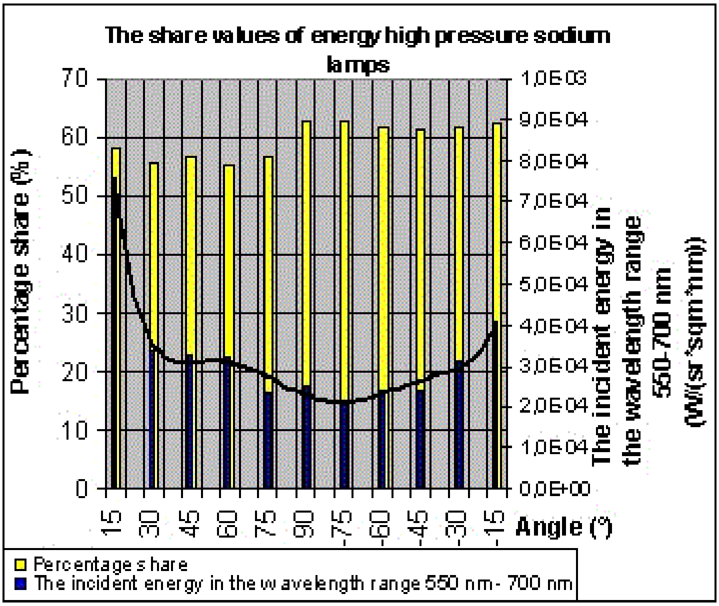


Fig. 5. The share of the high pressure sodium lamps power values in the total reflected radiation from the night sky

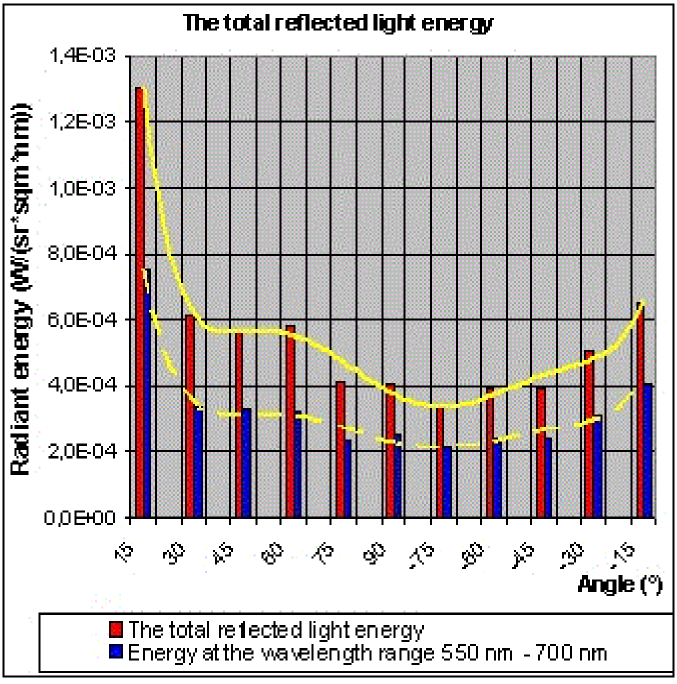


Fig. 6. The decrease of the measured values of the radiation power

4 Conclusion

The current measurements show that the influence of high pressure sodium lamps used in public lighting systems, to the night sky in the Czech and Slovak regions is significant. It is valid both in the metropolitan areas and countryside. It could be expected that more frequent usage of newer and other light sources (LEDs, metal halide lamps, induction lamps) will decrease the influence of high pressure sodium lamps in the night sky brightness. We hope that this contribution will start a discussion about the impact of high pressure sodium lamps from the measured spectra. This includes particularly a defining their wavelength limits of their radiation. The next step of research resulting from these experiments will concern the influence of other light sources on the obtrusive light, which was briefly shown in Fig. 3. It is primarily their specification in the measurement and quantification of visible radiation reflected back by atmosphere to the Earth.

Acknowledgment

The article was created by the project KONTAKT - project number is MEB 0810007

Reference

1. Diviš, D., Novák, T., Sokanský, K.: Měření spektrálních vlastností rušivého světla. In sborníku EPE 2011, Ostrava:VSB - Technical University of Ostrava, 2011, 3 s., 978-80-248-2393-5

2. Diviš, D., Novák, T., Sokanský, K.: The measurement of the spectral qualities of obtrusive light. In sborníku IEEE 2011. Ed. M. Caciotta, Z. Leonowicz, Wroclaw:Wroclaw University of Technology, 2011, 4 s., 978-1-4244-8781-3

3. Diviš, D., Dostál, F., Šnobl, J., Mišák, S., Šnobl, J.: Hybrid power system of public lighting in smaller villages. In časopis Przeglad Elektrotechniczny, Issue 04/2011, page 53, Warszawa, 2011, ISSN 0033-2097

4. Studentská grantová soutěž - SGS - SP2011/135 - Optimalizace spotřeby elektrické energie využitím moderních světelných zdrojů.

5. Philips product. URL: <http://www.ecat.lighting.philips.com> [cit 2011-6-7].

Filtering-Compensating Devices at Traction Transformer Stations of the Czech Railways

Bogdan Fukala

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
bogdan.fukala@cbox.cz

Abstract. Generation of harmonic currents towards the power supply network is one of the most relevant influences of AC electric railway operation. A filtering-compensating device aims to alleviate this disturbing effect. The purpose of this paper is to approach the function of this device and assess its benefit on the basis of measured values.

Keywords: harmonic distortion, traction transformer station, power factor, *THD* factor, filtering-compensating device.

1 Introduction

At its beginning, a single-phase electrification system was considered advantageous for less loaded railway lines. This idea was based on the fact that the traction contact line requires less conductor size and traction transformer stations consisting of only one traction transformer can be located at larger distances from each other than traction converter stations for a DC traction current system. Power utilities required, at that time, in terms of EMC (electromagnetic compatibility) of this traction current system only one condition relating to permitted voltage unbalance in the 110kV power supply network.

2 Power Factor and Harmonic Distortion in Traction Operation of AC Electric Railways

Gradual development of modernization of drive vehicles was driven by an effort to efficient and smooth operation and non-contact drive control. Implementation of power diodes and thyristors met these requirements in the early stages of the application of power electronics to drive vehicles. Using these elements, however, negative interference began to appear which started bringing a number of problems in other areas such as power networks, communications and security technology to such an extent that there could have been direct or indirect threat of security not only for the railway service. This circumstance requested a quick solution and gave the space for creation and implementation of such measures so that the semiconductor control would lose none of their technical and economic importance and yet would not influence other devices.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 7–12.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2.1 Origins of Harmonic Distortion

The Czech Railways still use drive vehicles of older concepts. These drive vehicles use vehicle traction transformers with tapping regulation and uncontrolled single-phase diode rectifiers in a bridge design. This circuit is applied in locomotives of the 230 (S 489.0) and 240 (S 499.0) classes. Shunting locomotives of the 210 (S 458.0) class and electric pantograph units 560 (SM 488.0) were equipped with the thyristor half-controlled rectifier bridges and fixed-ratio transformers.

In terms of today's EMC requirements vehicles designed in this manner exhibit two major imperfections: power factor of the fundamental of the traction take-off varies around the value 0.84 (in the case of thyristor control it reaches a lower value) and the drawn traction current is greatly distorted by higher order harmonics. The factual numerical values of these two quantities are dependent on own concepts of traction circuits and an instantaneous traction mode. [1]

These vehicles do not draw sinusoidal currents from the power supply network and constitute non-linear loads. Harmonic current generated by the load must flow through the line and source impedances thus flows back to the energy source. Distorted load currents then cause distorted voltage drops across network impedances and distorted voltages are applied to all other connected loads and cause that harmonic currents flow even in the case of linear loads.

2.2 Power Factor Improvement and Harmonic Distortion Mitigation

As already mentioned in the introduction, at the beginning of the AC electric traction, power utilities only demanded that the traction power would not have exceeded 2% of short-circuit power of 110kV power supply network at the connection point of the traction transformer station (2% voltage unbalance). Later it has also been required that the lagging power factor of the fundamental has ranged from 0.95 to 1. Therefore the Czech Railways had to accede to its compensation. Technically, it has been tested several variants of compensation devices (the location of the compensation capacitor inside drive vehicles, on the open track, at the station) but their results were not satisfactory. The most serious problem was control of reactive power compensation. In addition, efforts of power utilities to limit the voltage distortion at connection points of traction transformer stations had also begun to appear so that it was convenient to solve these two requirements comprehensively. A project of a filtering-compensating device was the only way at that time.

2.3 Filtering-Compensating Device (FCD)

A FCD at a traction transformer station (TTS) has to provide the following tasks [1]:

- To compensate the power factor of the fundamental of traction demand for an inductive value of 0.95 to 1 and prevent its value to cross to the capacitive area.
- To limit the transfer of the 3rd and 5th eventually 7th current harmonics so that the corresponding voltage components at the connection point of the TTS will lay below the limiting values prescribed by power utilities. Limiting values for 10-

minute averages are: 2.5% for voltage total harmonic distortion THD_u and 2% for each individual harmonic Uh/Un. [2]

- To ensure that TTS input impedance as a whole (*i.e.* including traction contact line capacitance) for the operating frequency of a centralized ripple control system used by power utilities will meet the desired value, *i.e.* to prevent the level degradation of the 216.67Hz control frequency at TTS connection points.

The FCD operation is based on a series resonance of *LC* elements connected between a 27kV bus bar and a return line. The *LC* branches serve to reduce the majority of tuned harmonic current components so that they do not penetrate through the traction transformer into the power supply network, where they would cause creation of voltage harmonic distortion. A FCD diagram is in Fig. 1A.

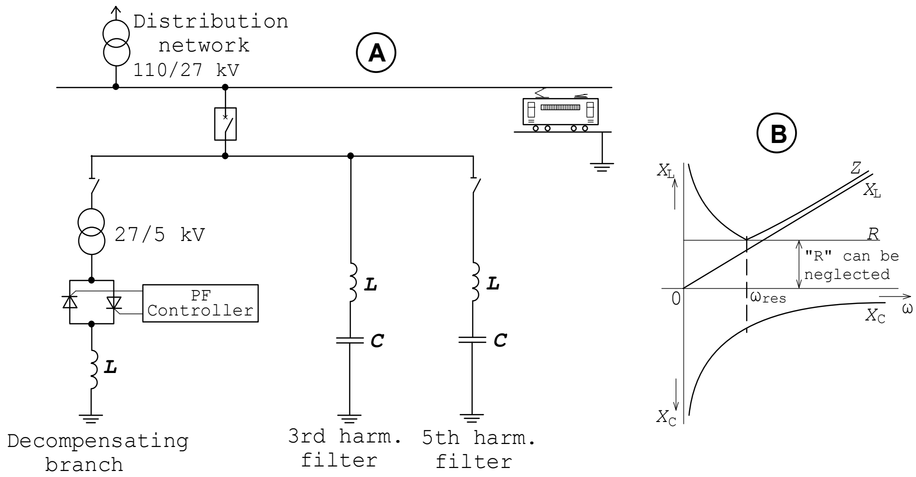


Fig. 1. Filtering-compensating device diagram (A) and graph of frequency dependence of series resonant circuit elements (B)

The FCD now used on the Czech Railways contains two series *LC* branches for the 3rd and 5th harmonics ranked in parallel. The FCD configuration allows supplementation of *LC* branches for the 7th harmonic with the condition that, when designing, the decompensating branch had been designed for this addition. If an applied frequency is higher than the resonant frequency of an *LC* branch then this branch acts as an inductor. If an applied frequency is below the resonant frequency then this branch acts as a capacitor. These facts are used for power factor correction of the fundamental because for the fundamental frequency these branches act as capacitors. This phenomenon is evident in Fig. 1B

Compensating power of both branches is constant (assuming constant voltage) and is independent of the traction take-off thus the need for the compensating power of instantaneous traction take-off. As mentioned above, it is necessary to compensate for relatively strict power factor in real time. Therefore the FCD involves another branch called "decompensating" which contains a 27/5 kV or 27/6 kV step-down transformer whose secondary winding is connected through a thyristor phase controller to a decompensating reactor which "consumes" excessive reactive (capacitive) power from both *LC* branches and thus prevents undesirable capacitive reactive power supply to the utility network. The decompensating branch must absorb capacitive reactive energy coming not only from the FCD but also from the traction contact line and 110kV power supply line.

When designing the *LC* resonant branches the issue of their accurate tuning arose. In terms of mitigation of harmonic currents coming from the traction contact line side it would seem to be the best to tune both of these branches "sharply", *i.e.* at 150 Hz and 250 Hz. This tuning, however, could result in the need for uneconomical sizing of compensating capacitors because the existence of harmonic voltages of the 3rd and 5th orders in the power supply line would lead to impose excessive voltage stresses on capacitors. Therefore, the *LC* resonant branches are tuned at frequencies just below their resonant frequencies. [1]

To avoid the signal damping of the centralized ripple control system at the TTS power utilities require a certain minimum value of TTS input impedance (500 Ω to 900 Ω). The fact that a parallel resonant frequency always exists between two neighbouring series resonant frequencies is used to achieve this value. In this way, the high value of input impedance can be reached. In the south of the Czech Republic, where the AC electric railways operate, the operating frequency of the centralized ripple control system has a value of 216,67 Hz. An appropriate choice of *LC* power components allows to set up the parallel resonant frequency close to 216.67 Hz. [1]

3 Measurement at TTS

The measurement [3] was carried out at the TTS in Blansko from 20 January (10 a.m.) to 21 January 2010 (11 a.m.) thus for 25 hours. Considering the purpose of this paper, to analyze the FCD operation, graphic presentations of measured values display shorter time frames that better approach relevancy of the FCD.

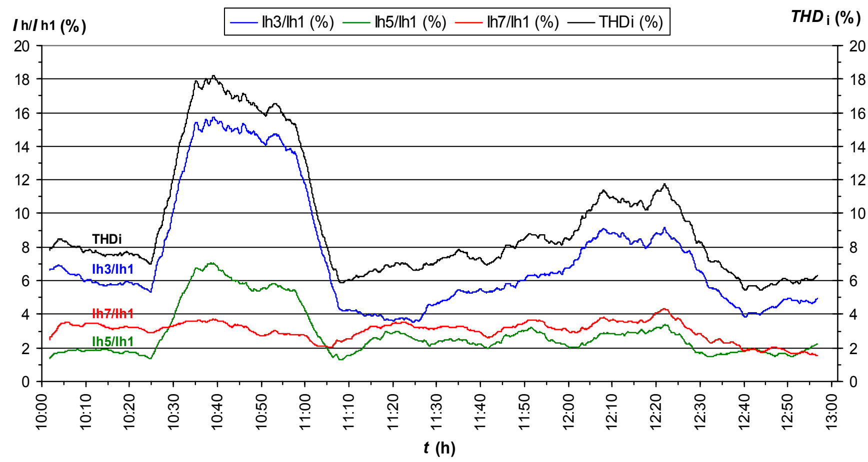


Fig. 2. Time diagram of 10 minutes rolling averages of harmonic current ratios Ihx/Ih1 and 10 minutes rolling averages of current total harmonic distortion THD_i in the 110kV supply line

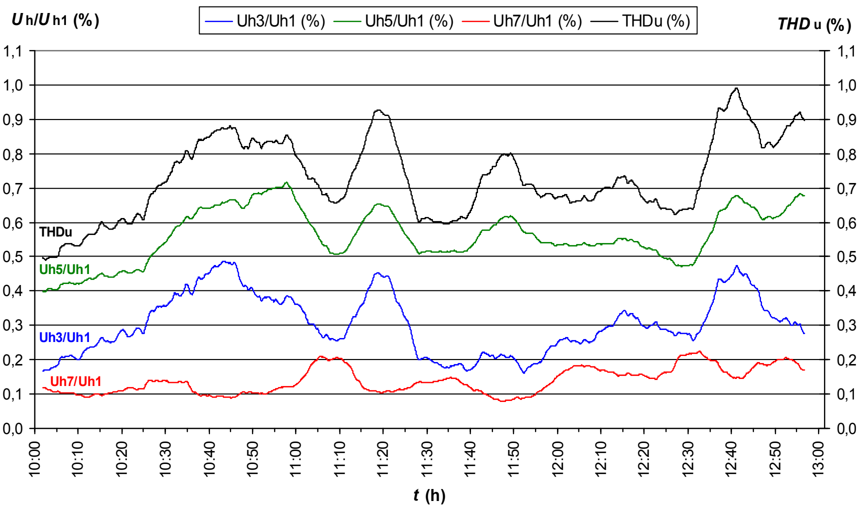


Fig. 3. Time diagram of 10 minutes rolling averages of harmonic voltage ratios *U*hx/*U*h1 and 10 minutes rolling averages of voltage total harmonic distortion THD_u in the 110kV supply line

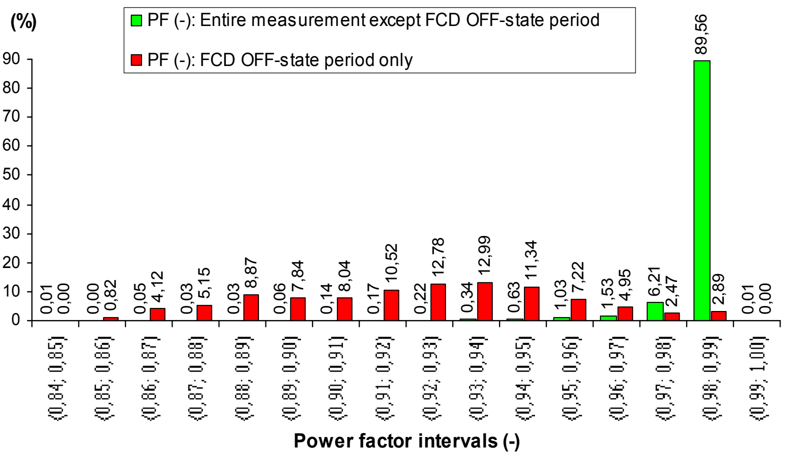


Fig. 4. Histogram of occurrences of instantaneous values of power factor for the entire measurement period (25 hours)

4 Discussion

The measurement provided the sufficient view of traction take-off nature and its influences on the power supply network. The importance and useful effect of harmonic filters for the 3rd and 5th current harmonics are evident from the time diagram of these harmonics in Fig. 2 accurately between 10:28 and 11:08 a.m. when the FCD was switched off. At that time, there was a sharp rise in THD_i which is understandable, considering a significant contribution of these harmonics to THD_i.

The time diagram of harmonic voltage ratios Uh_x/Uh_1 and THD_u in Fig. 3 shows that the occurrence of voltage harmonics in the power supply line does not always track harmonic currents in traction contact lines. This is particularly evident during the FCD off-state period when, otherwise, a percentage of THD_u levels was increased but not as much as current harmonics. In addition, during the entire measurement period some intervals when THD_u values rise with no evident linking to traction take-off and its THD_i can be registered. Voltage harmonic distortion in the power supply line thus is influenced by other energy consumers in the distribution network.

The histogram in Fig. 4 depicts the percentage of power factor values belonging to particular intervals. The evaluation of power factor values is divided into two parts: the entire measurement period, except the time interval from 10:28 till 11:08 a.m. (FCD switched off) and measurement only in this interval. Finally, this separation allows to get an idea of importance of power factor compensation and about how low power factor values would be achieved with no compensation. It is evident that during the FCD off-state period the bulk of instantaneous values (82.47%) do not meet the prescribed interval for power factor of the fundamental (inductive 0.95 to 1). On the other hand, when the FCD works, only 1.67% of values do not fit into the prescribed interval. This percentage is a representation of instantaneous values that only occur during rapid dynamic changes of traction current.

5 Conclusion

FCD installations at TTSs bring acceptable solutions for power utilities' requirements for "clean power take-off". In this way, satisfactory conditions have been found which were confirmed by the results of measurements. The FCD, however, does not solve the essence of the problem. Recently, due to the development of power electronics, the change of concept of drive vehicles is applied. This ensures that unfavourable effects are immediately eliminated at their source – a drive vehicle.

References

1. HLAVA, Karel. *Elektromagnetická kompatibilita (EMC) drážních zařízení* [*Electromagnetic compatibility (EMC) of railway facilities*]. 1st edition. Univerzita Pardubice, 2004. pp. 124. ISBN 80-7194-637-0.
2. ČEZ Distribuce, a. s. *Pravidla provozování distribuční soustavy 2009 – Příloha č.3* [*Rules of distribution network operation 2009 - Appendix 3*] [online]. Access date [2011-07-03]. Available: http://www.cezdistribuce.cz/cs/energeticka-legislativa/pravidla-provozovani-ds/ppds-2009.html
3. FUKALA, Bogdan. *Influence of Electric Railways on Power Supply System.* Ostrava, 2010. Thesis. VSB-TU Ostrava, Faculty of Electrical Engineering and Computer Science, Department of Electrical Power Engineering, 2010-04-20.

Prototype Detector of Covered Conductor Faults

Štefan Hamacek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
stefan.hamacek@seznam.cz

Abstract. Lines with bare conductor have rebuilt lines with cover conductor in the Czech Republic. Lines with cover conductor have many advantages. Advantages and disadvantages of PAS system are described here. This paper describes the issue of fault detection for lines with cover conductors and also describes the dependence of partial discharges on fault. The next section provides a design and realization fault detection of cover conductors. This is a prototype which will be tested and optimized. The system will be used for different voltage levels and different types of outdoor lines.

Keywords: partial discharge detector failures, cover conductor,

1 Introduction

The system PAS of cover conductors began to build in Finland. It was used in Europe and the Czech Republic later.
This system has a lot of advantages over lines conductors with without insulation:

- reduce the failure rate caused by falling trees in to the line or touch wet branches,
- fallen trees on the line can be removed later under favorable conditions,
- there is no short circuit when conductors touch each other,
- conductor insulation reduces the corrosion in extreme outdoor conditions,
- the distance between the phases reduce by 1/3 length versus bare conductors.

The biggest problem is the inability to detect the fault today. There are types of fault: *fallen branch on the isolated conductor or broken conductor and its subsequent fall to the ground.* Standard digital protection is unable to detect these types of faults. It is not a standard ground fault see. **Fig. 1**.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 13–17.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Fig. 1. The consequence of failure - Conductor fell on the ground

2 Development of detector failures.

The primary purpose of dissertation is to create a measurement system for fault detection cover conductors.
Partial discharge has effects of degradation on the cover conductor. Climatic conditions may amplify these degradative conditions. The emergence of fault has resulted in increasing voltage U_{RMS} and frequency of partial discharges. These values are measured by suitable measuring system. The voltages of partial discharge are from µV to mV even so can degrade the cover conductors. Everything is dependent on the intensity of the electric field. Electrons and ions can bombing of wall cavities. Ozone is created in the cavities at the presence of oxygen. Ozone has erosive effects on cover conductors. Cavity is gradually increased and continued erosion. This can cause of electrical breakdown of the dielectric. Conductors are exposed to the partial discharges which dried cover conductors and then are created surface cracks. Effect of partial discharges can lead to such a short circuit between phases or ground fault cause. I will be able to measure the parameters of partial discharges. I will be able to evaluate the emergence of fault of the measured parameters on the overhead lines. I will use the parameters of partial discharges for the evaluation different faults on the overhead lines.

The emergence of fault "branches fall on the phase" is possible to see. **Fig. 2**. Faults will result in the occurrence of partial discharges as can be seen from the voltage signal. The voltage signal No. 1 is fault-free state and the signal No. 2 is a state of failure. At the No. 2 shows a partial discharge which are modulated on the voltage signal. So dependency was found between the occurrence of partial discharges and failure. This methodology is currently protected by patent P 2008-647.
Partial discharges have their measurable parameters - U_{RMS} (V, effective voltage partial discharge) and n (s^{-1}, frequency).
Fault indicators were selected:

- U_{RMS} value of the partial discharge,
- the frequency of partial discharges,
- FFT analysis in the frequency domain,
- the median value of partial discharge.

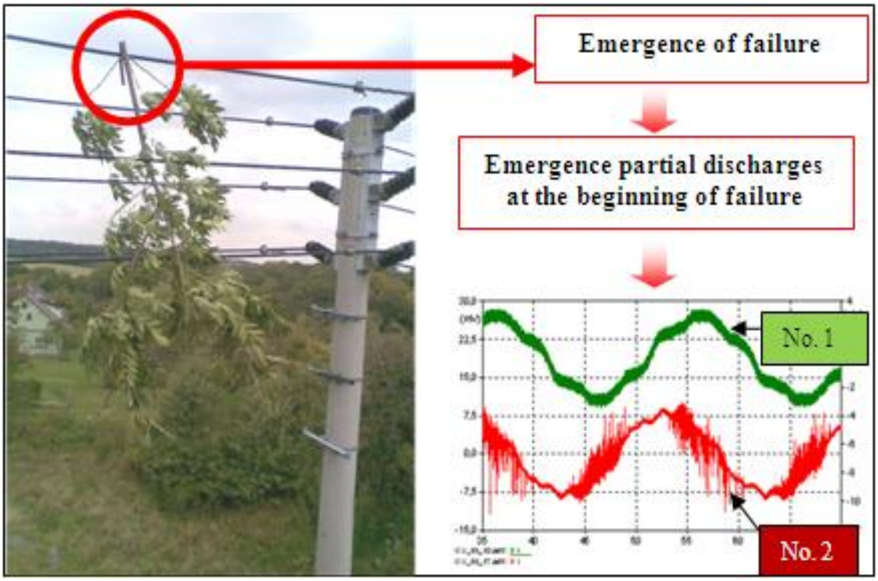


Fig. 2. The dependence of partial discharges on fault

Detector was designed and implemented. Detector can monitor the discharge activity of partial discharges and monitors weather conditions in the long term period of time.

Climatic variables are involved in the degradation of cover conductor. Measured parameters are therefore - pressure, temperature, dew point, humidity and global radiation.

The detector of failure of the covered conductors is placed on the mast medium voltage. Parts of the detector failures are in **Fig. 3**.

Primary requirement for the measuring system is oscilloscope's measurement of the timing voltage signal. This signal is scanned around the covered conductor hanging on the medium voltage lines. Measurement card connected via PCI interface is used for this purpose. Other measured parameters are temperature, pressure, humidity and light radiation in the vicinity. Two sensors are therefore used. The first sensor measures temperature, pressure and humidity. The sensor has an interface RS485. The second sensor is pyranometer.

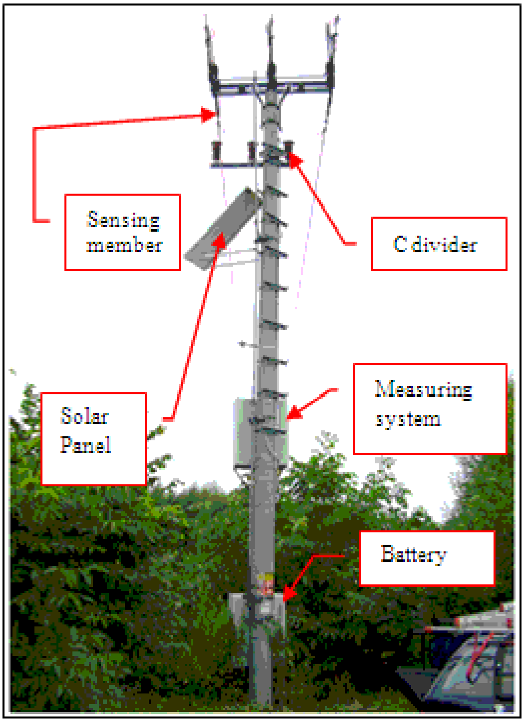


Fig. 3. The prototype detector failures of covered conductors

Measured data will be sent to an external computer via the GSM network, where will be processed. Modem for data transmission in the GSM network will be connected directly to the PC via Ethernet. We designed control electronic circuits for automatic control and switch on and off the computer. Entire supply measurement system will be provided by batteries charged by photovoltaic panels.

3 Conclusion

The proposed system to detect failures of covered conductor will be checked on real medium voltage lines. The detector is placed in the selected area and is exposed to real climatic conditions in the long time. Sensitivity and selectivity of the prototype detector will be determined by measurement. The whole system will be optimized to the needs of fault detection on overhead lines covered conductor. This prototype is developed based on methodology of patent P 2008-647. No more information can be given for this reason.

Grants and projects:

- SP2011/130 Energeticky soběstačný „SMART“ dům

Reference

1. Mišák, S., Hamacek S., "Utilization of the Finite Element Method For Optimizing of Overhead Covered Conductors", In Conference Proceedings DAAAM 2010, VIENNA:DAAAM INTERNATIONAL, 2010, vol. 21, 1111-1112, ISBN 978-3-901509-73-3.
2. Mišák, S., Hamacek, Š., "Utilization of the ANSYS Program Environment For Optimizing of Overhead Covered Conductors", Brno:MSD, s.r.o., 2010, 859, ISBN 978-80-214-4094-4.
3. Hamacek, Š., Mišák, S., "Analysis of Stress Applied On Insulation of AYKCY and NAKBY Type Cables In Fault and Operating Conditions Using the Modelling Options In the ANSYS Program", Ostrava, 15, 2010, ISSN 1210 – 311X.
4. Hamacek, Š, "EXPERIMENTAL ANALYSIS OF BREAKDOWN STRENGTH OF TYPES CYKY CABLES IN DEPENDENCE ON THE TEMPERATURE", Conference Proceedings ISEM 2009, CTU Prague, ISBN 978-80-01-04417-9.
5. Hamacek, Š., "Experimental Analysis of Breakdown Strength of Types CYKY Cables In Dependence On the Temperature With Electrostatic Field Simulation", Conference Proceedings ANSYS 2009, Pilsner, ISBN-978-80-254-5437-4.
6. Hamacek, Š., "Experimental Analysis of Breakdown Strength - Fixed Insulators", Thesis, VŠB-TU Ostrava, FEI, Kat-453, 2009
7. web pages, ELCOM, a.s. [quot. 2011-05-11], Available at WWW: http://www.elcom.cz/virtualni-instrumentace/produkty/bk-elcom2/clanek/366-ena440.html.
8. Kolcunova, I., Marci, M., "Measurement of Surface Discharge in Electroinsulating Liquids" 11TH INTERNATIONAL SCIENTIFIC CONFERENCE ELECTRIC POWER ENGINEERING 2010, PROCEEDINGS, Brno,CZECH REPUBLIC, ISBN: 978-80-214-4094-4.
9. Kurimsky, J., Cimbala, R., Kolcunova, I., "Multi-scale decomposition for partial discharge analysis" 13th International Power Electronics and Motion Control Conference, SEP 01-03, 2008 Poznan, POLAND, ISSN: 0033-2097.

Annotation:

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (ENET No. CZ.1.05/2.1.00/03.0069) and project SP2011/130.

Hydrogen Storage System – Electrolyser Supplied from PV Panels

Roman Chválek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
roman.chvalek@vsb.cz

Abstract. The fuel cells laboratories at the VSB-TUO have recently finished the realisation of the laboratory energetic system for accumulation of electric power from RES (with varying and unreliable production process). The system comprises of two parts. The first part of the system deals with production hydrogen using the electric power from photovoltaic panels and its storage. The other part of the system uses fuel cells to transform the energy from hydrogen into electric power. The system shall serve for analysis of various methods of accumulation under conditions within the Czech Republic and their utilisation in the insular operation mode.

1 Introduction

The current conditions associated with the increasing extent of non-controlled renewable energy sources (RES) within the electric power supply system assumes the accumulation of energy from such sources as an important prerequisite for an uninterrupted supply of electric power to customers. The problem with such sources, as the photovoltaic power plants are, lies in their backup for circumstances in case of sunlight insufficiency, as they are dependent on it. Accumulation of the energy produced might be one of the solutions, using hydrogen to accumulate the electric power with the return of energy into the power supply system in such periods, when these plants are not operable due to the lack of sunlight. There is a significant opportunity to use the RES system with accumulation facilities in the field of power supply backup under insular operation mode conditions. The said accumulation into hydrogen, being the carrier for energy, is still subject to intense research. In spite of the latter, there have been some examples of such accumulation methods applied in practice already around the globe.

2 Energy Accumulation System

The main parts are represented by the electrolyser and a fuel cell. These affect the resultant technical parameters of the hydrogen accumulation system. The output from both devices is directly proportional to the surface of electrodes, which play the

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 18–23.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

most important role in determining their final price. The total volume of energy accumulated depends on the hydrogen tank size and the magnitude of input from a renewable energy source, i.e. the photovoltaic panels.

The laboratory energy accumulation system (Fig. 1) consists of:

- 12 pcs of polycrystalline photovoltaic panels with the total nominal output of 2 kWp,
- semi-conductor DC/DC inverters, DC/AC convertors with a control system,
- 4 pcs of batteries - lead gel, type FLB 300,
- 2 pcs of hydrogen generator – Hogen electrolyser GC600 producing H_2: 1.2 l/min,
- hydrogen tanks – metal-hydride or pressure vessels,
- 2 pcs of low temperature cell modules type PEM with the output of 1.2 kW/1 NEXA module.

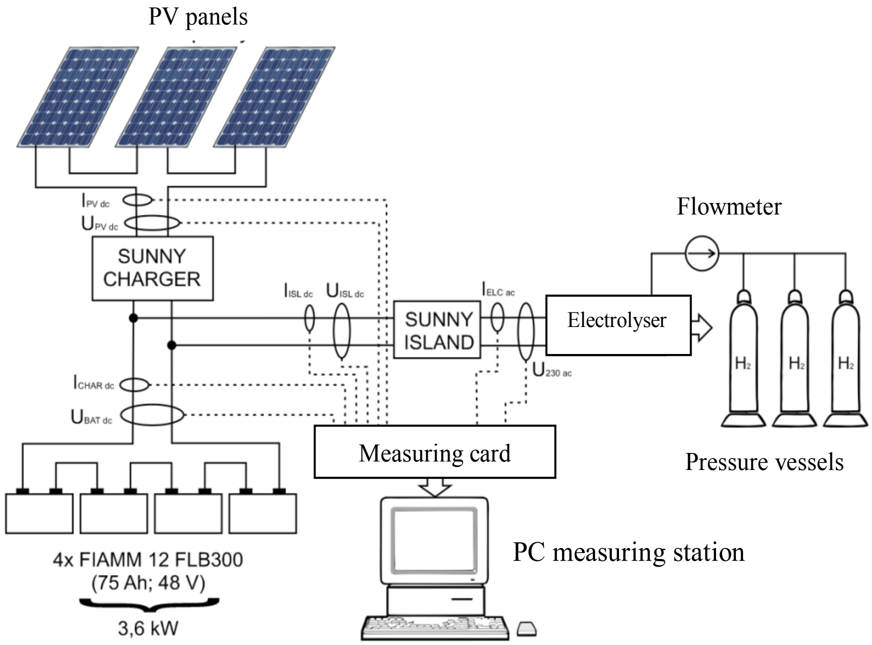


Fig. 1 Laboratory energetic system for accumulation of electric power from PV panels

The electrolyser supplied with electric power produced using photovoltaic panels decomposes de-mineralised water to hydrogen and oxygen by passing electric current. The compressed hydrogen is stored in a tank for further use. Oxygen generated during decomposition of water is either released into the atmosphere or stored within a separate vessel. Both of these gases can be further utilised to supply the fuel cell. Fuel cells are supplied with hydrogen from storage and oxygen from the atmosphere or a vessel. When using pure oxygen from a vessel, the efficiency of fuel cells increases resulting in overall improvement of the entire system, unlike using the oxygen from atmosphere. For a simplified diagram showing accumulation of electric power from photovoltaic panels into hydrogen and its subsequent utilisation in fuel cells see Fig. 1.

3 Measurements

Measurement was taken (for about 8 h) on the electrolyser Hogen GC600 (see Fig. 2) which was supplied by the energy of polycrystalline solar panels Schott Poly 165 (see Fig. 3) with a total installed capacity of 1980 Wp. The measurement was recorded by program Datalogger. The measured data were then evaluated by using program DIAdem2010 and MS Excel.

Fig. 2 Electrolyser Hogen GC600

Fig. 3 PV panel Schott Poly 165

From the measured values of currents and voltages in a group of solar panels and electrolyser were compiled following graphs (Fig. 4 and 5). From Fig. 6 is evident that the power supplied from PV panels was not enough after 3.pm. The control system changed power supply of electrolyser from batteries, which allowed its operation at least 30 minutes, until the voltage fallen down to about 52 V, then has been electrolyser switched off. The total amount of produced hydrogen was about 316 l.

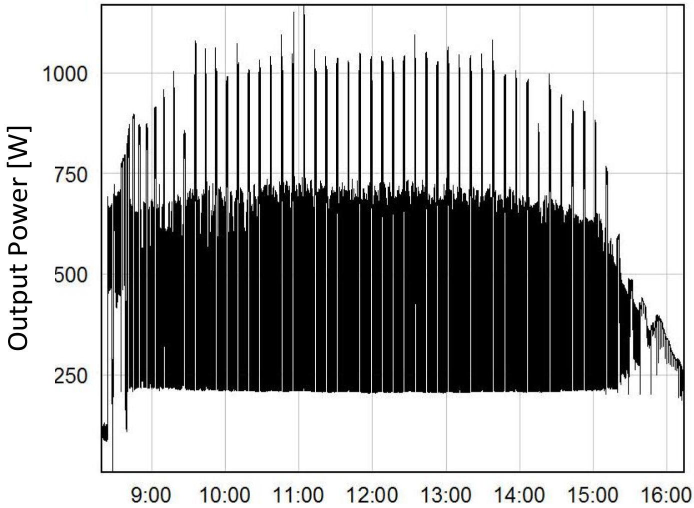


Fig. 4 Output power of PV panels

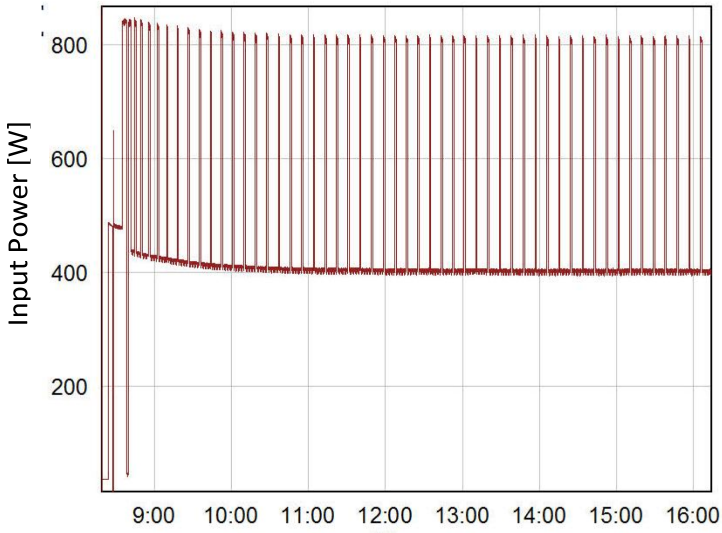


Fig. 5 Input power of electrolyser Hogen GC600

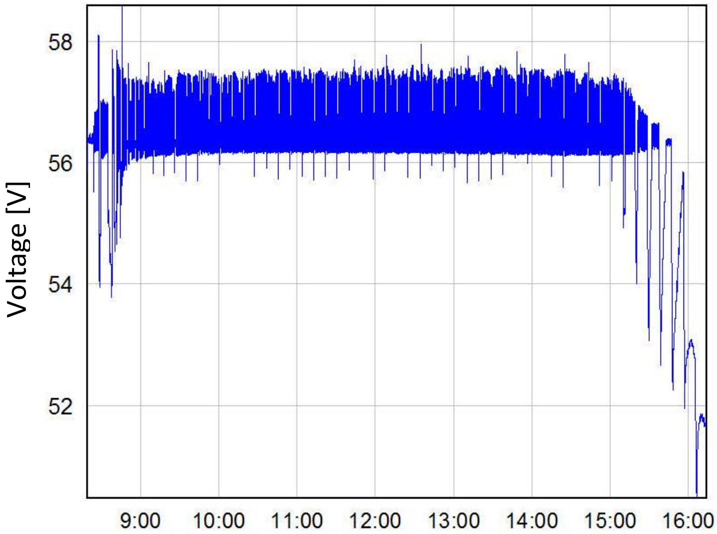


Fig. 6 Battery voltage

4 Conclusion

The use of solar energy in the PV power plants is an important way of generating electricity from RES. Its potential options are extensive and technically easy to use. Construction of hydrogen energy sector will also be motivated by the potential for hydrogen to become a fuel for stationary power sources. Secluded facilities using renewable energy sources, installed in areas with high potential solar energy may become a device for mass production of hydrogen. Accumulation of electrical energy is also important for Smart Grids and Micro Grids systems. It is a small island network, where you can compose multiple sources. An important parameter to decide on future implementations of commercial hydrogen storage system, in addition to the investment cost, will also determine the overall energy conversion efficiency of the system. Laboratory system powered by PV panels will be compared with the hydrogen production of electrolyser powered from the grid. This will need to perform additional measurements, which show the efficiency of individual parts of the storage system. The essentials values for evaluation will be voltage and current on polycrystalline panels Schott Poly165, Sunny Island Charger, Sunny Island 4282 and electrolyser Hogen GC600. To calculate the efficiency of PV panels will be needed also (in the cycle of hydrogen production) to measure the intensity of irradiation. From the obtained values will be computed power and efficiency of individual devices. In conclusion will be compared the overall efficiency and economical evaluation. From the results will be also possible graphically show and describe the processes (taking place directly in the electrolyser) such as pressurization and recovery of reaction water inside the electrolyser. Further continuation of the work will also look at the increasing efficiency of storage system because the limiting factor is used electrolyser, its efficiency ranges from about 19% to 27% depending on the time of its operation and especially on its switching frequency, which can cause losses from the necessary heating and pressurization. This type of electrolyser produces high purity hydrogen (99.9999%), which is particularly suitable for diagnostic purposes. Module(s) Nexa used in the system require hydrogen purity "only" 99.99%. From previous measurements on storage system as well as that in the control system is certain gaps manifested adverse operating conditions of the system. In the next period will therefore need to analyze in detail the operation of the system and optimize it in such a way that the aforementioned operating conditions completely eliminated. Last but not least, it will also need to be resolved storage of hydrogen, namely choice between the pressure bottles and metal-hydride vessels.

Acknowledgement

This work is devoted to experimental intention SP2011/193.

References/Publications

1. Hradílek, Z., Moldřík, P., Chválek, R.: *Electrical Energy Storage Systems*. In Impaktovaný zahraniční časopis Przeglad Elektrotechniczny, **2011**, vol. 87., čis. 2., p. 105-110
2. Moldřík, P., Chválek, R.: *Akumulace energie z fotovoltaiky do vodíku*. In časopis Elektrorevue, **2011**, vol. 13., 5 s.
3. Moldřík, P., Chválek, R.: Fuel cells reliability in energy storage system. In *sborník konference EPE*. Ed. Stanislav Rusek, Radomír Goňo, Ostrava:VSB - Technical University of Ostrava, **2011**, vol. 12., čis. 1., p. 472-474, VŠB - TU Ostrava, ISBN 978-80-248-2393-5
4. Moldřík, P., Chválek, R.: Hydrogen Storage Methods in the Fuel Cells Laboratory. In *sborník konference EEEIC*. Ed. M. Caciotta, Z. Leonowicz, Roma:IEEE Xplore, **2011**, vol. 10., čis. 1., 8-11 May 2011, 3 s., ISBN 978-1-4244-8779-0
5. Moldřík, P., Chválek, R.: PEM Fuel Cells - The Basic Characteristics. In *sborník konference EEEIC*. Ed. M. Caciotta, Z. Leonowicz, Roma:IEEE Xplore, **2011**, vol. 10., čis. 1., 8-11 May 2011, 4 s., ISBN 978-1-4244-8779-0
6. Moldřík, P., Chválek, R.: PEM Fuel Cells - The Effect of Fuel Parameters on Efficiency and Quality of Electric Power Supply. In *sborník konference EEEIC*. Ed. M. Caciotta, Z. Leonowicz, Roma:IEEE Xplore, **2011**, vol. 10., čis. 1., 8-11 May 2011, 4 s., ISBN 978-1-4244-8779-0
7. Moldřík, P., Chválek, R., Procházka, O.: Production of Electric Power Using Photovoltaic Systems in the Czech Republic. In *sborník konference EEEIC*. Ed. M. Caciotta, Z. Leonowicz, Roma:IEEE Xplore,**2011**, vol. 10., čis. 1., 8-11 May 2011, 4 s., ISBN 978-1-4244-8779-0

Grants

SP2011/193, GA ČR 102/09/1842

Simulation of the Electrical Field Between Parallel Plates

Lukáš Paulinský

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ok2pzp@centrum.cz

Abstract. Computer simulation has become an essential part of science and engineering. Digital analysis of components, in particular, is important when developing new products or optimizing designs. Today a large number of options for simulation are available. We have available basic programming languages for various levels of implementing advanced methods of simulation. A computer simulation environment is simply a translation of real physical laws into their virtual computer form. We can measure and then process simulated phenomena.

Keywords: electrical field, paralel plate, Poisson equation, charge density.

1 Introduction

Computer simulation now gives us the point of view about many physical laws. With Comsol Multiphysics there will be simulated – modeled electrical field between two parallel large plates. Gradual integration specifies and draws curves E and φ between two plates with a voltage U with different charge density.

Everything will be applied with different charge density and:

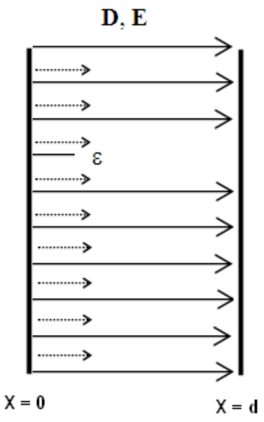
a) non-charge density: $\rho = 0$
b) constant charge density: ρ_0 = konst.
c) variable charge density: $\rho = \rho_0 \cdot \frac{x}{d}$

Between paralel plates there is no non-uniform dielectric. The dielectric is still linear, in other words:

$$P = \varepsilon_0\, x_e E$$

Let us assume that the area of the plates is A and their separation is d. as shown in figure 1.

VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.



Between the plates is dielectric elected air. Therefore, its permittivity is $\varepsilon_r = 1$. It may also be subsequently calculated, using this formula:

$$\varepsilon_r = D / E$$

It should be noted that the whole simulating object is set to the environment where the permittivity $\varepsilon_r = 1$, and so is the air. The boards + a – are chosen as material copper.

Fig 1. Paralel plates and non-uniform dielectric.

2 Non-charge density

By gradual adjustment of boundary conditions and substituting into equations, we get the equation for calculating the electric potential and intensity of electric field:

$$\varphi(\mathrm{x}) = -\frac{U}{d}x + U \quad \mathrm{E}(\mathrm{x}) = -\frac{\mathrm{d}\varphi}{\mathrm{dx}} = \frac{\mathrm{U}}{\mathrm{d}}$$

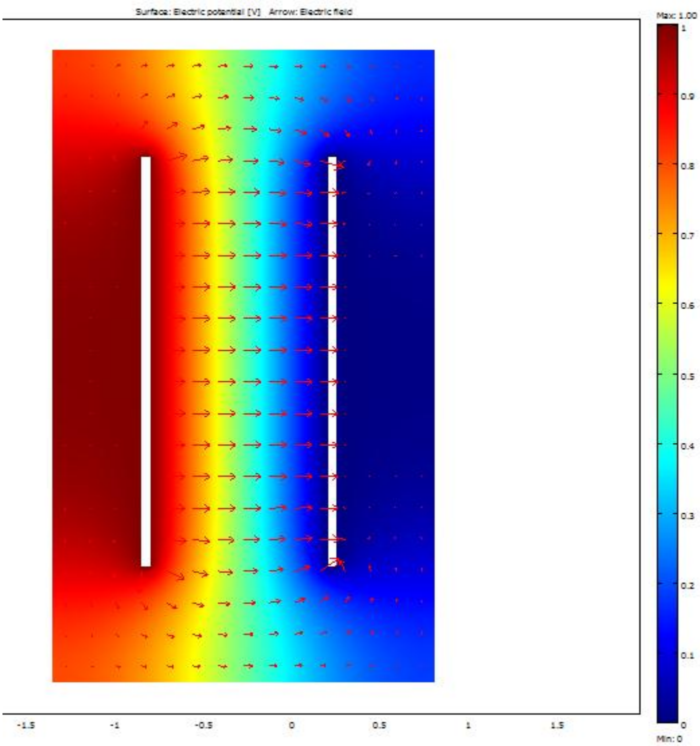


Fig 2. Surface – Electric potential [V].

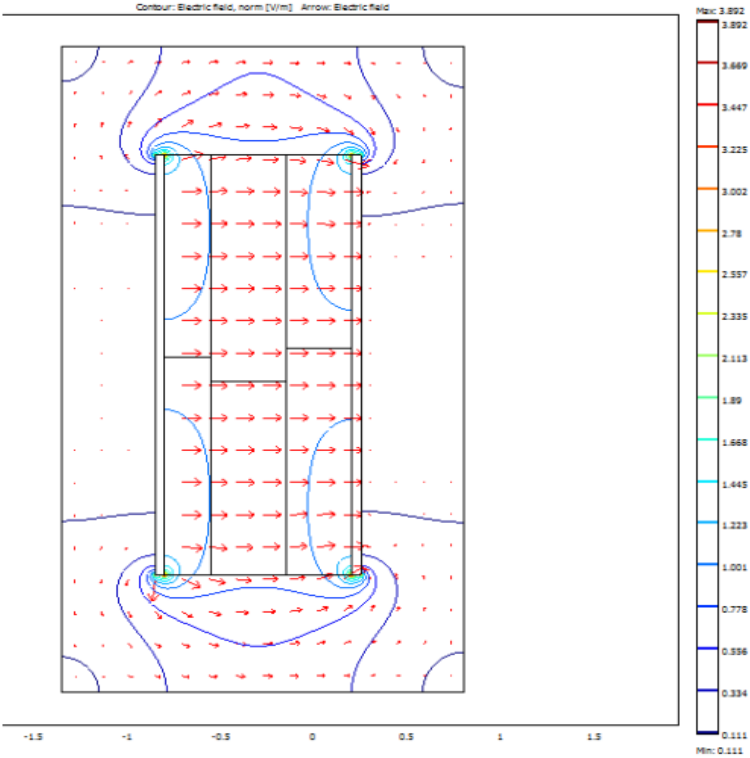


Fig 3. Surface – Electric field [V/m].

In figures 2 and 3, we can see the surficial simulation of electric potential and the electric field. Charging density between the plates was chosen here 0 C/m^3. Electric potential with red color is maximum value and it is 1 V. Electric field has a maximum value of 3.8 V/m and is located at the corners of the plates.

3 Constant charge density

This situation is similarly solved by dual integrating equations. The constants A and B are determined by the boundary conditions. The resulting equations are:

$$\varphi = -\frac{\rho_0}{2\varepsilon_0}x^2 + \left(\frac{\rho_0}{2\varepsilon_0}d - \frac{U}{d}\right)x + U\ ; \quad E = -\frac{d\varphi}{dx} = \frac{\rho_0}{\varepsilon_0}x - \left(\frac{\rho_0}{2\varepsilon_0}d - \frac{U}{d}\right)$$

,

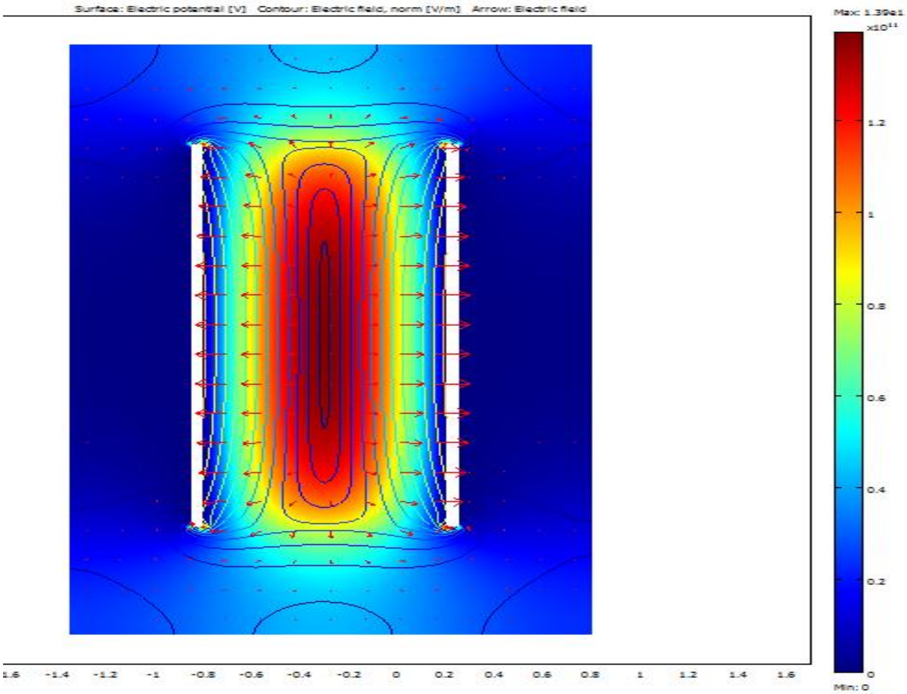


Fig 4. Surface – Electric potential [V].

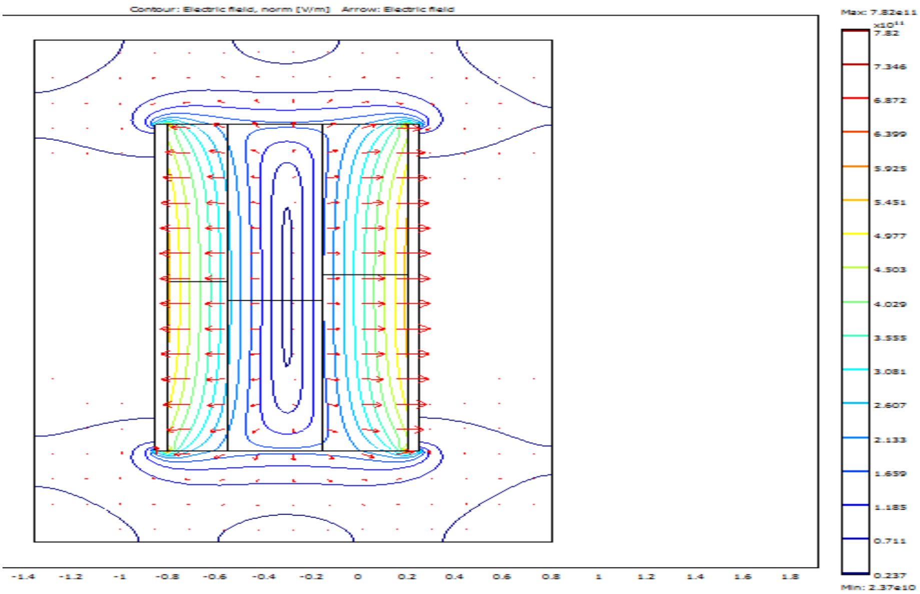


Fig 5. Surface – Electric field [V/m].

In figures 4 and 5, we can see the surficial simulation of electric potential and the electric field. Charging density between the plates was chosen here 10C/m^3. Electric potential with red color is maximum value and it is 1.39e11 V. Electric field has a maximum value of 7.8e11 V/m and is located at the corners of the plates. Compared with the previous models, the increased intensity appears on the inner surfaces of both boards.

4 Variable charge density

This situation with a variable charge density is similarly solved again by dual integrating equations. The constants A and B are determined by the boundary conditions again. After editing, the resulting equation equals to:

$$\varphi = -\frac{\rho_0}{6\varepsilon_0 d}x^3 + \left(\frac{\rho_0}{6\varepsilon_0}d - \frac{U}{d}\right)x + U \quad ; \quad E = -\frac{d\varphi}{dx} = \frac{\rho_0}{2\varepsilon_0 d}x^2 + \frac{U}{d} - \frac{\rho_0}{6\varepsilon_0}d$$

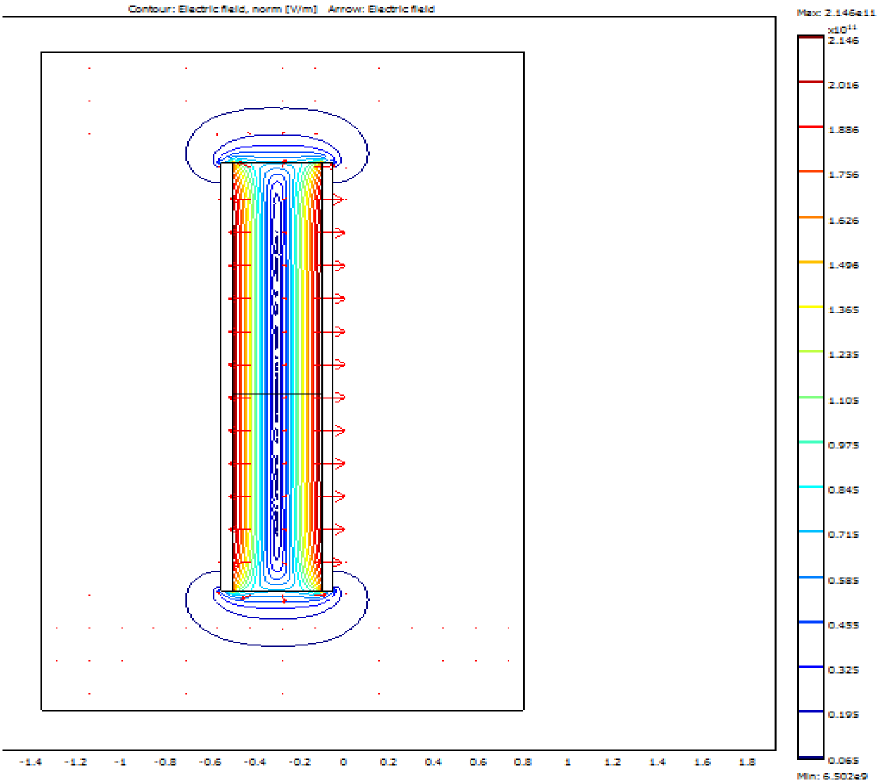


Fig 6. Surface –Electric field [V/m].

In this case, the volume density of electric charge between the plates is 10 C/m^3. The dependence E and φ on pattern plates is examined between them. It also reflected the reduction in electric potential 2.3e10 and 2.1e11 electric field V/m.

As you can see from the pictures, the electric field, which concentrated more on the space between the plates, altered the most compared to the previous case which spread out into space around the boards.

5 Conclusion

The entire simulation took place in the program of Comsol Multiphisic. First, it was necessary to declare the input conditions which the simulation took place. As already mentioned, the dielectric and the surroundings of the plates took place in an air environment. Bulk density of electric charge was set at a point of 10 C/m^3. As you can see from the attached pictures, variable E and φ grew the most with constant charge density which reached its greatest values more than in other cases. It is also well evident from Figure 5, where the electric potential was at its maximum value, which is greater than the intensity. Everything was also verified, using the resulting equations listed in each chapter. For the lack of space the resulting equations and pictures of input conditions were given. As already mentioned above, COMSOL is a very powerful tool to simulate various physical phenomena. The great advantage is that COMSOL itself is a graphical representation, from which you can make an image of a simulated problem.

References

1. Ivánek, L.: Elektromagnetismus – učební texty, v rámci operačního programu Rozvoj lidských zdrojů spolufinancováno z prostředků ESF a státního rozpočtu ČR: Ediční středisko VŠB – TUO Ostrava, 2007.

2. COMSOL AB, *Comsol Multiphysics.* USA: Comsol Publishing, October 2010.

3. Kulda, J.: Magnetické pole v silnoproudé elektrotechnice: ACADEMIA Praha, 1974.

4. Shepherd, M. R.: Electricity and Magnetism, October 2008.

Using the Spectral Analysis in the System of Elimination the Noise by Antiphase

Viktor Pokorný

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
viktor.pokorny.st@vsb.cz

Abstrakt. The basic issue of our solution is a noise generated by the transformer operation. If we ignore the constructional intervention into the transformers, we can proceed to the problem with active and passive solution. The passive solution can be covers, rubber silent blocks and elimination due to construction of the building. Our proposed solution for this problem is on active base by sending an identical acoustic signal turned to anti-phase, to the machinery causing an undesirable noise effects. During the optimal setting of this anti-phase there is a conflict of the wave and acoustic shunt fault which should cancel all noise activities expanding by the air during the optimum conditions.

Key words: Anti-phase, transformer, noise, acoustic signal

1 Introduction

Currently, the health of human population in any activity is a priority of human pursuit. Indivisible factor having an adverse impact on the human organism is a parasitic noise around us. Human physiology is not adapted to deal with noise emissions, which currently man produces, therefore are issued orders, regulations and directives governing the noise level to human in various activities in various areas.

2 Parasitic noise of electrical device

Methods of elimination can be divided into three basic systems:

1) constructional,
2) passive systems,
3) active systems.

Ad1) Constructional: systems heading to basic essence of parasitic sounds for each electrical device. Interventions for eliminate itself relate to construction and development groups and projectors.
Ad2) Passive solutions are covered, the rubber dampers and elimination using the building structure. Machine noise can be reduced by consolidating all the proper parts, different coatings, insulation covers.
Ad3) Active systems are devices capable of eliminating noise, such as using anti-phase. [Chmelík at al., 2003]

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 30–36.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

3 Antiphase

To understand the system is important to know basic, namely waves alone. Phase of the wave is a dimensionless quantity that determines the relation variable waves, (e.g. displacement noise) at that place and time and to the state variables characteristic waves in temporal and spatial origin. Dependence characteristic of variables determines the shape of „waves" regardless of its dissemination. Phase is a parameter, which depends on the timing characteristic values in a fixed location, which the wave passes, respectively spatial field characteristic values for a fixed moment in time. Noise so we can capture by the curve demonstrate displacement in time. Anti-phase is turning the current signal about 180°.
The result of the exact anti-phase is an absolute deduction of both signals and therefore their complete elimination. [Chu at al., 2001]

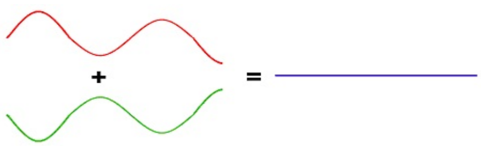


Fig. 1. Generating of antiphase wave

3.1 Generating the antiphase wave

This needs to be done as quickly as possible, and ideally output as soon as the input wave is received. This is largely a computing problem, and as the power of processors increases, this process can be tackled a lot quicker.

4 System elimination of noise using antiphase

System consists of measuring microphone that its high-quality electro-acoustical converter receives the noise signal and changing it to a sinusoidal voltage curve, it is received by the system, which inverts the signal. Inverted signal is moved exactly by half period. This part is the most important and also must be made with great accuracy because phase shift means the misoperation of equipment. This just turned and shifted signal is amplified by an amplifier and it is radiated to signal of noise, Fig.1 and Fig. 2. [Tomlinson at al., 2009]

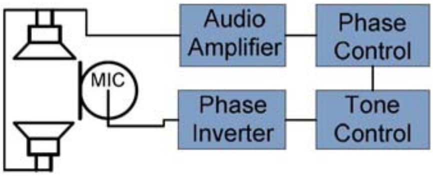


Fig. 2. The block diagram of noise-elimination by antiphase.

5 Purpose of the noise-elimination of the electrical equipment.

The initial equipment will be tested transformers to eliminate by using anti-phase. Their constant frequency noise is very ideal for sampling and subsequent rotation the signal to the anti-phase. The result will be possible limitations of a majority in the meantime use the anti-noise measures and restrictions of noise enclosure of buildings. There will be able to use materials of electrical sheets embodying increasing noise of transformers, but have better magnetic properties. The overall construction will be not pretty much influenced by the requirements on noise and thus increases the effectiveness of itself transformers. Construction of the transformer will also be economically favorable.

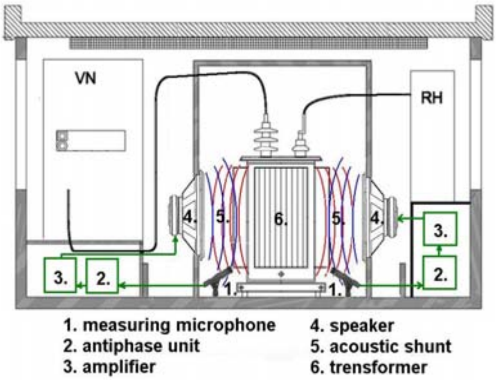


Fig. 3. Elimination of the transformer noise by anti phase.

After refinement of the system may be subject to certain physical conditions applied to almost any electrical equipment. Everything depends on the speed of sampling and capabilities of the right anti-phase waves. It depends to some extent on the frequency diversity of parasitic noise. The aim of our effort is to use this application to eliminate the noise of wind power plants.

6 Measurement and acoustic testing

The measurements carried out on four transformers installed at the VSB – Technical University of Ostrava. This is the core three-phase transformers old design. These transformers are not equipped with active cooling – fans that would participate in making some noise. Transformer cover is opened and the microphone scans the noise that is reduced to the closest possible distance in order to eliminate environmental influences. For this reason it is also used for directional microphone. Scanned signal by quality microphone is then sent to quality sound card that it digitizes and this signal is sent to a computer thru USB interface. It works only during the measurement as a digital recording device. The actual records are stored in WAV format, the sampling frequency of 48kHz and 24bit quality. This ensures the highest possible quality of recording and thus the quality of the result of subsequent processing of the acquired signal. Fig.4. shown a simple block diagram of the recording of measuring device, which is the source of the analyzed signals.

Fig. 4. Block diagram of recording devices for spectral analysis

For measurements were selected randomly three trans-formers, which are long-term measurement of noise intensity depending on load. Therefore, the results are taken as in no-load state and also in loaded state.

Subsequent analysis of the transformers seem numbered T615/1 of its parameters are 1000kVA , 10kV/400V, 50Hz.

Simulations were carried out in anti-phase action in software environment. The entire test is designed of a two mono tracks which were imported into the audio signals with frequency of 1kHz.

In the first case, the transformer worked to 7% of maximal load = 70A. In this demand can be considered, that transformer worked in unload state. The measured noise reached 74,7dB when the housing is closed and 75,65dB for open.

The latter was loaded transformer 480A per phase, it means 48% load of the total power. Measured noise reached very similar values as the unload state, 76,8dB for closed housing and 78,3dB for open.

6.1 Spectral analysis of transformer T615 in no-load state

Fig. 5. shows the FFT analysis of captured noise signal transformer T615 in the range 0-1500Hz. The transformer operates in the unload state with a large saturation magnetic circuit is expected to be significantly reflected by the frequency 100Hz causing electromagnetic noise, which also corresponds to twice the fundamental harmonic excitation current f=50Hz. There are also manifested in the excitation of upper harmonic due to deformation of the excitation current owing to saturation and the

shape of the magnetic circuit of the transformer. The images are then measured upper harmonic on audio track and basically equivalent to frequencies 200, 300, 400, 500, 600 and 800Hz. The magnetic flux is closed by air, housing of transformer and other design components and causing the vibration manifesting additional noise.

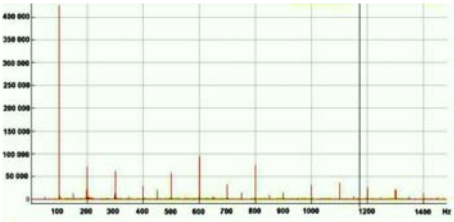


Fig. 5. Spectral analysis of transformer T615 unloaded state

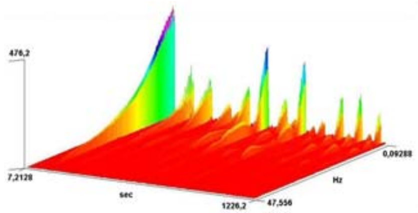


Fig. 6. Spectrogram transformer T615 in unloaded state

6.2 Spectral analysis of transformer T615 in the loaded state

Whole system is spectrally ease after loading the transformer and carrier frequency noise remains 100Hz. There are core to de-saturate the load transformer and thus also to mention upper harmonic. This state war confirmed by spectral analysis of the transformer in loaded state, see Fig. 7.

The size of the transformer noise depends on the magnetic flux, it depends on the voltage and then idle current. The noise of the transformer is strongly associated with the magnetization process. Saturation of the transformer core idling is higher and is given to the design, construction and the type metal. As can be seen from the performed measurements are an essential component of the noise frequency element 200Hz. This frequency band is due to the magnetostrictive effect of sheets, which will be most pronounced in the area of maximum permeability. This point is illustrated in Fig. 9. (this figure show teoretic current look) and Fig. 10. (figure show the current look on osciloscope). Upper harmonic to attenuation the exponential load transformer, which is tied to the decline in core saturation. The decay rate of component noise 200Hz is given by current course, this is why the more meaningful representation of frequency 200Hz, which is an even-numbered multiple of the fundamental power frequency of 50Hz. Even-numbered multiples of the fundamental frequency are given to the nature of magnetic induction.

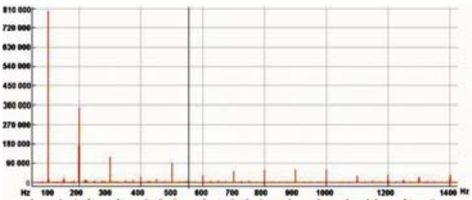


Fig. 7. Spectral analysis of the loaded transformer T615

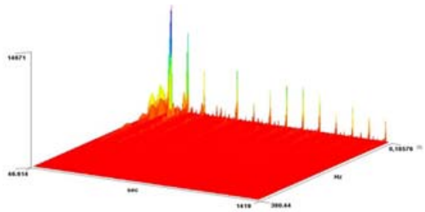


Fig. 8. Spectrogram loaded transformer T615

It is clear from these measurements, that entire system to eliminate noise using antiphase transformer id designed entirely to eliminate the noise around 100Hz. This applies to the measurement of transformers, in the case of transformers using active cooling would be difficult to analyze the situation of the fan noise.

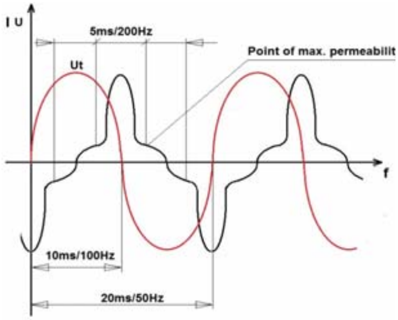


Fig. 9. Figure show teoretic look of current

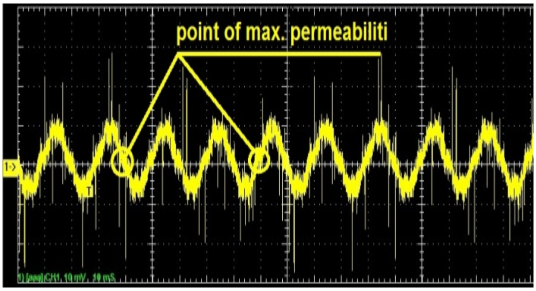


Fig 10. Figure show the current look on osciloscope Simulations

7 Presentation of results

The results of acoustic tests and measurements were presented at these conferences:

1. WOFEX 2010 PhD Workshop - Elimination the Noise of Electrical Transformers by Antiphase

2. Electric power engineering 2011 - Elimination the Noise of Electrical Transformers by Antiphase

3. 10 th EEEIC International Conference - Elimination the Noise of Electrical Transformers by Antiphase

8 Conclusion

If the certain physical conditions are kept, the system after improvement could be applied to any electrical equipment. This process is depended on a speed of sampling and ability to apply the right antiphase wave. There could be also influence of frequency heterogeneity of parasitic noise. The ambition of the project is to use this application for an elimination of noise various electrical equipments.

9 References

1. Chmelík, K.. (2003*) DC and commutator electrical machinery, general machinery*, VŠB – Technical University of Ostrava., ISBN 80-248-0314-3

2. Chu M.; (2001) http://gilmore2.chem.northwestern.edu/tech/anr_tech.htm, *Accessed on: 2010-04-19*

3. Krejci, P.; Prokop, L. & Misak, S.(2009) *Analysis of Long Time WPP Power Flows Measurement. In RUSEK, Stanislav, GONO, Radomir (ed.).*10th International Scientific Conference Electric Power E ngineering 2009., ISBN 978-80-248-1947-1, p. 373-378

4. Manwell J.F ;McGowan J.G & Rogers A.L, (2002) *Wind energy explained, design and application,* University of Massachusetts, Amberst, USA., ISBN 13: 978-0-471-49972-5 (H/B)

5. Tomlinson H.; (2009) http://www.mefeedia.com/– Anti-noise by Tomlinson Holman, *Accessed on: 2010-05-15*

Monitoring Power Quality Parameters in the Ostrava

Petr Rozehnal

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.rozehnal.st1@vsb.cz

Abstract. The monitoring of the quality of electricity distribution network and subsequent compliance with the parameters of power quality in distribution networks is very useful in terms of reducing transmission costs of electricity, followed by transformation, supply to the consumer and the subsequent correct functioning of electrical appliances. The quality of voltage is often reduced electrical appliances connected to the electricity network, as these appliances operate retroactively to the mains. Harmonic voltage distortion caused by the sine wave in the network.

Keywords: Monitoring, harmonics, flicker, unbalance

1 Introduction

Electrical power is considered to be the most noble from the energies used. This statement is backed-up mainly by advantages, among which there are such elements as electrical power production from various primary energy sources, possibility of long distance transport, easy backwards transformation to other types of energy (heat, light, mechanical energy, and to many other forms, which are foundation of modern telecommunications, information technologies, and entertainment), used in people's everyday lives, as well as in various areas of production and consumption. [2]

The electrical power quality is assessed by the CSN EN 50 160 standard via 13 parameters of voltage. Extensive use of appliances with non-linear characteristics in LV distribution networks (appliances such as TVs, computers, copy machines, compact lights, etc.) is substantially deteriorating to quality parameters of electrical energy supplied. [2]

The monitored parameters include pure sine wave with constant frequency and symmetrical three-phase system, which has constant amplitude. All these parameters

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 37–42.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

should be observed, but in the distribution system are several factors that deteriorate the quality of electricity and will thus cause interference in the electricity system and the possibility of unwanted impact. Failure to comply with power quality parameter gives rise to major problems or damage and that both the customer and the supplier. These problems include interference signals, warming devices, and the emergence of harmonics, voltage sags, surges, and flicker. All these issues cause operational restrictions of production, higher losses, disconnecting devices, etc. For this reason you need a device that will monitor the quality of voltage and then analyze these values. Power quality also addresses a number of standards that are used in the code transmission system, of which the core standard is a standard DIN EN50160. Among other standards, which seeks to ensure quality supply of electricity, EN 61000-2-2 include, EN 61000-2-4, EN 61000-3-2, and EN 61000-3-12.

2 Method of monitoring power quality

Monitoring of the power quality was gradually done in individual parts of the company. The measuring was done in a complex way within the LV distribution - mixed development consisting of family houses and apartment buildings. [1]

In accordance with the Standard CSN EN 50 160, the measuring and evaluation of the power quality of single points was done in one week intervals, while the parameters of quality were evaluated for 10 minute intervals in the course of measuring.

In single feeder points they evaluate measured data in all phases and on all voltage levels:

- Selected voltage harmonics (3., 5., 7., 9., 11.)
- Flicker
    - Unbalance

3 The content of harmonics

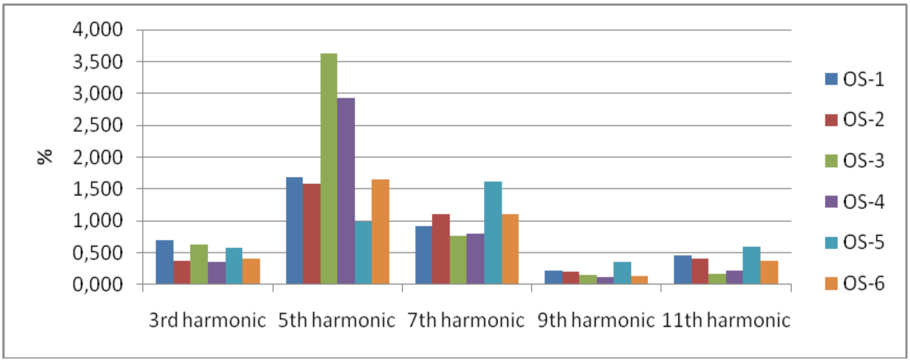
With the increasing use of electronic devices, the incidence is increasing harmonic voltage. Suppliers of electricity should be supplied to homes and businesses clean sinusoidal voltagewith constant frequency and amplitude. Nonlinear appliances connecting to the mains mean that these appliances are taken from the network of non-linear current. This current causes the impedance of the network voltage drops and the distortion sine wave. In the case of households it can cause, computers, televisions, fluorescent lamps and network adapters. In industrial enterprises is the creation of harmonic voltage caused most converters, network resources, climate, etc. Most occur frequencies 150, 250 and 350 Hz, i.e. 3rd, 5th, and 7th harmonic.

When monitoring power quality are monitored values up to 50. harmonics, which is 2,5kHz. Electrical network, transformers and appliances are designed for power frequency of 50 Hz. High frequency shares, which causes distortion sine wave results in a malfunction or some appliances, damage, higher operating costs and the higher price of electricity. For the harmonic tension is true that under normal operating conditions does not exceed 95% of the 10-minute averages of rms values of harmonics in any one-week period size. These values are given in the following table.

Table 1 - Values of the harmonic voltage in Un% for the 25th harmonic orders including

Odd harmonics Not multiples of 3		Odd harmonics Multiples of 3		Even harmonics	
Order harmonics h	Voltage harmonics %	Order harmonics h	Voltage harmonics %	Order harmonics h	Voltage harmonics %
5	6	3	5	2	2
7	5	9	1,5	4	1
11	3,5	15	0,5	6...24	0,5
13	3	21	0,5		
17	2				
19	1,5				
23	1,5				
25	1,5				

The following chart shows the measured values of harmonic content in parts of Ostrava in the low voltage level.



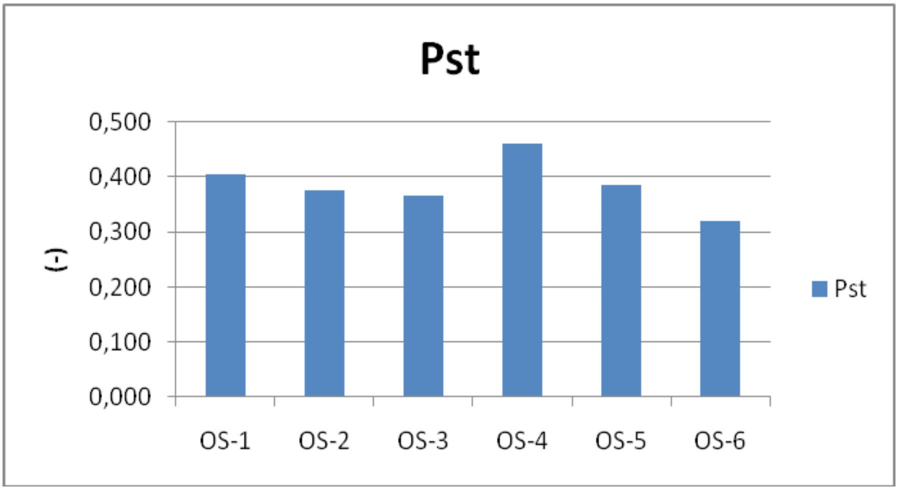
Graph 1 - Comparison of individual harmonics in various parts of Ostrava in the low Voltage

The graph above shows that the highest incidence is at the 5th harmonic components, as well as at the 7th third harmonic, 11 and 9 harmonic voltages. The above graph shows that none of the readings does not exceed the limits specified in EN 50160 the most 5th harmonic effect the OS-3.

4 Flicker

Flicker is the sensation of instability of visual perception induced by light stimulus whose luminance or spectral distribution fluctuates with time. This phenomenon is a major problem in the case of light sources, it adversely affects the human eye and the use of light sources in an environment where there are rotating electrical machines, could cause serious injury.

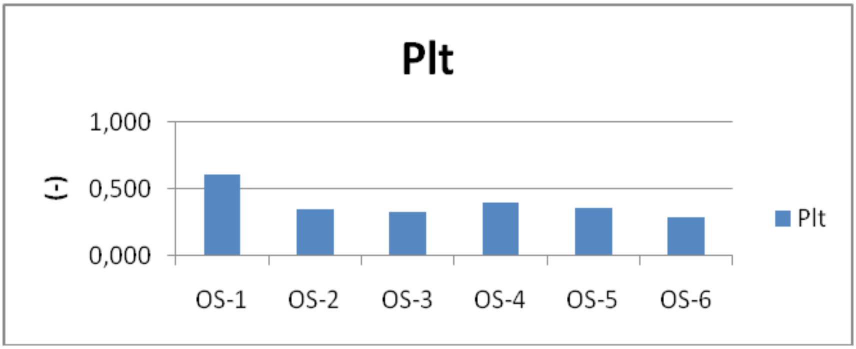
During the period in Ostrava was carried out short-term rate monitoring parameters flicker severity P_{st}. This monitoring was carried out at low voltage level.



Graph 2 - Comparison of individual values short-term flicker severity level in the areas of Ostrava.

The graph 2 shows that the largest value of short-tem flicker severity level of the OS-4.

During the period in Ostrava were also carried out monitoring the parameters of long-term rate of flicker severity P_{st}. This monitoring was carried out again at the low voltage.



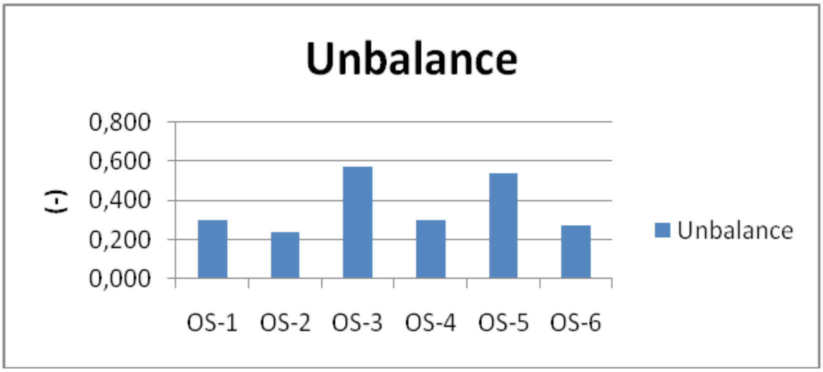
Graph 3 - Comparison of individual values of long-term rate of flicker severity P_{lt}

From Graph 3 can be seen that the greatest value of long-term flicker severity level in the OS-1. The highest measured value exceeds the limit value 1, which is given in EN 50160.

5 Unbalance

Three-phase voltage system is called symmetric, if the voltage of each phase phase-shifted by 120 ° and has the same voltage amplitude. In the event that one of these conditions is not met, the system is called unbalanced. Voltage unbalance is caused by uneven loading of each phase.

For the asymmetry is true that under normal operating conditions, the ratio is not 10-minute RMS negative sequence voltage and the corresponding size of the direct component exceed 2% in 95% of the measurement interval week. In Ostrava was done at several sites tracking asymmetry parameters at the low voltage an are shown in Figure 4.



Graph 4 - Comparison of individual values of unbalance in the Ostrava

In Graph 4 shows the values of asymmetry in different parts of Ostrava. The graph shows that the largest value of asymmetry has been made in OS-3 and OS-5.

6 Conclusion

In areas of Ostrava measurements were taken of selected power quality parameters. Measured values are given tables, from which it is obvious that all values are measured in accordance with EN 50160 ed.2, which is a key standard for the reference power quality parameters.

References

1. Santarius P., Krejci P., Chmelikova Z., Ciganek J. Long-term Monitoring of Power Quality Parameters in Regional Distribution Networks in the Czech Republic. Conference “ICHQP 2008”, Wollongong
2. Rozehnal P., Cumpelik R., Krejci P., Evaluation of selected parameters of quality of electric energy distribution network. Conference “EPE 2011”, Kouty nad Desnou

Application of Statistical Methods on Power Lines 110 kV Failure Database

Jaroslav Sojka

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jaroslav.sojka1@vsb.cz

Abstract: This paper presents a procedure for processing data from database failures 110 kV lines, using the selected statistical methods. The database is analyzed according to the causes of failures. The most common causes of failure include *Natural influences* (94), and disorders caused by the O*peration and maintenance* (55). The use of selected statistical methods aims to determine that this group of disorders is within their durations individual failures more significant. As part of this work was necessary to familiarize yourself with the program for statistics - Statgraphics [2].

Keywords: Failure database, exploratory analysis, Kruskal-Wallis test, Box with whisker plot

1 Introduction

Electricity is still among one of the most important needs for human life on Earth. The issue of quality and reliability of energy supply is therefore an important parameter. Interruption of power supply and energy at all is a big problem in terms of both production and financial compensation. Decree 41/2010 Coll. The quality of electricity supplies and related services in the electricity industry, set standards, the interruption of transmission or distribution of electricity [3]:
a) 18 hours on the grid with voltage levels up to 1 kV and 12 hours on the grid with voltage levels up to 1 kV in the capital city of Prague.
b) 12 hours in grid networks with voltage levels above 1 kV and 8 hours on the grid with voltage levels above 1 kV in the capital city of Prague.
For non-standard termination of interruption of electricity distribution in accordance with paragraph 1 provides the distribution system to the customer a refund of 10% of its annual payments for distribution, maximum:
a) 6 000 CZK in networks up to 1 kV
b) 12 000 CZK in the networks of over 1 kV to 52 kV
c) 120 000 CZK in the networks above 52 kV

Careful management and processing of information resulting from failure of the power line, helps us to address system reliability and supply electric power customers. To process data from the database we can not do without knowledge and use of statistical methods.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 43–48.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Description of dataset

For statistical analysis I used the data as data on failure 110 kV power distribution company. These data describe the start and end of the failure, date of creation and removal of failure, each failure type, failure location and the name and number of high voltage lines, from 2000 to the beginning of 2011. The following table shows sample input data (332 entries) with the imputed value of the duration of disturbances reported in minutes.

Year	Month	Time of failure	Start date of failures	End time of failure	End date of failures	Duration of failures (min)	Cause
2000	2	14:00:00	28.02.2000	13:04:00	29.02.2000	**1384**	Foreign influences
2000	2	13:47:00	28.02.2000	13:53:00	28.02.2000	**6**	Foreign influences
2000	2	11:29:00	28.02.2000	11:33:00	28.02.2000	**4**	Foreign influences
2000	2	09:03:00	28.02.2000	09:28:00	28.02.2000	**25**	Foreign influences
2000	3	14:59:00	09.03.2000	16:35:00	09.03.2000	**96**	Operation and maintenance
2000	3	14:56:00	09.03.2000	15:24:00	09.03.2000	**28**	Natural influences
2000	3	22:08:00	10.03.2000	22:17:00	10.03.2000	**9**	Operation and maintenance
2000	3	12:46:00	24.03.2000	13:15:00	24.03.2000	**29**	Cause unknown
2000	4	05:43:00	23.04.2000	13:09:00	23.04.2000	**446**	Abnormalities of the power system

Fig. 1 Failure database

3 Exploratory analysis of fault length t [min]

Exploratory data analysis in statistics is a summary of the methods used for data exploration and the search for hypotheses that are worth testing. The processing of file obtained statistical variable must organize to form and illuminating description is a few values that would contain the greatest amount of information contained in the original file. This part of data processing is called exploratory (descriptive) statistics and is the first step to reveal information hidden in a large number of variables and their variants. As input data for the exploratory analysis, I used the failure length (in minutes) at 110 kV overhead lines (332 entries) for the years 2000 to 2011.

3.1 Statistical characteristics

	Natural influences	Operation and maintenace
Lower quartile	3,000	9,000
Upper quartile	29,000	350,000
	Natural influences	**Operation and maintenace**
Average	233,085	233,127
Std. Error	110,665	46,440
Median	6,000	83,000
Mode	3,000	9,000
Std. Deviation	782,518	328,381
Variance	618918,143	109830,965
Kurtosis	27,483	2,949
Skewness	5,046	1,778
Range	5413,000	1440,000
Minimum	0,000	0,000
Maximum	5413,000	1440,000
Sum	21910,000	12822,000
Count	94,000	55,000

Mode $\hat{x}$*:* is defined as the most numerous variant variables.
Median $x_{0,5}$: dividing a data file so that half (50%) values are smaller than the median, and half (50%) values larger (or equal).
Selection slant $\alpha > 0 \Rightarrow$ variable values less prevalent than the average.
Selection spikiness $\beta > 0 \Rightarrow$ corresponds špičatému division variables

Fig.2 Statistical characteristics

3.2 Graphic data processing, Histograms and Box with whisker plot

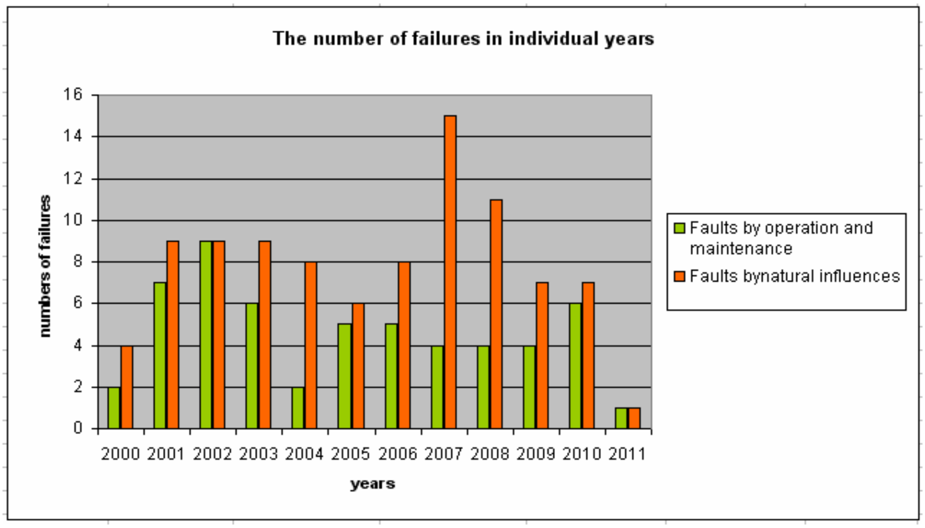


Fig. 3 Number of failures in individual years

The histogram shows the frequency (number of) in a given class of data. To illustrate the histogram created in the program Statgraphics. It displays the number of individual variables in the given limits.

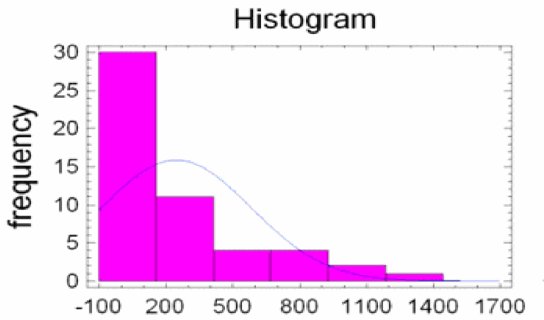


Fig. 4 Histogram Operation and Maintenance category

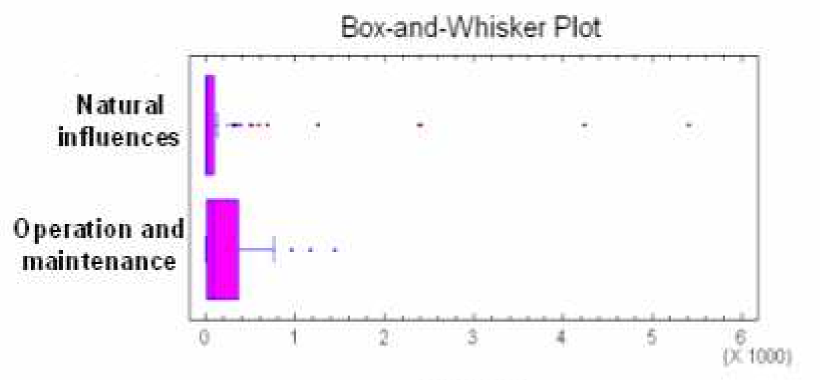


Fig. 5 Box with whisker plot

3.4 Identification of outliers

When evaluating the obtained data we will need to identify whether the data are not so-called outliers. The method of identification:
For outliers can be considered a value x_i, the absolute value of the z-coordinate is greater than 3:

$$z - coordinates._i = \frac{x_i - \bar{x}}{s} \qquad \left(\left|z - coordinates._i\right| > 3\right)$$

Z-coordinate is "less stringent" to outlying observations than the median coordinates. This is because the z-coordinate is determined based on the average and sample standard deviation, which are strongly influenced by the values of outliers. [1]

Outliers (min)	
Operation and maintenance	Natural influences
964	5413
1157	4236
1440	2403
	2379
	1250
	690
	587
	516
	496
	491
	385
	347
	342
	329
	319
	311
	302

In the data file, for example, there are eight failure from category of *Natural influences*, and three failure from category of *Operation and maintenance*, which have a duration of 0 min. Unfortunately, from the data provider has not explained this matter. The data also occur due to disturbances such as natural disasters or states of acute symptoms of this disorder and the time is over 16 hours. In terms of next steps I take these data to be statistically important, so leave them in a file. The reason is that data are measured in practice and the final form distorted data file.

Fig. 6 Outliers

4 Analysis of variance

Analysis of variance tests the hypothesis of several selections (H0), which assumes that the individual selections come from the same underlying probability distribution and, therefore, that the average duration of failures in these two categories are equal (H0: $\mu1 = \mu2 = ... = \mu k = \mu$). Compares it with the alternative hypothesis (HA), which assumes that the average duration of failures in these two categories Animals Art is not the same (HA: not H0).
This analysis can be performed in two ways, using the table using ANOVA and Kruskal-Wallis test. [1]

4.1 Testing normality of data

The analysis of variance, I ascertained, how far apart from each period of failures in the categories. But I had to choose the appropriate method (ANOVA or Kruskal-Wallis), based on testing the normality distribution of these random selections.

4.1.1 Kolmogorov – Smirnov test

For this test I chose the Kolmogorov-Smirnov test, using Statgraphics software. Later is the procedure for this test.

Determination of hypotheses:

H0: differences between the empirical distribution function Fn (x) and distribution function F (x) are statistically insignificant and have a normal distribution.

HA: the differences are statistically significant and have a normal distribution.

The test criterion is a random variable: $D = \sup_x |F_n(x) - F(x)|$

Critical field for this test is: $W_\alpha = \{D : D > D_\alpha(n)\}$

The output value P-value of Kolmogorov - Smirnov test (from Statgraphics) categories of causes for failure are listed below.

Natural factors: Approximate P-Value = 0.0.
Operation and maintenance: Approximate P-value = 0.00000163138.

For both types of causes failures based on the P-value <0.01⇒ reject the null hypothesis H0.

Based on the Kolmogorov-Smirnov test, function Fn (x) have not normal distribution. Analysis of variance I made by Kruskal-Wallis test.

4.2 Kruskal-Wallis test

Kruskal-Wallis test is similar to the non-parametric analysis of variance of simple classification. Not normality distribution. Its disadvantage is lower sensitivity. As published, the data are not normally distributed, then doing an analysis of variance using Kruskal-Wallis test.

Determination of hypotheses:
H0: median duration of failures of these two categories of causes are equal.
HA: median duration of failures of these two categories of causes are not equal, the differences are statistically significant.

Sample Size	Average Rank
Causes Nature influences	**60,5575**
Causes Operation and maintenance	**85,7981**
Test statistic = 12,8032	**P-Value = 0,000345782**

Tab.1 Output table from Statgraphics:

P-Value < 0,01 ⇒ rejection of the null hypothesis H0

Based on the Kruskal-Wallis test, the median duration of failures of these two categories of failures are not equal, the difference is statistically significant selections. The above conclusion is confirmed by the comparison on the Box with whisker plot Fig. 5.

6. Conclusions

Statistically disorders were evaluated length of database failure VVN 110 kV lines in years 2000 - 2011. From this database were selected two largest categories of causes of failures (*Natural influences* and *Operation and maintenance.* Detected outliers are reported in Chapter 3.4. The resulting values were retained in the data file, as they correspond to problems that occur less frequently. Were also created Box with whisker plot in the program Statgraphics, from which we get a good overview of the center and dispersion of variable durations failures. Data according to the Kolmogorov-Smirnov test did not have a normal distribution of values (for both types of disorders based on the P-value <0.01 $\Rightarrow$ reject the null hypothesis H0). Based on the Kolmogorov-Smirnov test for analysis of variance, I used the Kruskall-Wallis test. Based on the Kruskal-Wallis test (PVALUE <0.01 $\Rightarrow$ reject the null hypothesis H0) the median duration of failures of these two categories of causes are not equal, the difference is statistically significant selections. The analysis of data shows that the duration of failures caused by the *Operation and maintenance* is statistically more significant than the duration of failuress caused by *Natural influences*. And despite the fact that *Natural influences* are found errors that last longer than 10 hours, these disorders are caused by natural disasters, which in ten years, there were very few.

Acknowledgement

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2011/193).

References:

[1] Bris, R. - Litschmannová, M.: Statistics I, subject to the Syllabus 2004
[2] Litschmannova, M.: Statistics I. Exercise
[3] ERÚ: Decree 41/2010 Coll. the quality of electricity supplies and related services in the electricity, the Act 458/2000 Coll.

Automatic Starting of Plasma Furnace

Petr Stoklas[1], Martin Marek[1], Daniel Petlák[2], and Peter Staš[1]

[1]Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.stoklas@vsb.cz
[2]Department of Non-ferrous Metals, Refining Processes and Materials Recycling,
Faculty of Metallurgy and Materials Engineering,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba

Abstract. The paper deals with the commissioning of modern concept of the plasma furnace management. At the beginning of paper is illustrated and described solved furnace and its individual parts. The following chapter describes design and realization of proposed plasmatron starting automat. The next chapter describes analysis of real plasma furnace design. Part of this thesis is comparison of new and old power supply, and conclusion, where is summarizing the results of achieved work.

Keywords: Plasma re-crystallizing furnace, plasma, electric arc, plasma burners, power systems, automatic starter.

1 Introduction

Plasma metallurgy was begun by the need for smelting of metals and alloys of high purity. For smelting and heating in plasma furnaces the low-temperature plasma is used at temperatures up to value 30 000 ° C. The plasma is generated by a plasma burner, which uses electromagnetic controlled arc used with inert gases. According the involvement of plasma burners we distinguish the burners with an independent electric arc, burners with dependent electric arc and burners with a combined involvement. This paper deals to the upgrading of specific plasma furnace and especially the design and realization of plasmatron starting automat.

Detail of this issue is described in works [1] and [2].

2 Plasma furnace

Discussed recrystallization plasma furnace with horizontal crystallizer is operated by the Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering in Laboratory of Plasma Metallurgy. The furnace is used for smelting and preparation of pure metals and various alloys. The furnace is shown in Fig. 1.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 49–54.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Fig. 1. Plasma furnace

Plasma furnace consists of these basic parts: /1/ plasma burner; /2/ outer shell of furnace; /3/ cooling circuit; /4/ process indicators of voltage and current; /5/ control panel; /6/ panel indicators and measuring instruments; /7/ control and signalization of starting automat; /8/ engine for shifting material in a boat under the cover; /9/ control electronics under the cover.

The essential part of the furnace is a plasma burner (Fig. 3), which has the task of concentrating the thermal energy and stabilizes the electric arc. It consists of cathode and anode. In the case of the combined involvement, we talk about auxiliary and working anode. Auxiliary anode serves to ionize the space between the cathode and working anode, on which there is the heated (smelted) subject.

The second essential part is the crystallizer (Fig. 5), which is located inside the furnace (Fig. 4). For controlling the furnace serves the control panel where you can control engine for shifting crystallizer (boat), to set the value of electric current for setting a suitable temperature of plasma, and then manually launch the furnace. Starting of the furnace is now realized using the modernized control of furnace (Fig. 2). All you need for starting the furnace is to press a single button. If necessary, the furnace may be started by the old method.

Near the furnace there is cooling device (water tank). This tank serves as a water reserve in the event of closure of water pipe, which is located outside the room where the furnace is located. Operator of the furnace has still some time for smelting and in the absence of water in the tank he must immediately turn off the furnace.

The crystallizer has the shape of a trough and for the need of different sizes of ingot it can have different shapes, for example, e.g. see Fig. 5.

Fig. 2. Photography of the electronic furnace after modernization

Fig. 3. Detail view of the plasma burner

Fig. 4. Detail view of the location of the crystallizer and nozzle burner

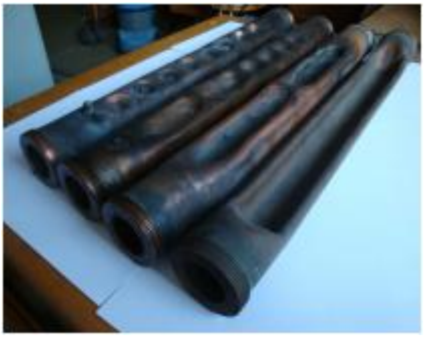


Fig. 5. Samples of copper crystallizers [2]

3 Design and realization of plasmatron starting automat

The main task was to replace the outdated furnace control. The newly proposed method of managing arc furnace eliminates the need for manual ignition. Manual mode ignition of the arc was causing arcing between the cathode and the auxiliary anode for longer time than is was necessary for ignition of the arc. That led to great (mechanical) stress of the nozzle (diffuser), therefore it was necessary to change the nozzle (diffuser) quite often, which was time costing and expensive.

Power diagram of the furnace is shown in Fig. 6. Starting resistor which is connected in series with the auxiliary anode is used for the controlled increase in current passing through the arch, which is only connected at startup. Low resistance of electric arc could cause a short circuit, and therefore it is necessary to limit current value using resistor. The arc can then be reliably ignited, without giving the emergence of a narrow unmanageable arc, given the sharp increase in current.

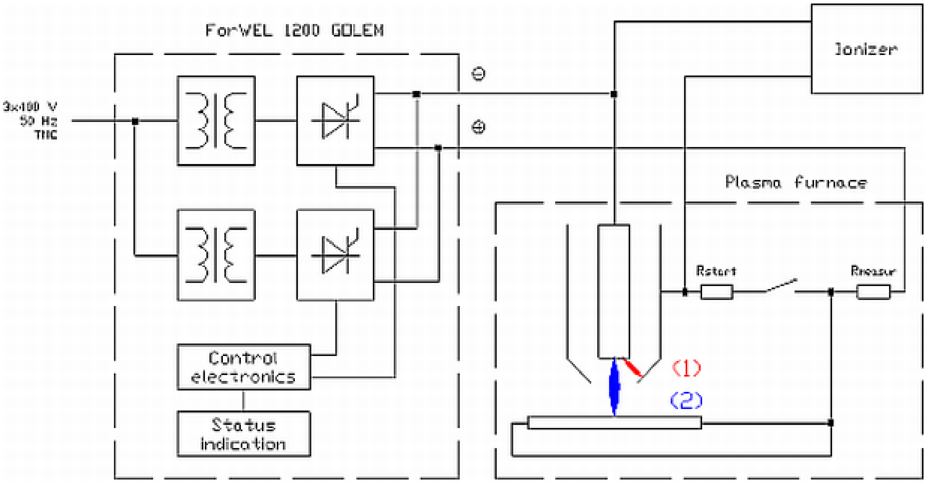


Fig. 6. Power diagram of the furnace

The electric arc is first created using the ionizer between a cathode and an auxiliary anode (Fig. 6 /1/). After ignition of the arc the contactor disconnects the auxiliary anode. Subsequently, the arc is transferred to the working anode (Fig. 6 /2/). More details about starting the process described below.

When designing a new controlling of furnace the effort was devoted to the easiest operation and so that the operator of furnace needn't perform redundant operations (e.g.: the need for switching contactors to connect the auxiliary anode, subsequently ignition of the arc, and disconnecting the contactor after a suitable period of time).

The old procedure of starting the furnace

1. Turn on power supply.
2. Toggle the switch for connecting the auxiliary anode.
3. Connect the auxiliary anode by switch and keep the switch in position "ON".
4. Press the "START" button.
5. Check electric arc in furnace through the peephole.
6. Release the "START" button.
7. Return the auxiliary anode connecting switch to its original position "OFF".

A new procedure of starting the furnace

1. Turn on power supply.
2. Turn on circuit breaker of new control electronics (if it was off).
3. Press the start button.

After this procedure the electric arc is ignited. It should be kept in mind that before running the process itself the furnace must be checked to determine whether the water supply is not closed to the cooling system, to vacuum the space in furnace, to set the appropriate flow of gas or its mixture.

After igniting the arc it is necessary to set suitable current for the meltdown and to regulate it according to further needs. For shifting crystallizer inside the furnace the operator sets the proper shift speed.

The modernized design allows igniting the arc by pressing a single button. Two different timers are used. After pressing the first button the first timer is activated, switching the first contactor, which connects the auxiliary anode contactor, and subsequently there is a delay to activate the second timer. The second timer activates the second contactor which ensures ignition of the arc. This process is indicated by three indicators. The first indicator indicates the presence of the control voltage. The second indicator indicates the auxiliary anode contactor activation. The third indicator indicates connection of high voltage transformer, which causes a spark igniting of the arc in the space between the cathode and the auxiliary anode.

Electric arc is ignited by pressing a single button. By using two timers and two contactors the automatically connection of the auxiliary anode is ensured, then the electric arc is ignited and then the auxiliary anode is disconnected.

4 The analysis of real plasma furnace design

Power supply for plasma furnaces must be current-adjustable. Value of current mustn't fall to zero, because it would cause extinction of the electric arc.

Parameters for simulation are based on the real values of the solution furnace.

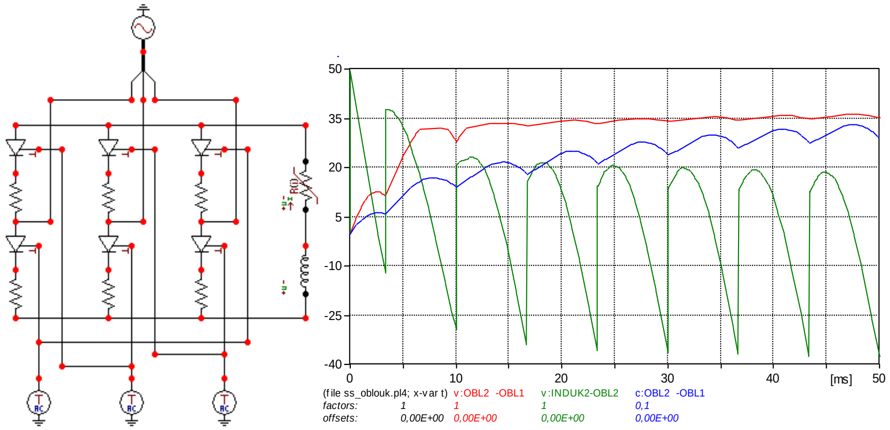


Fig. 7. Scheme of controlled rectifier with the model of electric arc, and simulation waveform of current and voltage in discussed furnace for the value of a circuit inductance 1 mH

The voltage of source has a value 33 V. Resistors are connected in series with the rectifier diodes and have a value of resistance 5 mΩ. Nonlinear resistor has values which were acquired from the measured volt-ampere characteristics of the discussed furnace. Factor for the parallel resistance of the component showing the inductance of the circuit is kp = 0 Ω.

5 Comparison of new and old power supply

Comparison of measurements made in the furnace using old 6-pulsed power source [3] with a new 6-pulsed power source.

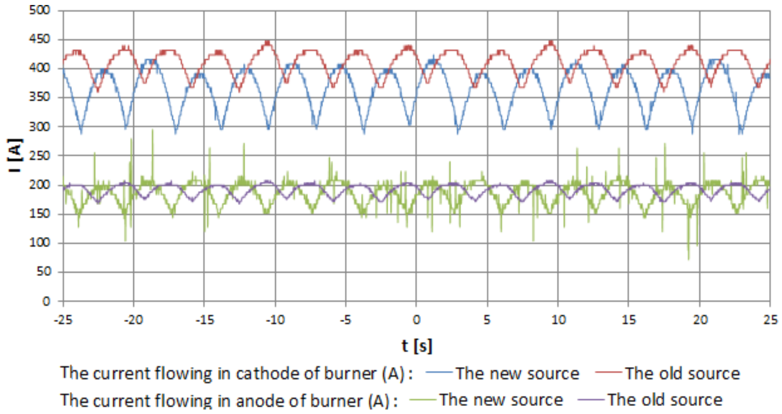


Fig. 8. Comparison of currents when using the old and the new power source

It can be seen from the graphs that the voltage of old source is less wavy than in a new voltage source, because older source has larger inductance of coils. A new source has the advantage of better regulation than older source. Better regulation is achieved by using thyristor control instead of controlled magnetic circuit transformer (transformer has primary, secondary and control windings [3]) in the old source.

Conclusion and Acknowledgment

The main contribution of the thesis is increasing the success of ignition of the electric arc in the work area and therefore reducing the excessive wear on the nozzle of plasma burner. Also operator of the furnace is satisfied with the current situation, because it is no longer need to difficult manually start the furnace and also it is no longer need to time-consuming exchange of damaged diffuser. Because of this the costs of operating the furnace were reduced as well.

This work was supported by the Ministry of Industry and Trade of the Czech Republic, MPO-CR, project No. FR-TI1/334.

The results of the thesis are published in conferences proceedings:

- ELNET 2010, Ostrava: VŠB – TU Ostrava, ISBN 978-80-248-2349-2,
- Ph.D. Workshop 2011, Ostrava: VŠB – TU Ostrava, ISBN 978-80-248-2378-2,
- EPE 2011, Ostrava: VŠB – TU Ostrava, ISBN 978-80-248-2393-5.

References

1. Dembovský, V. Plazmová metalurgie. 1. vydání. Praha: SNTL. 264 p. typové číslo L14-B3-V-41f/41744. (1978)
2. Hradílek, Z., a kol. Elektrotepelná zařízení. 1. vydání. IN-EL. 170 p. ISBN 80-902333-2-5. (1997)
3. Voldán, P. Silový napájecí systém plazmového generátoru rekrystalizační pece. 80 p. Master's thesis at VŠB-TUO. FEI. Leader of thesis is Martin Marek. (2008)

Power Supply of the Electric Traction System and Innovation of the Devices for Testing of the Contact Stress

Peter Staš[1], Martin Marek[1], Radim Halama[2], and Petr Stoklas[1]

[1]Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
peter.stas@vsb.cz
[2]Department of Mechanics of Materials, Faculty of Mechanical Engineering,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
radim.halama@vsb.cz

Abstract. Work deals with a proposal of eligible power supply of electric traction systems for steel torsion loading diagnostic equipment. With respect to equipment functionality demands it is necessary to project a conception, which allows setting up of wide spectrum of equipment parameters for different types of examined specimens. Project is realised in close cooperation of Department of mechanical engineering and Department of electrical engineering and informatics.

Keywords: Diagnostic equipment, regulation of rotation speed, frequency converter.

1 Introduction

Increasing demand for transport requires development of advanced technological products for the transport industry. Contemporary railway transport requires for this purpose a combination of carriages connected to a locomotive. Mechanical rolling parts exposed to torsion loading during the whole lifetime. This becomes a key issue of safeguard of reliability and safety of transport itself. To realize above described, it is necessary to effectively test the all mechanical parts of locomotive traction systems. Testing cannot be realized without superior steel torsion testing device which one of key parts is reliable and stable power supply and regulation unit. Task of the Department of electrical engineering and informatics is to develop optimal power supply and regulation unit for above mentioned diagnostic equipment.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 55–60.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Steel material failures in railway transport

Steel material failures in railway transport (taking into account only wheels defects) are consequence of various reasons. Extensive force acting on wheels causes deformations. Micro structural changes induced by cyclic strain causes changes of physical properties and deformation response of the material. Discontinuities and stress fractures are created. In praxis is this situation represented by combination of above described defects created e.g. during locomotives run up and slowing down. Quality and particularly purity of materials plays key role in lifetime and durability. These defects are more detailed described in theory material cyclic deformation.

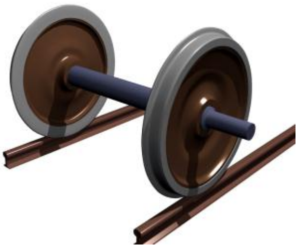


Fig. 1. Wheels used in railway carriages and its cross section [6]

2.1 General principle of railway wheels and lines diagnostic

Contact stress testing equipment device has significant place in examining of steel stress test and these test have a great importance for railway transport.

Great amount of mechanical parts is during the work and also during rest physically strained. This may cause changes in material's structure and its physical properties. In technical praxis this problem is represents contact of the wheel and the line surface.

With regard to above mentioned material failures are tests of these materials necessary especially with respect to safety of people and to reliability and durability of transport. For correct realization of tests exact setting up of parameters of diagnostic equipment is necessary, most important is set up of exact revolution speed and set up of dimension parameters of examined specimens.

Diagnostic by itself is performed using circle shaped specimens with the same or different diameter. Specimens surfaces touch each other and spin at the same or different speed (this depends on the test type).

One specimen represents the wheel, the second represents the line. Specimens are pushed together under specific pressure. As soon as specific number of cycles is achieved, one measures mass decay of specimens, diameter, hardness and abrasiveness changes. Detail of this issue is described in the literature [3] and [5].

2.2 Schematic drawing and principle of work of TUORS material strain testing device

Traction of both specimens is realised using asynchronous engine (2.2 kW). Construction of equipment limits the diameter of used specimen on 85 mm.

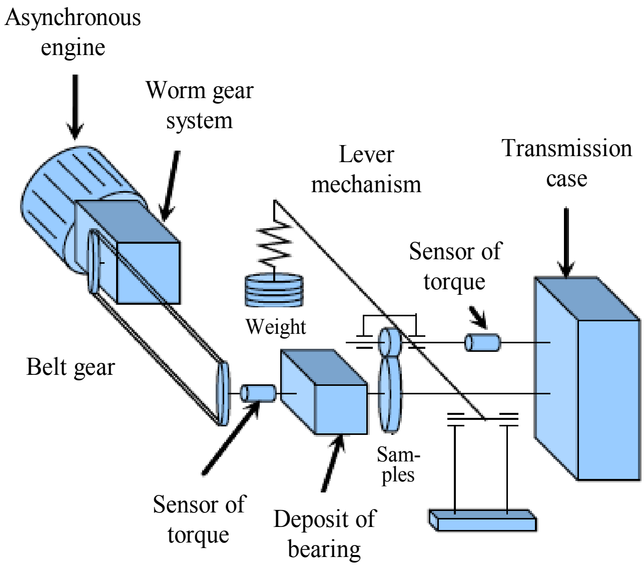


Fig. 2. Diagram of the testing machine TOURS [4]

Slip realization. On this equipment it is possible to perform tests with forced slipping using different specimens tip speed or using different wheel diameter or using gearing inserted between output spindle of specimen representing wheel and output spindle of specimen representing line.

Realization of pushing force. Radial force is created statically using combination of weight and lever. Force should not exceed 900 N in case of free slip, in case of forced slip it should be lower.

Innovation. In frame of cooperation of faculties is subject also further innovation of equipment, namely innovation of electric break, which allows more effective testing. Other parameters, which should be innovated are more effective realization of pushing force and realization of independent traction of both specimens. These aspects are not flexible at nowadays TOURS equipment.

2.3 New concept of testing device

New concept of diagnostic equipment is shown in Fig. 3. It contains independent left and right traction system which allows realizing required slipping tests.

Construction is flexible so it allows diagnosing specimens with different diameters and with various combination of slipping parameters.

Aggregate is not energetically demanding. One of traction engines will work in generator mode. This mode is known also as recuperation (recovery) mode.

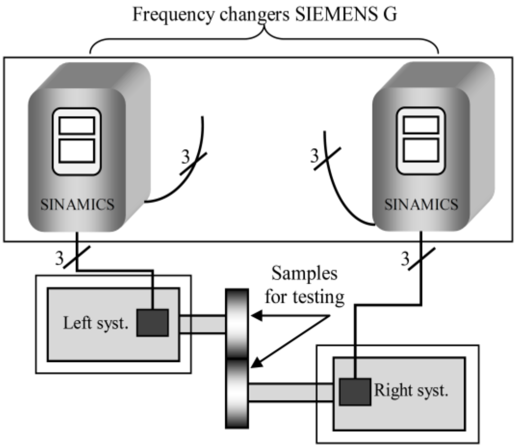


Fig. 3. New concept drive

Realization of slip. As most convenient solution a use of asynchronous electromotor has been chosen (further only AM). Realization of slip is not depended on the mechanical parts, namely on diameter of examined specimens or on gear ratio as in former case. Another advantage of this concept is higher quality of tests and wider spectrum of equipment's applications. On the other hand, disadvantages are higher costs of traction and control unit block systems.

Realization of pushing force. Pushing force between specimens will be realized only for one traction system what fulfils all requirements of tests. To achieve this, the traction unit will be pushed to second unit using hydraulic arm. This also allows real-time control and correction of pushing force value. Size of pushing force necessary to create maximal pressure (1200 MPa) in the place of contact and depends on both specimen's diameter. Detail of this issue is described in the literature [5].

Selection of optimal power supply and selection of eligible asynchronous engines. Power supply unit have to fulfil wide spectrum of requirements. One should to judge requirements like independent regulation of traction electromotors, adequate economical costs, and high level of functionality, long term stability and durability. The most optimal solution with regard to spatial disposability and construction of used motors is to use frequency regulation. Chosen frequency changers have modular concept. Energy recuperation is solved using eligible type of power supply unit which is capable to provide this function. Frequency changer Sinamics G120 is used as a final technical project solution.

Let's assume following parameters: diameter of specimen: d = 85 mm; height: h = 10 mm; modulus of elasticity: E = 21000 MPa; friction coefficient: f_t = 0,45; maximum revolution: n = 1500 rpm a contact pressure: p_o = 1200MPa. Taking into account these parameters, we will calculate power P = 30 kW and driving moment M_H = 160 N.m. The calculations of parameters are in the literature [1], [2] and [5].

Using these values, electromotors of type Siemens 1LG4 with power equal to 30 kW and driving moment equal to 196 Nm were selected.

3 Revolution speed regulation requirements

Requirements on speed regulation are specific for every type of exam. In general it is a question of determination of speed regulation by the frequency change for particular test (e.g. vector regulation) or it is a question of set up of concrete working frequency and current. Exact requirements are lay on behaviour of both traction units. In preference parameters like maximum work revolution, continuousness of start and start revolutions are taking into account. Programmable control units make use of equipment flexible and perspective for use in other types of tests in the future.

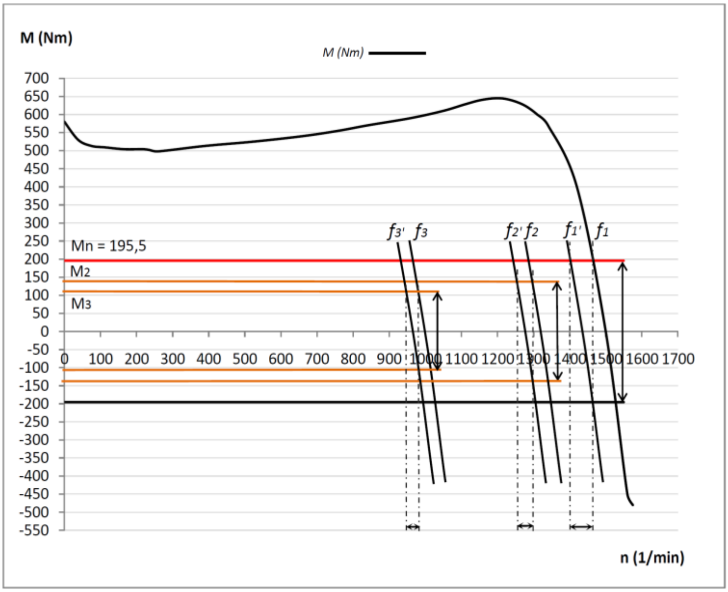


Fig. 4. Momentum characteristics of asynchronous electromotor [3]

Conclusion

This paper refers partial results from development of unique testing device for examining of contact strain of steel materials in transport.

Czech Republic belongs to leaders in area of construction of railway wheels (company Bonatrans with which we cooperate on development of testing device). Goal is to achieve world standards and to continue with further research.

The results of work are published in conference proceedings:

- ELNET 2010, Ostrava: VŠB – TU Ostrava, ISBN 978-80-248-2349-2
- EPE 2011, Ostrava: VŠB – TU Ostrava, ISBN 978-80-248-2393-5
- Ph.D. Workshop 2011, Ostrava: VŠB – TU Ostrava, ISBN 978-80-248-2378-2

Acknowledgements

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (ENET No. CZ.1.05/2.1.00/03.0069).

References

1. Koan-Son, B., Keiji, K., Tsunamitsu, N.: An experimental investigation of transient traction characteristics in rolling-sliding wheel/rail contacts under dry-wet conditions Wear, (2007)
2. Tyfour, W. R., Beynon, J. H., Kapoor, A.: The steady state wear behavior of pearlitic rail steel under dry rolling-sliding contact condition.Wear, (1995)
3. Staš, P.: Design of electric drive for testing system of steel materials, Master's thesis, VŠB – TU Ostrava (2010)
4. Halama, R.: Experimentální poznatky a fenologické modelování cyklické plasticity kovů, Habilitační práce, VŠB – TU Ostrava (2009)
5. Šofer, M.: Concept of Modification of Test Machine for Rolling Contact Fatigue Research, Master's thesis, VŠB – TU Ostrava (2009)
6. Figure [online]. [cit. 2011-08-04]. WWW: < http://www.starmans.net/en/magnetic-particle-testing-of-railway-wheels.html >

Disturbances in AC Traction System

Pavel Svoboda

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavel.svoboda1@vsb.cz

Abstract: Electric traction is in aspect of electricity consumption great consumer and also a significant source of interference effects that adversely affect the power grid. These effects are especially non-symetricity and higher harmonicss.
There is evaluation in this article of measuring alternating current traction system 25 kV, 50 Hz under normal operating conditions at Czech Railways.

Keywords: voltage asymmetry, electric traction, mensurement

1 Introduction

One of the densest rail networks is located in the Czech Republic. On electrified lines in the CR, two current systems are mainly used: 3 kV DC system (1705 km) and 25 kV AC system, 50Hz (1307 km). Electric traction is characterized by the great dynamism, which places high demands on the power system. The energy ratio with AC traction system of 25 kV and 50 Hz has these benefits:

- greater distance between the traction converter stations (approx. 50 km)
- traction substations are technologically simpler and cheaper in terms of construction
- active transformer losses are low.

AC traction, however, has its disadvantages, which include:

- Power Station is a necessary supply of 110 kV power system
- causes distortion sine wave current drawn
- compliance with the minimum power factor
- traction sampling variation in time leads to voltage fluctuation, which causes a "flicker-effect"
- high voltage power network must have a sufficiently high short-circuit power, does not give rise to unbalanced load asymmetry between phases illegal.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 61–66.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

In this article we will discuss about voltage asymmetry in the traction transformer Svitavy.

2 Asymmetrical classification

For classification of the asymmetry in power, the exclusively Fortescueova method is used. According to the aforementioned method, n-phase asymmetrical system can replace the (n-1) symmetrical systems, which creates a rotating magnetic field (positive and negative sequence) and one non-rotating, which creates a vibrant field. The system that creates a pulsating field is called zero as it occurs only in systems with grounded star point.
Asymmetry in three-phase systems classifies the coefficient of voltage and the current asymmetry factor:

$$\rho_u = \left| \frac{\underline{U}_{(2)}}{\underline{U}_{(1)}} \right| .100\,(\%),\ \ \rho_i = \left| \frac{\underline{I}_2}{\underline{I}_1} \right| \cdot 100\,(\%) \tag{1}$$

2.1 Evaluation asymmetry

Standard ČSN EN 50160 specifies us closer that under normal operating conditions in any one-week period, 95% of ten-minute high rms-sequence voltage in the range 0% to 2% will be direct components. In certain areas where customers are partially connected to single-phase or two phases, in the supply occur the points of asymmetry to 3%. The above-mentioned standard, however, applies only to the power system to 35 kV.
In the international law, it is not to the level of vvn set vacation, or planning level for backward components. As a compromise, it was therefore agreed with the PPS value of 1.5%, which represents a reserve for an increase in return for component toward its source, i.e. the traction transformers that ČD railway networks are connected to 110 kV distribution systems [1], [2].

3 Measurement of the traction transformer

Measurement of the traction transformer Svitavy conducted for 48 hours a week because the organizational measurement could not take place although the measured data are sufficient for understanding the relationships in the traction transformer.
For realization of measurements, the measurement system developes the enabled multi-channel measurement and data storage. The system consists of the following elements:

- laptop dual-core processor running Windows

- Measurement card with a bandwidth of 450 kHz, the maximum sampling rate of 250 kS / measurement on one channel and input ranges from ± 0.2 V to ± 10 V
- voltage sensors LEM LV25-P
- current sensors Clamp MN38
- virtual measuring environment LabWiev

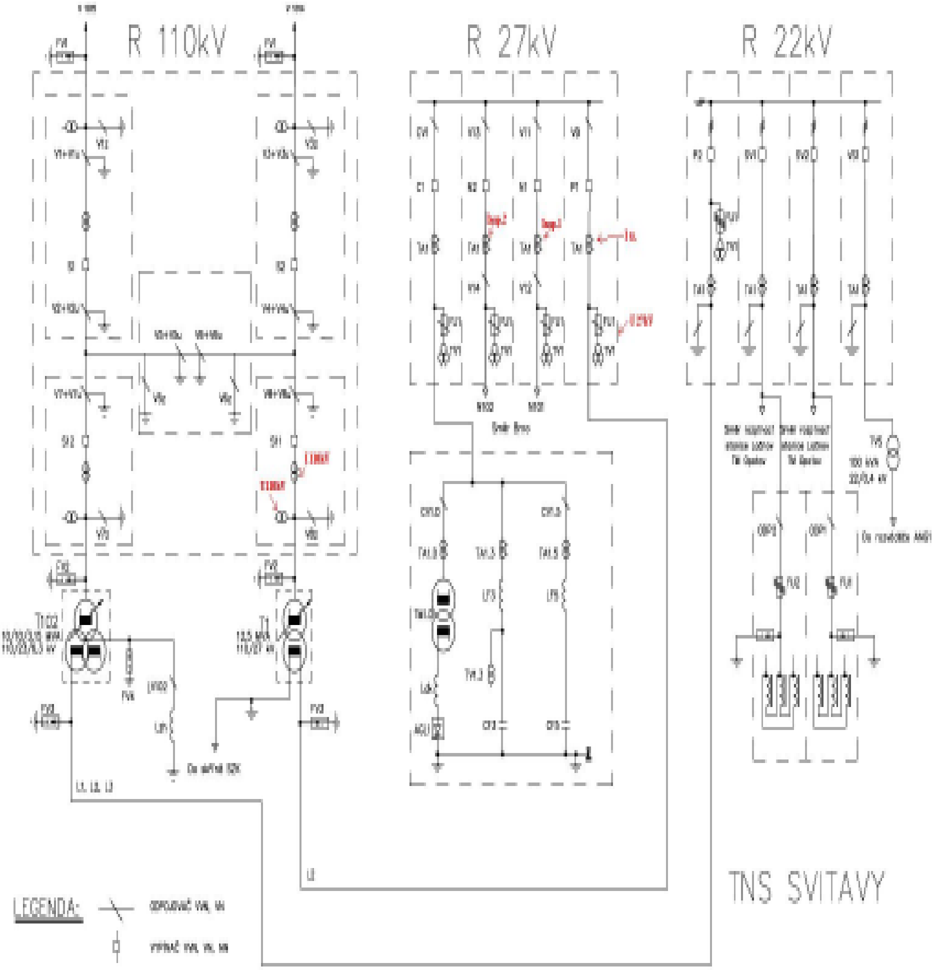


Fig.1. The connection of the measuring apparatus in TNS Svitavy

Fig.2 View of one of the measuring sites

3.1 Evaluation of measurement

When evaluating the measurements, they chose the method of moving averages 10 minutes. The method chosen is a compromise between immediate values and fixed 10 minute averages into a timeline. Moving averages are 10 minutes compared to a solid 10 minute averages, accurate, and after every 120 samples containing these fixed diameters, it satisfies the conditions PPDS.

Moving average forecasts of future values of variables is based on the average values of variables from the previous 120 values. From 120th line for any other moving average value, 120 previous values are always available. The first 119 values of the moving average is calculated by the same logic with the difference that is only available averaged number of values ranging from 1 to 119th [3].

The measurement data were stored every 5 seconds. Over a period of 10 minutes, 120 values were imposed. During the entire measurement was obtained 34 560 values for each observed variable.

In the picture 3, 10 minutes are plotted moving averages of the 110 kV line and a 10-minute moving averages of voltage asymmetry. This chart depicts the analogy between voltage and unbalance. The biggest differences between strains can be observed during the peak traffic on the monitored line. The picture 4 shows a direct correlation between the size and power drawn by the voltage asymmetry factor.

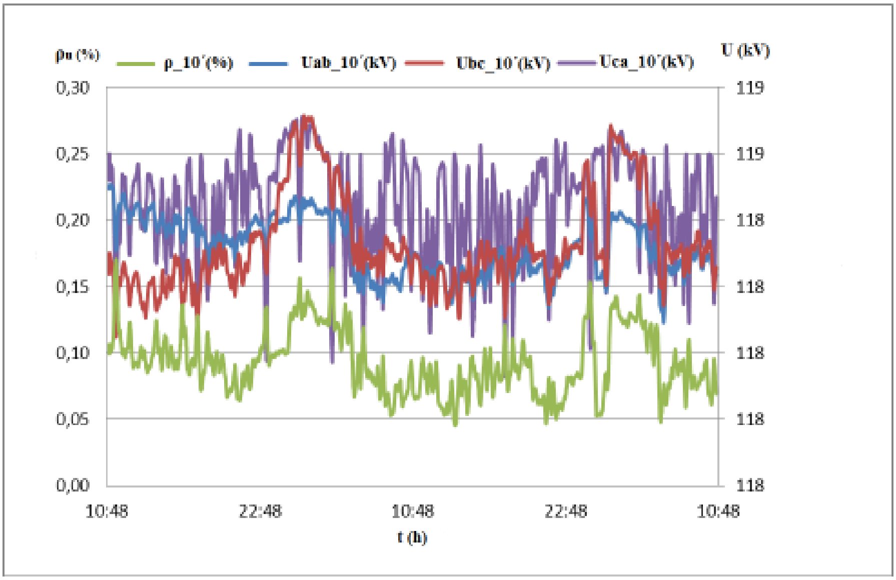


Fig.3. 10.minutové moving average line voltage and voltage asymmetry ρu

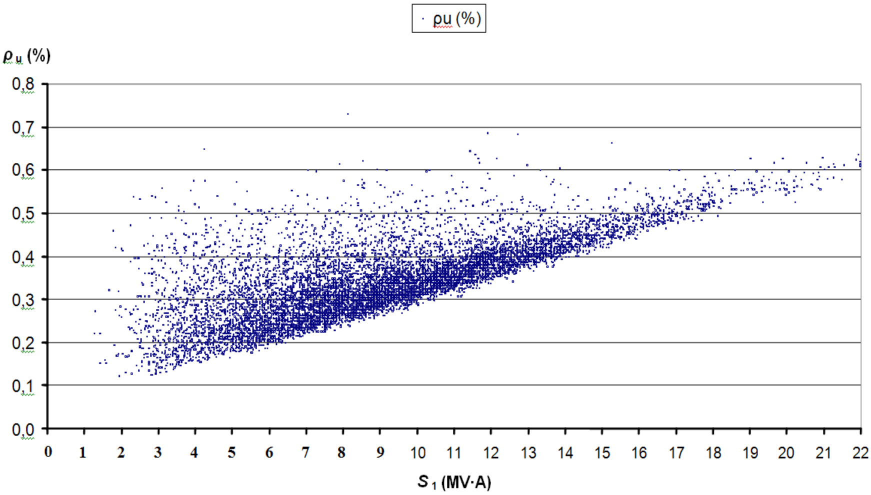


Fig.4 Dependence of voltage asymmetry factor ρu on the size of consumed output

4 Conclusion

After analyzing this measurement we can conclude that the chariots AC 25 kV, 50 Hz, their collection cause a voltage imbalance in the supply network. The measured values are within the range of measurement moved with sufficient margin below the limits defined in the PPDS. The final value of the asymmetry will

also affect the distribution of consumption between the two transformers and using equipment symmetry.
In the event of an increase traction, reasonable reserve lines and equipment can also take in, and values should not exceed a set limit.

References

[1] Norm ČSN EN 50 160 on the quality of electricity

[2] Rules for the operation of distribution systems, Annex 3

[3.] FUKALA, B.: Influence of electric power system on the traction , thesis FEI, VŠB-TU Ostrava 2010

Possibility of Using of New Technologies in the Street Lighting

Jaroslav Šnobl

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jaroslav.snobl@vsb.cz

Abstract. This article assesses the current ways how to reduce energy consumption in public lighting sector. One possible way is using lights with LED technology, which is still in early development, or to continue in using of ordinary sodium lamp fixtures and regulate it.

1 Introduction

This article was based on a study that was made for Ostravské komunikace, a.s. and the study is used as a base for deciding whether the current LED lighting technology is in technical and economic terms suitable for application to road. The biggest problem who hinder their deployment is the purchase price which is still substantially higher in comparison with purchase price of ordinary sodium lamp fixtures. Here are stated calculations of the individually tested lights and calculations of return and current prices estimate calculations as well.

2 Specification of LED lights available in 2011

For adequate comparison of current LED technologies on the Ostrava roads there had been specified values of parameters for the most commonly used high-pressure sodium lamps of 70 W. This type of lamp is used on more than a twenty thousand points of light. The main part of these points of light on the road is classified in class S4. Attention is drawn to the fact that more than 50 % of installed fixtures of 70 W exceeded its recommended useful life (approx. 12 years). The next most represented class roads in Ostrava are classes S5, ME2 and ME3. Other classes of roads are only minimally represented and for those marginal classes can be implemented technologies-used in the above-specified classes.

Below are average values based on the above specifications for fixtures available in 2011. Those fixtures are tested and produced by manufacturers in more then ten thousand series.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 67–70.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Table 1. Comparison of parameters of LED lamps and ordinary sodium lamps

Parameters	LED lamp	Ordinary lamp
power	30 to 60 W	83 W
efficiency	85 %	75 %
specific output	110 lm / W	95 lm / W
distance of pylon	40 meters	40 meters
life time of lamp / source	12 years / 12 years	12 years / 4 years
maintenance	cleaning once per 4 years	cleaning and exchange resource once per 4 years
maintenance factor	0,8	0,8
price	8000 CZK	4000 CZK
color rendering index	70 to 80	25
color temperature	4000 K	2200 K
regulation	0 to 100 %	0 and 50 to 100 %

3 Economic comparison of LED lights available on the market in 2011

3.1 Type road

As a basis for light-technical calculation model must be defined the appropriate road parameters. In Ostrava the most represented road class is S4 according to EN 13201-2 (Lighting of roads - Part 2:) and as well very significant is the number of points of light placed on the class road.

Table 2. Parameters of a model road S4

Parameters	Specification
width of road	7 meters
lanes	2
lighting system	one-side
high of light location	8 meters
fixtures offset from the road	0,5 meters
required maintained illuminance	5 lx
minimum required illumination	1 lx
distance of pylon	30 meters
maintenance factor	0,8

3.2 The individual calculated parameters of compared fixtures

The following table shows the light-technical calculations LED lights from four reputable manufacturers that are already on the market. The last column gives the values of the sodium lamp lights which are mostly on the roads of class S4.

Table 3. Lighting and technical calculations of selected LED lamps and ordinary sodium lamp

Parameters	LED lamp 1	LED lamp 2	LED lamp 3	LED lamp 4	Ordinary lamp
average	5,1 lx	5,9 lx	5,6 lx	5,3 lx	7 lx
min	2,7 lx	2,6 lx	1,9 lx	2,2 lx	2,2 lx
min / average	0,53	0,44	0,34	0,42	0,32
power	45 W	54 W	50 W	63 W	83 W
light flux	4530 lm	5400 lm	5040 lm	5860 lm	6600 lm
color temperature	6000 K	4000 K	3500 K	3500 K	2200 K

3.3 The economic balance sheet and return

Table 4 shows the comparison of mutual investment and operating costs of LED lamp and ordinary sodium lamp; it is clear from values stated below that after 12 years (the lamp useful life) that the total cost are nearly identical. The biggest difference is the initial investment, purchase price of LED lamp is doubled in comparison to ordinary sodium lamp. However, this cost is compensated by the low power consumption of LED lamp (power 50W) in comparision to ordinary sodium lamp (power 83W). Gradually, the initial investment in LED lamp is compensated.

Table 4. The summary of the total cost of LED lamp and ordinary sodium lamp and their comparison

Parameters	LED lamp	Ordinary lamp	Difference
average lighting power (kW)	0,050	0,083	0,033
annual operating time (hours)	4150	4150	-
annual consumption of lighting (kW)	207,5	344,45	136,95
the current price of power. energy per kWh (CZK)	2,01	2,01	-
price for annual consumption of electric energy (CZK)	417,5	693,0	275,5
average price fixtures (CZK)	8 000	4 000	-4000
lights maintenance for 12 years (CZK	900	1 566	666
the total cost for the lifetime of lamps (CZK)	8 900	5 566	-3334
the cost of lighting for 12 years (CZK)	**13 910**	**13 882**	**-27**

4 Conclusion

As the conclusion, the market available LED lamps are able to compete with ordinary sodium lamp fixtures, which are still majority-used. The calculations for LED lamps selected from reputable companies and sufficiently tested in practice to fully meet the requirements of EN 13201-2 for class S4. The content of originally prepared study is also assessing the possibility to use of LED lamps on different road classes. As to road classes with lower illuminance requirements than the S4 the LED lamps can compete in total cost parameter to ordinary sodium lamps. It should be noted that the price of LED lights is related to a larger volume of lamps purchased (the price of individual items can vary a lot). For classes with higher road requirements for illuminance than the S4 class, it means especially ME, LED lights are still in development and especially the cost of such lamps is significantly higher and return in 12 years is not possible. The current LED lamps can therefore be recommended for road class S4 and lower.

References

1. ČSN EN 13201-2 - Osvětlení pozemních komunikací – Část 2: Požadavky
2. Sokanský, K., Novák, T., Šnobl, J.: Využití nových technologií ve veřejném osvětlení města Ostrava, study for Ostravské komunikace a.s., Ostrava 2011
3. Sokanský, K., Novák, T., Šnobl, J.: Hodnocení VO z pohledu výkonnostních parametrů, The conference proceedings Kurz osvětlovací techniky XXVIII, Dlouhé stráně 2010
4. CIE Pub.115 Berlin Editorial Version II (draft) (2007), Recommendations for the Lighting of Roads for Motor and Pedestrian Traffic

Heat Pumps, Improvement of Utilisability, Differences in the Calculation of Heat Loss

Jan Šrámek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.sramek.st2@vsb.cz

Abstract. This article strives towards and summarizes issues of new standard for calculation of heat loss. There are shown results by old and new standard for calculation heat loss and heat demand. Next part describes existing trends and options for accumulation of the heat. It provides explanations for reasons and advantages of accumulation not only for heat pumps, but also describes deals with concepts of heat pumps of smart grid area type and utilisation of accumulation for improvement of effectiveness in running of air/water heat pump.

Key words: heat pump, heating, heat loss, heat accumulation

1 Differences in the calculation of heat loss

For calculation of heat loss was used standard CSN 06 0210 calculation of heat loss for central heating. This one was replaced by Czech version of European standard EN 12831 heating systems in buildings – Method for calculation of the design heat load.

1.1 Comparing calculation of both methods

The comparison of both methods was made at chosen house, which is situated in town Bruntál. Total build up area of this house is 239m^2 and external designed temperature is -18 °C. This house will be also used for research work in the future.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 71–76.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

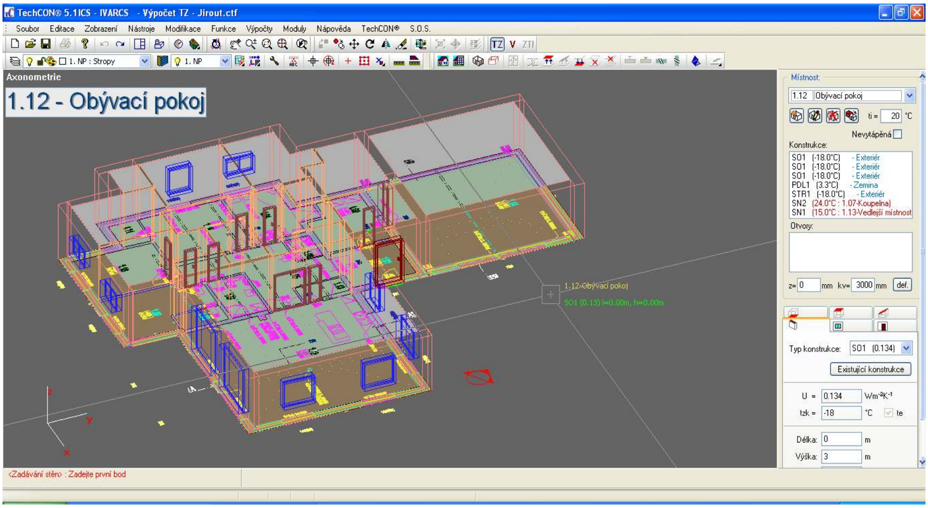


Fig. 1. 3D image of the chosen house showed in software Techcon IVAR CS.

1.1.1 Result of standard CSN 06 0210

Result of old Czech standard norm for chosen house is 5733 kW.

1.1.2 Result of standard and latest news in CSN EN 12831

The major news in European standard EN 12 831 is definition of heat bridge. Heat bridge can be described as place of crossing two different constructions e.g. isolation and brick wall.

There are two types of results. First result includes heat bridges 7159 kW and second result 6185 kW without them.

1.2 Summary of standards

Next figure shows obtained result by calculation software for heat loss named Techcon IVAR CS.

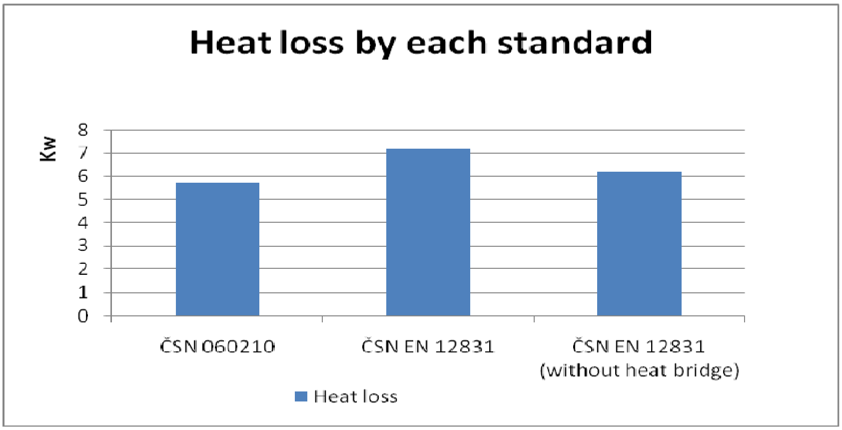


Fig. 2. Calculation heat loss results of chosen house.

1.3 Comparison of thermal demand

We have to design heating very precisely as is shown on table 1, because differences in result of heat loss and also thermal demands are very large, nearly 20%.

Table 1. Differences of thermal needs of the chosen house by standards listed above.

	ČSN 060210	ČSN EN 12831	ČSN EN 12831 (without heat bridge)
Thermal demands (GJ/year)	39,53	49,36	42,64

2 Heat pumps, accumulation of the heat and improvement of efficiency

In correct layouts, the heat pumps possess fairly stable parameters throughout the entire heating season. There is an exception represented by the air/air heat pumps and especially the air/water type pump. Their output is very variable depending on the ambient air temperature and its humidity.

2.1 Heat pump vs. electric boiler

Investment returns to heat pump in comparison with electric hot water boilers is about 13 years. There is no consideration with electricity price change.

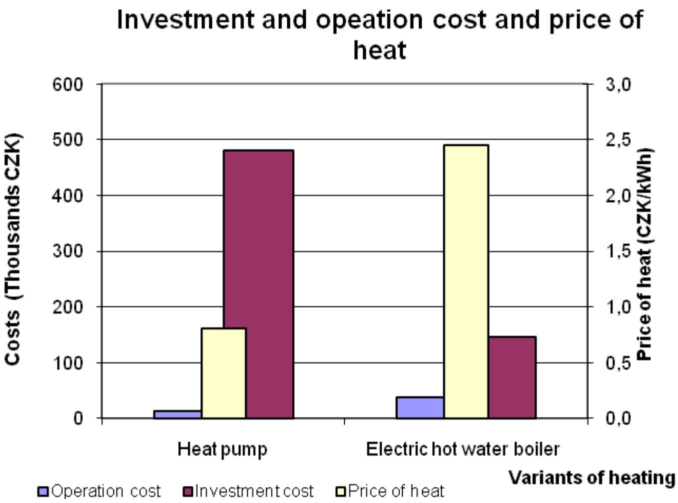


Fig. 3. Investment and operation cost of comparing systems.

Calculation of operation cost is very accurate as shows table 2. Heat pump with nominal heat power 6 kW supplied heat and domestic hot water for house during whole year and it was not necessary to use an additional source of heat.

Table 2. Differences of electricity consumption by heat pumps in a real and calculated operation.

Consumption of real operation (kWh/per year)	6013
Consumption of calculated operation (kWh/per year)	5931

Data from real operation of heat pump were obtained from house owner.

2.2 Heat accumulation

Accumulation of heat energy obtained either from the Sun or even a boiler can be conducted by either of the two essential processes described below:

- Into structures
- Into accumulation media (mostly water)

Accumulation into structures is of great significance especially in case of houses built in accordance with low-energy.

Accumulation of heat into water finds massive range of applications at heat pumps, biomass boilers with output control, solar collectors and other sources for the following reasons:

- Transmission of generated heat over time for the maximum utilization
- Enhancement of efficiency and life of the heat source

Transmission of heat over time enables utilization of irregular gains (e.g. from solar sources) with the greatest efficiency possible over time, exactly in periods associated with the need for heating. This accumulation can be defined as short-term (several days, but the accumulation of heat gains can be also accumulated in a long-term, especially during the summer season, to be utilized during the heating season

2.3 Improvement of heat pump efficiency by means of accumulation

Accumulation with heat pumps provides for continuous operation, i.e. reduction of the frequent switching. However, as far as the air/water is concerned, its greatest disadvantage (decreasing output depending on the ambient temperature) can be utilized upon implementation of suitable accumulation features. The Fig. 4 shows the performance diagram of the air/water heat pump and substantial decrease in output after the 0°C mark. The purpose of this idea deals with suitable design of heat accumulation system to operate the heat pump within such area, where the same amount of electric power yields more heat energy to be accumulated for further use during frosty days.

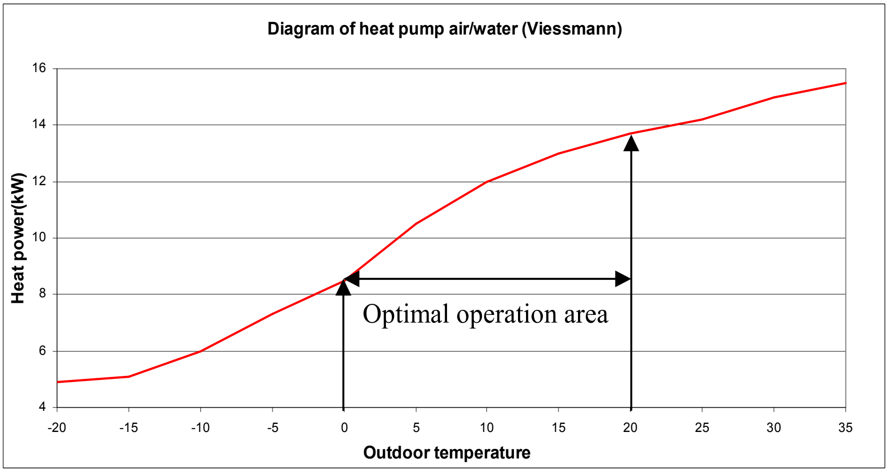


Fig. 4. Output diagram of the air/water heat pump with the nominal output of 9kW at A2/W35.

Accumulation of heat allows improvement efficiency and also quickest return of investment cost into heat pumps in comparison with others methods of heating. Next step of this research will determine sufficient capacity of accumulation tank and possibility of the prediction of outside temperature for optimal operation of air/water heat pumps.

Conclusion

Result by different standard for calculation of heat loss gives us different numbers. Therefore it is necessary precise design and calculation heat demand by as much as possible relevant data.

To enhance the competitive strength of heat pumps in comparison with other heating sources, it is necessary to seek method towards improvement of their effectiveness.

Such opportunity is granted by the methods of accumulation, which can reduce the variable availability of low-potential energy contained in the air.
Other parameters to be considered include the method of output of the air/water heat pump to enable its optimal operation as depicted in the Fig. 4 to produce sufficient energy for further days. The ideal layout will require correct selection of the accumulation tank. However, all these steps need to be taken with reference to the financial exigency in order to ensure that the effect required, i.e. reduction of the return on investment, is not the contrary.

Publications

ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. Accumulation of Heat Energy from Heat Pump. In. Proceedings of the 12th International Scientific Conference EPE 2011 May 17-19, 2011 Kouty nad Desnou, Czech Republic. ISBN: 978-80-248-2393-5

References

[1] HRADÍLEK, Zdeněk. Energetika průmyslových a distribučních sítí. 1. vyd. 2008. ISBN 987-80-7225-291-6.
[2] ČSN EN 12831 Heating systems in buildings – Method for calculation of the design heat load
[3] ČSN 06 0210 Calculation of heat losses in building with central heating
[4] ŠRÁMEK, Jan. Návrh vytápění rodinného domu tepelným čerpadlem. Ostrava 2010, Diplomová práce
[5] ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. Accumulation of Heat Energy from Heat Pump. In. Proceedings of the 12th International Scientific Conference EPE 2011 May 17-19, 2011 Kouty nad Desnou, Czech Republic. ISBN: 978-80-248-2393-5

Evaluation of Cooperation WPP and PV Connected Through the Shared Transformer to the Network 22kV

Tomáš Šumbera and František Střída

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
tomas.sumbera@vsb.cz

Abstract. The article deals with the evaluation of the possibilities of cooperation of wind and solar power plant connected through a shared transformer to the network 22kV. Evaluation of the cooperation is based on measured data for wind power plant with installed capacity of 2 MW and solar power plant with installed capacity of 1.1 MWp. In other parts of the article is an analysis of variations in active power of resources in comparison with the system of WPP + PV In the last part is evaluating the effects after connecting of WPP + PV to the network 22 kV.

Keywords: wind power plant, photovoltaic power plant, utilization coefficient, fluctuations of active power

1 Introduction

Wind and solar power stations have a great advantage against ordinary sources of electricity, because energy production requires no fuel and don't produce hazardous waste and fouling, especially CO_2. Conversion efficiency of wind energy and solar radiation into electrical energy is however very small. Effectiveness of wind power stations vary from round about 50% and solar power plants from 15% to 20%, depending on the type of photovoltaic panel. Significant disadvantage of wind and solar power stations is unstable and stochastic supply of electricity, which is influenced by a large number of factors such as seasons, day-time, wind speed, solar Significant disadvantage of wind and solar power stations is unstable and stochastic supply of electricity, which is influenced by a large number of factors such as seasons, day-time, wind speed, solar radiation intensity, etc.

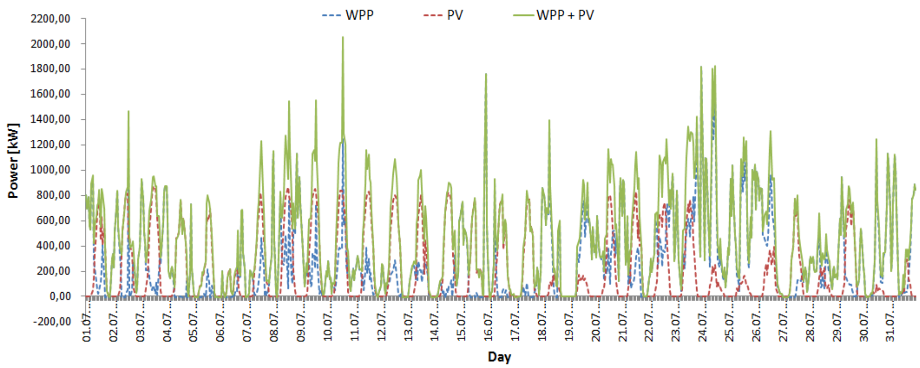
In the article is reviewed cooperation of wind and solar power stations, connected via a common transformer to 22kV network. Evaluation is done in July and October. The data are obtained from measurements of wind power installed capacity of 2 MW and solar power installed capacity of 1.1 MWp installed in the Moravian-Silesian region.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 77–82.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Evaluation of cooperation WPP + PV

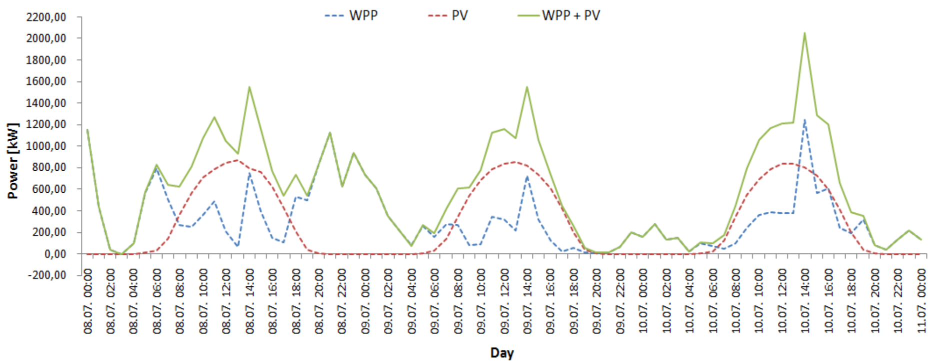
Supplies of electrical energy from wind power plants aren't dependent on day-time, as is the case with solar power, but only on wind speed. In cooperation WPP + PV, wind power plants are able to deliver electricity at night, while the solar power plant is out of operation.

In graph 1 there are presented courses of instantaneous power at hourly intervals for the month of July. Utilization coefficient of wind power station was in July at average 14% and utilization coefficient of solar power station was at average 18%.



Graph. 1. The course of instantaneous power of WPP, PV and WPP + PV - July

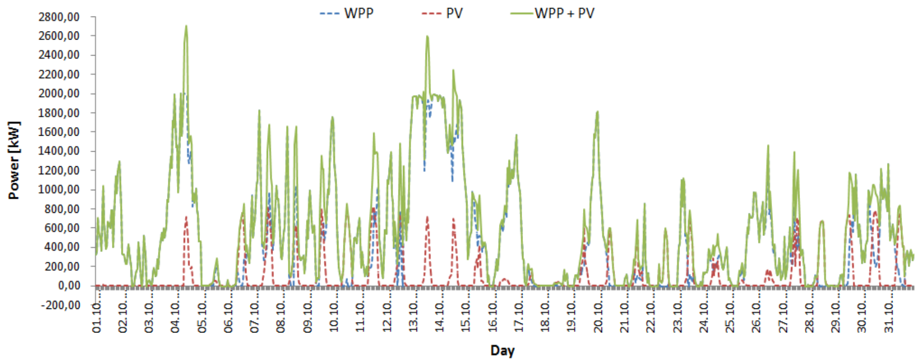
Graph 2 shows the courses of instantaneous power in selected days. From the graph we can see that 8.7., at 19:00, solar power plant stopped to supply electricity to the grid and the delivery was covered only from wind power station. On the 10.7., between 13:00 and 14:00, there came to accentuated growth of supplied output, on wind power station, from value 1215 kW to 2050 kW.



Graph. 2. The course of instantaneous power of WPP, PV and WPP + PV – from 8.7. to 10.7.

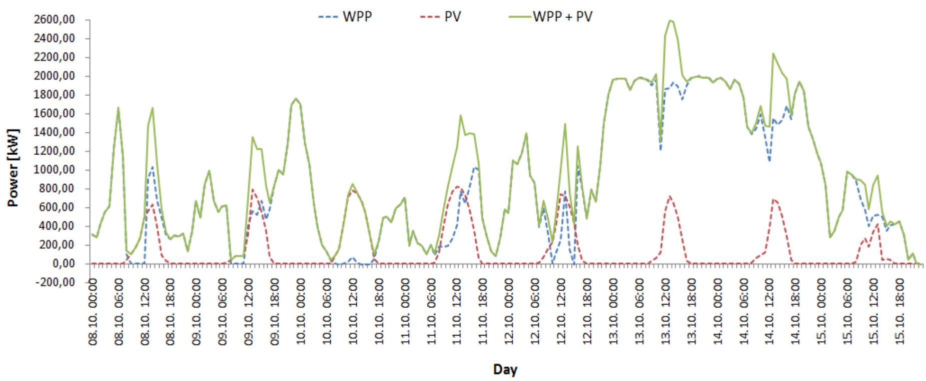
Utilization coefficient of WPP + PV would range around 15% in July. The system would supply 355 MWh, 209 MWh from wind power station and 145 MWh from solar power station.

In graph 3 there are presented courses of instantaneous power in October. Utilization coefficient of solar power station in this month was average 9% and the coefficient of wind power station in this month was significantly higher than in July and its average value was 25%. From the waveform is evident that the electricity supply was mainly from wind power station.



Graph. 3. The course of instantaneous power of WPP, PV and WPP + PV - October

Graph 4 presents waveforms of instantaneous power at selected days. From the waveform is evident that the output supplies from solar power plants are about 4 hours shorter than in July.

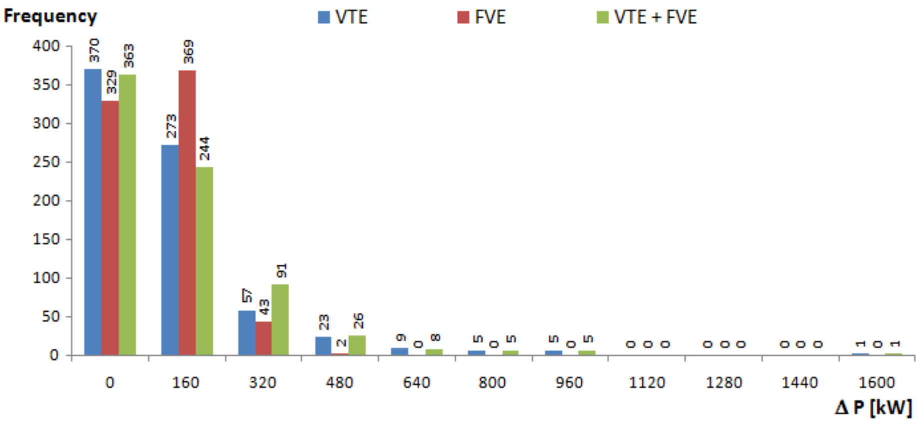


Graph. 4. The course of instantaneous power of WPP, PV and WPP + PV – from 8.10. to 15.10.

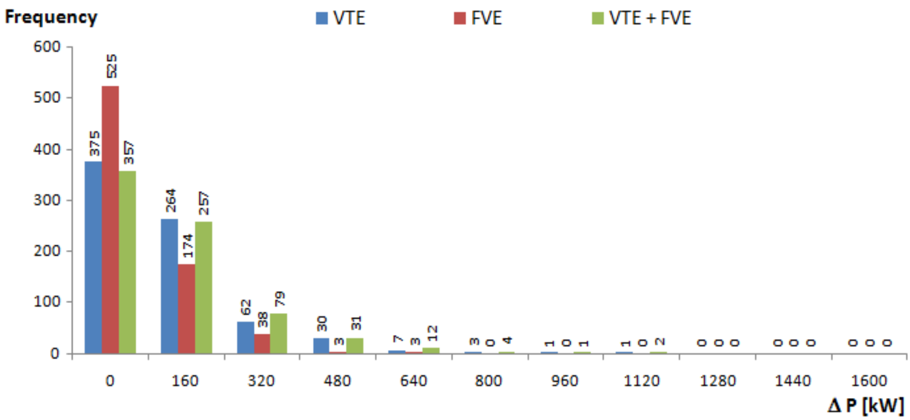
Value of WPP + PV utilization coefficient would be around 19% in October. The system would supply 440 MWh, 366 MWh from wind power station and 73 MWh from solar power station.

3 Evaluation performance fluctuations

Variation of active power in wind power station is caused by wind speed change. Wind power station works at intervals of wind speed from 3 m/s to 25 m/s. Wind speed is dependent on terrain segmentation (meadows, forests, cities). Variation of active power in solar power station is caused by change intensity of solar radiation. The intensity of solar radiation varies during the day and during the seasons. An important factor that affects the power of solar power station, is rotation of solar cells, the influence of cloudiness, panels pollution by dust or shading caused by snow.
Difference frequency of active power between individual clocks is shown on the graph 5 and graph 6. Fluctuation size of active power is most frequently below 480 kW, during July and below 320 kW in October.



Graph. 5. Histogram ΔP - July



Graph. 6. Histogram ΔP - October

The graphs show that power variation is lower than compared with individual sources. It is important to aware that installed power of system would be 3.1 MW, 2 MW from WPP and 1.1 MWp from PV.

4 Influence of connection system to the 22kV network

Connection of sources was evaluated on a real network in the Moravian-Silesian region. There were chosen maximum voltage peaks for evaluation of influence on the network and for ratios in switchgear. In table 1 is set percentage change of voltage after connecting to the PV, WPP or WPP + PV system. At wind power plant was calculated with cos φ = 0.98 inductive and at solar power station was cos φ = 1. In July, there was exceeded the value of allowed voltage change ΔU= ±2 % at the connection point, which is in contradiction of the operation rules of distribution system.

Tab. 1.

	dU [%] – 10th of July			**dU [%] – 4th of October**		
	PV	**WPP**	**WPP + PV**	**PV**	**WPP**	**WPP + PV**
Connection point	0.50	1.78	2.25	0.57	1.12	1.67
Bus bar 110 kV	0.09	0.46	0.54	0.10	0.30	0.38
Bus bar 22 kV	0.04	0.17	0.21	0.05	0.11	0.16

5 Conclusion

WPP + PV system, so-called the hybrid system, would be suitable to operate in conjunction with appropriate accumulation. Change of power fluctuations is not so frequent and expressive in this system, in regard to installed power, than it is known at separately workung sources. The big advantage of this system is the supply possibility of electricity at night hours, in which is the solar power plant out of operation.

WPP + PV system during operation violates operation rules of distribution system due to increase of tension in the connection point where it is allowed voltage change ΔU= ±2 %. In the case that acumulation would be connected to the system, first part of energy could be applied to battery charging and the other part would cover power supplies to the network.

Acknowledgment:

This contribution was prepared in the framework of research project MSM6198910007 and SP2011/130.

Reference

1. Hradílek Z; Šumbera T. Reliability of electrical power supplies from renewable sources. Electric power engineering 2010. Brno 2010.
2. Prokop L.; Hradílek Z.; Šumbera T. Application of WPP Power Flows Measurement Methodology. Prognozawanie w elektroenergetyce PE 2010. Polsko, Wisla 2010.
3. Hradílek Z.; Šumbera T. Simulator of power forecasting gained from wind power plants. Prognozawanie w elektroenergetyce PE 2010. Polsko, Wisla 2010.
4. Mišák S.; Šumbera T. Folvarčný A.; Holčáková R. Assessment usage energetic potential from renewable sources. Prognozawanie w elektroenergetyce PE 2010. Polsko, Wisla 2010.
5. Hradílek Z; Šumbera T. Reliability of electricity supply from renewable sources and possibilities predicting power. Konference WOFEX. VŠB-TU Ostrava, 2010.
6. Rusek S.; Střída F.; Šumbera T. The Issue of Voltage Changes in Electricity Network in Relation to Operation of Wind Power Plants. Konference ELNET, Ostrava 2010.
7. Hradílek Z; Šumbera T. Reliability and Predictions of Power Supplied by Wind Power Plants. International Conference on Renewable Energies and Power Quality 2011. Las Palmas de Gran Canaria, Španělsko.
8. Hradílek Z; Šumbera T. Simulator for Prediction of Energy Obtained from Wind Power Plants. Electric power engineering 2011. Kouty nad Desnou 2011.

Determination of Daylight Illuminance under Uniformly Overcast Sky in order to Reduce Energy Consumption in Buildings

Jan Šumpich, Zbyněk Carbol, and Karel Sokanský

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jan.sumpich, zbynek.carbol, karel.sokansky}@vsb.cz

Abstract. This paper describes modeling of possible savings in lighting systems of artificial lighting, working in combination with daylight. The proposed model works with overcast sky daylight (CIE). Based on knowledge of geographical position, day and time, the calculations are done in the conditions of unshadowed outdoor illuminance levels. From this figure (when daylight factor is known) illuminance caused by daylight can be gain in the differents parts of the room. By controlling the artificial lighting (dimming) to constant cumulative value (daylight plus artificial light) in the areas of visual tasks, then it is possible to determine potential of savings in areas with controlled lighting system and without it. According to this model (calculation), we can consider a suitability of the dimmable lighting systems in the specific work areas and make recommendations to reduce energy requirements of buildings in energy audits.

1 Introduction

The current economic requirements lead to the reduction of energy consumption in the buildings. In the field of interior lighting these requirements lead to greater usage of daylight in combination with artificial light. The technical level of the current lighting systems enables the regulation of the flux in the lighting control systems at a constant level of illumination due to artificial lighting dimming systems. This can put into practice a reduction of electricity consumption.

1.1 Daylight factor

Contribution of daylight in buildings in areas of visual tasks, is the share of internal illuminance (direct and reflected light) and the illuminance of an outdoor unshad-

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 83–88.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

owed flatland in condition of uniformly overcast sky. The contributions of direct sunlight are excluded from the both illuminances. The value of daylight factor includes the glazing effects, pollution, indoor and outdoor shading, etc. On the contrary, due to using uniformly overcast sky is eliminated the influence of windows position in various points of the compass. Daylight factor D is expressed as a percentage and is calculated according to equation:

$$D = \frac{E}{E_V}.100[\%]$$

E - is the illuminance of the point of internal comparison flatland in lx;
E_V - a comparison illuminance in the point of outdoor unshadowed flatland in lx.[1]

1.2 Uniformly overcast sky

Uniformly overcast sky is a condition, that in reality occurs only a few times a year. However, this situation enables to create a model of the sky which we can work with. This model sets clearly how sky luminances behave from horizon to zenith. Our model uses uniformly overcast sky on a dark area considering the sky with brighteness gradation from horizon to zenith 1:3 (according to CIE). Sky brightness of any point is then given by:

$$L_\gamma = L_Z / 3.(1 + 2.\sin\gamma)\ [\mathrm{cd.m - 2}]$$

$L\gamma$ - the brightness of the sky at an angle above the horizon
γ
L_Z - the brightness of the sky at the zenith [1]

2. Analytical calculations of Sun coordinates

The calculation of the apparent position of the sun in the sky is determined by the geographic location of the specific site on the planet, Earth's rotation and mutual position of the sun and the Earth in space. For calculations of the daylight and insolation of buildings there is a sufficient substitution the irregular shape of globe. A position of any point on Earth's surface is usually determined by latitude φ and longitude λz, see Figure 1.

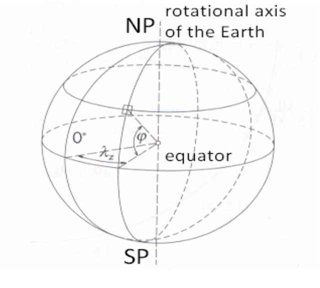


Fig.1 - Coordinates of the point on the Earth's surface. [1]

During the motion of Earth around the sun in slightly elliptical orbit the Earth's equator plane is inclined to the ecliptic plane tipping by 23.45 · See Figure 2.

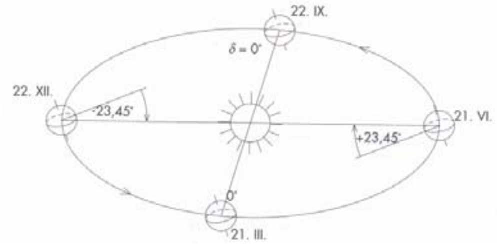


Fig. 2 - Declination for four typical days in a year. [1]

2.1 Solar Declination

For our purposes the declination is determined by mathematical expressions. We consider that the zero declination occurs at midnight before the day of spring equinox on the 21st March. We consider the equation.

$$\delta = 23{,}45° \sin\left[\frac{360°}{365.(J-81)}\right] \text{[rad]}$$

J - is the ordinal day of the year [2]

2.2 Time and the hour angle of the sun

Central European Time CET is the same as the central solar time on λZ = 15 ° east longitude, and as a zone watch time is used in the range of longitudes 7.5 ° ≤ λZ ≤ 22.5 °. The CET determines the genuine solar time according to the relation (3).

$$GST = CET + \frac{\lambda_Z - 15°}{15°} + ET \text{ [hrs]}$$

λ_Z -the longitude in degrees
GST -genuine solar time in hours
CET -Central European Time in hours
ET - the time difference between GST - CET in hours [1,2]

2.3 Height of the sun

Height of hte sun γS is the angle formed by the solar ray with a horizontal plane. To calculate the height of the sun is used the known relation:

$$\gamma_S = \arcsin\left[\sin\varphi.\sin\delta - \cos\varphi.\cos\gamma.\cos(15°GST)\right] \text{ [°][1,2]}$$

3. Calculations of the daily illuminance

Based on the knowledge of the position of the sun and the brightness distribution of the uniformly overcast sky it can be proceeded to deal with specific illuminance levels of comparison outdoor unshadowed flatland in specific locations and specific days and hours.

3.1 Horizontal illuminance

Earth's surface.Extraterestrial horizontal illuminance E_V will be different every day. Therefore, the luminous solar constant EV0 for the equinox must be corrected for the daily distances from the sun to the Earth's using agents of excentricity ε.

$$\varepsilon = 1 + 0{,}034\cos\left[\frac{360°}{365}.(J - 2°)\right] \text{ [-]}$$

$$E_V = E_{V0}.\varepsilon.\sin\gamma_S \text{ [lx]}$$

E_{V0} - light solar constant of 133 334 lux[1,2]

3.2 Diffuse illuminance

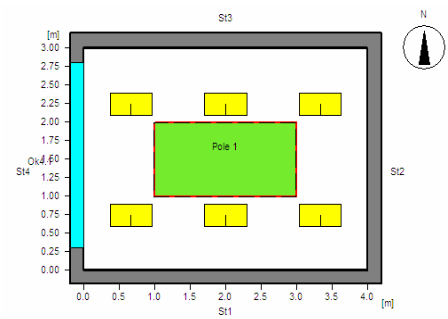
Diffuse illuminance Dv, ie illuminance of outdoor unshadowed horizontal plane under uniformly overcast sky with knowledge of specific geographical and temporal coordinates can be determined based on the following relation.

$$D_V = \left(\frac{D_{Vm}}{E_V}\right).E_{V0}.\varepsilon.\sin\gamma_S \text{ [lx]}$$

D_{Vm} / E_V-permeability coefficient of the sky light [1]

4. Calculation of savings of the dimmable lighting system in cooperation with daylight

Based on the knowledge of the diffuse illumination of the unshadowed outdoor flatland and the daylight factor of specific point of the visual task, we can calculate the internal illuminance at every moment by the relation (1) . As an example of the calculation there was chosen one of the offices in the VSB-TU Ostrava.



Average daylight factor	Dav=4.9
Minimal daylight factor	Dmin=2.1
Maximal daylight factor	Dmax=10.2

Fig. 3 - Values of daylight factor in the comparion plane in office at VŠB-TUO

As a requirement to calculate the potential savings for this particular case was taken a minimum illuminance 500 lux at any point of the visual task. Illuminance caused by the daylight is the value of illuminance, which can dim the installed artificial lighting system, when the minimum illuminance does not fall below the required value in place of the visual task.

The office is illuminated by a 6-lighting pieces (3 rows of two pieces from windows) mounted linear fluorescent T5 - 3x14W. Total power of the office lighting system (including mature ballasts) is 277W. In this case we considered the value 95% of lights power at a luminous flux (100%). The usage of the room is every day from 7 am. to 3 pm.

There was chosen the worst situation for non-dimmable lighting system. It means that all the lamps were connected to the only one electrical circuit and to achieve the 500 lux illuminance at every point of visual task the lighting system must be switched on throughout the whole time of office usage.

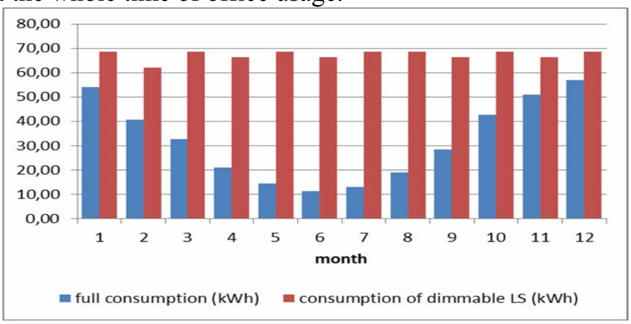


Fig. 4 – Consumption comparison of the standard and dimmable LS

Power consumption of lighting systems (see Fig. 4) and the resulting savings (see Fig. 5) can always be modeled only for specific situations when in the mode space the usage and the required illuminance are set. Then savings can be modeled for different types of lighting systems with different levels of their operation system from non-dimmable lighting system switched at once, to the lighting system, which is able to be controlled through each lamp of the lighting system according to the share of the extra daylight with using sensors for the presence of the people and the overall level of programming choices lighting according to specific visual activities..

Fig. 5 - Savings which are achievable by the controlled lighting system

According to the above mentioned specific example in the model office would be saved 424 kWh of electricity in one year, which is about 50% of the power of the original system.

It can be expected that in the real life will be achieved even greater savings, because the possibility of the sky modeling and the elimination of the influence of the compass can be calculated only in conditions of an uniformly overcast sky, which absorbs a substantial portion of light energy, which would fall to comparison unshadowed outdoor flatland. [3,4,5]

5. Conclusion

The objective of the authors of the article is making SW to determine the potential savings in the combined artificial and daylight lighting systems. Its expected usage will be in the design and renovation of the lighting systems, energy audits of buildings and in recommendations of reduction of their energy consumption.

References

1. 1. RYBÁR, P. a kol.: Denní osvětlení a oslunění budov, ERA 2001.ISBN 80-86517-33-0

2. DARULA, S. a kol.: Osvětlování světlovody, Grada Publishing 2009, ISBN 978-80-247-2459-1

3. Norma ČSN 73 0580-1 Denní osvětlení budov,

4. Norma ČSN 73 0580-2 Denní osvětlení obytných budov

5. Norma ČSN EN 15 193 Energetická náročnost budov – Energetické požadavky na osvětlení

Optimizing Power Flow in the Smart Region

Marián Uher

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
marian.uher@vsb.cz

Abstract. Efforts to maximize use of electric power from renewable sources necessarily leads to change the current conception of electrical distribution networks. This requires knowledge of the behavior of existing distribution network. For this purpose, the Department of Electrical Power developed the Dynamic Model.

1 Introduction

Connecting constantly large quantities of renewable sources into the network is damaging the conception main idea which is traffic based on a balance of energy consumed and produced from one central source. This fact in the near future will necessarily lead to the implementation of smaller units of networks with decentralized sources of electricity, because the existing conceptually outdated distribution network with this problem do not even.

In Czech Republic group CEZ deals this problem. Its intention is to create in eastern Bohemia first functional Smart Grid, which is self-regulating network, capable of island mode that meets the requirements of production and transmission of electricity from renewable sources.

To be implemented this pilot project it was necessary to know the conditions in existing networks and response to different operating conditions. Therefore was on the Department of Electrical Power developed a simulation program called Dynamic Model. This is unique software which simulates the behavior of the distribution system in the investigated network. In Department of Electrical Power was for this model created a team of workers which this model assembled and is working hard on its innovations.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 89–94.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Dynamic Model

2.1 Introduction to Dynamic Model

The Dynamic Model is a software, which mathematically calculates a behavior of the devote distribution network on a level of medium and low voltage under the conditions which specifies the user. It is possible to change the network configuration as a result of the handling activities in the network or failure conditions, such as short circuit.

The task of the Dynamic Model is therefore to provide a vision of distribution network behavior at different operating situations. This is one of the fundamental prerequisites for building modern self-regulating networks called smart grids.

2.2 Dynamic Model Structure

The basic structural element is computational core of the software LabWIEV and to the processing operations required by the user using other preprocessors, such as preprocessors of software group EMTP-ATP. The main preprocessor is ATP-Draw, where is created the basic mathematical model of the whole analyzed network.

This software has conditionally open structure which means that the user can change some parameters from the default elements palette elements on the users interface. It is possible to change the cross section of some lines or change some of the apparent power transformers etc. The user can also change some critical value such as maximum current of line, under-voltage levels etc. If these limits are exceeded the user sees the warning triangle which follows relevant warning message on the users interface.

When talking about devoted distribution network in a case of Dynamic Model, we refer to the specific existing network of eastern Bohemia, which was this software developed for. The examining of the operating states in this area leads to development of the first smart grid in the Czech Republic.

2.3 Dynamic Model Development Phases

In the first phases of development of Dynamic Model the emphasis was placed on steady state simulation of the investigated network. This preceded the long-term measurement installation of electric energy consumption on the different types of donations (flat units while respecting tariffs, houses units, industrial units). The data thus obtained was used not only as the default information to simulate the current load of line, but also for design of cogeneration units, which are already located in examined area.

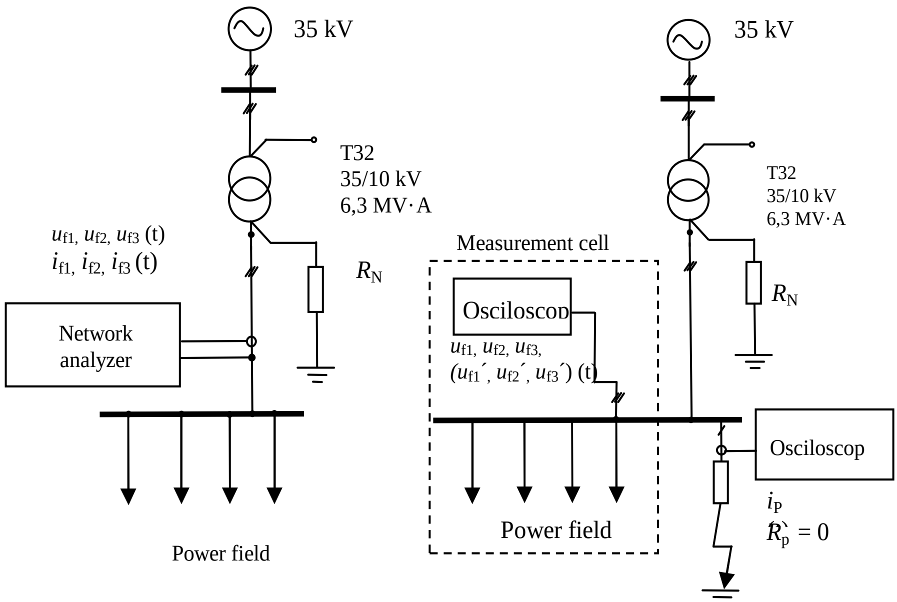


Fig. 1. Connection diagram of power measurement on the secondary side of transformer 35/10 kV and connection diagram of the experimental measurement on the secondary side of the same transformer.

To verify the correct functioning of the Dynamic Model was performed some experimental measurements. These were mainly the metal ground connecting which run on the secondary side of transformers T101 and T102; 110/35 kV, always in the different tuning of the compensation coils.

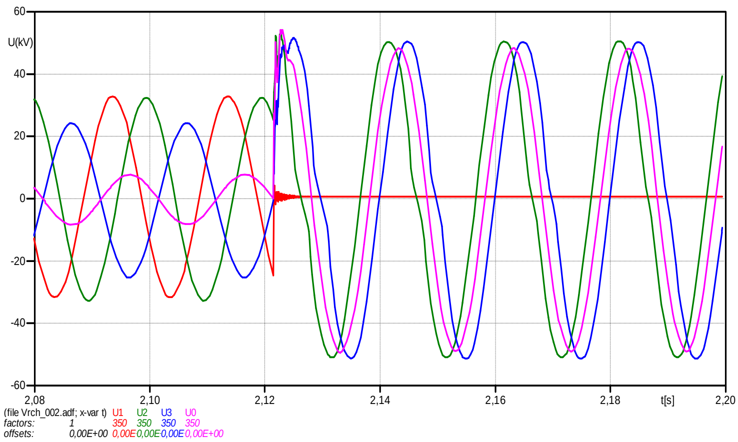


Fig. 2. The course of the phase voltage $U_{1f}´$, $U_{2f}´$, $U_{3f}´$ and the network node voltage U_0 of the ground connection on the secondary side of transformer T 101; 35/10 kV, coil tuned.

Fig. 2 shows a progress of the phase voltage and neutral network node voltage. At the time of 2.12 s starts a metal ground connection. On this Fig. 2 can be seen that the voltage of the neutral network node is triple than the real one because of the connection of voltage measuring transformers in open delta.

Further experimental measurements also consisted of the metal ground connection as before but at the 35/10 kV transformer.

2.4 Cogeneration Unit Model

As a part of the Dynamic Model innovation is currently processed the synchronous machine model as a model of an existing cogeneration unit just situated in studied power system. It is the four-pole synchronous machine operating at 230V voltage level, output 1.5 MV·A, powered by gas internal combustion engine.

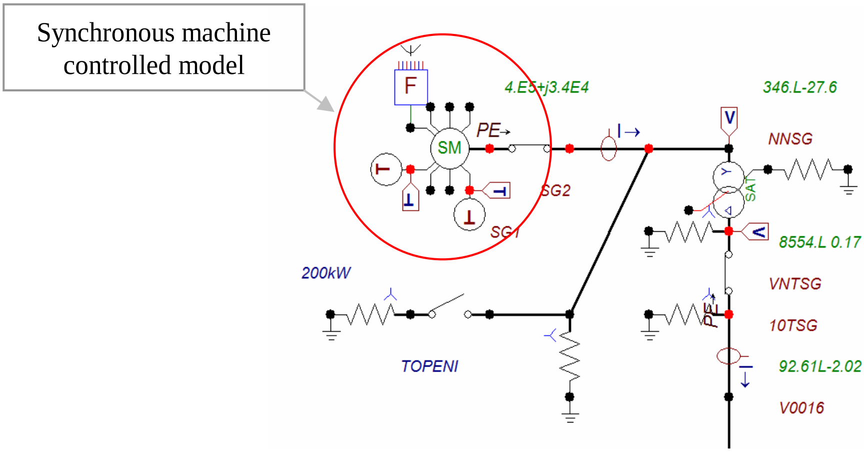


Fig.3. Part of the mathematical model of the distribution network with the cogeneration unit model highlighting

The cogeneration unit model is in software ATP-Draw realized by the synchronous machine control model. This model, compared to a uncontrolled model, which has only 3phase outlet of the stator winding, can be controlled by eight in/out signals TACS (Transient Analysis of Control System).

In our case it is the synchronous machine model which can regulated output power and stabilize output voltage under varying loads. At present runs testing of the correct cogeneration unit model function under variable load of different characters (capacitive, inductive, resistive) and in a fault conditions as short circuits.

The electric machine modeling is one of the most difficult modeling industry in general. In this case the problem of cogeneration units modeling is difficult because the behavior of the model must exactly match the real one, especially during transitional storylines and island mode.

3 Transients investigating

Investigating the cogeneration unit model is very important because in real it is possible to operate a whole examined part of the electric system in island mode. Therefore it is necessary to know behavior of the machine under the changing load of the machine and under varying operating conditions of the network.

At present is examining the steady state power system as well as transients such as short circuits. The Dynamic Model was extended for this option now by the special transient simulation function. Specifically there can be simulated symmetrical and asymmetrical failure (3ph short circuit, 3ph ground fault, 2ph short circuit and 2ph ground fault). For all this cases the user is informed about both electrical and mechanical variables, such as a voltage, current, frequency, machine speed, power flow etc.

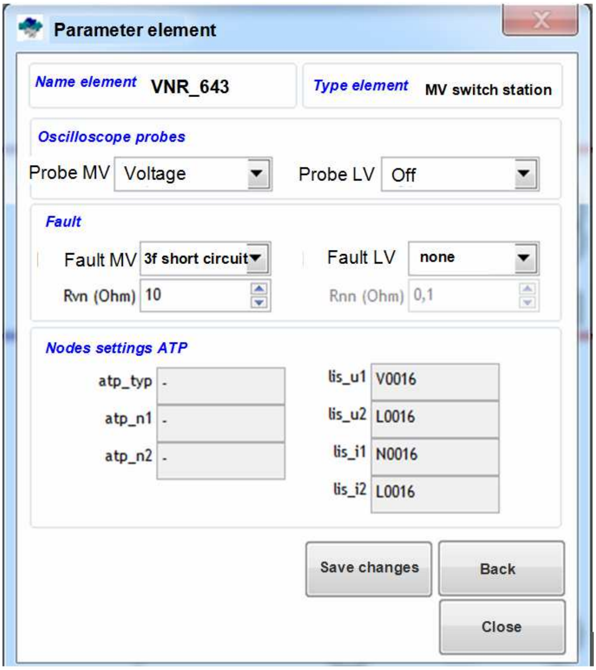


Fig. 4. Bookmark for settings faults conditions in an environment of Dynamic Model

The Fig. 5 shows the stator winding current waveform of cogeneration unit at 35 kV network failure, which is in a respect to cogeneration unit power state near symmetrical short circuit. In the time t = 0.1 s there will be a short circuit and in the time t = 0.280 s start reconnection occurs and this leads to relief of cogeneration units. This simulation used energy workers in eastern Bohemia to settings the protecting system in the devoted network.

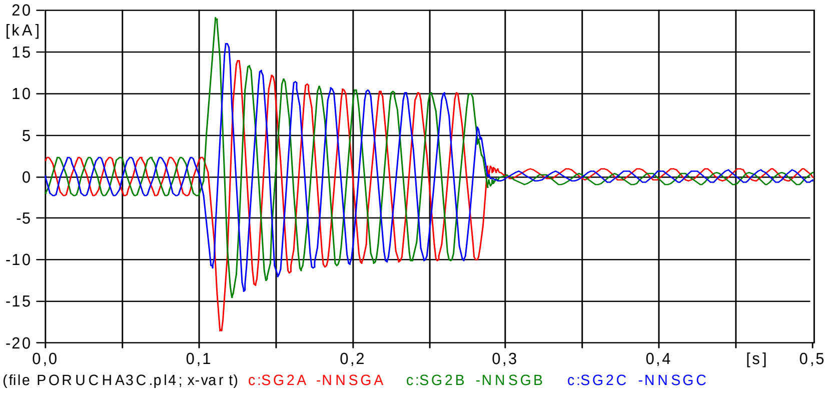


Fig. 5. The stator winding current waveform of the synchronous machine in case of simulation short circuit at the 35 kV voltage level

3 Conclusion

The Dynamic model development is very important part in building of the pilot project called Smart Region, whose realization means the emergence of the first smart grid in the Czech Republic. As mentioned above, the Dynamic Model is used by energy workers to understanding to the devote distribution network behavior. Based on these findings, energy workers can react to events in the devoted network.

Already now is the Dynamic Model used for the setting of the distribution system protect conception on the medium and low voltage level with respect that the whole studied network can be operation as an island mode powered by the cogeneration unit, or in parallel co-operation with the major electrical system.

The Dynamic Model development is far from over. The first task as the steady state simulations, the possibility of transient simulations were completed, but before us there are stages of development as connecting of the whole system of the Dynamic Model on the system of smart metering to possibility of monitoring network behavior in Online mode.

References

1. Uher M., Mišák S. Optimalizace provozu sítí s obnovitelnými zdroji s využitím dynamického modelu. XI Workshop. Ostrava 2011. ISBN 978-80-248-2378-2.

2. Uher M., Mišák S. Optimalizace provozu sítí s obnovitelnými zdroji s využitím dynamického modelu. EPE 2011. Dlouhé Stráně 2011. ISBN 978-80-248-2393-5.

Higher Current Harmonics on the Level 27 kV Caused by Traction Transport

Jan Unger

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.unger@vsb.cz

Abstract. Electrical circuits are active in two main voltage systems. The first one and from the point of EMC view incident-free one is direct current system 3 kV. The second one is alternating system 25 kV, 50 Hz, that shows unfavourable effects in the field of asymmetry and higher current harmonics. Spectrum disturbing frequency components in the reverse traction current are dependent on power system, traction power station (TPS) and driving vehicle, this means on the type of traction engine and the way of his regulation.

1 Introduction

Locomotives which are operated in alternate system shows whole spectrum of odd order harmonic currents on collector, starting with order of third harmonics (150 Hz). Another disadvantage of this system is small value of effect, which is changing according the dependence on the ride of locomotives through the powered section. This is why the compensation filtering equipments are located in ČD traction substations. There are graph in this article that displays progresses of odd harmonics in distribution point 27 kV. Because this distribution point is powered by the second phase, there are also compared odd harmonics of currents and voltages of individual phases between themselves. Measuring took place in traction power substation in Svitavy.

2 Influence of harmonics

Extensive using of semiconductor devices with considerable power implicates distortion of voltage and current harmonic progresses in power system and therefore distracts the quality of supplied energy.

Harmonic currents and voltages are in electricity system sources of:

- additional operating losses,
- malfunction of protection and automations

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 95–100.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

And causes:

- deterioration of appliances temperature regime, because their warming increases,
- speed-up of isolation aging process, which decreases lifespan of appliances,
- decrease of supplied active power from generator and therefore lower effectiveness

Usually are harmonics spread by electromagnetic way also to appliances, which are not related with electromagnetic way directly and therefore they cause malfunction e.g. telecommunication appliances.

3 Suppress of harmonics

The most common way how to decrease the level of harmonics in electrical net, which powers appliances with semiconductor converter, is use of multipulse connection. It is possible to reach similar results when are connected two and more rectifiers on one distributor.

Restriction of harmonic influence in the electrical net is possible to reach also by its appropriate connection. Usually are appliances, which generate harmonics, connected to power transformers or at least to individual section of three-reel transformer.

To appliances, which have compensatory effectiveness also belongs:

- condensers,,
- synchronic machines,
- serial resonant filters.

3.1 Compensation filtering equipments (CFE)

Driving vehicles of single-phase system 25 kV, 50 Hz which are kept by ČD until now, use on its input circuit diode traction rectifier. This is the reason why elementary harmonics shows effectiveness approximately 0,84. Use of diode traction rectifier leads to origin of odd harmonics spectrum in supply current starting with component 150 Hz. These are the reasons why it was necessary to add compensation filtering equipments CFE, which have two functions:

- adjust effectiveness TPS as a whole unit on the value which is requested by supplier of electric energy (0,95),
- restrict penetration of current harmonics in to the net 110 kV with a goal to decrease distortion of voltage course in connection point TPS to the net 110 kV. [2]

4 Measuring places

Measuring took place in traction power system in Svitavy from 2.2.2011 to 4.2.2011. The secondary voltage, current of transformer and current on charger 1 and 2 in terminal to traction lines N101 and N102 direction to Brno were measured on distribution point 27 kV.

For all currents and voltages were taken values until the order 40 of harmonics in time intervals 5 second.

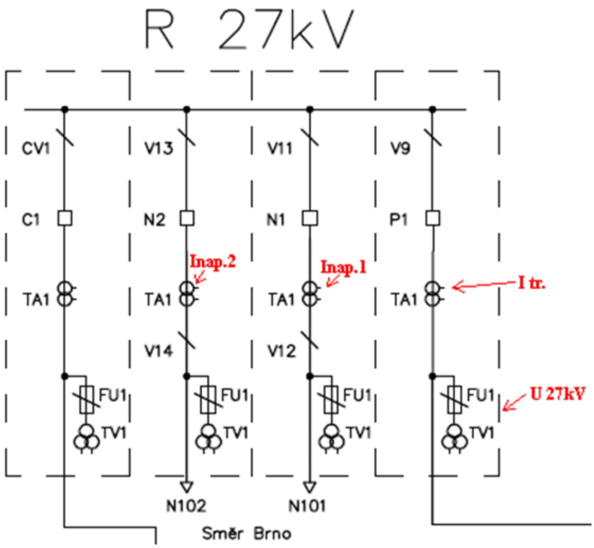


Fig.1. The connection of the measuring apparatus in TPS Svitavy

5 Graphs

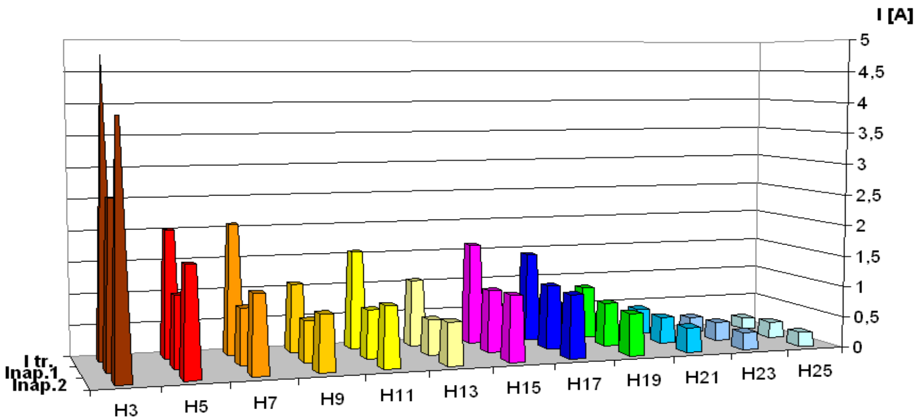


Fig.2. Current Harmoninc in traction terminals

We can see comparsion of individual harmonics between themselves and comparsion of every single harmonics in different measuring places on the previous graph. From this comparsion are evident higher values of current on supply lateral.

There are displayed average values of current harmonics in distribution point 27 kV on the first graph. Based on this graph we can say, that the smallest values of currents were in charger 1, where 3rd harmonics gained average value 2,63 A. Contrariwise the highest values were measured on transformer in supply lateral, where 3rd harmonics gained average value 4,76 A. This current harmonics were definitely the highest. Other odd harmonics significantly decreased and fluctuated in interval 0,7 – 2 A. From 21st harmonics was current moving under 0,4 A.

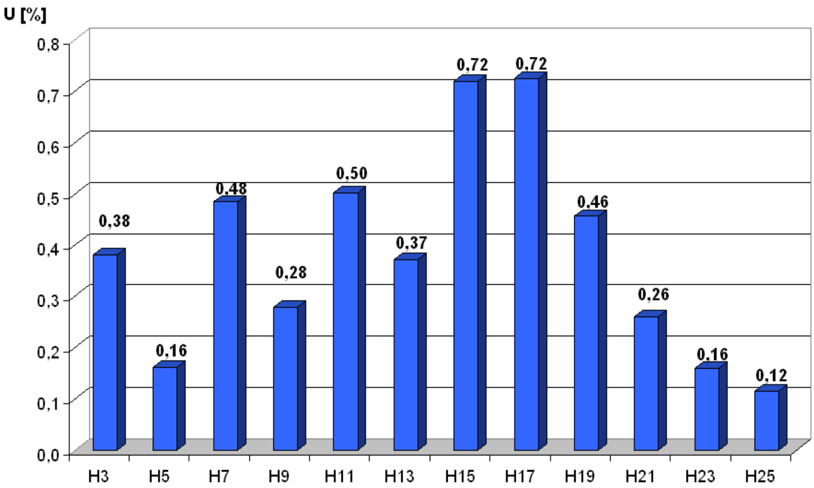


Fig.3. Avarage values of voltages

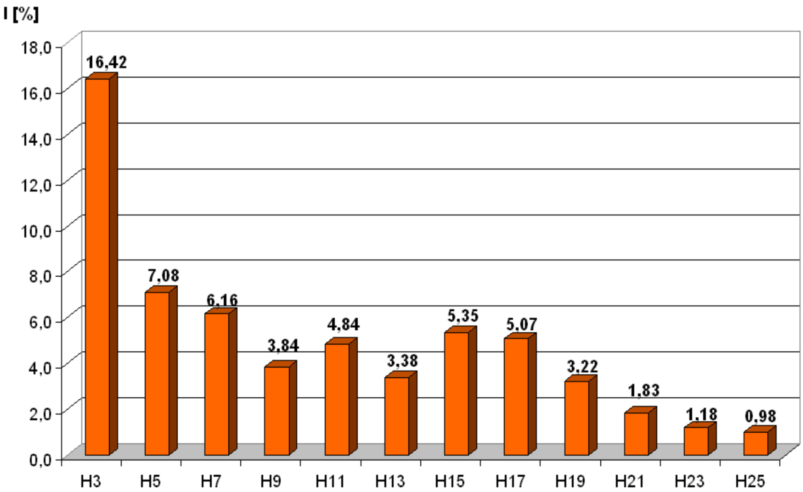


Fig.4. Avarage values of currents

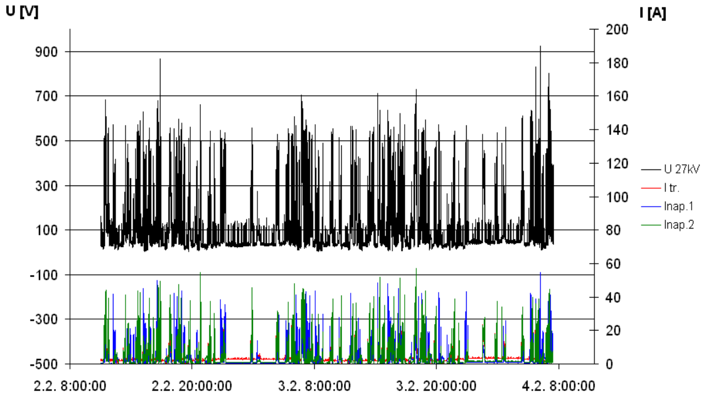


Fig.5. 3rd harmonics

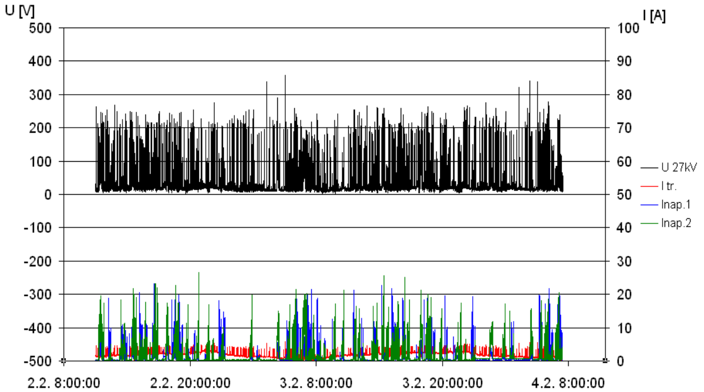


Fig.6. 5th harmonics

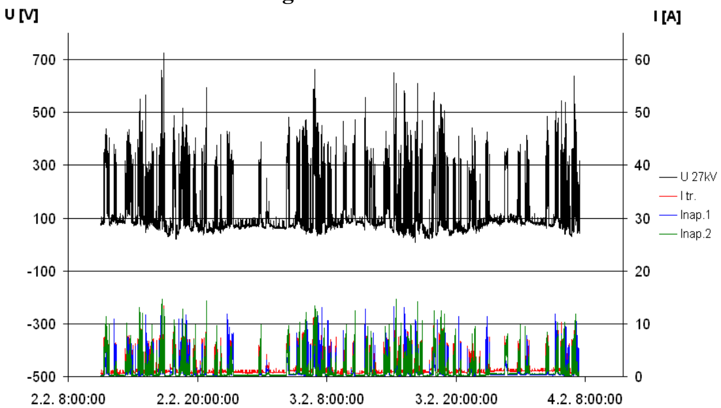


Fig.7. 7th harmonics

6 Conclusion

There is significant decrease of all harmonics in minimum limit on all progresses, particularly during night hours, when there was not any driving vehicle operating on railways. There are regular increases and decreases of harmonics voltages and currents during night and day. The issue is about regularity of traction traffic particularly the passenger one. Minor deflection from this regularity can be caused by freight transport, which is not operated with 24-hours intervals.

Voltages of 3rd, 5th and partly 7th harmonics are successfully suppressed with CFE, which decreased it's value in distribution point 27 kV. We can see high values of 15th and 17th harmonics in graph.

Regarding currents any harmonics did not exceed percentage orientational limit of Amplitude law (In = 100/n).

In bar charts are average values of harmonics during the whole measuring period.

References

1. Rules for the operation of distribution systems, Annex 3
2. Šolc P.: Diagnosis of electromagnetic phenomena, dissertation FEI, VŠB – TU Ostrava 2005

Appendix: Springer-Author Discount

Jan UNGER is alumnus of master's degree study FEI VŠB-TUO in the field of energetics. He is internal postgraduate student at the Department of Electrical Engineering of FEI VŠB-TUO, under the guidance of prof. Ing. Josef Paleček, CSc.

Effect on Power Quality of the Power Electric Railway Traction

Tien Trung Vo

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vo.tientrung@yahoo.com

Abstract. This paper presents a systematic approach to estimate effect of harmonic in AC electric railways, which incorporates a novel directional track shift method to deal with the train. On the basis of the united model, the transfer characteristics of harmonic currents and voltages are simulated and harmonic distortion is calculated while the locomotive is situated at different positions. The railway simulation is performed in the ATP version of the Electromagnetic Transient Program (EMTP) and the results of the harmonic simulation are compared for different positions of locomotive on a railway system.

Keywords: Power quality, Total harmonic distortion (THD), Electric railway, Traction network.

1 Introduction

The electric railway is in a competition with other possibilities of transportation. The locomotives became faster and faster and in consequence their power is growing as well. The high power locomotives and transmission pollute the supply system with harmonics. In order to ensure the power quality of the whole energy system, it is necessary to reduce this pollution. Strict requirements were established on the power quality of the electric supply network in the past. One of these requirements is in connection with the harmonic distortion of the voltage. The voltage distortion is caused by the non-linear loads connected to the network on different voltage levels. It can also due to effect of overvoltage and overcurrent consequences of transmission mode.

More kind of simulation models has been investigated calculating the harmonic effect of AC electric railway traction. In a former paper a detailed model of the traction system was discussed in the ATP version of the Electromagnetic Transient Program (EMTP).

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 101–106.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Theory

2.1 Harmonics

Harmonics are steady-state components of the voltage or current which have a frequency. They should not be confused with inter-harmonics or transients. The magnitudes of individual harmonics are often expressed as a percentage of the fundamental component, or as a percentage of the Root Mean Squared (RMS) magnitude of the overall voltage or current. Harmonic distortion in the power supply is an increasingly important concern for power utilities because of the negative consequences that it produces. These consequences include decreased efficiency of the power system, increased neutral currents, resonant overvoltage, and the accelerated aging or malfunction of sensitive equipment. The increasing sensitivity of many appliances to harmonics is further complicated by the increased use of harmonic producing power electric. As harmonics propagate deeper into the distribution and transmission systems, they affect a larger number of customers over a wider area. Therefore, power utilities are keen to limit the propagation of harmonics as much as possible.

Total Harmonic Distortion (THD) is often used as an overall measure of the harmonic distortion. The equation to calculate the THD is indicated below:

$$THD_X = \frac{\sqrt{\sum_{k>1}^{k_{\max}} X_k}}{X_1}$$

Where X_k is the RMS of k^{th} harmonic component of the current or voltage.

2.2 Electric railway traction system

Traction systems supply electricity to trains as they move along the track. Electric circuit that feeds electric traction consists of:

- The locomotive.
- The contact line system.
- The feeding transformer.
- The high voltage supply network.

The locomotives are running under the contact line system, dividing it into two parts. At the contact point the locomotives could be represented by a base frequency consumer and harmonic generator.

The contact line system should be considered as a multiconductor system with earth return that is composed of the contact wire(s), suspension wire(s), and the return rails. It can be reduced to a two-wire-system that leads the current.

Diagram of the electric railway system is shown in Figure. 1

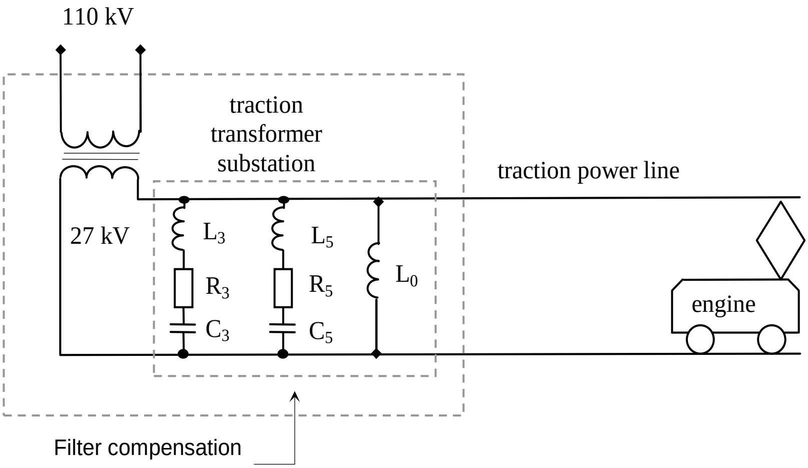


Fig.1. Simplified circuit for calculation of electric railway traction

3 Models

The computer simulation model of traction system is given on Fig. 2.

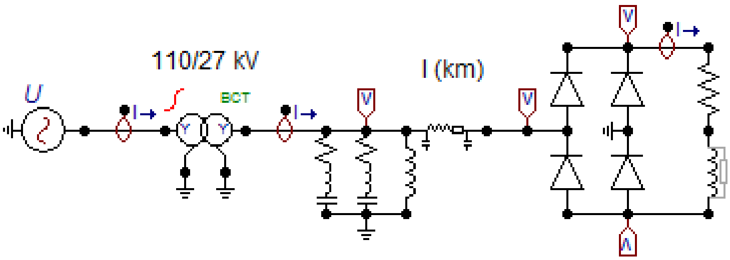


Fig. 2. Electric traction system model

The following models are applied for the calculations:

The locomotive model: It could be considered by Rectifier Bridge, the rectifier units are single phase double-way diode bridges. The traction motor has got series excitation. It could be represented with a simple R-L. The values of R, L as:

$R = 365\ \Omega$; $L = 0{,}6$ H.

The traction power line system has given by π parameter mode. The length of the system is an average of 20 km, where the feeding transformer is contacted to one of the ends. Parameters of contact line could be determined as:

$R_k = 0{,}2\ \Omega/\text{km}$; $L_k = 1{,}43$ mH/km; $C_k = 20{,}5$ nF/km.

High voltage power is represented by AC source 1 phase voltage 110kV.

Transformer traction model: In the ATP modeling software it could be represented by one BCTRAN 1 phase 110/27 kV.

Filter-compensation unit parameters:

$R_3 = 0{,}98\ \Omega$; $L_3 = 111$ mH; $C_3 = 10{,}5\ \mu$F;

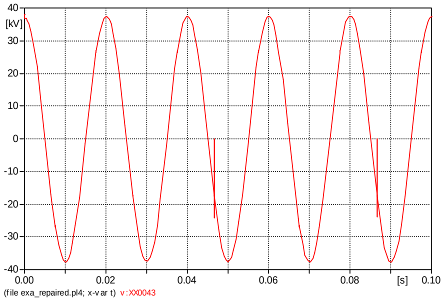
$R_5 = 0{,}91\ \Omega$; $L_5 = 116$ mH; $C_5 = 3{,}5\mu$F;

$L_0 = 533$ mH.

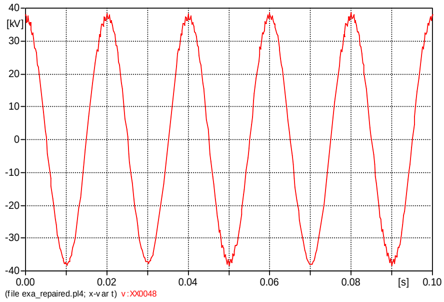
4 Simulation results

The simulation model of traction system in the ATP-EMTP is given on Fig. 2, the result presents calculation of harmonic of voltage and current along the 27 kV supply line for three locomotive positions, that locomotive at substation, 20 km and 40 km distance from the substation.

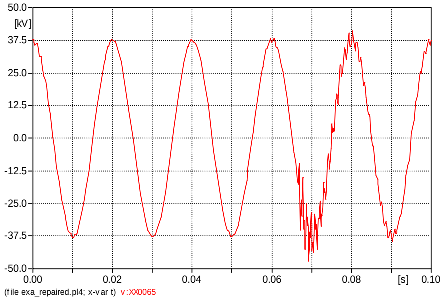
Voltages at substation.



a) Locomotive at substation



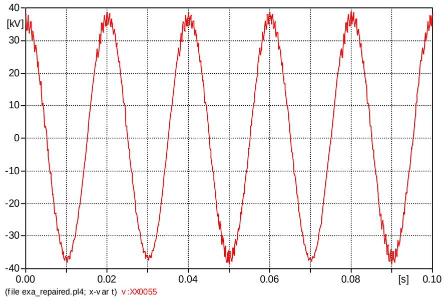
b) Locomotive at 20 km



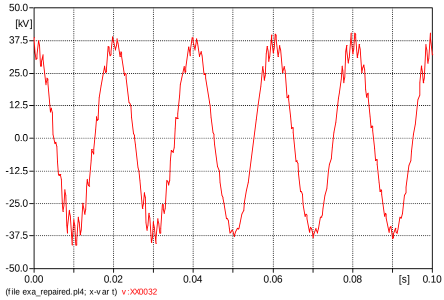
c) Locomotive at 20 km

Fig. 3. Comparison of harmonic voltages at substation when locomotive at different positions in the traction network

Voltages at locomotive positions in the traction network.



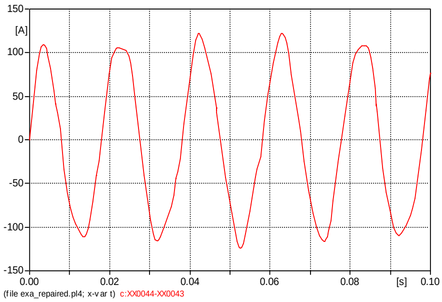
a) Locomotive at 20 km



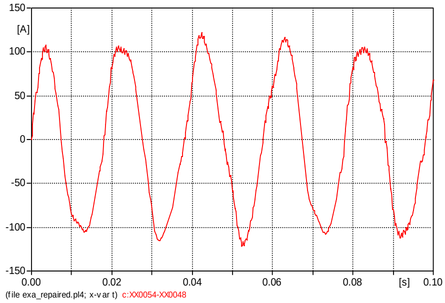
b) Locomotive at 40 km

Fig. 4. Comparison of harmonic voltages at substation when locomotive at different positions in the traction network

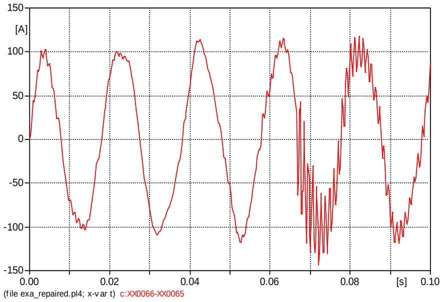
Currents on the contact line system in the traction network.



a) Locomotive at substation



b) Locomotive at 20 km



c) Locomotive at 40 km

Fig. 5. Comparison of harmonic currents on the contact line system in the traction network

Using the harmonic components of the currents and voltages, the THD values can be calculated along the supply line. The Table 1 shows the value of the voltages at the substation and currents by the substation on the contact line.

Table 1. Comparison of total harmonic distortions when locomotive at different positions

Harmonic \ Positions		Substation	20 km	40 km
Voltages at substaion	Without filter	0.9329%	1.821%	2.2904%
	Filter	0.55805%	0.77757%	1.0171%
Voltages at locomotive position	Without filter		3.5562%	6.9182%
	Filter		1.8021%	4.2838%
Currents at 27kV side	Without filter	7.1504%	10.154%	11.179%
	Filter	8.6492%	10.79%	11.192%
Currents at 110kV side	Without filter	6.6351%	9.3761%	10.314%
	Filter	8.1593%	10.155%	10.499%

From results in the table, we can see that the harmonic values varied significantly when the train was at different locations along the traction system. The longer distance between locomotive and traction substation, the bigger harmonic distortion value of voltage and current.

5 Conclusion

This paper has used the Electromagnetic Transient Program (EMTP) to establish the general united simulation model of the traction network. Base on modeling the simulations through harmonic transmission characteristic of traction network are simulated and analyzed. The harmonic distortions are also calculated while the locomotive is situated at different positions. When in the same condition of traction network, the vary position of the locomotive may cause the different harmonic distortions. The longer distance between locomotive and traction substation, the bigger harmonic distortion value of voltage and current. The system impendence, the distributed parameter of traction network, the location of locomotive, etc may cause harmonic distortion. However, the transmission characteristic and resonance on the traction network of the injected harmonic currents from the locomotive are mainly under the influence of the length of traction network.

References

1. Edward W.C. Lo, T.M. Lai: Harmonic Analysis of EMUs in Railway Systems. IEEE Conference, pp 142-148.
2. Andras Dan, Peter Kiss: Modeling of High Power Locomotive Drives for Harmonic Penetration Studies. IEEE Conference, pp 1-5.

Heat Pumps for Heating Aula at VŠB - TU Ostrava – Evaluation of This Year's Heating Season

Petr Zach

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.zach@vsb.cz

Abstrakt. The article deals with assessment of the heating season 2010/2011 using heat pumps at the Assembly Hall of VŠB – Technical University of Ostrava. Heating within the newly built Assembly Hall building involves 10 heat pumps manufactured by the IVT company from Sweden with the total output of 700 kW. The source of low-potential heat is ensured with the system of 110 boreholes down to the depth of 140 m. The bivalent, i.e. auxiliary, source is represented by the interchanger station within the central heating supply. The heat loss of this Assembly Hall – the heating output required to maintain the Assembly Hall with the necessary environmental temperature, with the ambient air temperature of -15 °C, is equal to approximately 1.200 kW.

Keywords: Heat pumps, geothermal energy, heating season.

1 Introduction

This article deals with assessment of heating in the Assembly Hall of VŠB-Technical University Ostrava. The heating in Assembly Hall is provided using heat pumps of ground-water type. The section of primary heat source is of essential importance to the design and characteristics of a heat pump. The ground-water heat pump draws heat from the depth of ground´s surface using interchangers, the so called "collectors". The primary circuit of the heat pump is closed. The heat transfer fluid used must be anti-freeze and environmentally friendly material. The heat collected is then transferred into the heating water. Geothermal heat is accumulated within the ground´s mass. Its collection causes gradual and uneven drop of temperature within the layer around the collector with subsequent broadening of the layer. That depends mainly on the specific output of the collector, the method of collector installation, the ground´s mass material and the heat collection transit time during heating season.

2 Description of the assembly hall

The building that hosts the Assembly Hall and the Information Technologies Centre is heated using the ground-water heat pumps. The source of low-potential heat is represented by boreholes. Heat pumps provide approximately 82 – 85 % of the total heat supply to the building during the year. The bivalent, i.e. auxiliary, source is represented by the interchanger station within the central heating supply. The building

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 107–112.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

comprises of the ceremonial Assembly Hall, lecture rooms, the multipurpose hall, the Information Technologies Centre with computer classrooms, development laboratories and office premises. The Assembly Hall building comprises four floors with basement, featuring a reinforced concrete monolithic structure and lightweight glass-metal shell. The total built-up area equals to 3.917 m^2. The basement of the building is dedicated for underground parking with the capacity for 40 vehicles. This area also contains the building´s utilities: the transformer station, the heating and air conditioning generator room, the power distribution centre and storage premises.

3 Technical data

The system of heat pumps installed within is the largest one to have been established in Czech Republic and the whole of Central Europe. The source of heat comprises 10 heat pumps made by the IVT company from Sweden with the total output of 700 kW using the system of 110 boreholes drilled to the depth of 140 m (the total aggregate length of boreholes is 15.400 m). There boreholes are located within the parking space area of the Assembly Hall of VŠB – TUO and the parking adjacent to the library of VŠB – TUO. The fittings on these boreholes include four-pipe arrangement with two loops of PE piping DN 32 mm with special joint mount. The total length of PE piping installed within the boreholes is approximately 70.000 m. The boreholes are injected with cement-concrete mixture. The primary circuit is filled with the total amount of 18.000 litres of anti-free heat transfer fluid. The system of 110 boreholes is run into five collection shafts, which means there are 22 boreholes per 1 shaft. Every shaft is then connection to a separate pipe with an independent circulation pump. The heat loss of this Assembly Hall – the heating output required, to be determined in order to ensure the necessary envi-ronmental temperature within the Assembly Hall, with the ambient air temperature of -15 °C, is approximately 1.200 kW. The heating system is formed by the under floor heating, heating bodies and air conditioning. Particular heating systems have been designed for low-temperature descent. The preparation of domestic hot water is ensured by means of plate exchanger within accumulation magazines of hot water.

Tab. 1. Performance parameters of heat pump in accordance with EN255:

Temperature (°C)	Heating output (kW)	Electric power input (kW)	Performance factor (-)
0/35 °C	67.8	16.7	4.06
0/50 °C	69.8	22.3	3.13

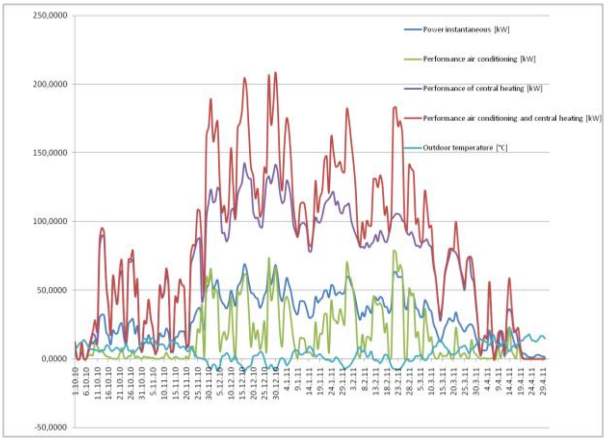
Fig. 1. Heat pumps, 10 units
Fig. 2. Circulation pump in the primary circuit, 5 units

4 Database

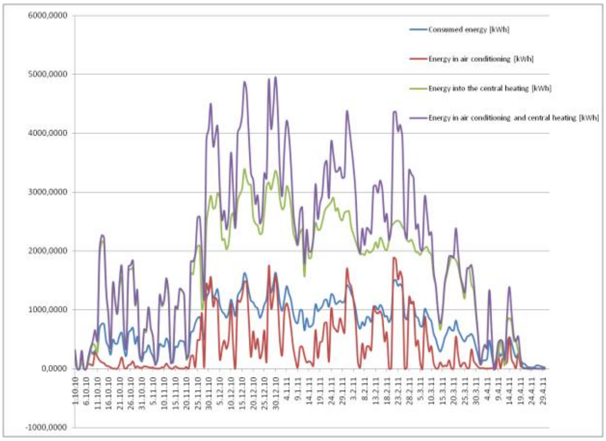
The heat pumps are controlled using the ProCop Monitor programme. The programme runs automatic scans within 10-minute intervals to monitor all the values and performs continuous saving of the data into the database. The ProCop monitoring and visualisation system is distributed by the Siemens company under the trade name Visonik Alfa in the Czech Republic. The data used for assessment was collected during the heating season 2010/2011.

5 Heating season assessment

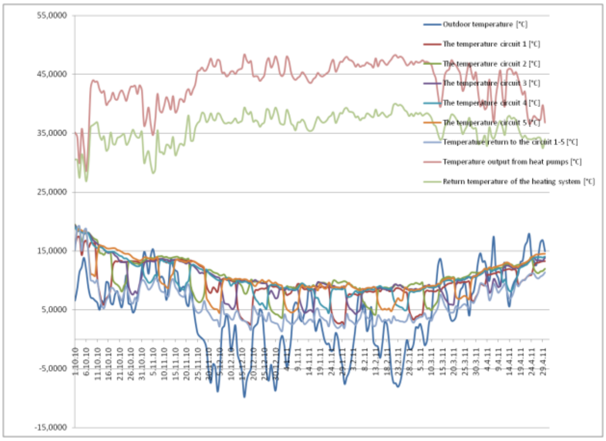
The heating season 2010/2011 subject to assessment started on 1 October 2010 and ended on 30 April 2011. The amount of energy supplied by heat pumps over the said season was 1524.20 GJ, out of which the amount of energy supplied to the air conditioning system – hot air heating – is 291.65 GJ and the amount supplied to the central heating system – under floor heating + radiators – is equal to 1232.55 GJ. Throughout the entire heating season, the heat pumps consumed 150239.75 kWh. These values provide for the final performance factor for the heating seasons at the level of 2.82. The performance factor is the basic parameter showing the efficiency of heat pumps, it is calculated as the ratio of the heat produced and the electric power consumed. During the heating season, the average outdoor temperature of nearly all months of the values higher than the previous year. The coldest day was 16 December 2010, when average daily temperature reached -9.79 °C.



Graph. 1. Behaviour of output and input of heat pumps



Graph. 2. Behaviour of energy produced and consumed by heat pumps



Graph. 3. Behaviour of temperature within particular borehole circuit

6 Assessment of graphs

The Graph 1 shows behaviour of the following values: the immediate input, the output supplied into the air conditioning system, the output supplied into the central heating, the total output, which is the sum of output of air conditioning and central heating. This graph helps us determine the dependency of the total output on the ambient temperature as the decrease in ambient temperature will increase the heat loss of the building. When the ambient temperature is low, the output of heat pumps must be raised to keep the desired environment temperature within the building. The graphic depiction of output also shows the drops caused by decreasing ambient temperature on weekends, public holidays and those days, when no classes take place in this building. The graphic image of behaviour also shows that the supply of heat is ensured using rather the central heating than the air conditioning system. The latter is owing to the fact that the lower temperature of water from heat pumps is sufficient for the under floor heating system. If the output temperature is lower, then the heat pumps have a higher performance factor.

The Graph 2 shows behaviour of the following values: the energy consumed, the energy supplied into the air conditioning, the energy supplied into the central heating. This graph can help us determine the total amount of energy consumed and produced by heat pumps during particular months of the heating season.

The Graph 3 shows the behaviour of the following values: the ambient temperature, temperature in particular borehole circuits from 1 to 5, the temperature of output from heat pumps and reverse temperature from the heating system. This graph can be used to determine the dependency of the output temperature on the ambient temperature, as the decrease of ambient temperature invokes an increase in temperature of water supplied by the heat pumps. Further delineation represents particular behaviour of temper-ature around the boreholes. The low-potential source of heat comprises 110 boreholes connected to form 5 circuits, i.e. there are 22 boreholes per 1 circuit. Collection of heat from a particular circuit causes the drop in internal borehole temperature over certain period of time. He fully exploited circuit will be disconnected, the boreholes within shall be subject to regeneration and the system will use another unexpended circuit in the meantime. The period of exploitation of each circuit is equal to 7 days. If technically possible and cost efficient, the external exchangers shall be slightly oversized. The use of more boreholes is more expensive, yet that will allow the heat pumps draw more energy from the ground and there would be to threat of depletion of the low-potential energy.

7 Conclusion

The heating season 2010/2011 subject to assessment started on 1 October 2010 and ended on 30 April 2011. The amount of energy supplied by heat pumps over the said season was 1524.20 GJ, out of which the amount of energy supplied to the air conditioning system – hot air heating – is 291.65 GJ and the amount supplied to the central heating system – under floor heating + radiators – is equal to 1232.55 GJ. Throughout the entire heating season, the heat pumps consumed 150239.75 kWh. These values provide for the final performance factor for the heating seasons at the level of 2.82. In this heating season, enough to heat the hall only heat pumps, even when outdoor temperatures were below -9.79 °C and did not have to use supplemental heat source.

Acknowledgment:

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2011/193).

Reference

1. Hradílek, Z., Zach, P.: Heat pumps, renewable source of energy for Aulum at VŠB - TU Ostrava. *EPE 2010*. Brno, 2010. ISBN 978-80-214-4094-4.
2. Hradílek, Z., Zach, P.: Heat pumps, renewable source of energy for Aulum at VŠB - TU Ostrava - Analysis of histograms of heating factor. *ELEN 2010*. Praha, 2010. ISBN 978-80-254-8089-2.
3. Hradílek, Z., Zach, P.: Heat pumps, renewable source of energy for Aulum at VŠB - TU Ostrava - Evaluation of the heating factor. *ELNET 2010*. Ostrava, 2010. ISBN 978-80-248-2349-2.
4. Hradílek, Z., Zach, P.: Heat pumps, renewable source of energy for Aulum at VŠB - TU Ostrava - Evaluation of the heating season. *EPE 2011*. Dlouhé Stráně, 2011. ISBN 978-80-248-2393-5.
5. Zach, P.: Heat pumps, renewable source of energy for Aulum at VŠB - TU Ostrava - Evaluation of the heating factor. *WOFEX 2010*. Ostrava, 2010. ISBN 978-80-248-2276-1.
6. Zach, P.: Heat pumps, renewable source of energy for Aulum at VŠB - TU Ostrava - Evaluation of the heating season. *11.WORKSHOP 2011*. Ostrava, 2011. ISBN 978-80-248-2378-2.

Optimalisation Shape of the Air Gap of Mantle Electromagnet for Industry Application

Zdeněk Bača, Jan Otýpka

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{zdenek.baca, jan.otypka.st}@vsb.cz

Abstract: The article deals with finding an optimal shape of the air gap of the mantle cylindrical electromagnet. To find it finite element method was used, which is used by COMSOL program. This program is used to find the optimal solution.

Keywords: Electromagnet, skew keeper, skew pole shoe, air gap, maximum tensile force

1 Introduction

This article reassumes the previous publication presented at the conference called EPE 2010, during which we dealt with finding an optimal shape of electromagnet on the conditions of maximum tensile force and skew pole shoe of the electromagnet. Optimal values with skew keeper of the electromagnet were 45° and with skew pole shoe they were 0°. One of the solutions is shown in Fig. 1.

As in the previous publication this is the direct spin mantle electromagnet, the layout solution of which is shown in Fig. 2. We decided to come out of the previous shape for maximum tensile force and further explore what effect on the tensile characteristics the change of the air gap in the mantle of the electromagnet will have.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 113–118.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

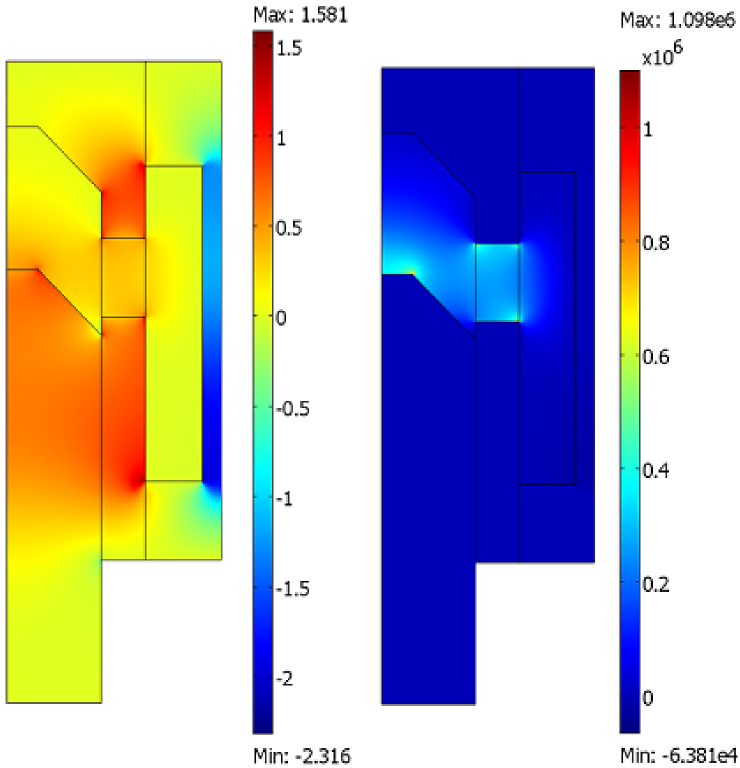
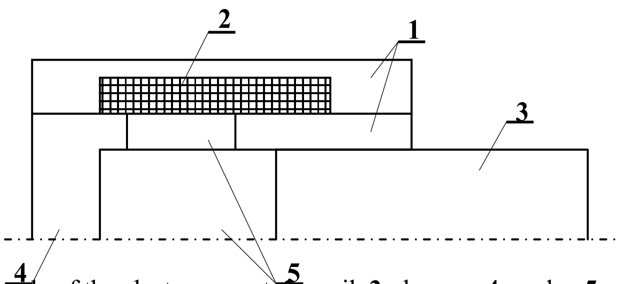


Fig. 1: Previous solution to the electromagnet



1 - mantle of the electromagnet, **2** - coil, **3** - keeper, **4** - yoke, **5** - air gap

Fig. 2: Layout solution to the electromagnet

2 Defining the parameters

The shape of the tensile characteristics is affected, among other things, by the mutual proportions of the electromagnet. In this particular case we are interested in changing the shape of the air gap in the mantle of the electromagnet. Therefore, we will examine the dependence of the tensile force of the electromagnet on the air gap scaling, which can be written as

$$F = \mathrm{f}\,(d) \qquad \mathrm{N;\ m} \tag{1}$$

F is a tensile force,
d is a size of the air gap in the mantle of the electromagnet.

In terms of optimization [3], we determined the parameters as follows: As an observable, we set the maximum tensile force of the electromagnet and as variable we determined the size of the air gap. (Fig. 3)

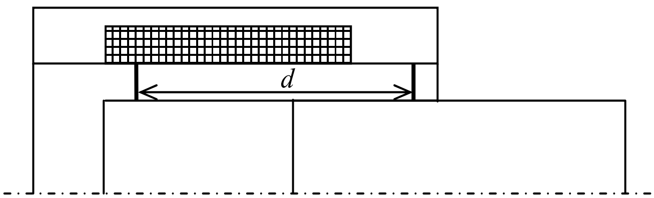


Fig. 3: The size of the air gap d in the mantle of the electro-magnet

The size of the air gap was changed from the size of d = 2 mm to the distance of d = 0,036 m with step of 0,002 m. This step was sufficient for our purposes, a more accurate value could be found, if we conducted a finer step around the extreme of the tension characteristics. The values, however, differed only minimally. We missed the case of d = 0,0 m, because the calculated values of the tensile force were mini-mal. Similarly, we limited the maximum size of the air gap value to d = 0,036 m. There were two reasons that lead us to the decision. The first one was the fact that values in this area had changed only minimally, and the second reason was technical parameters of the electromagnet. With further increase of the air gap the magnet would not work.

3 Solution

For solution COMSOL program is used here which al-lows to solve physical problems described by partial differential equations using the finite element method. The program can model multiphysical processes in engineering practice and in many developmental areas of scientific and technical disciplines.

COMSOL [2], as other similar programs, is divided into several parts. In the first part a geometric model is prepared, further material properties are defined and solution conditions are entered. Then covering the model with a mesh wiring (meshing), actual model solution and displaying results follows.

In this case it has been sufficient to create only half of the model (Fig. 2) because it is an axially symmetrical model. Material properties have been typed as such: for the mantle, the keeper and the pole shoe we defined *B-H*, *H-B* characteristics and dependence of relative permeability on magnetic induction, just as it was required by COMSOL program. For the air gaps we have defined only relative permeability which equals one and for coil, in addition, the value of current density equal to $8{,}426\cdot 10^6 \mathrm{A\cdot m^{-2}}$. After entering the boundary conditions we covered the model with mesh wiring and let it calculate. One of many results of the electromagnetic field lay-out is shown in Fig. 4.

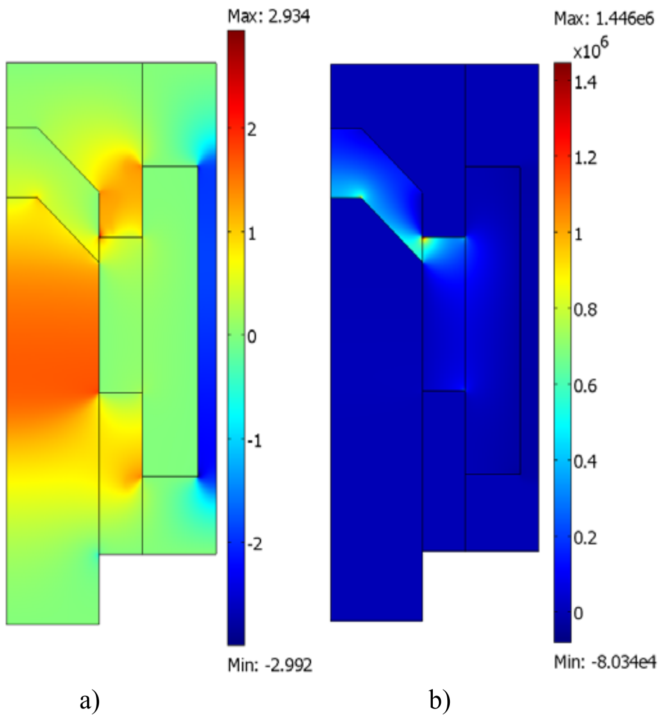


Fig. 4: Final field lay-out a) *electromagnetic induction (in T),*
b) *electromagnetic field intensity (in A/m*)

We get a tensile force so that first we mark an area in the pulling force of which we are interested, it means the air gap. In COMSOL [2] program this force is calculated using Maxwell tension tensor which is entered by unTz_qa+dnTz_qa. Similarly, it is carried out for each shift of the air gap. All results obtained are summarized in Fig. 5 which expresses the dependence of tensile force F on the distance of air gap d for ten distances of keeper from yoke of the electromagnet.

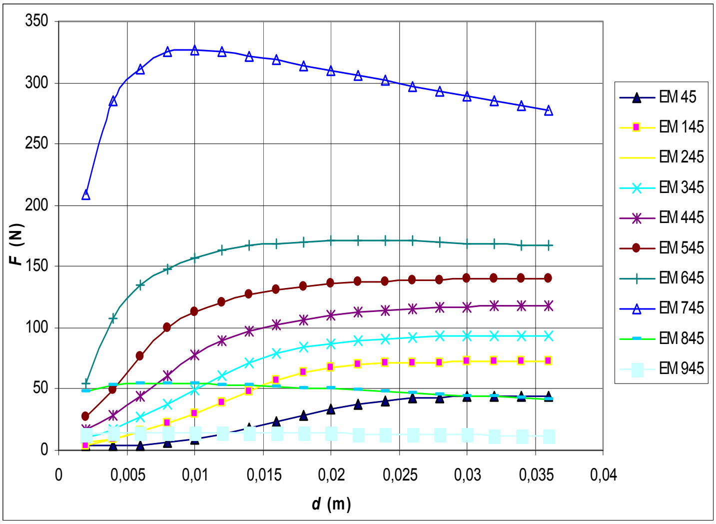


Fig. 5: Results of tensile force depending on air gap

4 Conclusion

As shown in Fig. 5, the maximum values of tensile force for most of the displayed dependences are almost constant with distance of 0,02 m from the beginning. Furthermore, it is evident that as the keeper gets too close to the yoke of the electromagnet (EM 845, EM945) the tensile force is almost constant throughout the changes in the air gap, so it does not depend too much on the measurement. What is atypical is the characteristics of EM 745, where there is a large increase of tensile force comparing to other characteristics. At distance of 0,01 m tensile force starts decreasing. This may be caused by super saturation of iron in the places where there are very small distances among different parts of the electromagnet.

The optimal solution, while maintaining maximum tensile force, is therefore an electromagnet, which will have an air gap of 0,02 m between the electromagnet mantle.

Acknowledgements

This article was supported by a grant SP2011/12 Adverse effects of electrified railways, magnetic fields, conductivity of nanomaterials and machine parts and TA0131231 of Technology Agency of the Czech Republic.

References

[1] Mayer, D., Ulrych, B.: Elektromagnetické aktuátory, Praha, 2008, ISBN 978-807300-216-9

[2] Promotional materials of company Comsol Multiphysics

[3] Zajaczek, S., Ivánek, L., Phan T. T. T.: Nalezení optimálního tvaru elektromagnetu, Electric Power Engineering 2010, Brno, ISBN 978-80-214-4094-4

Programming Vector Control for AC Induction Motor with DSC TMS320F28335

Pavel Bílek

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavel.bilek@vsb.cz

Abstract. In this paper, there is a description of programming Vector Control for AC Induction Motor used Digital Signal Controller TMS320F28335. It is a description of Vector Control and its result used motor with test stand. Here is a description of used hardware, drive and procedure for conceiving software for Vector Control.

Keywords. Vector Control, Alternating Current Induction Motor, Digital Signal Controller

1. Introduction

Vector Control provides management engine in optimum condition at steady over time and transient. It is characterized by linear behaviour and minimal delay.

1.1 Vector Control for Alternating Current Induction Motor

Vector Control for AC Induction Motor means the spatial distribution of stator current vector into two perpendicular components in the rotating coordinate system. Dimensional vector can be oriented to the resultant magnetic flux vector (oriented system of coordinates x, y) or the spatial vector of the excitation magnetic flux (rotor system of coordinates d, q). Control circuits are separated for the torque and magnetic flux of engine, so as not to affect to each other. More information are in [1].

The calculation is done in several steps :

- measure motor quantities (phase currents i_{1a}, i_{1b} and voltage)
- transform them into the 2-phase system ($i_{1\alpha}$, $i_{1\beta}$) using Transformation 3/2
- calculate the magnetic flux vector i_{1x}, and i_{1y} represented the torque of the machine
- calculate angle ε of rotor position, and their components sin ε and cos ε
- calculate oriented values sin γ and cos γ and magnetizing current i_{im}

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 119–123.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

- speed (R_ω) and magnetizing current controller (R_{im}) controlled their components
- the stator current flux (i_{1x}) and torque (i_{1y}) producing components are separately controlled
- stator voltage space vector is rotated by a Vector Rotation 1 from x,y reference frame into 2-phase system α, β.
- Transforming from 2-phase system to 3-phase system by Transformation 2/3 and using Pulse Wide Modulation

1.2 Block schematic for Vector Control with Alternating Current Induction Motor

Block diagram of Vector Control for AC Induction Motor is shown in Figure 1. There are designated places of software and hardware. The circuit is powered by power supply, there is also connected to a voltage-link, IGBT transistors with the respective drivers are used in the inverter.

Speed Reference input is selected command speed. Inputs to a Digital Signal Controller are measured currents i_a, i_b and incremental position sensor situated at shaft of asynchronous motor.

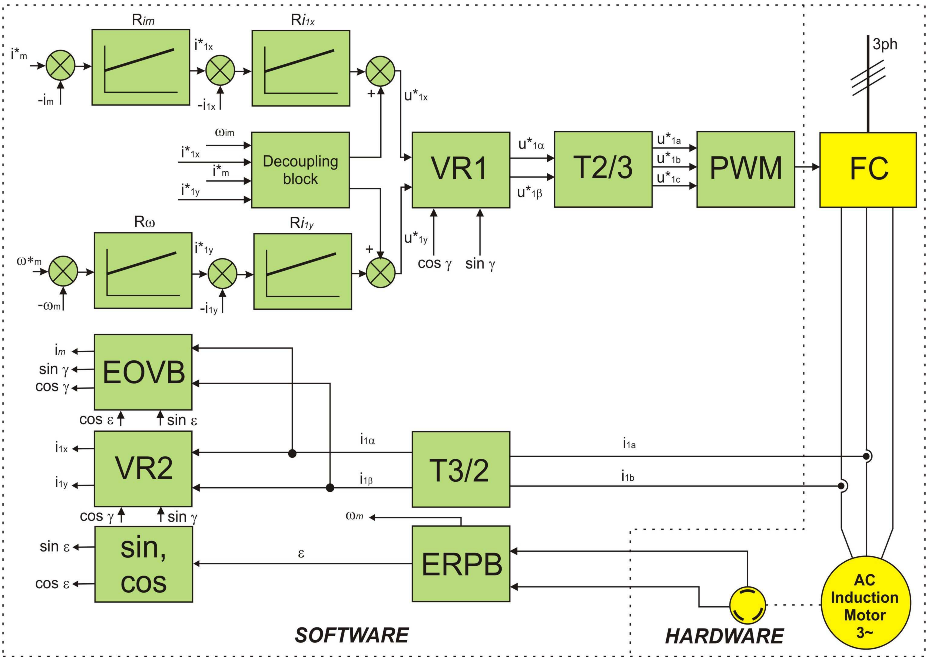


Fig 1. Block schematic for Vector control with AC Induction Motor

2. Practical realization

To program the Vector Control is not necessary to write a program in assembler or C language.

2.1 Presentation of test stand

In fig. 2 is a photo stand with AC Induction Motor, which was the vector control programmed. This is a combined induction AC and PMSM (load) motors connected with clutch. On fig. 3 shows the DSC control system TMS320F28335 with support board.

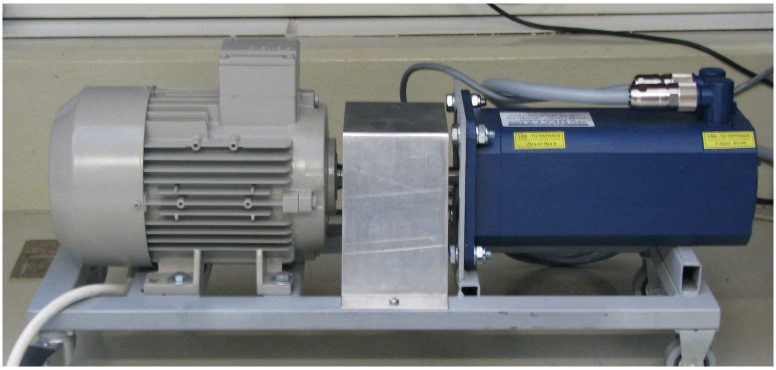


Fig 2. SMPM used in this programming Vector Control

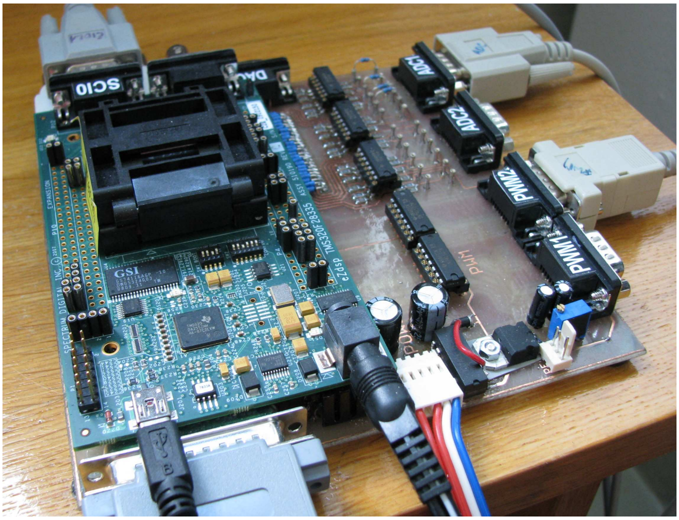


Fig 3. Kit with DSC TMS320F28335 with support board

2.2 Result

There are the results measured on the test stand. It was used induction motor: power $P = 2.9\ kW$, $p = 1$ *pole pare*, nominal speed $n = 2825\ rpm$. In first two graphs: command speed was $n^* = -600\ rpm$ and It was reverse to $n^* = 600\ rpm$. Command Magnetizing current i^*_{1x} was $i^*_{1x} = 3{,}5\ A$ and i^*_{1y} represented the torque of the machine was $i^*_{1y} = 2A$. In the middle graphs there are command speed $n^* = -500\ rpm$ and reverse to $n^* = 500\ rpm$, and $i^*_{1x} = 3{,}5\ A$, $i^*_{1y} = 2A$. In the last graphs, there are command speed $n^* = 300\ rpm$ and reverse to $n^* = -300\ rpm$, and $i^*_{1x} = 3{,}5\ A$, $i^*_{1y} = 3A$.

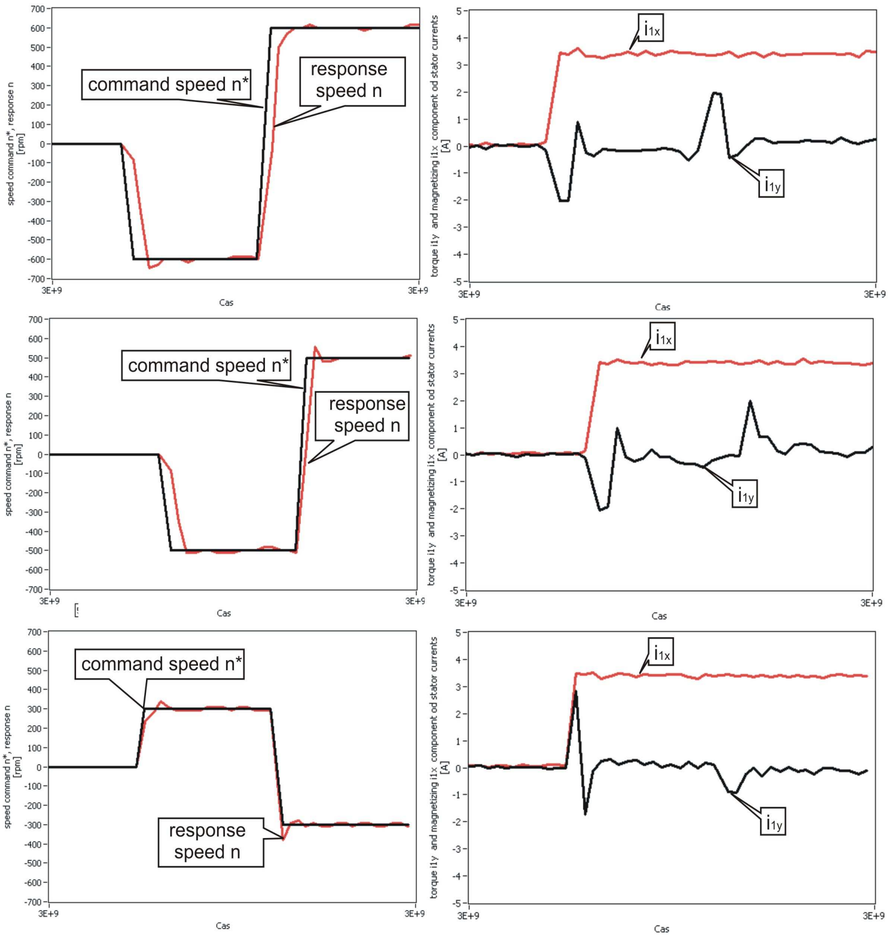


Fig 4. Results measured on test stand

3. Conclusion

In this article describes the basics of vector control AC Induction Motor and its implementation on the DSC TMS320F28335.

4. Acknowledgement

In the paper there are the results of the project SP2011/107 which was supported by Student Grant Competition of VSB-Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

5. References

1. Brandstetter, P.: *"A.C. Control Drivers, Modern Control Metod"*, VSB – Technical University of the Ostrava, 1999. ISBN 80-7078-668
2. Texax Instrument: *"TMS320F38335/28334/28332 Digital Signal Controllers"* (DSCs), Data manual, http://focus.ti.com/lit/ds/sprs439i/ sprs439i.pdf

Experimental Workplace for Teaching of Modern Control Algorithms Controlled AC Drives

Pavel Bílek and Josef Szotkowski

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{pavel.bilek, josef.szotkowski}@vsb.cz

Abstract. This paper attempts to describe an experimental workplace for designing of new control algorithms for electric regulated drives with induction motors. Laid out in few chapters, a reader is gradually introduced to the hardware to design these algorithms. First two chapters deals about a description of a microprocessor and starter kit from Texas Instruments. In other chapters there are description of indirect frequency converter, an incremental position sensor and support board for starter kit. In conclusion, there are final designs of the laboratory workplace.

1 Introduction

Experimental workplace need necessary hardware for creation of modern algorithms controlled AC drives, so these algorithms can be tested in practice. This experimental workplace should include next things:

- Indirect frequency converter with voltage fed
- Starter kit with microprocessor
- Electric motor with incremental sensor to evaluate the position of the rotor
- Support board for starter kit
- Modern current and voltage sensors
- Personal computers with necessary software for teaching

2 Processor TMS320F28335

The microprocessor from Texas instruments has been selected to realize an experimental workplace for teaching modern algorithms. This microprocessor was supplied like starter kit. There is TMS 320F2833x series. It is powerful processor with a wide

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 124–129.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

range of input/output peripherals, great computing power and good reliability. It work with floating point, which It simplifies the mathematics operation.

Microprocessor that contains:

- 256KB 3.3V on-chip flash (16-bit word)
- 34 KB Single-access RAM (16-bit word)
- 32-bit CPU Timers
- Floating-point Unit
- 12-bit ADC
- PWM output
- 8 external interrupts
- WDT (Watchdog Timer)
- SPI, SCI, CAN, I^2C
- 88 General Purpose I/O pins

3 Starter Kit eZdsp™ F28335

Starter kit eZdsp is a complete software development platform for the TMS320F2833x series of floating-point Digital Signal Controller. The eZdsp kit includes an F28335 target board with integrated JTAG emulation, 128kx16 asynchronous SRAM, clam-shell socket for the F28335 DSC, RS232 and CAN 2.0 interface with on-board transceiver and 9-pin DSUB connector and expansion headers that provide access to all I/O signals. One LED is added for user analysis, typical for actual program state. This LED is controlled by GPIOF output. DIP switches on the board select the boot mode and the processor configuration. Also included in the kit is the Code Composter Studio™ Integrated Development Environment, USB interface to the host PC and a universal power supply.

4 Indirect frequency converter with voltage fed

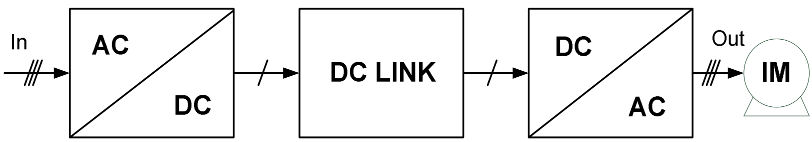


Figure. 1 Block diagram of the Indirect frequency converter

It needed Indirect frequency converter to control of Induction Motor. The block diagram of the Indirect frequency converter is in the Figure 1. This converter was made in the Department of Electronics. Parameters of converter are:

- Dimensions 330 x 240 x 240 mm
- Input supply voltage 400VAC
- Maximum voltage on DC link 600V
- Maximum output current 60A

The current and voltage sensors, switch to blocking function of converter, inputs connector for PWM modulation (canon 9) and output connector from sensors (canon 9) are implemented in converter. There are connector for supplying auxiliary source and power switch. In Figure 2, there is final version of the converter

Figure. 2 Final version of converter

5 Asynchronous motor with incremental sensor implemented

Motor to the experimental workplace was chosen asynchronous motor with encoder implemented on shaft. The parameters of this motor are:

- Power $P = 2{,}2\ kW$
- Nominal speed $n = 1425\ ot/min$
- Power factor $cos\varphi = 0.80$
- Efficiency $\eta = 80\ \%$

For measurement of the rotor position was used incremental encoder ERN 420. This encoder is mounted directly on the back of the motor shaft with screws and spacers. The encoder is equipped with cover against external or accidental damage. The parametrs of this encoder are:

- Power supply 5V ±10%
- Incremental signals TTL

- Line counts 2048
- Electrical connection Cable
- Mech. permissible speed n $\leq 6000\ min^{-1}$

6 Support board to connect starter kit and peripherals

The support board was made to connect and communication between starter kit and next peripherals. The block scheme of support board is in the Fig 3.

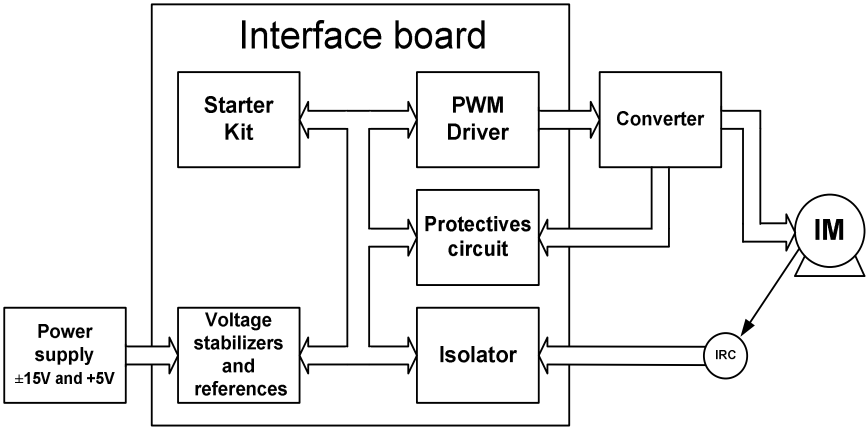


Figure 3: Block diagram for interface board

The PWM control is done with bus driver (7407) and then it is applied to 15-pins connector. The outputs from current and voltage sensors are applied to input connector of support board. Signals from incremental encoder are connect to support board with 9-pins connector. The supply of support board is done with external power supply (the value of voltage are ±15V and 5V).

7 Laboratory model

There is picture with all parts of laboratory model asynchronous motor drive. In Fig. 4 you can see indirect frequency converter, Texas Instruments signal processor system TMS320F28335, and asynchronous motor (produced by Cantoni) and incremental sensor.

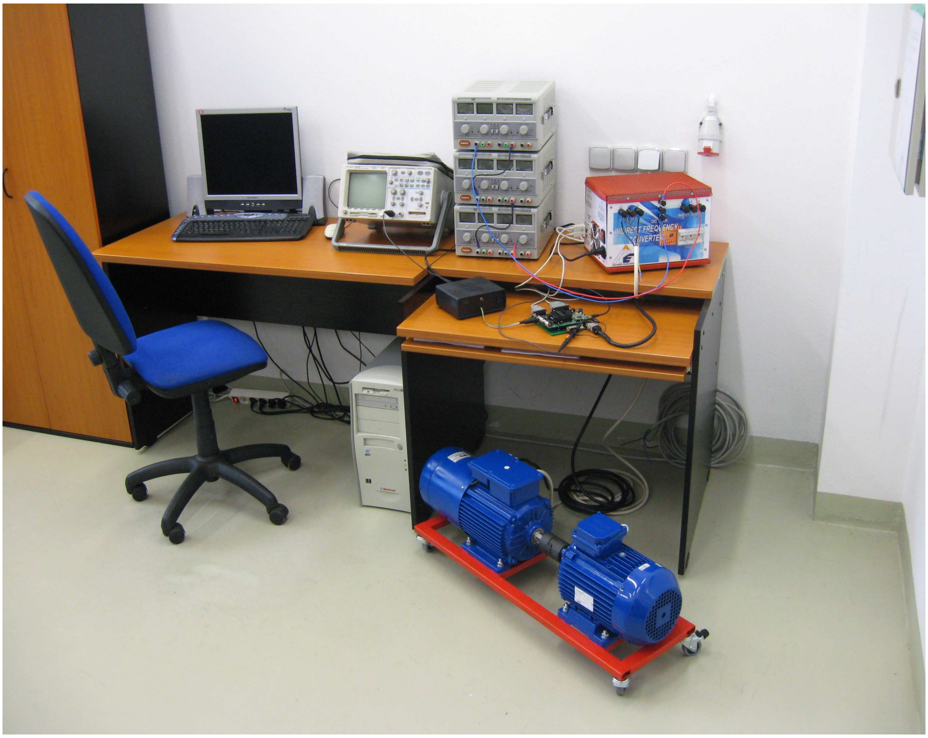


Figure 4: Laboratory model with asynchronous motor drive

7 Conclusion

This article describes the main hardware of experimental workplace for teaching modern algorithm controlled AC drives. This experimental workplace focuses mainly on asynchronous motor. Each chapter deals briefly with some hardware necessary for development of new control algorithms.

8 Acknowledgment

In the paper there are the results of the project G504 which was supported by The Univesity Development Fund (FRVŠ).

References

1. Brandštetter, P.: *"A.C. Control Drivers – Modern Control Method"*, VŠB Technical University of Ostrava, 1999. ISBN 80-7078-668

2. Neborák, I.: *“Střídavé regulační pohony – Moderní způsoby řízení”*, VŠB Technical University of Ostrava, 2002.
3. Getting Started with F28335 eZdsp [online]. Available from WWW: <http://focus.ti.com/lit/ds/symlink/tms320f28335.pdf>
4. TMS320F28335 eZdsp Starter Kit [online]. Available from WWW: <http://focus.ti.com/dsp/docs/thirdparty/catalog/devtoolsproductfolder.tsp?actionPerformed=productFolder&productId=2600#features>
5. TMS320F28335, Digital Signal Controller [online]. Available from WWW: < http://focus.ti.com/docs/prod/folders/print/tms320f28335.html>

Simulation of Sensorless Vector Control of Induction Motor Using HIL Method

Marek Dobrovsky

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
marek.dobrovsky@vsb.cz

Abstract. This work deals with simulation of sensorless vector control of induction motor using HIL method in Matlab-Simulink. The text in the first part deals with the description of vector control with the structure MRAS. The second part describes HIL method and then multifunction card MF 624. In the last section of this document are shown simulation results for changes engine speed from 0 to 500 rpm, respectively, from 500 to -500 rpm.

Keywords: Vector control, induction motor magnetizing flux, torque.

1 Observers using systems with reference model

MRAS observer structure is shown in Figure 1. In this system, induction motor state variables evaluated in the reference model based on the measured variables (stator voltages, stator currents). The reference model is independent of the speed and uses the voltage machine model. Adaptive model uses the current model of the machine and the mechanical angular speed of the machine is one of the input variables of this model. The difference between state variables are adaptive signal (AS) *Φ (e)*, which is evaluated and minimized most often by the PI regulator in the block adaptation mechanism, which performs the estimate value of the mechanical angular velocity ω_m and adapts adaptive model until the desired behavior. With feedback, the observer is able to limit the impact of changes in machine parameters on the accuracy of the calculation.

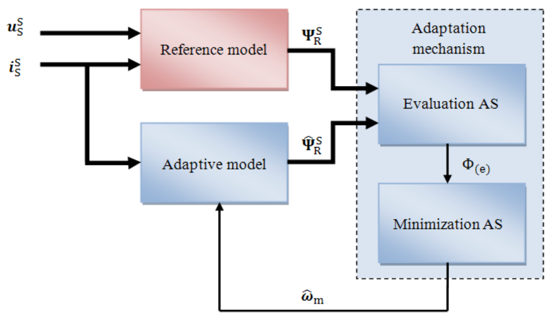


Fig. 1. MRAS - basic scheme for the estimation of mechanical angular velocity.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 130–135.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

1.1 Reference Frame MRAS

Reference frame MRAS (RF-MRAS) is the simplest variant of observers working on the principle of MRAS. For the estimation are used equations, which determining the rotor flux of machine.

1.2 Mathematical description RF-MRAS

The reference model of MRAS method is based on the application of voltage model rotor flux, which is in the stator coordinate system described by the following equation:

$$\boldsymbol{\Psi}_R^S = \frac{L_r}{L_m}\left[\int(\boldsymbol{u}_S^S - R_S\boldsymbol{i}_S^S)dt - \frac{L_S L_r - L_m^2}{L_r}\boldsymbol{i}_S^S\right] \quad \textbf{(1)}$$

This model requires, that equation (1) expand the component shape:

$$\Psi_{R\alpha} = \frac{L_r}{L_m}\left[\int(u_{S\alpha} - R_S i_{S\alpha})dt - \frac{L_S L_r - L_m^2}{L_r} i_{S\alpha}\right] \quad \textbf{(2)}$$

$$\Psi_{R\beta} = \frac{L_r}{L_m}\left[\int\left(u_{S\beta} - R_S i_{S\beta}\right)dt - \frac{L_S L_r - L_m^2}{L_r} i_{S\beta}\right] \quad \textbf{(3)}$$

Adaptive model of MRAS method is based on the application of voltage model rotor flux, which is in the stator coordinate system described by the following equation (4), depending on the mechanical angular velocity.

$$\widehat{\boldsymbol{\Psi}}_R^S = \int\left[\left(\boldsymbol{j}\widehat{\omega}_m - \frac{1}{T_R}\right)\widehat{\boldsymbol{\Psi}}_R^S + \frac{1}{T_R}L_m\boldsymbol{i}_S^S\right]dt \quad \textbf{(4)}$$

Equation (4) can be written in component form:

$$\widehat{\Psi}_{R\alpha} = \int\left[-\frac{1}{T_R}\widehat{\Psi}_{R\alpha} - \widehat{\omega}_m\widehat{\Psi}_{R\beta} + \frac{1}{T_R}L_m i_{S\alpha}\right]dt \quad \textbf{(5)}$$

$$\widehat{\Psi}_{R\beta} = \int\left[-\frac{1}{T_R}\widehat{\Psi}_{R\beta} + \widehat{\omega}_m\widehat{\Psi}_{R\alpha} + \frac{1}{T_R}L_m i_{S\beta}\right]dt \quad \textbf{(6)}$$

The error signal entering the controller corresponding to the deviation of rotor fluxs and is described by equation (7), after processing by PI controller, we get an estimate angular speed $\widehat{\omega}_m$ on output adaptation algorithm - equation (8).

$$\Phi(e) = \widehat{\boldsymbol{\Psi}}_R^S \text{ x } \boldsymbol{\Psi}_R^S = \widehat{\Psi}_{R\alpha}\Psi_{R\beta} - \widehat{\Psi}_{R\beta}\Psi_{R\alpha} \quad \textbf{(7)}$$

$$\widehat{\omega}_m = K_1\Phi(e) + K_2\int_0^t \Phi(e)dt \quad \textbf{(8)}$$

Equation (7) corresponds to the opening angle of both vectors. The sign of the error signal *Φ (e)* then determines the type of request to change speed. Positive error *Φ (e)*> 0 requires an increase in the estimated speed and negative error *Φ (e)* <0 requires a reduction.

2 Description of method HIL

Hardware in the Loop (HIL) simulation is a technique that is used in the development and testing of complex real-time systems. It is a tool that connects the hardware (controller) with a mathematical model (managed system) in a closed feedback loop. To simulate this method a real control system and a mathematical model is required. The mathematical model is used to configure the control system. The control system generates an actuator variable dependent on the control deviation, which is the difference between desired and actual quantity. The Control variable enters the model, then output variable (actual value) gets off from the mathematical model, and than gets back in to the control system. This simulation works in real time. Results of this simulation approaches to reality, because the control system with sensors and actuators is implemented as close as it would be in real. To the fact that the results would match the reality, we need a sufficiently accurate mathematical model, on which simulation accuracy dependents most. As soon as the control system is set up according to the requirements, a mathematical model can be replaced by the real system and the results can be compared.

Figure 2 is a general diagram of the simulation using method hardware in the loop. The meaning of values in this picture: y (t) ... process value, w (t) ... desired quantity, e (t) the..control. error, u (t) ... control variable v (t) ... fault value, reg ... regulator.

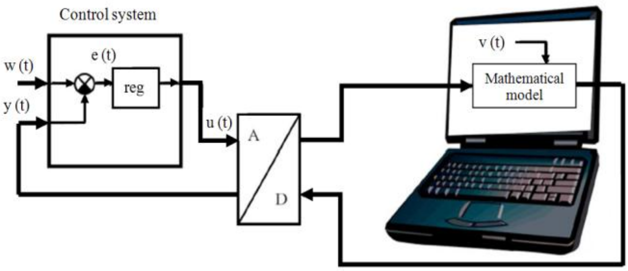


Fig. 2. Diagram of Hardware in the Loop Simulation.

3 Description of multi-card MF 624

MF 624 multifunction I/O card is designed for the needs of connection a PC compatible computer to the real signals from outside the world. MF 624 contains a fast 8-channel 14-bit A/D converter with simultaneous sampling of multiple channels and function of storage measured values from the A/D conversion, 8 independent 14-bit D/A converter, 8-bit digital inputs, 8 bit digital outputs and quintuple timer/counter. The card is designed for standard data acquisition and control applications and is optimized for use with the Real Time Toolbox for Simulink. MF 624 offers full 32-bit architecture for fast data processing.Multifunction I/O Card MF 624 has no switches or jumpers, and can be installed in any available PCI slot. Figure 3 shows the Multifunctional I/O card MF 624.

Fig. 3. Multifunction I/O Card MF 624.

4 Results of simulation sensorless vector control of AM using HIL method

This section shows simulation results of vector control, which were recorded by program for setting control variables in LabVIEW environment. Control variable magnetizing current is set to 4A and the desired speed, first at 500 rpm and then changed to -500 rpm.

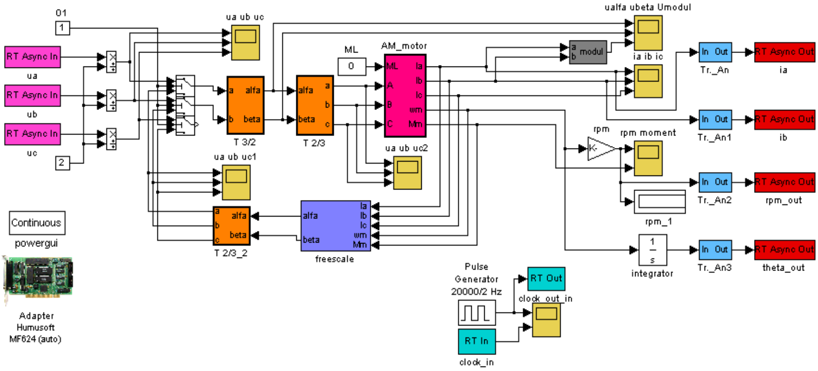


Fig. 4. AM in the simulated diagram, Matlab-Simulink.

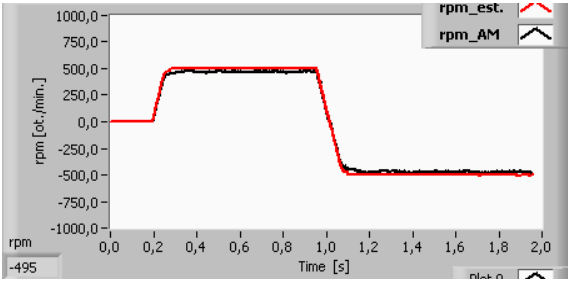


Fig.5. Change the engine speed [rpm] from 0rpm → 500rpm → -500rpm, the actual rpm (black points), estimate rpm (red points).

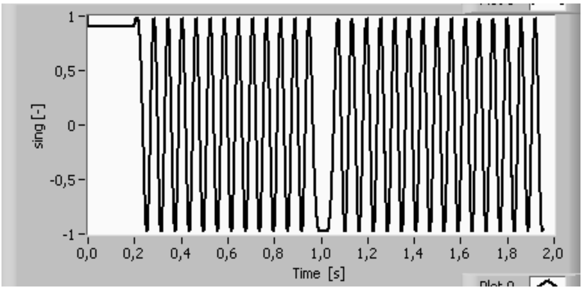


Fig.6. Sinγ [-].

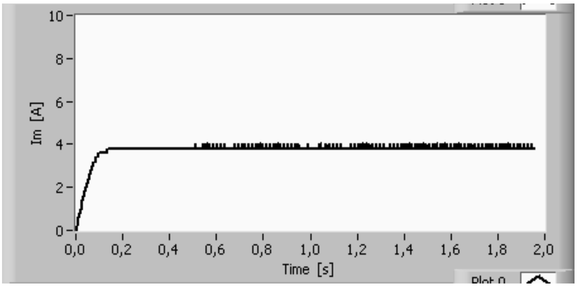


Fig.7. Magnetization current $i_m[A]$.

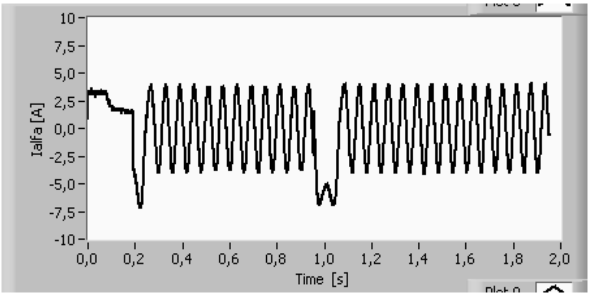


Fig.8. Stator current $i_\alpha[A]$.

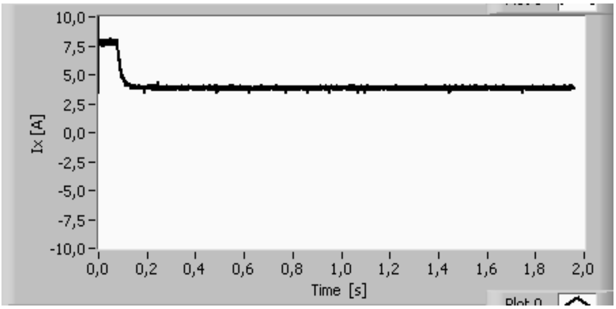


Fig.9. Magnetizing component of stator current $i_x[A]$.

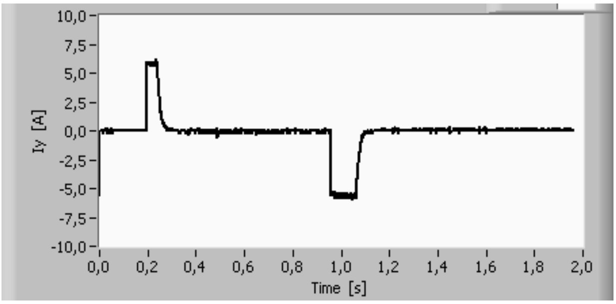


Fig.10. Torque component of stator current $i_y[A]$.

Acknowledgement

In the paper there are the results of the project SP2011/107 which was supported by Student Grant Competition of VSB-Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

References

1. Brandštetter, P.: Střídavé regulační pohony – Moderní způsoby řízení, Ostrava: VŠB – TU Ostrava, 1999. ISBN 80-7078-668-X
2. Neborák, I.: Modelování a simulace elektrických regulovaných pohonů, Ostrava: VŠB-TU Ostrava , 2002.

Magnetic Properties of Types of Core for Toroidal Transformers Made from Thin Low-Loss Sheets and Thermal Influence on Resulting Transformer Idle Current

Aleš Folvarčný

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ales.folvarcny@vsb.cz

Abstract: The paper presents the results of independent experimental analysis aimed at determining the magnetic properties of the selected type of toroidal transformer cores made from thin silicone alloyed sheets with a low loss number used in our case in the manufacture of toroidal transformers for higher class audio power amplifiers. Another part of the paper deals with the results of the experimental analysis aimed at determining the effect of temperature on size and type parameters of the current loading of small toroidal and conventional transformers. The aim of this measure was to assess the effect of temperature and voltage values at no load current of the selected type of toroidal and conventional transformers put in a climatic chamber. Furthermore, our attention was drawn to the observations of the effect of temperature and magnetic saturation on voltage transformer cores and to approximate formulation of magnetic properties of metal core type transformers in the form of BH hysteresis characteristics.

The samples were analysed and the waveforms U_{10}, I_{10}, U_{20} were measured and recorded. The measured values were then used for calculation of additional quantities of toroidal transformers (magnetic flux density, magnetic field intensity) that specify the magnetic properties of the measured samples. Finally, the measured results of both the toroidal and conventional transformers were compared.

The temporal analyses of saturation of cores are the outcome of extensive and time-consuming measurements and registration of signals in no-load state for various levels of supply voltage. The measurements were based on common knowledge of the properties of soft ferromagnetic materials.

Keywords: Toroid transformer, idly current, BH characteristics, climatic chamber, temperature.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 136–141.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

1 Introduction

The aim of this exploratory analysis was to assess the effect of temperature and voltage values of the selected type of toroidal and conventional transformers under no load condition, i.e. for no-load current. The influences of temperature and voltage values of magnetic toroidal transformers were then described and theoretical findings were verified.

The samples were analysed and waveforms U_{10}, I_{10}, U_{20} were measured and recorded. The measured values were then calculated for other quantities of the toroidal transformer characterizing the magnetic properties of measured samples (magnetic flux density, magnetic field intensity).

The final part deals with the implementation of the comparison of the measured results for the toroidal and current transformers.

During actual measurements, the knowledge of ferromagnetic properties of soft magnetic materials used for production of transformer cores was employed.

2 General and brief notes to the given issue

An idle transformer is characterized best by its idle current. The size and shape of the current describes substantially the resulting properties of design, structural and magnetic characteristics of the transformer core material. In the known geometric dimensions and parameters of the transformer windings, the timing of this current together with the induced voltage signal can then be used for the analysis of magnetic parameters and ratios of magnetic core transformer. Such an analysis can then be used both for determination of the magnetic properties of the metal core and for the control analysis of the resulting saturation and losses as a result of the design and construction of the transformer.

3 Experimental part

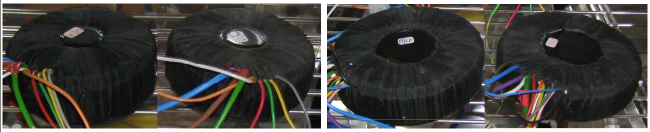
In all, 4 toroid transformer samples supplied by Talema s.r.o. were used for measurements.

These transformers were intended for power audio-amplifiers in higher classes.

These transformers were placed in climatic chamber Vötsch 4018 for measurement.

Following samples were used for measurement:

- 92 902 - P1S4 CD 5x, supply voltage 230V (small transformer)
- 92 166 - P2 S6 DGR 1053/3, supply voltage 230V (middle-size transformer).



a) b)

a) TR1A/B small, b) TR2A/B middle-size.

Fig. 1. Example of individual samples

The cores were made of the same material and with the same production technology. But they had different final technological finish.
The core transformer 1A was subject to a thermal finish and varnished, the other cores are not treated for final production.

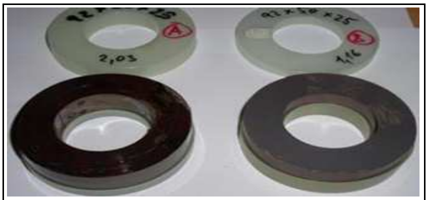
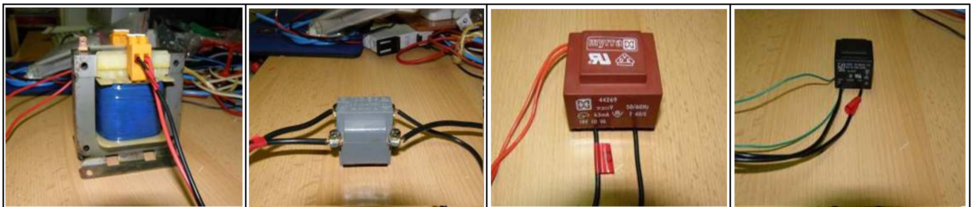


Fig. 2. Example of measured core samples 1A/1B

Both of these cores were winding of electrical sheet made by ORB Steel Cogent, type M111 with a thickness of 0.30 mm, max. loss of 1.11 W/kg.

The following samples were used for further measurement:



a) b) c) d)

Fig. 3. Example of individual samples

a) TR1-230/24V;160VA, b) TR2-230/24V;22VA, c) TR3-230/18V;10VA, d) TR4-230/15V;3VA

This is the standard employment of transformers for various applications. The objective of these measurements on the transformers is, in particular, from the current value of nothing. Subsequently, these values are compared with the results measured on toroidal transformers designed for power audio amplifiers of a higher class.

Design and preparation of the workplace

To analyse the effect of temperature on the idle current of selected transformer types, the measuring workplace has been prepared; refer to the block diagram in Fig.4

The transformers put in climatic chamber had primary and secondary windings brought out to the terminal placed outside the climatic chamber. The individual windings are successively connected to the measuring site, where the required quantities were registered.

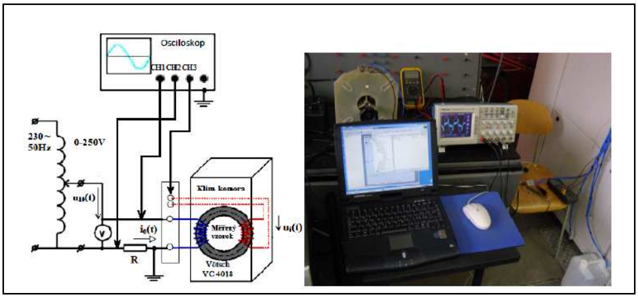


Fig. 4. Example of workplace

- Measuring resistor had a value of 1Ω.

The power for analysed samples was supplied by an autotransformer and the value of voltage for each type of transformer was adjusted in the range from 50 to 250 V. The individual toroidal transformers were put into the Vötsch VC 4018 climatic chamber.

The climatic chamber was always set to the required temperature before start of each measurement. The temperature set points for the climatic chamber: -40˚C, -35˚C, -30˚C, -20˚C, -10˚C, 0˚C, 10˚C, 20˚C, 30˚C, 40˚C, 50˚C, 60˚C, 70˚C, 80˚C, 90˚C, 100˚C and 120˚C.

After reaching the desired temperature in the chamber, the measured transformers were left at standstill for half an hour so as to reach the temperature stability of the measured transformers and all of their volumes.

The records of the idle current signal and the induced primary voltage were performed using a Tektronix digital oscilloscope with GPIB communication interface and PC.

WaweStar v.2.4 was used to convert the readings into a text format and these were then imported into Microsoft Excel to be processed further.

4 Measurement results

Table 1. Samples of resulting U_{10}, I_{10}, U_{20}

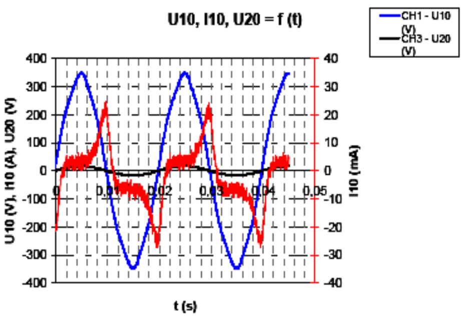
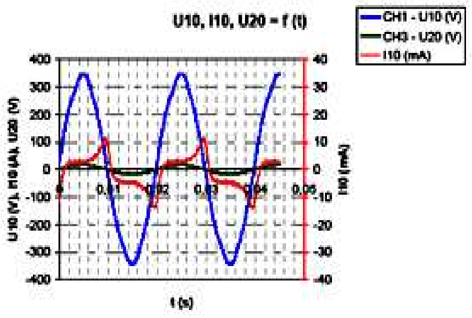
	
Samples of resulting U_{10}, I_{10}, U_{20} waveforms for measurements on toroid transformer, load -40ºC, TR1A	Samples of resulting waveforms U_{10}, I_{10}, U_{20} for measurements on toroid transformer at 120ºC, TR1A

Table 2. Samples of resulting U_{10}, I_{10}, U_{20}

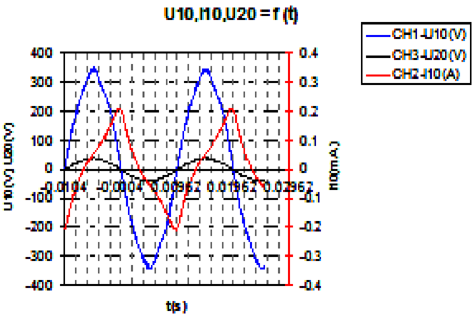
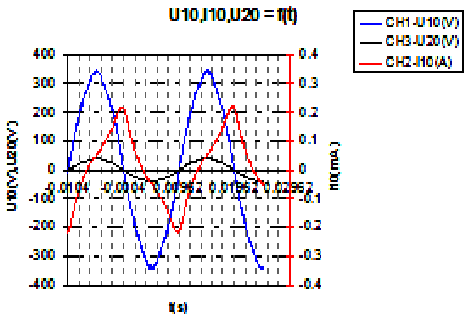
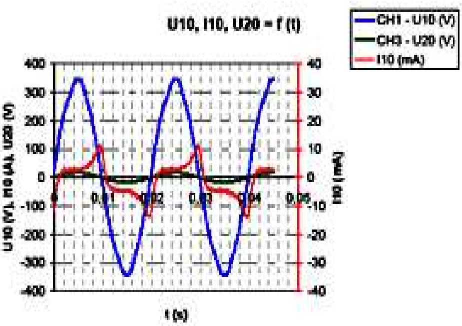
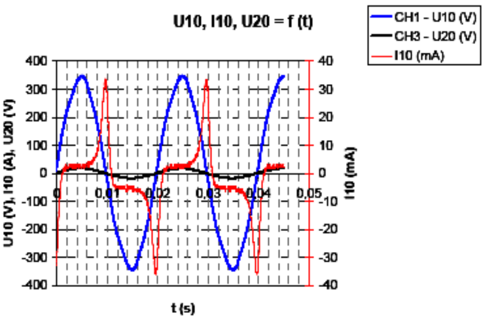
	
Samples of resulting U_{10}, I_{10}, U_{20} waveforms for measurements on conventional transformer at-40ºC, TR1	Samples of resulting U_{10}, I_{10}, U_{20} waveforms for measurements on conventional transformer at 100ºC, TR1

Table 3. Samples of resulting U_{10}, I_{10}, U_{20}

	
Unvarnished toroid core at 120ºC,TR1A	Varnished toroid core at 120ºC, TR1B

5 Conclusion

From the measured results of the analysed samples results that the current load value in the range of set temperature range -40ºC to 120ºC is lower temperature in growing, whilst this effect on conventional transformers was almost imperceptible.

The measured results of the analysed samples indicate that the value of idle current of toroidal transformers in the range of set temperatures -40ºC to 120ºC decreased whereas this effect was almost imperceptible in conventional transformers.

The magnetization of toroidal transformers made from very thin metal sheets with a high maximum permeability value decreased with increasing temperature, the hysteresis loop narrowed down proportionally along with the losses ensuing rom it. This phenomenon can only be explained by a more profound description of the magnetization processes and magnetostrictive effects which was omitted for the need of this paper.

Besides, it is evident from the measured results that the toroidal transformer denoted as Tr1A showed much smaller value than the sample 1B whereas both of them were made from the same material and with the same dimensions. The resulting difference in behaviour of otherwise identical transformers is given by intentionally different surface finish of the transformer cores. The magnetic core of sample 1A was unvarnished and thermally unhardened while the transformer core 1B was varnished and thermally hardened. The varnishing and hardening processes result in introduction of stress within the core and in enlargement of the transversal air gaps among individual layers of the transformer-core laminations and in deterioration of magnetic properties of all cores, making the idle current to rise.

Change of the transformer winding resistance by changing temperature has no effect on the resulting transformer idling current, because of insignificant value of the winding resistance with regard to high value of the inductive reactance.

As next it is obvious that the effect of temperature on the magnetic properties of the core is negligible with the normal quality of metal sheets. For very thin sheets of toroidal cores is the temperature effect significant due to their high permeability. At low temperatures magnetic properties of sheets turn to deteriorative, the current idling is increasing and at the same time the vibration and noise go up.

References:

1. Folvarčný A.: Typové transformátory-Vliv teploty na proud naprázdno, Diplomová práce. VŠB-TU Ostrava, 2009.
2. Smolka K.: Typové toroidní transformátory-experimentální analýza elektromagnetických poměrů, při různých provozních stavech, Diplomová práce. VŠB-TU Ostrava, 2008.
3. Draxler K., Kaspar P., Ripka P.: Magnetické prvky a měření, ČVUT Praha, 1999.

Publications:

1. Folvarčný A., Marek M.: Experimentální analýza vlivu teploty na parametry proudu naprázdno typových toroidních transformátorů, porovnání s konvenčními typy transformátorů, Scientific Journal TRILOBIT 2010, Zlín 2010.
2. Folvarčný A..: Magnetic properties of types of core for toroidal transformers made from thin low-loss sheets and thermal influence on resulting transformer idle current, Journal of Electrical engineering, Slovakia 2010. ISSN 1335-3632.

Summary of Fractional Order Operators Approximations and Simulation of Selected Methods

Jiří Friedrich

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jiri.friedrich@vsb.cz

Abstract. There are situations when technological process needs to work at low or almost zero speed. This causes problems with regulation because the measured signal changes very slowly. Because of this effect the demands on e.g. speed controller accuracy are more highly. Many different methods, how to solve this problem have been analyzed with various results. Heretofore little notice was taken of fractional calculus application. It was caused, among others, by problems with practical realization because the approximation processes brings considerable computing demands. The main aim of this paper is to compare and in time and frequency domain analyze possible models and approximations of fractional order controllers. From large number of different approaches only selected integer order continuous approximations were discussed.

1 Introduction

Standard elements of industrial control applications are proportional-integral-derivative (PID) controllers. The main benefit of using those types of controllers is the simplicity of design, because there are lots of suitable and very easy methods of design. The accuracy of this way controlled process is for major applications sufficient, however there are situations demanding better settling time respectively percentage overshot. As an example it can be used a technological process working at low or almost zero speed.

It was mathematically shown, that generalization and extending of the integer order PID to the plane is possible. To this purpose it is used the so-called theory of fraction-order calculus. The process of fractional integration and differentiation can be mathematical described in several manners. But mostly used is besides the Grünwald-Letnikov definition the so-called Riemann-Liouville definition. It is given as (1)

$$D^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau) d\tau \tag{1}$$

and for engineering praxis useable Laplace transform is shown in (2) as it is described e.g. in [3], [4].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 142–147.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

$$L\{D^{-\alpha} f(t)\} = s^{-\alpha} F(s) \tag{2}$$

In the similar way it is possible to obtain a definition of fraction-order derivative, expression in Laplace domain represents (3) [3].

$$L\{D^{\alpha} f(t)\} = s^{\alpha} F(s) - [D^{\alpha-1} f(t)]_{t=0} \tag{3}$$

If we go back to controllers, application of the mentioned theory causes following change of PID frequency response. Equation (4) represents the standard PID and equation (5) the fractional-order PID which is graphically expressed in fig.1.

$$G(s) = K_p + T_i s^{-1} + T_d s \tag{4}$$

$$G(s) = K_p + T_i s^{-\lambda} + T_d s^{\delta} \tag{5}$$

It is obvious from the frequency responses that the integer order PID requires designing of three parameters and fractional order PID five parameters. This expansion enables more flexibility to controller design, as it was desired.

On the one hand there are the mentioned benefits, but on the other hand it is necessary to mention, that this may also cause some problems. Firstly, standard design methods cannot solve this easily. Therefore it is necessary to use another one, e.g. particle swarm optimization technique. The next problem is selection of suitable FO operator approximation.

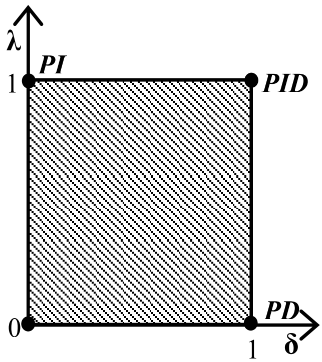


Fig. 1. Generalization of PID controller, where γ is the derivation order and δ is the integration order.

1.1. Integer-Order Continuous Models

The problem, how to obtain a realizable continuous model can be simplified to a problem of transforming the irrational transfer function which models the fractional order controller to a rational approximation. There are several possible methods however from control point of view two of them are suitable: the continued fraction expansion method and the rational approximation method used in interpolation of functions.

General CFE method

This method of approximation is based on fact that a rational approximation of fractional function G(s) = $s^{-\alpha}$, where 0< α <1, can be obtained using the CFE method of functions in following form:

$$G_h(s) = \frac{1}{(1+sT)^{\alpha}} \qquad \textbf{(6)}$$

$$G_l(s) = \left(1 + \frac{1}{s}\right)^{\alpha} \qquad \textbf{(7)}$$

Where $G_h(s)$ is the approximation for high frequencies (ωT >> 1), and approximation for low frequencies (ωT << 1) represents $G_l(s)$ [1].

Carlson's method

This method described in [2], is based on standard Newton process of iterative approximation. Using this method, it is necessary to specify the starting point of. Usually it is used following equations (8)

$$(H(s))^{1/\alpha} - (G(s)) = 0\,;\;\; H(s) = (G(s))^{\alpha} \qquad \textbf{(8)}$$

If in each iteration is α = 1/q, m = q/2, and the initial value $H_0(s) = 1$, it is possible to obtain following approximation (9):

$$H_i(s) = H_{i-1}(s)\frac{(q-m)(H_{i-1}(s))^2 + (q+m)G(s)}{(q+m)(H_{i-1}(s))^2 + (q-m)G(s)} - (G(s)) = 0 \qquad \textbf{(9)}$$

Matsuda's method

This method solves the approximation problem using mentioned CFE process and fits the original function in a set of logarithmically spaced points. Supposing that the selected points are s_k (where k = 0, 1, 2,…), we obtain a function which takes on the following form (10):

$$H(s) = a_0 + \frac{s-s_0}{a_1} + \frac{s-s_1}{a_2} + \frac{s-s_2}{a_3} + \ldots \qquad \textbf{(10)}$$

where

$$a_i = v_i(s_i),\ \ v_0(s) = H(s),\ \ v_{i+1}(s) = \frac{s - s_i}{v_i(s) - a_i} \tag{11}$$

3 Simulation results

To compare above mentioned methods of approximations a fractional order integrator of order 0.5 was simulated.

General CFE method

Performing the CFE process of the equations (6) and (7) we obtain following transfer functions:

$$H_1(s) = \frac{0.3514s^4 + 1.405s^3 + 0.8433s^2 + 0.1574s + 0.008995}{s^4 + 1.333s^3 + 0.478s^2 + 0.064s + 0.002844}$$

respectively

$$H_2(s) = \frac{s^4 + 4s^3 + 2.4s^2 + 0.448s + 0.0256}{9s^4 + 12s^3 + 4.32s^2 + 0.576s + 0.0256}$$

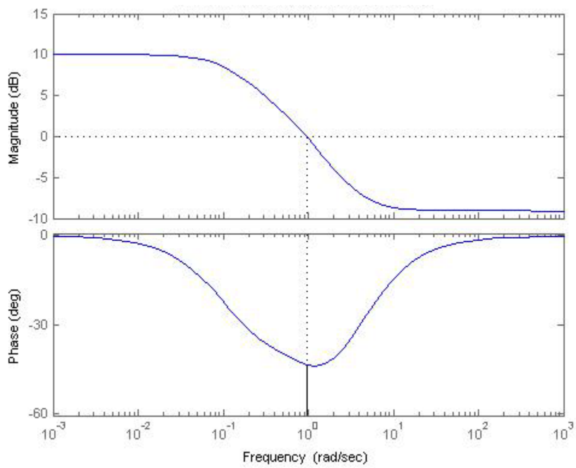


Fig. 2. Bode plot of $H_1(s)$.

Carlson's method

After, for simplicity, two iterations we obtain following transfer function:

$$H_3(s) = \frac{s^4 + 36s^3 + 126s^2 + 84s + 9}{9s^4 + 84s^3 + 126s^2 + 36s + 1}$$

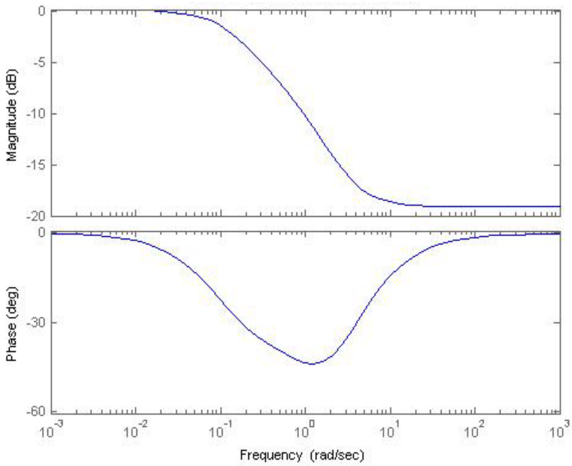


Fig. 3. Bode plot of $H_2(s)$.

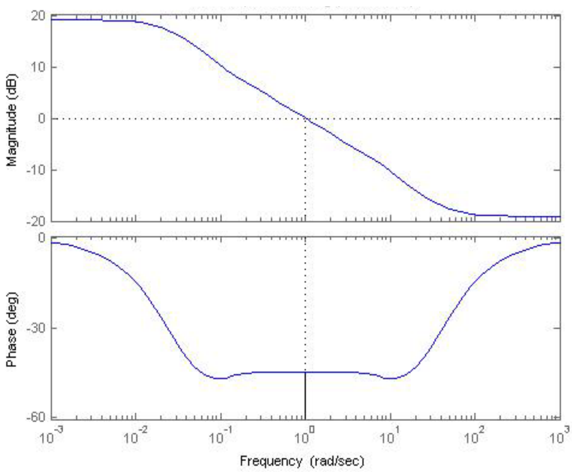


Fig. 4. Bode plot of $H_3(s)$.

Matsuda's method

If $G(s) = s^{-0.5}$, $f_{initial} = 1$ and $f_{final} = 100$, we obtain the approximated transfer function $H_4(s)$

$$H_4(s) = \frac{0.08549s^4 + 4.877s^3 + 20.84s^2 + 12.995s + 1}{s^4 + 13s^3 + 20.84s^2 + 4.876s + 0.08551}$$

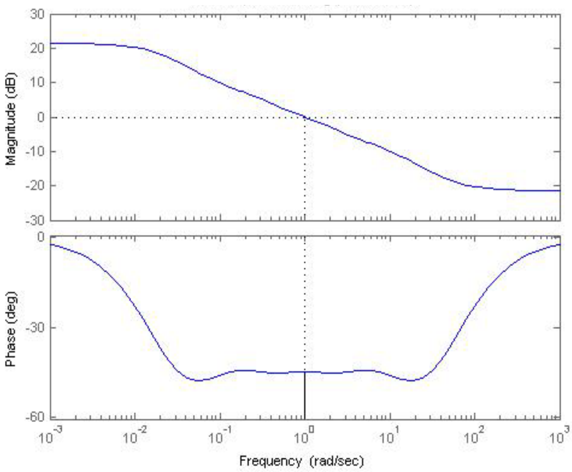


Fig. 5. Bode plot of $H_4(s)$.

Conclusions

The aim of this research was to prepare conditions for implementation fractional-order controllers in a structure of vector controlled induction motor. Simulation results fig. 2 – fig. 5 show that continuous approximations described above can be used for control process and bring in it very interesting parameters.

Acknowledgement

In the paper there are the results of the project SP2011/107 which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandštetter, CSc.

References

1. Press, W.H., Teukolsky, S.A., Vetterling, W. T., Flannery, B. P.: Numerical Recipes in C. The Art of Scientic Computing, 2nd edition, Cambridge University Press (1992).
2. Carlson, G.E., Halijak C.A.: Approximation of fractional capacitors $(1/s)^{1/n}$ by a regular Newton process. IRE Transactions on Circuit Theory, CT-11, No. 2 (1964), 210-213.
3. Podlubny, I.: Fractional-order systems and $PI^{\lambda}D^{\delta}$ controllers. IEEE Trans. On Automatic Control, vol.44, no. 1, (1999) 208-213
4. Petráš, I.: Fractional-order feedback control of a DC motor. Journal of Electrical Engineering, vol. 60, no. 3, 2009, 117-128

The Design of an Axial Flux Rotary Converter for Hybrid Electric Vehicles

Aleš Havel

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ales.havel@vsb.cz

Abstract. This paper describes an efficient axial flux arrangement of the four quadrant rotary converter for hybrid electric vehicles. Further design of the axial flux stator and double rotor machine appears from the radial air gap induction motor and permanent magnet synchronous motor, utilizing the method of constant magnetic circuit volume for dimensions conversion.

1 Introduction

The maximal efficiency of the internal combustion engine (ICE) used in conventional vehicles is due to Carnot's cycle approximately bounded above to 40%. Such efficiency can only be reached in a very small torque-speed area, which leads to the idea of full torque and speed control in hybrid electric vehicle (HEV) conceptions aimed at considerable power savings. Many of today's used HEV conceptions are based on the series-parallel hybrid arrangement, combining together the advantages of both essential arrangements [1].

The main task of here presented four quadrant axial flux rotary converter (AFRC) is to keep the ICE working in the area of its maximal efficiency during all possible driving conditions to satisfy minimal fuel consumption. In other words, the AFRC have to convert the optimal torque-speed operation point of the ICE on the input shaft to the required variable torque and speed on the output shaft, which is through the axle gearbox (AG) connected to the wheels.

The AFRC is supplied by the 3 phase voltage inverter (INV) from the DC power source (DCPS), which can be realized as a combination of two autonomous power sources – $LiFePO_4$ battery pack delivering nominal current and ultracapacitor pack for peak power supplying.

Whole system is controlled by the DSP TMS320F28335 with implemented direct torque control algorithm to satisfy sufficient computing power and dynamics of the four quadrant rotary converter (4QRC) drive system – see Fig.1. [2].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 148–153.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Axial Flux Rotary Converter

The proposed solution combines together two known principles of electric rotary machines in one unit, which is composed from synchronous and asynchronous machine having two rotors affixed to input and output shaft, one wound stator and two axial air gaps δ_1 and δ_2. Due to selected double rotor arrangement it is possible to vary both the speed and the torque between the input and output shaft, hence the transferred power. The input and output power difference is possible to supply or consume by the stator winding, so only a part of the total vehicle drive energy is transferred across the electromagnetic transfer.

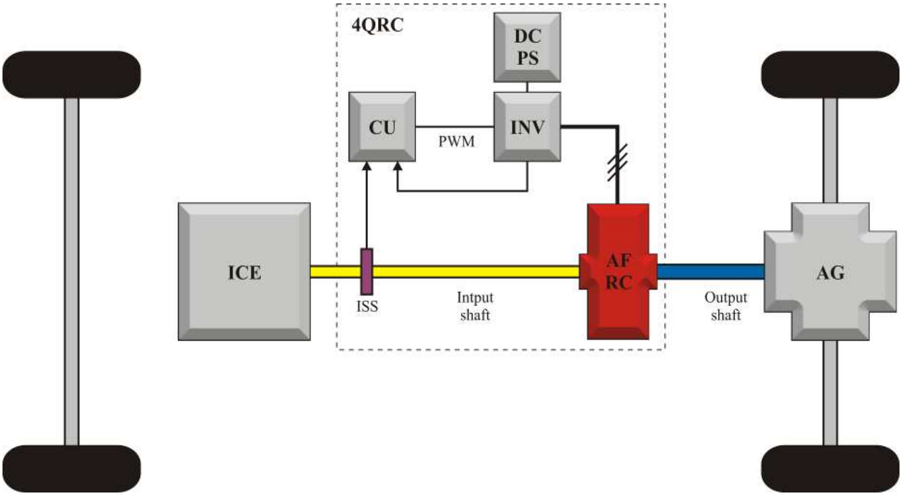


Fig. 1. Integration of the axial flux rotary converter into the HEV drive unit

The unneeded part of mechanical energy produced by the ICE or the vehicle braking energy can be transferred through the electromagnetic coupling between the outer rotor and stator and then stored into the rechargeable batteries or ultracapacitor pack. On the other hand, during the fluent change of the input and output shaft speed ratio it is possible to simultaneously compensate the peak power consumption needed for dynamic states during speed variation or vehicle loading [3].

Constructional simplicity of presented axial flux rotary converter is ensured by the disc conception with two axial air gaps and double squirrel cage inner rotor. The inner rotor can simultaneously interact with electromagnetic fields from permanent magnet outer rotor and wound stator to create two different asynchronous machines with them – see Fig. 2.

Another advantage of axial flux conception is in the possibility of air gap width adjustment, which allows the optimization of magnetic field depth penetration into the squirrel cage inner rotor. It also comes through that the AFRC can be run in pure mechanical or electrical mode when only the ICE or DCPS are producing the driving power.

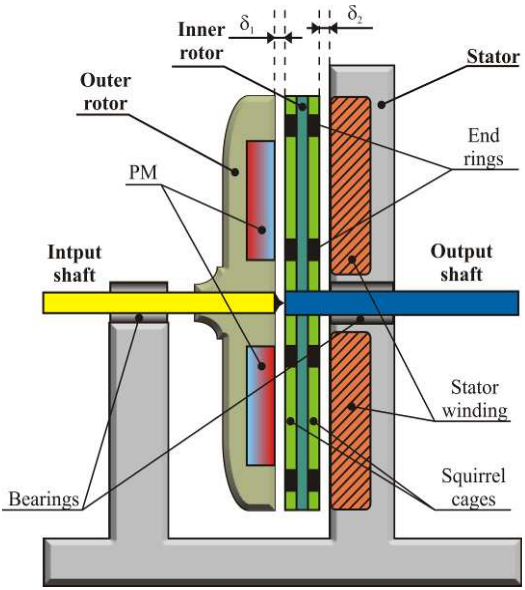


Fig. 2. The AFRC machine in axial cross section

3 Stator and Cage Rotor Design

The design of the axial flux stator and rotor is based on the design of standard world-wide used cylindrical conception of induction or synchronous motors. The resulting parameters are then recounted to the axial conception, more precisely to the inner and outer diameters and thicknesses of the axial stator and rotor utilizing the method of constant magnetic circuit volume.

The proportioning of the standard cylindrical motor usually starts with the determination of the machine diameters like outer stator diameter D_e, inner stator diameter D and the ideal air gap length l_i. The diameters D and l_i depend on the output motor power P_i, angular speed ω_S and electromagnetic loads A and B_δ. In the first stage of calculation, all the variables from (1) are unknown, except the synchronous angular speed, which is given by the optimal torque-speed operation point of the ICE, so the calculation is coming out from the recommended electromagnetic loads, coefficients $\alpha_{\delta,}$ k_B, k_V and approximately determined electromagnetic power [4].

$$\frac{D^2 l_i \omega_S}{P_i} = \frac{2}{\pi \alpha_\delta k_B k_V A B_\delta} \tag{1}$$

While the induction motor diameters are inversely proportional on the product of linear current density A and electromagnetic induction B_δ as mentioned before in (1), electromagnetic properties are affected by their mutual ratio. To calculate the ideal air gap length l_i, the electromagnetic induction in the air gap (Fig. 3) and the linear current density (Fig. 4) have to be empirically predetermined from the graphs below.

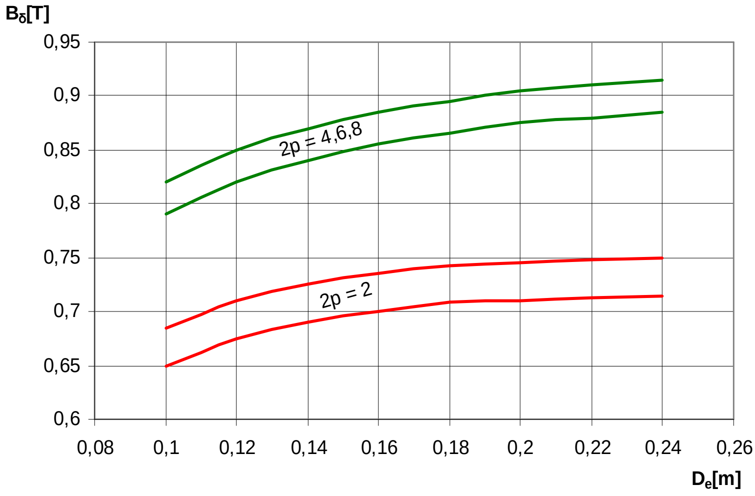


Fig. 3. The areas of optimal electromagnetic induction in the air gap of induction motor in dependency of the outer stator diameter for various pole numbers

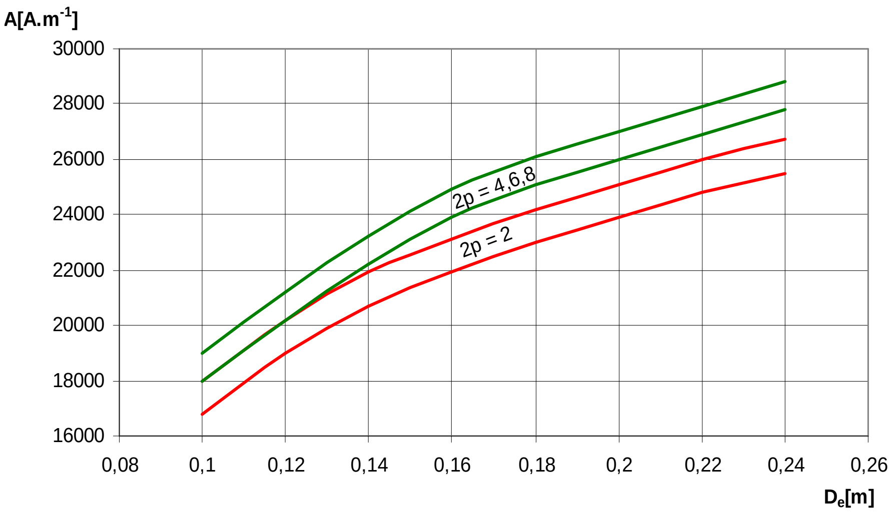


Fig. 4. The areas of optimal linear current densities in dependency of the outer stator diameter for various pole numbers

Last very important parameter, which is necessary for the ideal air gap length calculation is the inner motor power P_i. The coefficient k_E from (2) is the induced/rated voltage ratio. The optimal efficiency and power factor areas are possible to determine empirically from [4].

$$P_i = P_2 \frac{k_E}{\eta \cos\varphi} \quad \textbf{(2)}$$

Since all the main electromechanical parameters are calculated or empirically predetermined, the ideal air gap length l_i can be simply expressed from (1).

$$l_i = \frac{P_i}{k_B D^2 \omega_s k_{v1} A B_\delta} \tag{3}$$

4 Conversion to the Axial Conception

As mentioned above, the cylindrical stator and rotor dimensions including the ideal air gap length have to be converted into the axial flux conception using the method of the constant magnetic circuit volume. It also appears clearly that the inner and outer disc diameters should be the same for disc stator and disc rotor. The stator and rotor volume equations (4) and (5) are presented below.

$$\frac{\pi l_i}{4}\left[D_e^2 - D^2\right] = \frac{\pi h_S}{4}\left[D_{OUT}^2 - D_{IN}^2\right] \tag{4}$$

$$\frac{\pi l_i}{4}\left[D_2^2 - D_h^2\right] = \frac{\pi h_R}{4}\left[D_{OUT}^2 - D_{IN}^2\right] \tag{5}$$

Where are: h_S and h_R the disc stator/rotor thicknesses;
D_{IN} and D_{OUT} inner /outer disc diameters;

By the comparison of the equations (4) and (5) is possible to express the stator/rotor disc thicknesses ratio (6) and the outer disc diameter dependency (7), which finally leads to all main dimensions of the axial flux stator and rotor.

$$\frac{h_S}{h_R} = \frac{D_e^2 - D^2}{D_2^2 - D_i^2} \tag{6}$$

$$D_{OUT} = \sqrt{\frac{l_i (D_e^2 - D^2)}{h_S} + D_{IN}^2} \tag{7}$$

The right choice of the air gap thicknesses δ_1 and δ_2 has crucial influence on the energetic parameters of the AFRC drive unit. If the size of the air gap is reduced, also the magnetizing current reduces, which leads to better power factor and lower stator losses. However, making the air gap too thin may cause increasing of the amplitude of magnetic induction pulses in the air gap, hence increasing the surface and pulse losses [4].

Therefore the optimal air gap thickness has to be evaluated as a result of the 3D finite element method simulations.

5 Conclusion

The four quadrant AFRC drive unit may become a key component in hybrid and electric vehicles due to constructional simplicity and efficient realization of continuously variable electromagnetic and mechanic transmission. To utilize this technology in practical applications, the 3D finite element analysis will be crucial for better understanding of electromagnetic conditions such as electromagnetic flux and field behaviours in drive structures involving the axial flux based machines.

6 Acknowledgement

The research described in this paper was supervised by Prof. Ing. Petr Chlebis, CSc. and Ing. Petr Vaculik, Ph.D. on Department of Electronics of VSB - TU Ostrava and supported by SGS project SP2011/103: Research of accumulation subsystem structures.

References

1. Larminie J., Lowry, J. Electric Vehicle Technology Explained. John Wiley & Sons, Ltd, 2003, England, ISBN 0-470-85163-5
2. Brandstetter P., Chlebis P., Palacky P.: Direct Torque Control of Induction Motor with Direct Calculation of Voltage Vector. Advances in Electrical and Computer Engineering, vol. 10, n. 4, pp. 17-22, 2010, ISSN 1582-7445
3. Neborak, I., Skotnica, P.: Electric Car with Induction Motor Supplied by Voltage Inverter. University of Technology, Department of Electrical Power Engineering, Brno 2010, Ed. Jiri Drapela, Jan Machacek, vol. 11 th, ISBN 978-80-214-4094-4
4. Kopylov I. P. and collective: Construction of electric machines. First release, Praha: SNTL, 1988, 688 p.

2D Simulation of Electromagnetic Field Distribution in the Type Electron Microscope

Regina Holčáková and Martin Marek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
regina.holcakova@vsb.cz, martin.marek@vsb.cz

Abstract. The paper deals with 2D computer simulation of electromagnetic field distribution in the real type electron microscope with respect to real magnetic properties of steel and other ferromagnetic materials obtained from measurements. The computations were solved by finite element method using ANSYS software. This project and paper were created by financial support of state budget through the Ministry of Industry and Trade MPO-CR, project n. FR-TI1/334.

Keywords: electromagnetic field, lens, objective, electron microscope

1 Introduction

The electron microscopes rank among the versatile instruments used in many fields, such as biology, virology, electron tomography, in area of material science: metallography, materials qualification, materials preparation, nanometrology, in the area of electrical engineering: semiconductor device fabrication (chips), circuit edit, defect analysis and more other applications. In accordance with the purpose has been developed various types of electron microscopes. At present, the electron microscopes can be divided into basic types: TEM - Transmission Electron Microscope, SEM - Scanning Electron Microscope, and other special types.

The basic principle is similar to optical focusing. The electron beam is focused by a system of magnetic lenses - objectives. There are many types and versions of such lenses and objectives. These elements shape the electron beam and focus it in the appropriate position of the observed sample or specimen. The required position of the electron beam is given by accurately setting electromagnetic field in electron path. The accurate determination of distribution of electrostatic or magnetic fields in the objective lens area using a mere analytical computation is very complicated and sometimes practically impossible. However, with good knowledge of the actual magnetic field distribution of the lens many potential problems and possible defects in the lenses can be discovered. An available method how to determine the actual magnetic field distribution of the lens exactly with realistic consideration of magnetic properties of the used steel is presented in this paper.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 154–159.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 2D simulation of electromagnetic field distribution in the type electron microscope - solution

The knowledge of actual magnetic properties of ferromagnetic materials and the ability of FEM solvers using iterative methods to achieve convergence of the final solution is the basic element for the precise solution and determination of magnetic field distribution. In general terms, it is one of the most challenging computational procedures, the nonlinear electromagnetic analysis, based on the development of the Biot-Savart law and subsequent convergence of magnetic flux according to Maxwell's equations.. These advanced methods incorporated in FEM software along with high-quality hardware in the form of high-performance computing work stations make it possible to solve large and geometrically complicated 2D and 3D models.. The cross section of the solved electron microscope is shown in Fig.1.

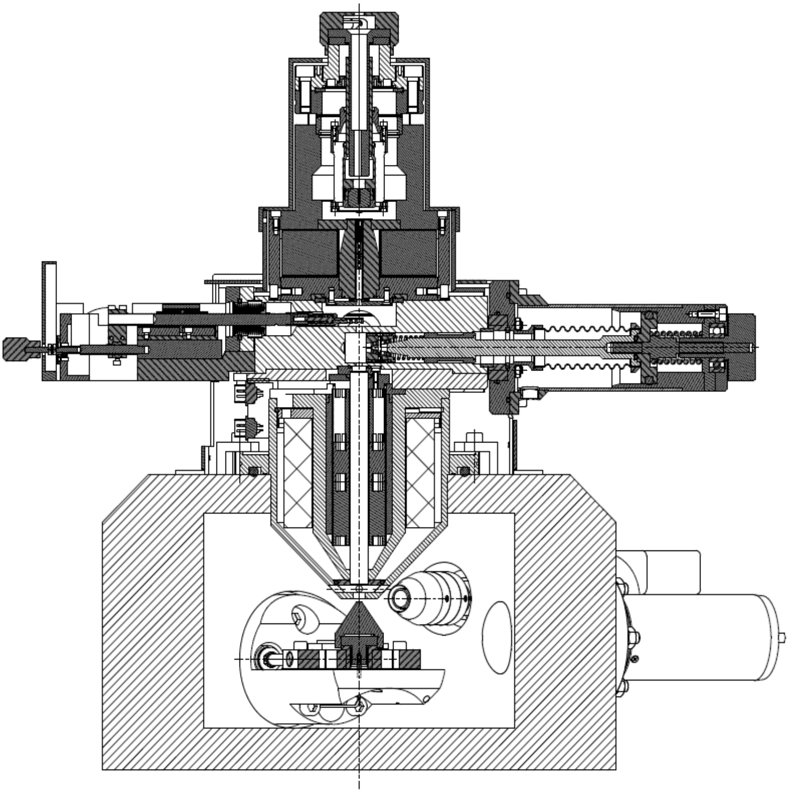


Fig. 1. Cross section of the solved electron microscope

The electromagnetic field distribution in the electron microscope was computed for 2 lenses: condenser lens and objective lens, for excitation currents of the coil corresponding to the magnetomotive force in the range NI=50-1000 Ampere-turns (condenser lens) and 100-2000 Ampere-turns (objective lens) . The magnetic properties of the used ferromagnetic materials were represented by the BH magnetization characteristics that were measured on toroidal samples.

The following Fig.2 illustrates the prepared FEM - 2D model of the electron microscope for computing of electromagnetic field distribution for various values of excitation current.

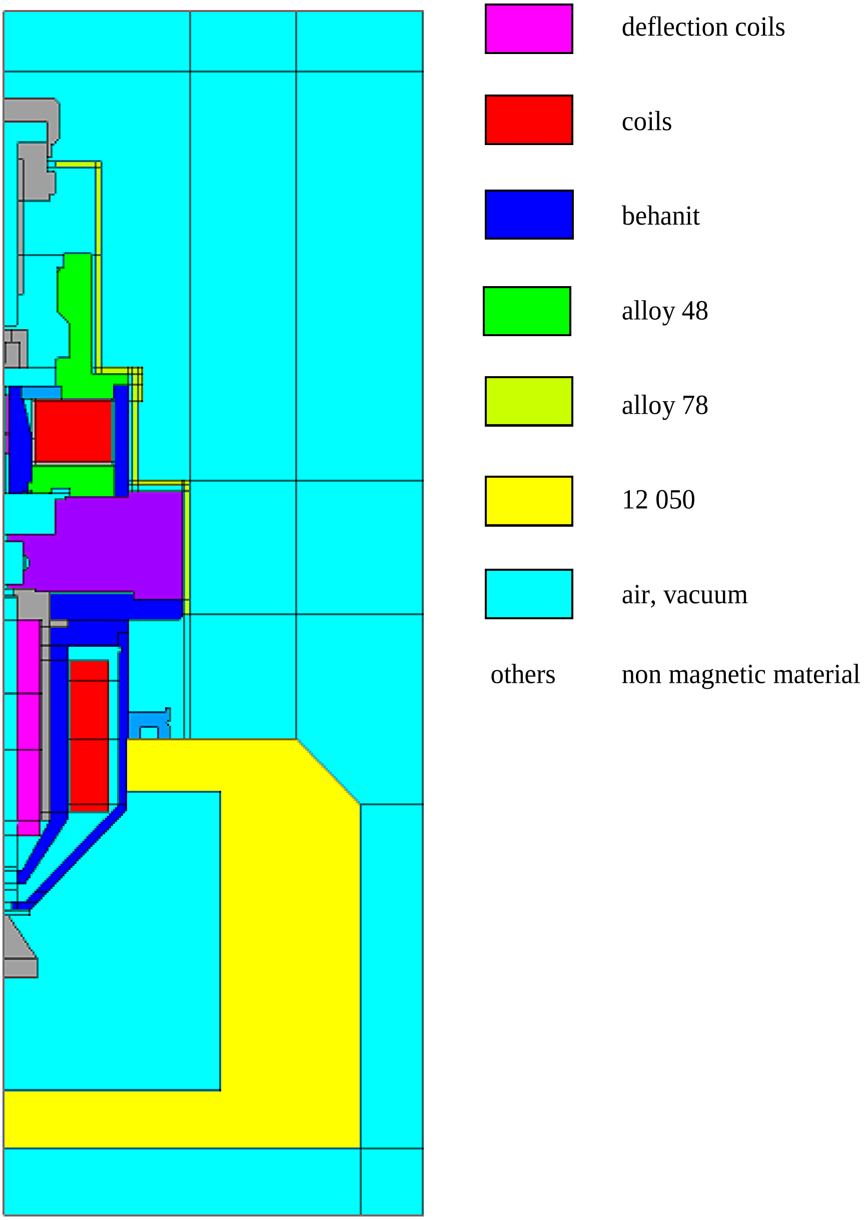


Fig. 2. 2D model of the electron microscope (ANSYS)

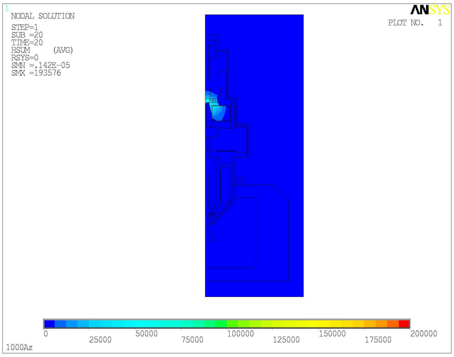
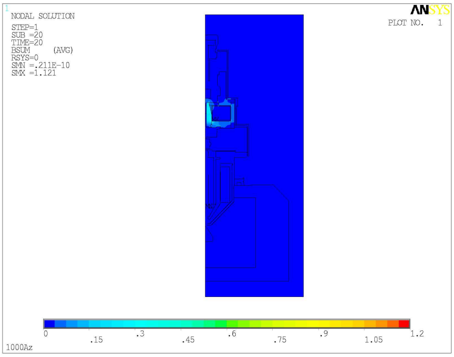
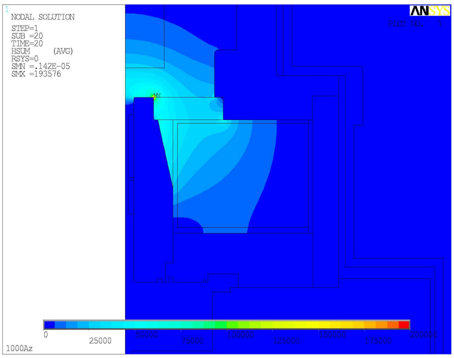
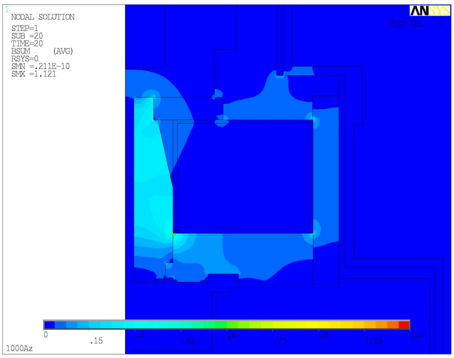
3 2D simulation of electromagnetic field distribution in the type electron microscope - results

Results of computer simulations describe the actual electromagnetic field distribution in the type electron microscope with high accuracy. Another objective of these simulations is to find the critical points in the geometry, which could cause a defects of the lens.

Tab.1, Tab.2 and Tab.3 illustrate the results of computed electromagnetic field distribution in the type electron microscope for one step of the excitation current corresponding to NI=1000 Ampere-turns (condenser lens) and NI=2000 Ampere-turns (objective lens). The results are indicated in various views of the solved 2D model of the electron microscope for.

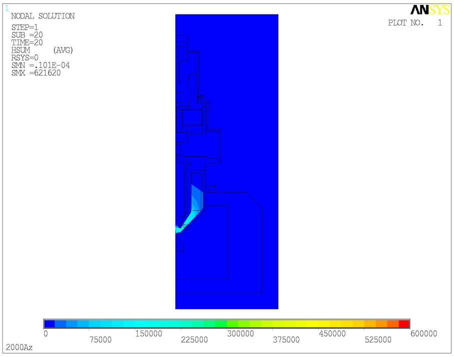
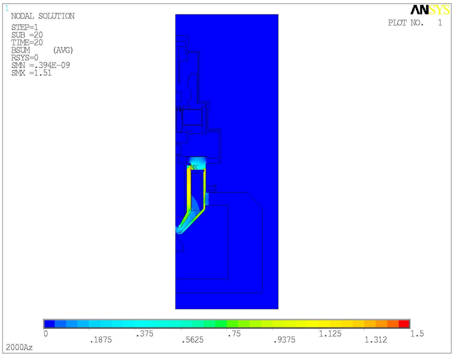
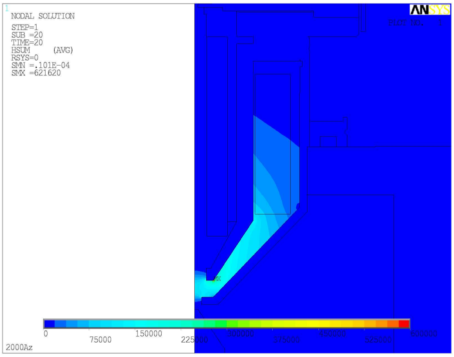
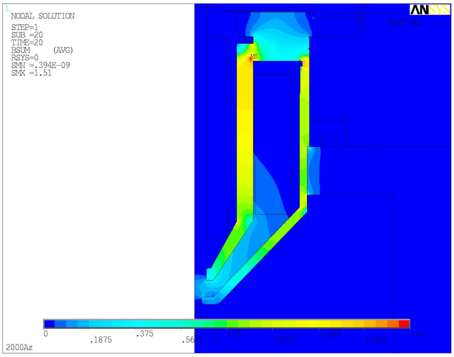
Tab. 1. Condenser lens

Magnetic intensity *Hsum* [A/m] and Flux density *Bsum* [T] for *NI*= 1000 Az

H **[A/m] - for** ***NI*****=1000Az**	***B*** **[T] - for** ***NI*****=1000Az**
	
	

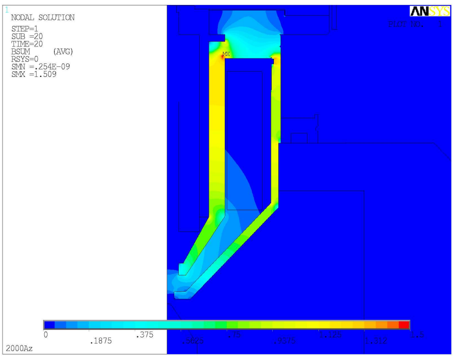
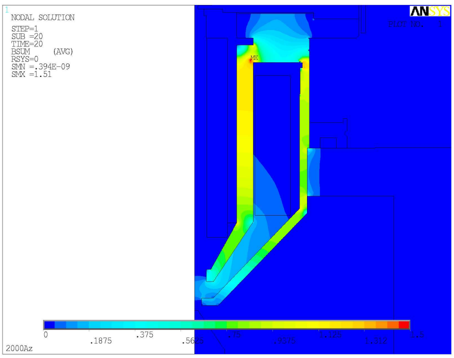
Tab. 2. Objective lens

Magnetic intensity *Hsum* [A/m] and Flux density *Bsum* [T] for *NI*= 2000 Az

<table>
<tr><th>H [A/m] - for NI=2000Az</th><th>B [T] - for NI=2000Az</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Tab. 3. Objective lens - comparison between model with and without magnetic chamber

Magnetic intensity *Hsum* [A/m] and Flux density *Bsum* [T] for *NI*= 2000 Az

<table>
<tr><th>H [A/m] - for NI=2000Az</th><th>B [T] - for NI=2000Az</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

4 Conclusion

This paper presents the employment of results of magnetic measurements for simulation of electromagnetic field distribution of the type electron microscope .

This way, you can very accurately determine the electromagnetic field distribution in the magnetic circuit and in the active vacuum space of the lens and also find critical points in the geometry which could cause a defects of the lenses. The results achieved in this way can be used for optimization of the geometry of lenses, objectives or magnetic shielding and for other analysis. Further, there are being prepared other points of solution. Firstly, measuring of electromagnetic field in the lens and the comparison with computed simulation of magnetic field distribution and secondly, electron trajectory computing from results of simulation of electromagnetic field distribution in the lens.

References

1. Egerton, R., F.: Physical principles of electron microscopy: An introduction to TEM, SEM, and AEM, Springer, USA 2005
2. Holčáková, R., Folvarčný, A., Marek, M.: Innovative Research of Electron Microscopes, Optimization Possibilities of Modern Types Magnetic Lens and Chambers - part 2: Simulation of Layout Magnetic Field of Types Magnetic Lens, Society for Histochemistry, Praha 2010
3. Holčáková, R., Marek, M., Folvarčný, A., Urbánek, .P, Jelen L..: Innovative Research of Electron Microscopes, Problematic of Modern Types Magnetic Lens and Chambers - part 2: Simulation of Layout Magnetic Field of Types Magnetic Lens, Magnetic Measurements, Praha 2010
4. Marek, M., Holčáková, R.: Computation and simulation of electromagnetic field distribution in type lens of the electron microscope, EPE, Brno 2011
5. Holčáková, R., Marek, M., Folvarčný, A., Jelen, L., Režnar, J..: Innovative Research in Electron Microscopes, Analysis of Magnetic Field Distribution of some Types of Magnetic Lenses by FEM, Mikroskopie, Praha 2011
6. Holčáková, R., Marek, M.: Innovative Research in Electron Microscopes, Analysis of Magnetic Field Distribution of Some Types of Magnetic Lenses by FEM, EEEIC, Roma 2011

Grants

1. MPO: FR-TI1/334-*Implementation of the innovation from the research of modern types chambers for electron microscopes in order to strengthen the position and competitive advantage of the Czech Republic on the world market*

Control Unit for Parallel Active Power Filter

Michal Hromják

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.hromjak@vsb.cz

Abstract. This paper describes the basic principle of parallel active power filter and gives a basic description of its control unit based on kit eZdsp™ F28335. For this application was necessary to develop Interface-board, which provides correct adaptations of input and output signals for eZdsp™ F28335. The following is a brief description of Code Composer Studio v4 which is an integrated development environment for Texas Instruments DSP and MCU devices.

Keywords: Active power filter, DSP, eZdspF28335

1 Introduction

Currently, there is a great use of appliances that are nonlinear or unbalanced draw of electricity and non-linear power network burden. They are mainly a variety of applications of semiconductor converters, controlled and uncontrolled rectifiers, switching power supplies, UPS, modern electronic lamps, etc. Application of these appliances cause distorted voltage and current, power supply network and also the power factor is worse. Harmonic current pollution of electrical power systems is most evident in overheating of transformers, electrical motors, capacitors for power factor correction, voltage waveform distortion, voltage flicker, interference with control and communication systems. The adverse effects on the electrical network can be limited by passive power filters. However, passive filters have many disadvantages such as large size, load dependent, their detuning associated with the aging of components, resonance problems, etc. These problems can be solved by using of active power filters. Although their basic compensation principles were described around 1970, their development and implementation is closely associated with the development of modern power switching devices (IGBT, GTO), and together with the availability of Digital Signal Processors (DSP).

The active filters can be classified by topology to parallel active power filter (PAPF), series active power filter (SAPF) and their combination known as unified power quality conditioner (UPQC). PAPF is most widely used to eliminate current harmonics, SAPF is preferable to eliminate voltage harmonics and UPQC allows a simultaneous compensation of the load currents and the supply voltages. There are also combinations of active power filter and passive filter called hybrid filters.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 160–165.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Basic principle of Parallel Active Power Filter

PAPF is connected in parallel with the nonlinear load. The filter is capable of removing the harmonic current contained in the load current i_L by injecting compensation current i_C into the ac power system. The final result of this means that the current drawn from the source i_S has a sinusoidal character, i.e. it doesn't contain higher harmonics. The basic configuration and compensating principle PAPF is shown on the next figure.

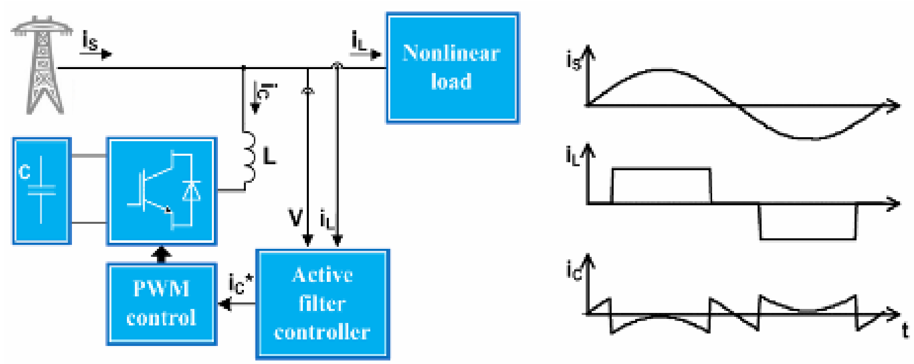


Fig. 1. Basic configuration of PAPF

PAPF consists primarily of a power PWM converter and active filter controller. The PWM converter is responsible for power processing in synthesizing the compensating current that should be drawn from the power system. The active filter controller is responsible for signal processing in determining real time instantaneous compensating current references, which are continuously passed to the PWM converter. It consists of a voltage-fed converter with a PWM current controller and an active filter controller that realizes almost instantaneous control algorithm. The parallel active filter controller works in a closed-loop manner, continuously sensing the load current i_L, and calculating the instantaneous values of the compensating current reference i^*_C for the PWM converter. [1]

APFs can be further divided by the type of converter for voltage source converter (VSC) or current source converter (CSC). Today most of the APFs are based on the VSC due to e.g. high efficiency, low cost, smaller size.

Figure 2. shows PAPF with VSC for three-wire supply system. This topology is considered for our purposes. The voltage source for the VSC is capacitor C. The current flowing through the capacitors can have both polarities, the voltage U_{DC} can only have the indicated polarity. The picture shows a marked polarity of the current i_{DC} at which energy flows from the compensated network to the inverter and the capacitor voltage increases. In the opposite direction of current flow, the energy flows from the capacitor to the compensated system.

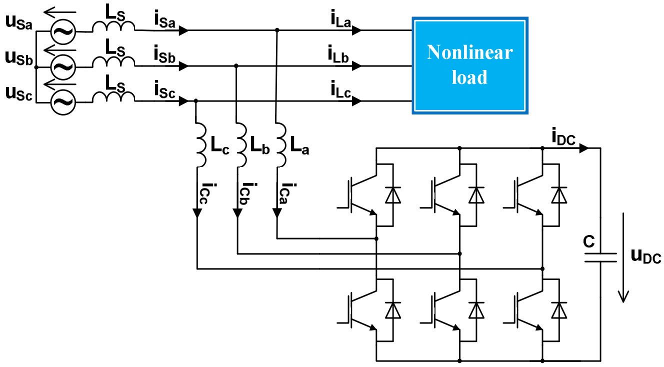


Fig. 2. Basic configuration of PAPF

3 Control unit

The control unit of PAPF is based on kit eZdsp^{TM} F28335 by Spectrum Digital Inc., which is intended for motor control development. The kit includes TMS320F28335 device which is a member of the C28x^{TM} family of Floating-point Microcontrollers from Texas Instruments and has strong management ability and embedded control functions. The floating-point device performs most of the math operations faster than fixed-point. The 32-bit Digital signal controller is clocked at 150 MHz and has an instruction cycle time 6,67 ns and can perform as many as 300 million floating-point operations per second (MFLOPS). Figure 3. shows features of the TMS320F28335 device.

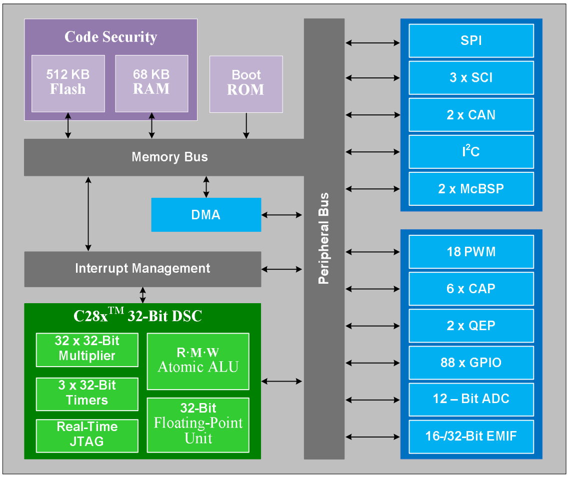


Fig. 3. Block diagram of TMS320F28335 DSC architecture

4 Development of Interface-board for eZdsp™ F28335

In our Department of electronics was developed interface for TMS320F28335. This interface provides basic functions which are necessary for controlling a PAPF. Figure 4. shows structure of this Interface-board.

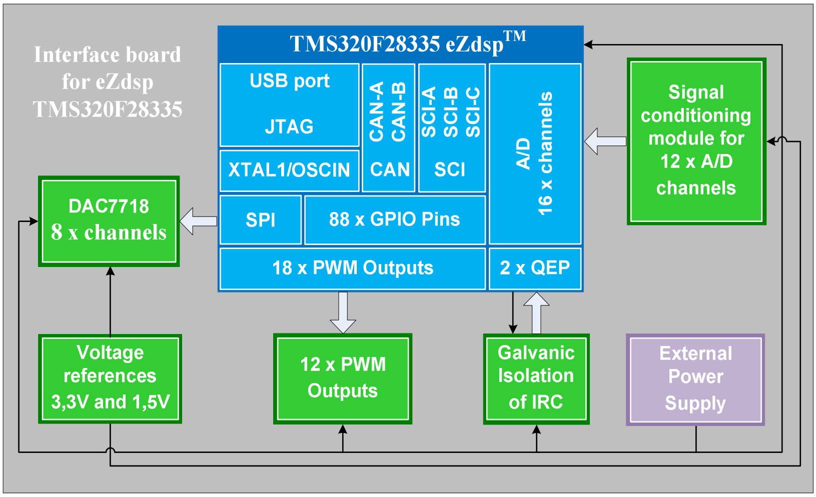


Fig. 4. Block diagram of Interface-board

The Interface-board is supplied from external source which contains ±15V and 5V. The sub-circuit ±15V supply the circuits for adaptations input signals of A/D converter (ADC). On the board are situated voltage references 3,3V for reference signal of DAC and 1,5V for offset factor of input circuits ADC.

The inputs of ADC must be properly limited because the range of input signals is only 0V to 3V. Therefore the Interface-board contains input circuits. These circuits are created by op-amps in the differential amplifier configuration, which provides possibility to connect input signals in voltage range -10V to +10V. All inputs are also secured against over-voltage above 3V.

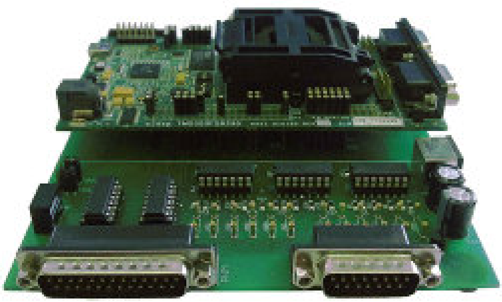


Fig. 5. Control unit consisting of Interface-board and kit eZdsp™ F28335

On the board is placed 12-bit D/A converter DAC7718 with 8 channels. The outputs of DAC are connected as a bipolar with range -10V to +10V. The communication between DAC and eZdsp™ F28335 is realized by SPI. All 12 PWM channels are boosted and isolated by two 74LS07 circuits with open-collector outputs.

The Interface-board also allows connecting signals from incremental speed sensor (motor control). The speed sensor can be supplied from external source or direct from Interface-board. These signals are electrical-isolated from the Enhanced Quadrature Encoder Pulse Module (eQEP). There is also SCI included on the Interface-board for communication with superior system or other systems.

5 Code Composer™ Studio v4

Code Composer Studio v4 (CCS4) is the development environment for family of Texas Instruments DSP and microcontrollers. The version 4 and higher is based on the Eclipse open source software framework. CCS4 has a lot of features which simply allow to develop many applications. It contains for example source code editor, compiler, project build environment, debugger, simulator, etc. The user can easy modify his workbench and switch between multiple perspectives. The perspective defines layout of views in the workbench. There are two default perspectives:
C/C++, Debug. The example of Debug perspective can be seen on the next picture.

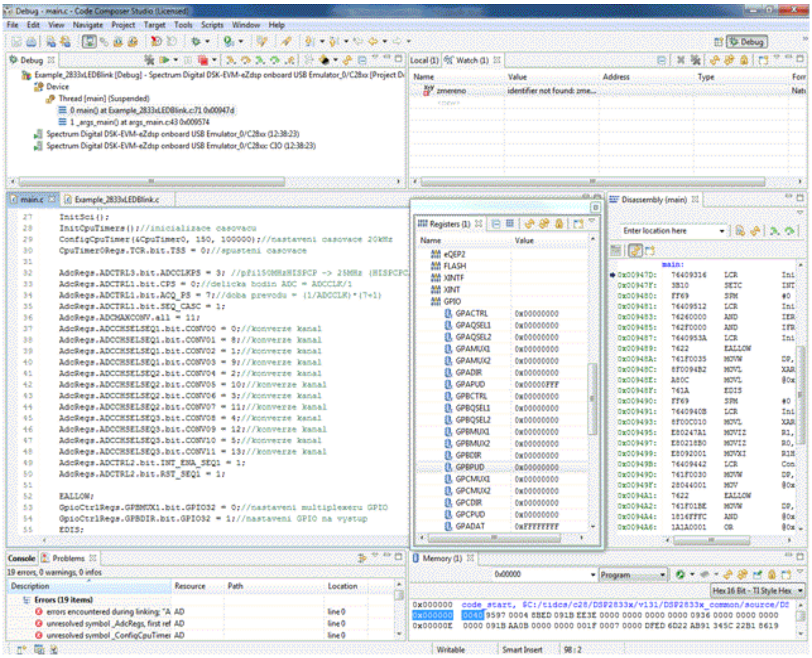


Fig. 6. Code Composer Studio v4 Workbench

CCS4 allows writing the code in C language. For time-consuming processes we can use writing in Assembler. It is possible to use combined writing too. The debugging process can work with a connected device in Real-Time Mode. Very useful are supported libraries, header files and peripheral examples for new users. The code can be used as a basis for development platform and new user can quickly experiment with different peripherals. Well-arranged bit field and register-file oriented structure approach is used for accessing to registers.

6 Conclusion

This paper shows basic information about parallel active power filter and its application range. The following is a description of basic properties of TMS28335, Interface-board and CCSv4. In the next stage of development it is necessary to realize experimental prototype of converter for PAPF and choose a suitable control algorithm based on the selected compensation strategy. The selected control algorithm will be implemented for TMS320F28335 and tested in real conditions.

References

1. Akagi, H., Watanabe, E. H., Aredes, M.: Instantaneous Power Theory and Applications to Power Conditioning, Wiley-Interscience a John Wiley & Sons, Inc., Hoboken, New Jersey, 2007, ISBN 978-0-470-10761-4.
2. Šimoník, P.: Nové přístupy k řešení aktivních výkonových filtrů., PhD thesis, VŠB-TU Ostrava, 2006.
3. Texas Instruments, TMS320F28335 Manual [online]. Av. from: <:http://focus.ti.com>.
4. Spectrum Digital, eZdsp™ F28335 Technical Reference [online]. Available from:<:http://c2000.spectrumdigital.com/ezf28335/docs/>.

Acknowledgments

This article has been supported by SGS Project No. SP2011/103 - Research of accumulation subsystem structures.

Real-time Estimation of Induction Motor Parameters Based on the Genetic Algorithms

Petr Hudeček

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.hudecek@vsb.cz

Abstract. This article shows one of many ways how to identify the parameters of the IM in real time. There is used the theory of genetic algorithms for IM parameters identification. The introduction describes why the problem is discussed. Next chapters show induction motors dynamic model and the principle and way how to implement the IM parameters identification. Theory of used genetic algorithm and experimental results are demonstrated in the end of this article. The conclusion describes the potential use of this method and discusses further developments in the real time estimation of induction motors parameters.

1 Introduction

The Induction Motors (IM) thanks to its well known advantages of simple construction, reliability, ruggedness and low cost, has found very wide industrial applications. The IMs are often supplied by frequency converters for better performance. Frequency converters commonly use complex control strategies like FOC (Field Oriented Control) or DTC (Direct Torque Control). Both of these control techniques are highly dependant on correct estimation of flux linkage of motor. It's particularly well known that FOC is very sensitive to variation of rotor time constant T_R while DTC is likewise sensitive to variation of stator resistance R_S during the estimation of flux linkage of IM. However in both of this control strategies any inaccuracy in evaluation of one parameter caused a wrong value of magnetic flux (both in amplitude and in angle) and of electromagnetic torque and it is not therefore possible to achieve a correct field orientation [1], the torque capability of drive deteriorates and instability phenomena can even occur.

The Genetic Algorithms (GA) is a search technique used in many fields, like computer science, to find accurate solutions to large optimization and search problems. The basic concept of GAs is to emulate evolution processes in natural system following the principles first laid down by Charles Darwin of survival of the fittest. The advantage of GAs is that it is a very flexible and intuitive approach to optimization and presents a higher probability of not converging to local optima solutions compared to traditional gradient based methods. More recently, research work has appeared in the sci-

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 166–171.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

entific literature about the use of GAs for control design in power Electronics and drives and general structure identification. This paper describes an automatic real-time estimation procedure of IM parameters based on the genetic algorithms.

2 Induction motors dynamic model

For the purpose of the present investigation the IM is assumed to be described in the stator reference frame by [2]

$$\frac{d\omega_m}{dt} = \frac{3}{2} \cdot \frac{p \cdot L_h}{J_C \cdot L_R} \cdot \left(i_{S\beta} \cdot \psi_{R\alpha} - i_{S\alpha} \cdot \psi_{R\beta}\right) - \frac{M}{J_C} \tag{1}$$

$$\psi_{R\alpha} = \int \left(\frac{M}{T_R} \cdot i_{S\alpha} - \frac{\psi_{R\alpha}}{T_R} - p \cdot \omega_m \cdot \psi_{Rb} \right) dt \tag{2}$$

$$\psi_{R\beta} = \int \left(\frac{M}{T_R} \cdot i_{S\beta} - \frac{\psi_{R\beta}}{T_R} + p \cdot \omega \cdot \psi_{R\alpha} \right) dt \tag{3}$$

$$\frac{di_{S\alpha}}{dt} = \frac{1}{\sigma \cdot L_S} \cdot u_{s\alpha} - \gamma \cdot i_{S\alpha} + \frac{\beta}{T_R} \cdot \psi_{R\alpha} + p \cdot \beta \cdot \omega_m \cdot \psi_{R\beta} \tag{4}$$

$$\frac{di_{S\beta}}{dt} = \frac{1}{\sigma \cdot L_S} \cdot u_{s\beta} - \gamma \cdot i_{S\beta} + \frac{\beta}{T_R} \cdot \psi_{R\beta} - p \cdot \beta \cdot \omega_m \cdot \psi_{R\alpha} \tag{5}$$

$$\beta = \frac{M}{\sigma \cdot L_S \cdot L_R} \tag{6}$$

$$\gamma = \frac{R_S}{\sigma \cdot L_S} + \frac{M^2 \cdot R_R}{\sigma \cdot L_S \cdot L_R^2} \tag{7}$$

$\psi_{R\alpha}, u_{S\alpha}, i_{S\alpha}$ - are alpha components of rotor flux, stator voltage and current.
$\psi_{R\beta}, u_{S\beta}, i_{S\beta}$ - are beta components of rotor flux, stator voltage and current.
L_S, L_R, L_h - is stator, rotor and main inductance.
ω_m, σ, p - is mechanical angular speed, leakage factor and number of pole pairs.
J_C, T_R, M - total moment of inertia, the mechanical time constant, moment of IM.
R_S, R_R - is stator and rotor resistance (rotor resistance is oriented to stator).

3 Principle of parameter identification

The method of IM parameter estimation in real time presented here is a method when thanks to modification of previous equations (1) to (7) we get the new relations:

$$K_1 i_{S\alpha} + K_2 u_{S\alpha} + K_3 p\omega_m i_{S\beta} + K_4\left(\frac{du_{S\alpha}}{dt} + p\omega_m u_{S\beta}\right) + K_5 \frac{di_{S\alpha}}{dt} = \frac{d^2 i_{S\alpha}}{dt^2} + p\omega_m \frac{di_{S\beta}}{dt} \qquad \textbf{(8)}$$

$$K_1 = -\frac{R_S}{\sigma L_S T_R},\ K_2 = \frac{1}{\sigma L_S T_R},\ K_3 = -\frac{R_S}{\sigma L_S},\ K_4 = \frac{1}{\sigma L_S},\ K_5 = -\frac{R_S L_R + R_R L_S}{\sigma \cdot L_S \cdot L_R} \qquad \textbf{(9)}$$

Unknown parameters are than:

$$R_S = -\frac{K_3}{K_4},\ R_R = \frac{K_3 - K_5}{K_4},\ L_S = \frac{K_3 - K_5}{K_2},\ L_h = L_S \cdot \sqrt{1 - \frac{1}{K_4} \cdot L_S} \qquad \textbf{(10)}$$

4 Implement of parameter identification

Required values of $u_{S\alpha}$, $u_{S\beta}$, $i_{S\alpha}$, $i_{S\beta}$ we can obtain in SW part from a vector control structure, which is shown on Fig.1.

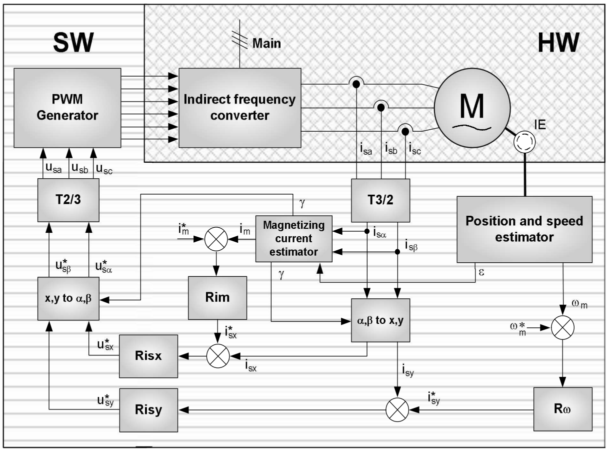


Fig. 1. Block diagram of the vector control used.

On Fig.1. the **Indirect frequency converter** supplies an induction motor (**M**). Phase currents are measured by current sensors. Position of the rotor is measured at the **Incremental encoder** (IE). True value of mechanical speed ω_m and the rotor angle ε are than evaluated in the **Position and speed estimator** block. The values in the three-phase stator coordinate system [a, b, c] are transformed into two-axis stator coordinate system [α, β] in the **T3/2** block. In the block **α,β to x,y**, vector components are transformed to the oriented two-axis rotating coordinate system [x, y]. For vector rotation of components of stator current $i_{s\alpha}$, $i_{s\beta}$ to the oriented coordinate system [x, y] is used the variables γ, which is calculated in the **Magnetizing current estimator** block. The components i_{sx}, i_{sy} serve as feedback variables for the current PI controllers $\mathbf{R_{isx}}$, $\mathbf{R_{isy}}$. Control of magnetizing current (or magnetic flux) is realized by a PI controller $\mathbf{R_{im}}$. The block processes a deviation between the desired magnetizing current i_m* and value i_m calculated in the magnetizing current estimator block. Angular speed control provides PI regulator $\mathbf{R}_\omega$, which handles difference between a speed command ω_m* and speed ω_m, which is evaluated as derivation of rotor angle epsilon. The control voltages u_{sa}, u_{sb}, u_{sc} for **PWM generator** block are obtained from the components u_{sx}*, u_{sy}* by means of the **x,y to α, β** block and subsequent transformation block **T2/3**. The PWM modulation generates pulses for IGBT power transistors in the indirect frequency converter.

4 Genetic algorithm

To describe how the GA actually works is used the relation (8). Consider the following simplistic rule. We know that IM parameters vary only in a certain range of values such as:

$$1\Omega \le R_S \le 10\Omega,\ \ 1\Omega \le R_R \le 10\Omega,\ \ 0.1H \le L_S \le 1H,\ \ 0.1H \le L_h \le 1H\ldots$$

Ranges of course can also expand or use other parameters, if needed. Substituting the upper and lower limit in relations (9) we find the boundaries. Between these boundaries the K unknown could fluctuate. The next procedure can be summarized in four points.

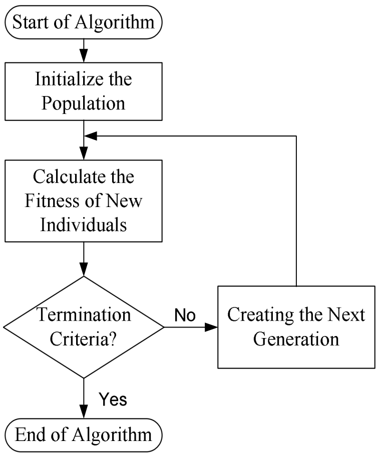


Fig. 2. Flowchart of the genetic algorithm used.

1) **Initialize the population** - Creates a zero population which is composed from randomly generated individuals. Specifically, an individual is composed of five unknowns K_1 to K_5 randomly generated within the range specified above.

2) **Calculate the Fitness of New Individuals** - Is calculated the fitness of new individuals. Than is selected a few individuals with high fitness from a population. In this case, the best marked individual is the individual whose unknowns K have the difference between right and left side of the equation (8) as much as possible close to zero.

3) **Termination Criteria** – If the stop condition isn't fulfilled, continue from point 4 again. Stopping condition may be a number of passes through the loop. This condition affects the speed and accuracy of determination unknown K. Individual with highest fitness is the main algorithm output and represents the best founded solution.

4) **Creating the Next Generation** - To create the next generation are used two basic genetic operators. The mutation and the crossing. In some cases it may be useful to keep copies of the parents for the next generation unchanged. Specifically new population we generates as follows:

   - *crossing* - swap parts (unknown K) of a few individuals among them. It is necessary to ensure that unknown through the swapping must have the same index (not for example $K_1$with K_5).
   - *mutation* – is a random change of parts of a few individuals or individual. Again, it is necessary to ensure that mutations for example in the K_1 individual were in appropriate range.
   - *reproduction* - the remaining individuals are copied unchanged.

5. Experiments and results

In all simulations was used model of real three-phase IM type CANTONI Sg100L-4A which has the value of stator resistance $R_S = 2.78\Omega$ and value of rotor resistance $R_R = 2.84\Omega$. The IM was supplied by a model of indirect frequency converter in connection consisting of a DC source voltage at the input and the voltage inverter with IGBT transistors on the output. DC-link voltage was set to $U_d = 300V$. Control of output voltage was made by comparing PWM. The frequency and amplitude of the saw tooth signal was $f_p = 5kHz$, $U_{pmax} = \pm 1V$. The incremental encoder model is based on the real encoder type ERN 420/TTL with 2048 pulses/rev. The total moment of inertia was set at $J_C = 0.043kgm^2$.

From fig.3 is evident that the value of R_S and R_R has the true values. Identification is not accurate in area of excitation of IM (time 0 to 0.2s) and during the start of IM (time 0.2 to 0.8s). This behavior is caused by neglecting of $d\omega/dt$ member during the modification of equations (1) to (7). This phenomenon can be mitigated. The solution can be found in the literature [3].

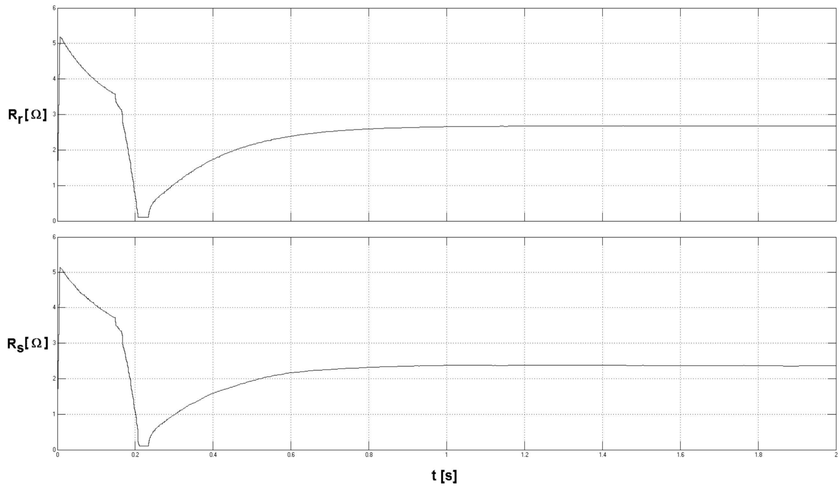


Fig. 3. Transients of R_S and R_R estimation using genetic algorithms

6. Conclusion

Here described method can be used in modern high performance electric drives. This is evident from fig.3 and chapter 5. Reasons why it is necessary to estimate IM parameters in real time are described in the introduction. Furthermore for identifying unknowns with GA, it is possible to detect unknown with using other methods such is least squares methods (Total Least Squares, Ordinary Least Squares, Recursive Least Squares) or by using theory of neural networks (see literature [2, 3]).

Publication in which the results were presented

1. Palacký, P., Hudeček, P., Slivka, D., Sobek, M.: Využití TMS320F28335 pro online estimaci parametrů asynchronního motoru napájeného z frekvenčního měniče. In International Scientific Conference EPE 2011, Ostrava, 2011, vol. 12, ISBN 978-80-248-2393-5

References

1. Leonhard, W.: Control of Electrical Drives, Springer –Verlag, Berlin, Germany, **2001**, ISBN 3-540-41820-2
2. Cincirone, M., Pucci, M., Cincirone, G., Capolino, G. A.: A New Experimental Application of Least-Squares Techniques for the Estimation of the Parameter of the Induction Motor, IEEE Trans. on Ind. Applications, vol. 39, No. 5, September **2003**, pp. 1247-1255.
3. Chiasson, J.: „Modeling and High-Performance Control of Electric Machines“, IEEE Press, USA, **2005**, ISBN 0-471-68449-X.

Sensor Control Simulation of Switched Reluctance Motor Using HIL Method

Petr Krňa

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.krna@vsb.cz

Abstract. This work deals with simulation of sensor control switched reluctance motor using a HIL metod. Area of this method is very wide. Using this method is in aviation, military, automotive etc., ie wherever is a complex control systems. HIL simulation method extends the classical simulation to simulation, where control is represented by the real control system and the simulation is close to the real world. Controlled system is a simulation model of SRM, which was created in Matlab-Simulink and therefore we can simulate error conditions, which we could not test on a real machine. HIL test method in this work is performed by the multifunction card MF 624, which connects the electronic device and the simulation model of the controlled system.

Keywords: SRM, HIL, magnetic flux, torque.

1 Introduction

Switched reluctance motor is very similar to stepper electric motor with variable reluctance and this similarity is topological and electromagnetical. The main difference between these two engines is based on design, control methods, the performance and application character. Switched reluctance motor is usually controlled by feedback with shaft position to synchronize occurred commutation phases currents with rotor position. Stepper motor usually operates in open loop control, ie without feedback from the shaft position. Switched reluctance motor has several advantages:

- Simply wound stator.
- A simple rotor, that requires little manufacturing steps. Moment of inertia is usually small.
- Good efficiency over a wide speed range and torque.
- It works at very high speeds (up to 100 000 min^{-1}).

The main disadvantages are:

- Output torque is fairly curled, and it may contribute to the acoustic noise.
- Very non-linear magnetizing characteristics of engine.
- Larger number of wires than induction motor.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 172–177.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

1.1 Mathematical model of SRM

The basic voltage equation:

$$u = Ri + \frac{d\psi(\vartheta,i)}{dt} \tag{1}$$

$$u = Ri + \frac{d\psi}{di}\frac{di}{dt} + \frac{d\psi}{d\vartheta}\frac{d\vartheta}{dt} \tag{2}$$

$$u = Ri + L(\vartheta,i)\frac{di}{dt} + D(\vartheta,i)\omega \tag{3}$$

The equation for the moment of the machine:

$$T(\vartheta,i) = \frac{\partial}{\partial\vartheta}\int_0^i \psi(\vartheta,i)di \tag{4}$$

For the mechanical motor speed applies:

$$\omega = \frac{d\vartheta}{dt}\left[\frac{rad}{s}\right] \tag{5}$$

Magnetic flux:

$$\psi = \int_{ton}^{tc} (u - Ri)dt\ [Wb] \tag{6}$$

1.2 The simulation model of SRM

Switched reluctance motor was designed according to the mathematical model of engine with its parameters close to the real engine SRM90 from company Magna Physics. Block diagram of one phase of SRM is shown in Figure 1.

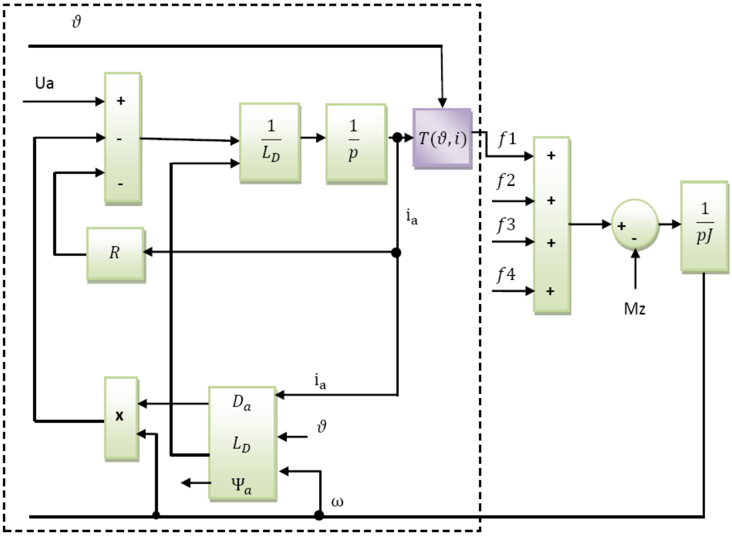


Fig. 1 Block diagram one phase of SRM

Simulated model uses a simplified function coupled magnetic flux. Torque equation is replaced with graphical function.

2 Current control of SRM using the method HIL

HIL simulation technique is used for testing and development of complex simulation models with real control system in real time. This method generally facilitates the transition to the real controlled system, the real control system is used for control. The simulation results are very close to reality, if the simulation model of the controlled system is realized properly. For current control of SRM is important a hysteresis current regulator, which compares the actual current value of the phase current with reference current. This regulator is implemented in the control processor and evaluates the actual current phase in real time. The communication cannel between controll and controlled system is created by using the A/D converter. Real control system is the signal processor from Freescale MC53F8037. General block diagram of HIL test method is shown in Figure 2. Individual variables are: y (t) ... process value, w (t) ... reference value, e (t) ... the error, u (t) ... action variable, v (t) ... fault value, reg ... regulator.

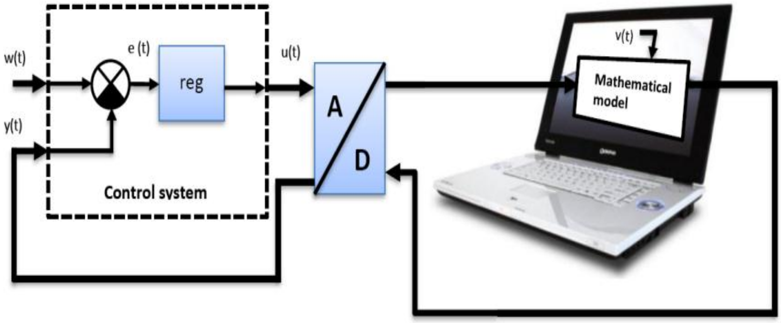


Fig. 2. Diagram of Hardware in the Loop Simulation.

3 Simulation results of sensor control SRM by method HIL

Simulation diagram for control of SRM is shown in Figure 3 and it was created in program Matlab-Simulink. The following waveforms show the behavior of a simulation model for reversing rotation.

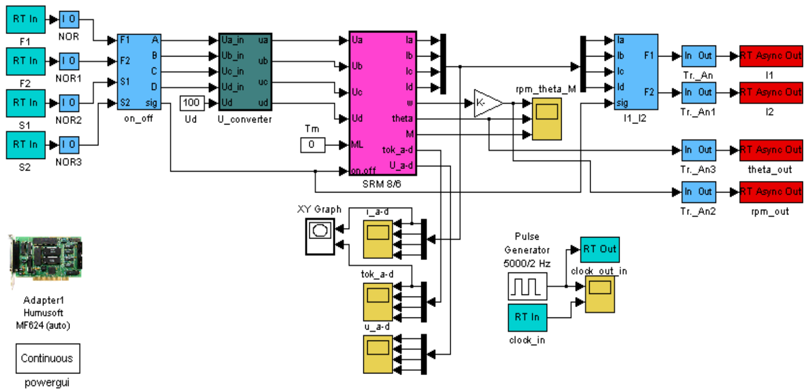
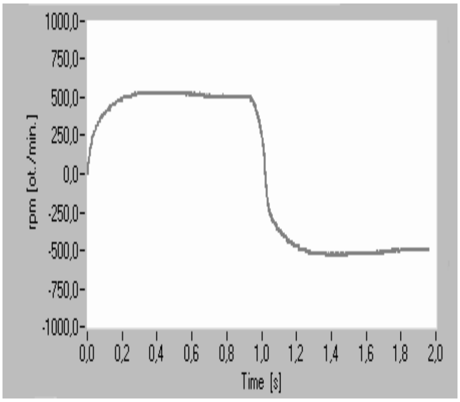


Fig. 3. Simulation structure of the SRM in Matlab



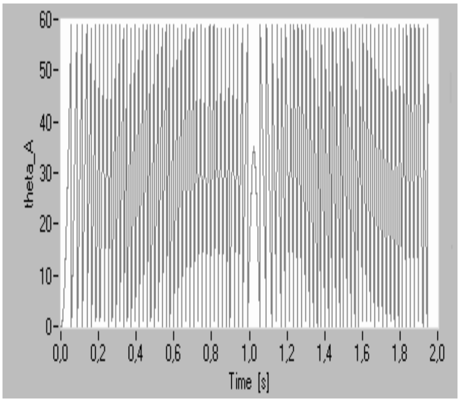


Fig. 4. Change the engine speed 0/500/-500 [rpm].

Fig. 5. Theta [ ].

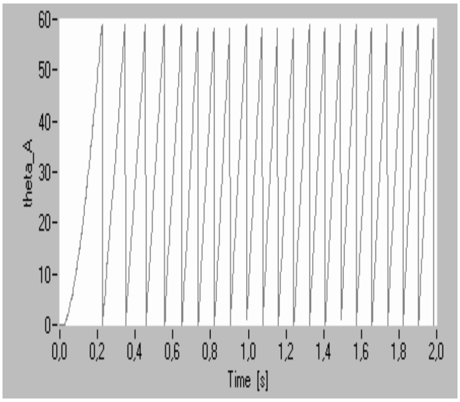


Fig. 6. Detail of theta [ ].

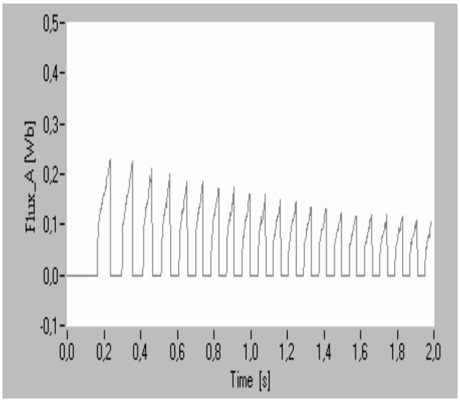


Fig. 7. Magnetic flux of phase a [Wb].

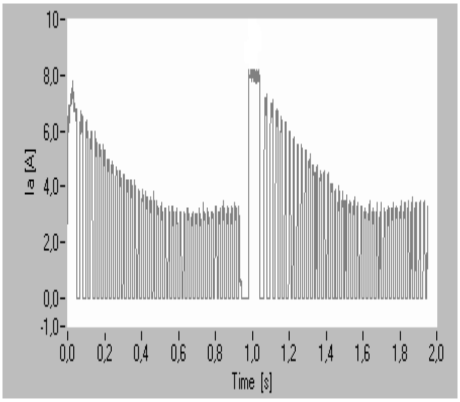


Fig. 8. Current of phase a [A].

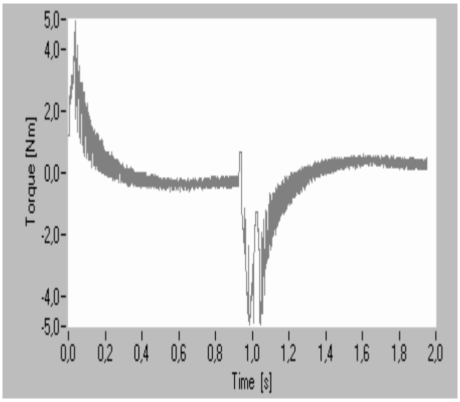


Fig. 9. Total torque [Nm].

4 Conclusion

This work was created by the structure of HIL test methods for control simulation model of SRM. Applied Card MF 624 was used for communication and signal processor MC56F8037 and mathematical model of motor was created in Matlab-Simulink. Control took place in real time. Measurements of currents were performed using the A/D converters, implemented in the control processor. The data for the display of the measured waveform was sent to PC and processed in Labview.

Acknowledgement

In the paper there are the results of the project SP2011/107 which was supported by Student Grant Competition of VSB-Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

References

1. Brandštetter, P.: AC variable speed drives - Modern methods of management, Ostrava: VŠB - TU Ostrava, 1999. ISBN 80-7078-668-X
2. TJE Miller, Electronic control of switched reluctance machines. UK: Newness, 2001.
3. T. Miller, Switched reluctance motors and their control. Oxford, UK: Magna Physics . Publishing/Clarendon Press, 1993

Control Methods of Series Active Power Filters

Lukáš Odlevák

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.odlevak@vsb.cz

Abstract. Active filters are very important to mitigate harmonic pollution due to the widely used nonlinear loads both in home appliances and in industry, therefore are mentioned in Introduction. However, main focus is devoted to series active power filter (SAPF), with a focus on the control methods. Finally, simulation of SAPF is shown to understand its function.

1 Introduction

The proliferation of nonlinear loads caused by more and more modern electronic equipments results in deterioration of power quality in power transmission or distribution systems. Harmonic, reactive, negative sequence and flickers are the reasons of various undesirable phenomenon in the operation of power system. In order to solve these problems, the concept of Active power filters (APF) was presented. APF's, which compensate harmonic and reactive current component for the power supplies, can improve the power qualities and enhance the reliabilities and stabilities on power utility.

APF's can be classified based on the topology used as series or shunt filters, and unified power quality conditioners use a combination of both. Combinations of active series and passive shunt filtering are known as hybrid filters.

Active shunt filter is most widely used to eliminate current harmonics, reactive power compensation, and balancing unbalanced currents. It is mainly used at the load end, because current harmonics are injected by nonlinear loads. It injects equal compensating currents, opposite in phase, to cancel harmonics and/or reactive components of the nonlinear load current at the point of connection. It can also be used as a static var generator in the power system network for stabilizing and improving the voltage profile.

Active series filter is connected before the load in series with the mains, using a matching transformer, to eliminate voltage harmonics, and to balance and regulate the terminal voltage of the load or line. It has been used to reduce negative-sequence voltage and regulate the voltage on three-phase systems. It can be installed by electric utilities to compensate voltage harmonics and to damp out harmonic propagation caused by resonance with line impedances and passive shunt compensators.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 178–183.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Unified power quality conditioner (also known as a universal APF) is a combination of active shunt and active series filters. The dc-link storage element (either inductor or dc-bus capacitor) is shared between two current-source or voltage-source bridges operating as active series and active shunt compensators. It is used in single-phase as well as three-phase configurations. It is considered an ideal APF which eliminates voltage and current harmonics and is capable of giving clean power to critical and harmonic-prone loads, such as computers, medical equipment, etc. It can balance and regulate terminal voltage and eliminate negative-sequence currents. Its main drawbacks are its large cost and control complexity because of the large number of solid-state devices involved [2], [5].

Hybrid filter, which is a combination of an active series filter and passive shunt filter is quite popular because the solid-state devices used in the active series part can be of reduced size and cost (about 5% of the load size) and a major part of the hybrid filter is made of the passive shunt L–C filter used to eliminate lower order harmonics. It has the capability of reducing voltage and current harmonics at a reasonable cost.

2 Basic Series Active Filter

Fig. 1 shows an example of a basic series active filter in three-phase, three-wire systems. The series active filter consists of either a three-phase voltage-source PWM inverter (VSI) or three single-phase VSI, and it is connected in series with the power lines through either a three-phase transformer or three single-phase transformers. Unlike the shunt active filter, the series active filter acts as a controllable voltage source and, therefore, the VSI has no current minor loop. This makes the series active filter suitable for compensation of a harmonic voltage source such as a three-phase diode rectifier with a capacitive load.

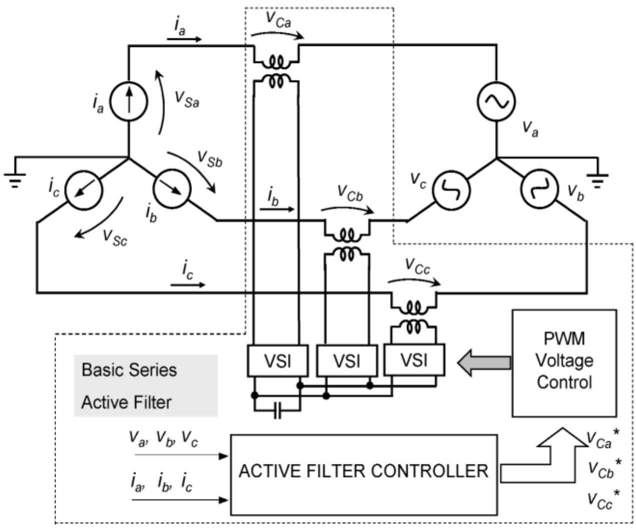


Fig. 1. Basic series active filter

For simplicity in Fig. 1, system “A” represents the source side, which has been represented by balanced sinusoidal current sources, and system “B” represents the load side with its voltage sources containing harmonic components. Moreover, it is assumed that there is no zero-sequence current. The voltages on the current sources are given by V_{Sa}, V_{Sb}, and V_{Sc}. The basic series active filter voltages are synthesized by three single-phase inverters with a common dc-bus capacitor. The reference voltage for these inverters is calculated by the “active filter controller” shown in Fig. 1.

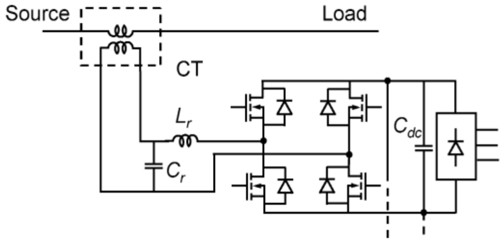


Fig. 2. Single-phase detailed circuit of the voltage source PWM inverter (VSI)

In Fig. 2, the inductance Lr and capacitance Cr in this circuit form a small-rated passive filter to suppress switching ripples generated by PWM [1].

3 Control Methods

Control strategy is the heart of the APF and is implemented in three stages. In the first stage, the essential voltage and current signals are sensed using power transformers, Hall-effect sensors, and isolation amplifiers to gather accurate system information. In the second stage, compensating commands in terms of current or voltage levels are derived based on control methods and APF configurations. In the third stage of control, the gating signals for the solid-state devices of the APF are generated using PWM, hysteresis, sliding-mode, or fuzzy-logic-based control techniques. The control of the APF’s is realized using discrete analog and digital devices or advanced microelectronic devices, such as single-chip microcomputers, DSP’s, etc. [5].

3.1 Filter-Based Method

Filter-based method is a control method of the APF’s in the time domain. Several control methods in the time domain have been developed. One of them is a synchronous d-q reference frame method. Its algorithm relies on the Park transformations to transform the three-phase system from a stationary reference frame into synchronously rotating direct, quadrature and zero-sequence components. These can easily be analysed since the fundamental-frequency component is transformed into DC quantities. The active and reactive components of the system are represented by the direct (d-axis) and quadrature (q-axis) components, respectively. The high-order harmonics still remain in the signal; however they are modulated at different frequencies.

These are the undesired components to be eliminated from the system and they represent the reference harmonic current. The system is very stable since the controller deals mainly with DC quantities. The computation is instantaneous but incurs time delays in filtering the DC quantities. This method is applicable only to three-phase systems [3], [4].

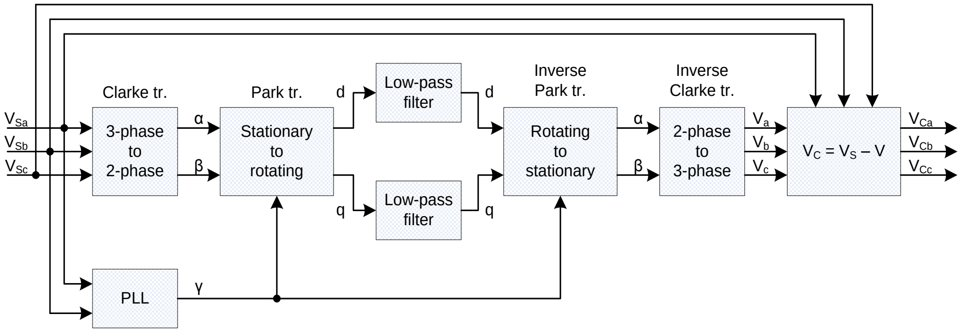


Fig. 3. Synchronous d-q reference frame method with low-pass filter

3.2 Frequency-Domain Method

Frequency-domain approaches are suitable for both single- and three-phase systems. They are mainly derived from the conventional Fourier analysis. Using fast Fourier transforms (FFT), the harmonic current can be reconstructed by eliminating the fundamental component from the transformed current signal and then the inverse transform is applied to obtain a time-domain signal. The main disadvantage of this system is the accompanying time delay. This technique needs to take samples of one complete cycle to generate the Fourier coefficients and it is therefore suitable for slowly varying load conditions [3], [4].

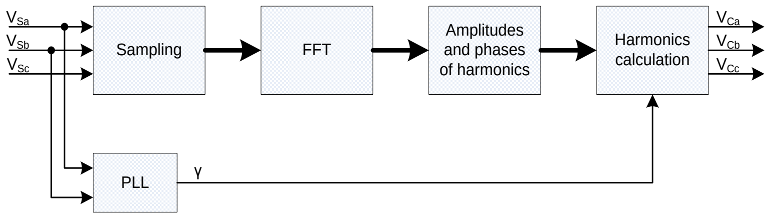


Fig. 4. FFT method

3.3 Comparative Method

The principle of this method is quite simple. An ideal sine wave is generated by DSP and synchronized with the source side. The desired sine wave voltage is then compared with the measured source voltage. Compensating commands are therefore derived [4].

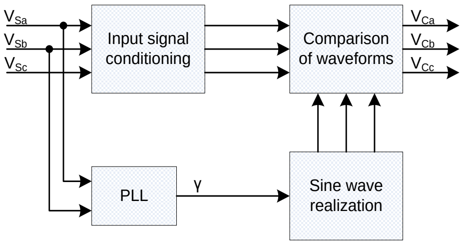


Fig. 5. Comparative method

4 Simulation

Series active power filter (SAPF) has been simulated using MATLAB/Simulink software. In this simulation, the source voltage V_{Sa} contains the following harmonics: fundamental 230V/50Hz, 20% of 3rd harmonic, 10% of 5th harmonic, 5% of 7th harmonic, 2% of 11th harmonic.

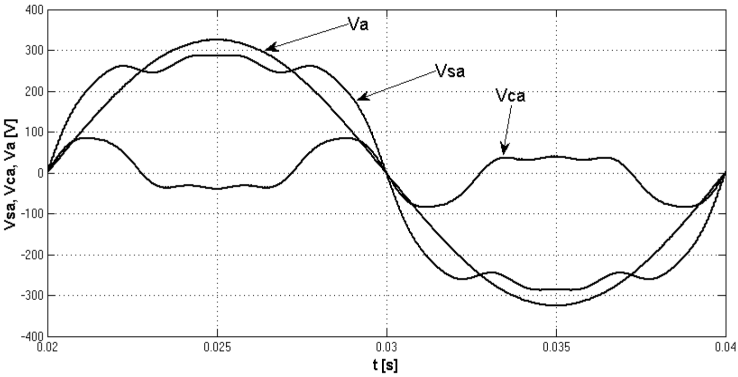


Fig. 6. Waveforms of voltages in Fig. 1

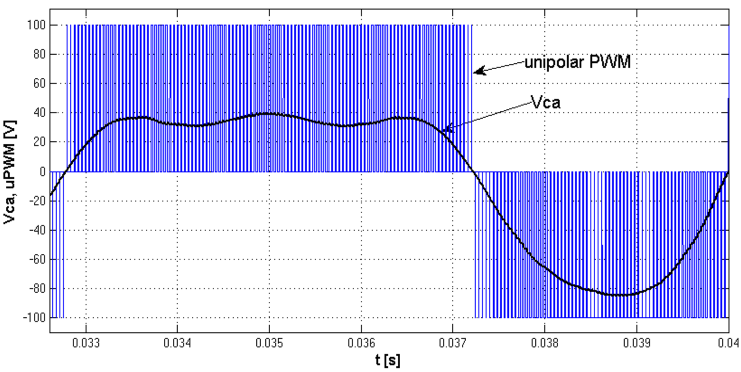


Fig. 7. Application of unipolar PWM

Fig. 6 depicts voltage waveforms of the deformed source voltage V_{Sa}, the load voltage V_a and the compensating voltage V_{Ca} in the a-phase. Fig. 7 shows the application of unipolar PWM to generate compensating voltage V_{Ca}. Small-rated passive LC filter have to be included in every phase to suppress switching ripples generated by PWM. This matter is related to Fig. 2.

5 Conclusion

The above mentioned control methods have been presented in [4]. The experimental and simulated results have proved that the comparative method is more convenient for controlling SAPF than filter-based or frequency-domain method. Comparative algorithm ensures low time delay as well as low total harmonic distortion (THD).

Essential task for correct operation of SAPF is the design of the small-rated LC filter eliminating switching ripples generated by PWM. Hence, LC filter determines switching PWM frequency.

Acknowledgement

This article has been supported by the project SGS No. SP2011/103: Research of accumulation subsystem structures.

References

1. Akagi, H., Watanabe, E.H., Aredes, M.: *Instantaneous Power Theory and Applications to Power Conditioning.* Wiley-Interscience a John Wiley & Sons, Inc., Hoboken, New Jersey, 2007, ISBN 978-0-470-10761-4.
2. Chen, D., Xie, S.: *Review of the control strategies applied to active power filters.* IEEE International Conference on Electric Utility DRPT, Hong Kong, 2004.
3. El-Habrouk, M., Darwish, M.K., Mehta, P.: *Active power filters: A review.* IEE Proceedings, 2000.
4. Mlčák, T.: *Sériové aktivní výkonové filtry.* Dissertation, VŠB-TU Ostrava, 2002.
5. Singh, B., Al-Haddad, K., Chandra, A.: *A Review of Active Filters for Power Quality Improvement.* IEEE Transactions on Industrial Electronics, 1999.

Electric Vehicle Control Units

David Slivka

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
david.slivka@vsb.cz

Abstract. This paper deals with an implementation of control units in the pure Electric vehicle, which has two independent voltage inverters and induction motors. The unit is equipped with Freescale DSC 56F8037. Each control unit drive own inverter and motor with using vector control algorithm with conception of current model. Computing structures from Freescale's motor control library are used in this algorithm with combination of fractional numbers implementation. Important part of whole system is data interchange between units, which is solved by CAN bus.

1 Conception of Electric Vehicle

Electric vehicle (EV) has two electric motors, where each has own inverter and control unit (CU). Each motor is 6 poles induction motor with rated power of 5,5 kW and maximum power of 16,5 kW. Rated torque is 55 Nm at speed of 950 rpm. The battery pack is created by LiFePo batteries which total voltage amount is 336 V. Arrangement of this EV can be seen on Fig. 1.

1.1 The control unit

DSC 56F8037 has been chosen due to its price, performances, quality and intuitive software supplied by manufacturer. Hybrid core of this DSC combines the Microcontroller unit (MCU) and the Digital signal processing unit (DSP) both based on 16 bits architecture. The DSC has internal relaxation oscillator, which makes possible to use frequency of 32 MHz. With one 1x instructions it has computing. The signals from DSC are converted and adapted, because the 56F8037 work only with 3,3V voltage. Unit's has implemented following blocks:

- Adaptation for A/D,
- Adaptation for D/A,
- Block of communication,
- Adaptation of PWM,
- Block of supply.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 184–189.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

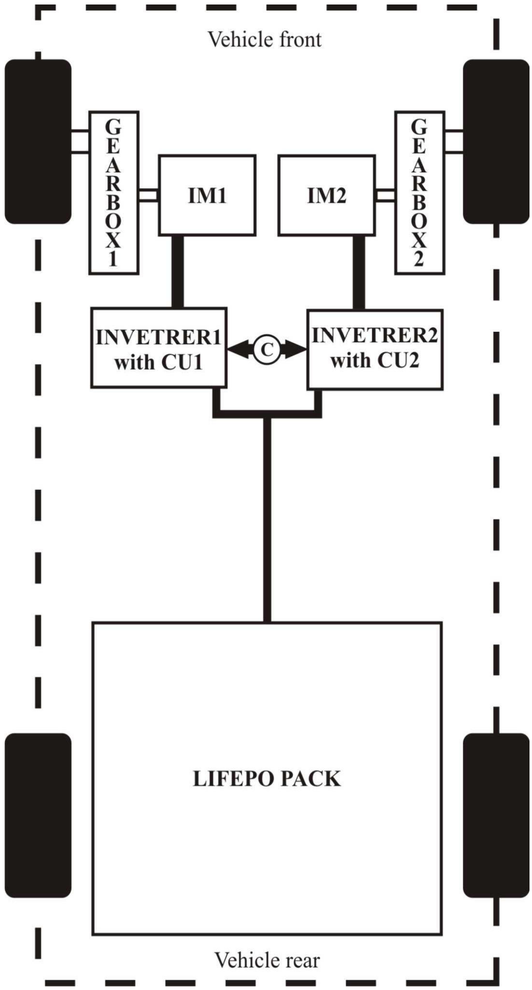


Fig. 1. Electric vehicle arrangement

1.2 Voltage Inverter

There is classic voltage inverter used. IGBT modules with ratings of 600V/200A have been chosen as main parts. Current transducers on each phase are integrated for using in various control algorithms.

1.3 Vector control algorithm

Conception, where the vector of stator current is decomposed into two perpendicular components in oriented coordinates, has been selected. The method of the current model is used to evaluate oriented values.

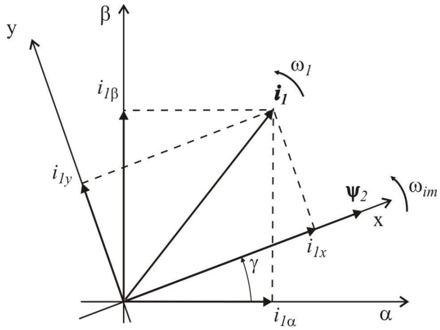


Fig. 2. Vector of stator current decomposition in oriented coordinates.

Stator system coordinates [α,β] and [x, y] are oriented system coordinates which real component is toward the rotor vector flux linkage ψ_2 can be seen on Fig. 2. Oriented system rotates with radial speed ω_{im}. Component i_{1x} represents magnetization of IM and i_{1y} represents torque of the machine [1].
Calculation of system coordinates [x, y] is represented by following relations [1]:

$$\boldsymbol{i_1^O} = i_{1x} + ji_{1y} \tag{1}$$

$$\boldsymbol{i_m^O} = i_{mx} = i_m \tag{2}$$

$$\boldsymbol{u_1^O} = u_{1x} + ju_{1y} \tag{3}$$

Block scheme of whole vector control algorithm can be seen on Fig. 4. Method of the current model is used in evaluation of oriented value block (EOVB). This solution can be seen on Fig. 3.

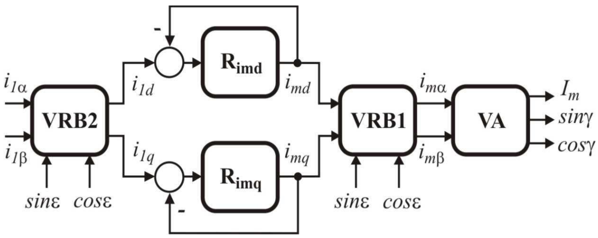


Fig. 3. Current model used in EOVB.

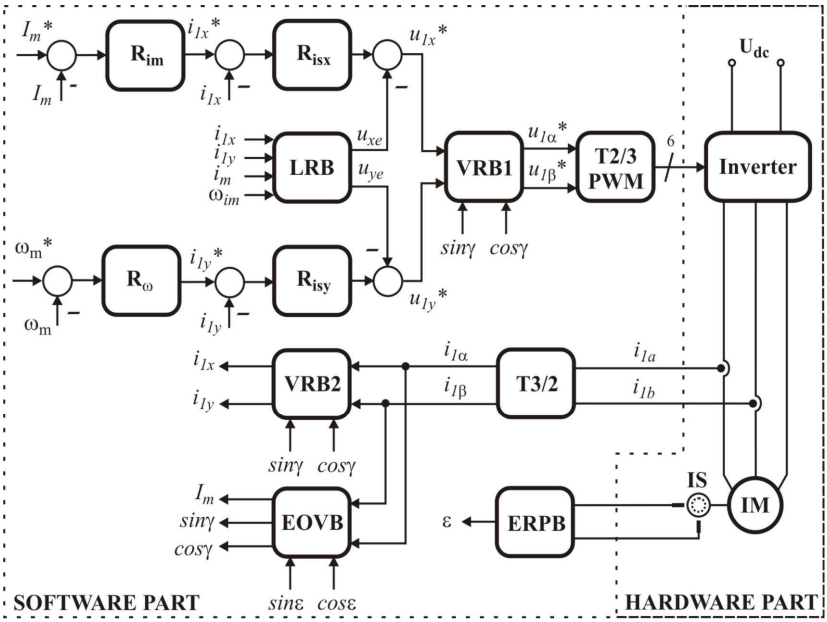


Fig. 4. Vector control of the IM in oriented system coordinates.

All structures have been created with help of Motor Control Library supplied by Freescale. This library is well debugged library with standard functions used in motor control area. In addition, fractional arithmetic is implemented. Signed 16 bits fractional numbers have range from -1 to 0,99997. Where number is represented as:

$$X_{F16} = \frac{X_{S16}}{32768} \tag{11}$$

Where X_{F16} is signed 16 bits fractional number representation and X_{S16} is signed 16 bits integer number.

2 Unit's communication

Very important part of whole system is communication between CUs. This communication can serve, for example, for exchange following data:

- Speed of rotor shaft,
- Torque current component of IM,
- Error state of IM,
- Error state of inverter.

2.1 Can BUS

CAN Bus serves for communication between unit's. This bus is widely used in vehicles, where all control parts of the vehicle are connected to this bus. DSC 56F8037 has implemented MSCAN, which offers [2]:

- CAN protocol–Version 2.0 A/B,
- Programmable bit rate up to 1 Mbps,
- Standard and extended data frames,
- 0-to-8 bytes data length.

Communication with CAN bus is realized with using various type of communication frames of programmable length [3]. It is multimaster bus type with arbitration, so when is bus free any unit can start transmitting. It supports error detecting too.

2.2 Data interchange process

Master-Slave configuration is very advantageous in this application, because offer simplicity of data interchange. One unit is selected as Master and second as Slave. Master will do main computations and decisions in coherence with control algorithm computations. Slave will only send the data to master and do control algorithm computations. Data interchange is in following of Fig. 5.

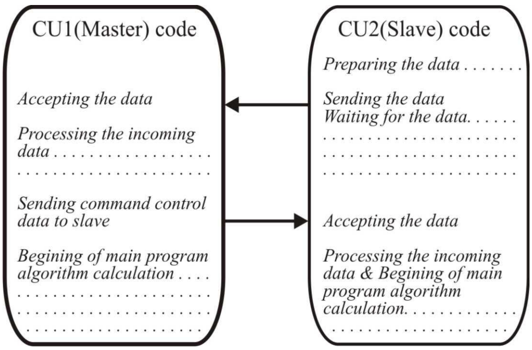


Fig. 5. Data interchange between units in Master-Slave configuration

3 Conclusion

Presented configuration can be used in electric vehicle with two independent IM and inverters with their control units. Used communication can lead to research in electronic differential. One disadvantage of Master-Slave configuration we see in the data processing and evaluate commands by only Master, but only two CUs are needed in comparison with alone master unit configuration.

Acknowledgements

This article was supported by Project No. SP2011/103 SGS.

References

1. Brandštetter, P. *Elektrické regulované pohony III.* Ostrava : VŠB-TU Ostrava, 2005.
2. Freescale. *56F802x and 56F803x Peripheral Reference Manual.* [Online] Texas : Freescale semiconductor, February 2007. http://www.freescale.com.
3. Robert Bosch GmbH. *CAN specification version 2.0.* Stuttgart : Robert Bosch GmbH, 1991. http://www.gaw.ru/.

Identification of Selected Parameters of Synchronous Machines with PM for the Purpose of Optimizing

Ladislav Slíva

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ladislav.sliva@vsb.cz

Abstract: This paper shows identification procedure selected parameters of synchronous machines with permanent magnet for optimization purposes. There is shown the steps of identification by analytical calculation, measurement, and also using the finite element method. There is also analysis of the magnetic induction in the air gap using FEM.

Keywords: synchronous machine, measurement, calculation, FEM, analysis, permanent, magnet, identification, parameters.

1 Introduction

Electrical machinery belongs to a group demanding disciplines of electrical engineering and are known are more than 100 years. Their theory was first written sometime around 1910 and so far the only rare exceptions, has not changed. Calculation methods are more precise, as well as calculations of cooling, ventilation calculations, mechanical calculations, and even machine parts. Using quality materials (particularly insulation), the performance of machines increase.

Not only at home but also abroad with both universities and private companies engaged in the improvement of various electric machines, especially the modern ones. In modern electric machine can be regarded as a synchronous machine with permanent magnet. Its use in practice gradually expanded and applied to the various electrical drives (trams, electric locomotives, etc.), among others, has a significant role in wind power as a synchronous generator with permanent magnet.

Before you start any engine optimization, you first need to understand its behavior in various configurations, and then find a possible way of optimizing. Can not only optimize the efficiency, which is very popular, but also as torque ripple or induced voltage. Very useful tool for optimization of electrical machines are used principally programs based on finite element method (FEM). Used as the 2D version, and the 3D version.

The actual work is concerned with optimization of synchronous machine with PM. In the first year of study I did identify the parameters of the replacement scheme (calculations, measurements, FEM) and also used the magnetostatic analysis in programming environments Ansys Workbench for the analysis of 3D magnetic circuit.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 190–195.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Analysis of Synchronous machine with PM

Analysis of electrical machines means to perform an analysis of its properties, ranging from analysis of parameters replacement scheme (Fig. 1). The analysis using analytical calculations, measurements, simulations and simulation programming languages, and more recently finite element method (FEM). Investigated engine is analyzed in the axis *d* and also in the axis *q,* because it has either expressed poles, or a magnet placed in the iron rotor (IPMS - Interior Permanent Magnet Synchronous Motor). Label real machines, including additional analysis of data in the following table (Tab. 1), while for obtaining the remaining data on the label usually are not trying to dismantle the machine thoroughly and measure the dimensions of magnetic circuit, including the size of the groove, count the number of wires in the groove etc.

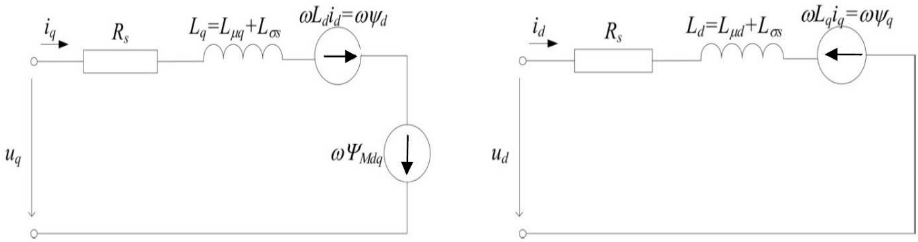


Fig. 1 - Equivalent circuit IPMS with fully internally stored magnets on the axis d and q

Tab. 1 - The label machines, including additional analysis of data identified

U_n	400 V	Rated voltage
I_n	8,3 A	Rated current
	Y	involvement
f_n	36 Hz	nominal frequency
n_n	360 1/min	Rated speed
Q_s	48	The total number of slots
Q	45	The total number of slots containing windings
q	1,25	Number of slots per pole and phase for Q
2p	12	number of poles
N_s	420	Number of turns of stator windings
l_i	0,14 m	Active length of iron
δ	0,8 mm	Air gap length
d	0,146 m	Outer rotor diameter
D	0,22 m	Outer diameter of stator
h_M	4,4 mm	magnet height
w_M	32,5 mm	Width of magnet
l_M	0,135 mm	Length of magnet

2.1 Analytical calculation

The analytical calculations can determine the following parameters: stator resistence R_s, leakage inductance $L_{\sigma s}$, magnetizing inductance $L_{\mu d}$ and $L_{\mu q}$, and while here I present only the final calculations. In fact, the analytical calculation is much more extensive. Were used known relations of the theory and construction of electric machines:

$$R_{s20^\circ} = \frac{l_c}{\sigma_{cu} \cdot a \cdot S_v} = \frac{205{,}268}{57 \cdot 10^6 \cdot 1 \cdot 0{,}88 \cdot 10^{-6}} = 4{,}07\Omega \tag{1}$$

$$R_{s75^\circ} = \frac{235+75}{235+20} \cdot R_{s20^\circ} = \frac{235+75}{235+20} \cdot 4{,}07 = 4{,}956\Omega \tag{2}$$

$$L_{\sigma s} = 2 \cdot \mu_0 \cdot \frac{N_s^2 \cdot l_i}{p \cdot q} \cdot \lambda = 2 \cdot 4 \cdot \pi \cdot 10^{-7} \cdot \frac{420^2 \cdot 0{,}14}{6 \cdot 1{,}25} \cdot 3{,}1519 = 0{,}026H \tag{3}$$

$$L_{\mu q} = \mu_0 \cdot \frac{2 \cdot m \cdot \tau_p}{p \cdot \pi^2 \cdot \delta_{qef}} \cdot l_i \cdot k_v^2 \cdot N_s^2 = \tag{4}$$
$$4 \cdot \pi \cdot 10^{-7} \cdot \frac{2 \cdot 3 \cdot 0{,}0386}{6 \cdot \pi^2 \cdot 0{,}00154} \cdot 0{,}14 \cdot 0{,}936^2 \cdot 420^2 = 0{,}0516H$$

$$L_{\mu d} = \mu_0 \cdot \frac{2 \cdot m \cdot \tau_p}{p \cdot \pi^2 \cdot \delta_{def}} \cdot l_i \cdot k_v^2 \cdot N_s^2 = \tag{5}$$
$$= 4 \cdot \pi \cdot 10^{-7} \cdot \frac{2 \cdot 3 \cdot 0{,}0386}{6 \cdot \pi^2 \cdot 0{,}005094} \cdot 0{,}14 \cdot 0{,}936^2 \cdot 420^2 = 0{,}02127H$$

2.2 Identification of the parameters measured

For the stator winding resistance measurement is appropriate to use the VA method. Each phase was measured separately. Measurements took place at a temperature t = 28°C. The median value of stator resistance R_s converted to 20°C is $R_{s20} = 3{,}93\ \Omega$ and at 75°C is converted to a value $R_{s75} = 4{,}78\ \Omega$.

Leakage reactance of stator winding $X_{\sigma s}$ (inductance $L_{\sigma s}$) can be measured using a method based on the general theory of electrical machines. Based on this theory it is possible to deduce that the non-rotating reactance $X_0 = X_{\sigma s}$ (Fig. 2).

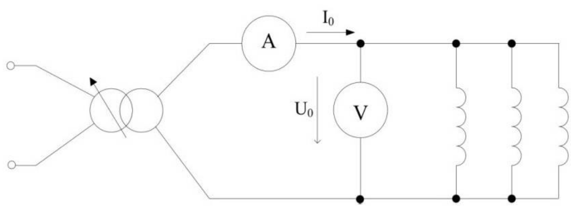


Fig. 2 - Circuit diagram for measuring non-rotating reactance IPMS

During the measurement of the rotor was braked, for power supply was used single-phase autotransformer. From the measured data of voltage U_0 and current I_0 can then calculate the leakage inductance by the relation

$$L_{\sigma s} = \frac{\sqrt{\left(3 \cdot \frac{U_0}{I_0}\right)^2 - R_s^2}}{2 \cdot \pi \cdot f} \tag{6}$$

The following table (Tab. 2) shows the measured and calculated values, where has the motor measured stator leakage inductance $L_{\sigma s}$ = 0,02897 H.

Tab. 2 - Measured and calculated values from measurement $L_{\sigma s}$

U_0 (V)	11.468	17.187	23.24	28.67	33.23
I_0 (A)	3.355	4.995	6.78	8.459	9.93
$L_{\sigma s}$ (H)	0.028649	0.029027	0.029089	0.028794	0.028419

When measuring the magnetizing inductance following the procedure, which is proved in [1]. In primarily on instead of AC in Figure 3a join DC source. Rotor synchronous machine is, aligns "with the axis. In this position, the rotor is braked SMPM (locks) to stayed be throughout this measurement in this particular position. Then perform the measurement in figure 3a.

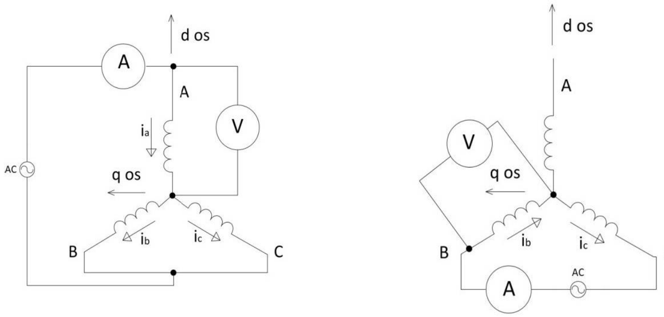


Fig. 3a, b - Schematic diagram for measuring $L_{\mu d}$ and $L_{\mu d}$

When investigating magnetizing inductance on the axis q is needed to implement according to diagram in fig. 3b. For measured parameters are needed oscilloscope, which displays the voltage and current and measuring instruments - ammeter, voltmeter. In the first phase corresponded to the measurement diagram in fig. 3a. The measured phase A was U_a = 142 V and current to phase A was I_a = 6,8 A . The oscilloscope was then determined by the phase shift between voltage and current, which was Δ_t = 3,88 ms, at a frequency 50 Hz which means voltage phase φ = 70°. $L_{\mu d}$ then will

$$L_{\mu d} = \frac{2}{3} \cdot \left(\frac{X_d}{2 \cdot \pi \cdot f} - L_{\sigma s} \right) = \frac{2}{3} \cdot \left(\frac{19{,}62}{2 \cdot \pi \cdot 50} - 0{,}02897 \right) = 0{,}02239 H \tag{7}$$

The following connection shown in Fig. 3b. Voltage phase B was U_b = 127,1 V and current was I_b = 3,95 A. Oscilloscope todetermine the shift between voltage and current that was Δ_t = 4,5 ms. This shift corresponds to the angle φ = 81°. $L_{\mu q}$ then will

$$L_{\mu q} = \frac{2}{3} \cdot \left(\frac{X_q}{2 \cdot \pi \cdot f} - L_{\sigma s} \right) = \frac{2}{3} \cdot \left(\frac{32{,}177}{2 \cdot \pi \cdot 50} - 0{,}02897 \right) = 0{,}04905 H \tag{8}$$

2.3 Analysis of machine parameters using FEM

Magnetizing inductance $L_{\mu d}$ and $L_{\mu q}$ can be easily analyzed using FEM. The following figures show the cross section of a synchronous machine with permanent magnets and on the axis d and q (Fig. 4a, b). Fig. 5a, b show the spatial distribution of magnetic induction B_δ in the air gap SMPM. The analysis was carried out in the FEMM 2D.

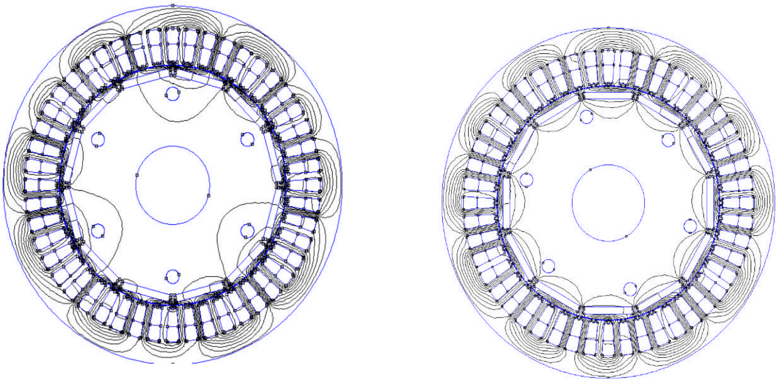


Fig. 4a, b - cross-section of synchronous machines with PM on the axis d and q

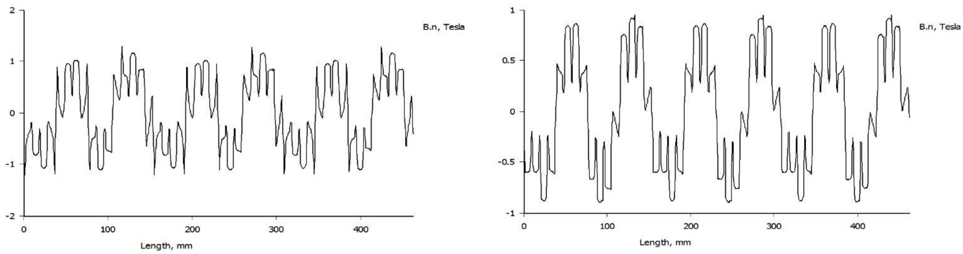


Fig. 5a, b - Spatial distribution of magnetic induction in the middle of the air gap on the axis d and q

Subsequently, on the program Ansys Workbench was constructed 3D model and solved in idle state for the purpose of analysis of a magnetic circuit.

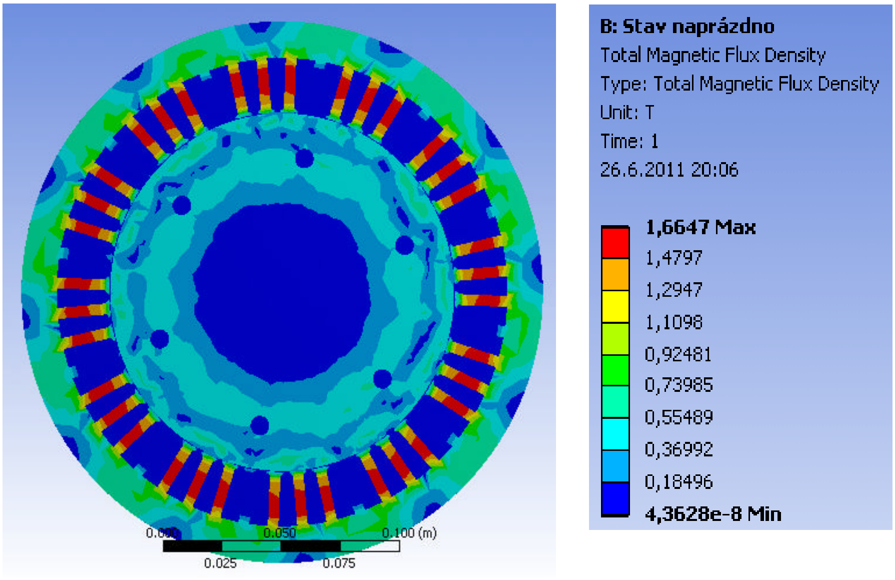


Fig. 6 - Status idle (3D model)

3 Conclusion

Parameters of the measurements obtained will be used in simulations to detect transient and steady states during operation SMPM. It will also serve as a reference to compare values when creating the different topology SMPM. From 3D model is also evident that the magnetic circuit could be better utilized.

References

1 Hrabovcová, V., Rafajdus, P., Franko, M., Hudák P.: *Measurements and modeling of electric machines*. 2nd. ed., Zilina, 2009, ISBN 978-80-8070-924-2.

2 Gieras, F.J., Wing, M.: *Permanent magnet motor technology*. 2nd ed., Dekker, 2002, ISBN: 0-8247-0739-7.

3 Pyrhonen, J.; Jokinen, T.; Hrabovcová, V.: *Design of Rotating Electrical Machines*. 1st ed., John Wiley & Sons Ltd., 2008, ISBN: 978-0-470-69516-6.

4 Ruzojic, B. ; Zarko, D.; Ban, D.: *Interior permanent-magnet motor for ship propulsion, design and testing*, Power Electronics and Applications. 2009. EPE 09. ISBN: 978-1-4244-4432-8.

Implementation of the Injection Method to Control Structure

Josef Szotkowski

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
josef.szotkowski@vsb.cz

Abstract. This paper describes the simulation procedure of sensorless control of induction machine using injection of high-frequency voltage signal. In several chapters, the description and design simulation control method is gradually introduced to a reader. The first two chapters deal with formation and classification of injection methods. Remaining chapters describe sensorless control with high-frequency signal injection method.

1 Introduction

Induction motor without speed sensor on the shaft does not cost much and ensures high reliableness. We have gained information of speed induction motor from measured stator voltage and currents in motor terminals. Injection methods have been developed not only for synchronous motor, but also for asynchronous motor. There exist two methods for injection of signals with asynchronous motors:

→ injection of auxiliary high-frequency signals
→ voltage testing impulses injection

Both methods follow specific motor anisotropy either to estimate the saturation magnetic flux or asymmetry of the folding rotor.

2 Distribution of injection method

As it is known, there are phenomena in the induction motor (asymmetry caused by saturation of magnetic cores, motor asymmetry caused by grooves, etc.) that are directly related to the position of the magnetic flux or rotor position. The main difference is the way of obtaining variable from the motor, so that we can use them for the estimation. All methods use some form of voltage or current signal injection. According to the type of injection, the methods can be divided into:

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 196–200.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

→ periodic signal injection
- synchronous pulsating signal injection
- high-frequency signal injection
- injection of synchronously pulsating high-frequency signal

→ discrete voltage pulses injection
- INFORM
- pulse injection test, so-called zero-sequence voltage
- pulse injection test, so-called zero-sequence current

The injection of discrete voltage pulses includes the use of switching inverter; we do not add any additional injected signal.

3 Voltage signal injection method

In figure 1 we can see the principle of injection of voltage signals. When injecting, symmetrical three-phase auxiliary voltage signal is used. The signal is modulated to the basic power supply. Frequency and amplitude of voltage signal must be chosen so that the problem with removing of the fundamental frequency in derivation of higher frequency current will not occur. Amplitude must be chosen so as not to influence the motor activity itself.

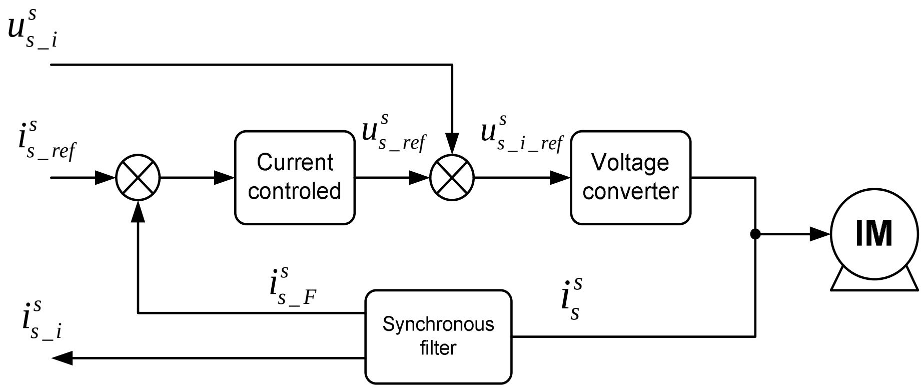


Figure 1: Voltage signal injection

4 Model of induction motor

The model of induction motor at higher frequencies can be derived from the motor voltage equations in the stator reference frame [α, β]:

$$\mathbf{u}_S{}^S = R_S \mathbf{i}_S{}^S + \frac{d\boldsymbol{\psi}_S{}^S}{dt} \qquad 0 = R_R \mathbf{i}_R{}^S + \frac{d\boldsymbol{\psi}_R{}^S}{dt} - j\omega_m \boldsymbol{\psi}_R{}^S \tag{1}$$

$$\boldsymbol{\psi}_S{}^S = L_S \mathbf{i}_S{}^S + L_h \mathbf{i}_R{}^S \qquad \boldsymbol{\psi}_R{}^S = L_h \mathbf{i}_S{}^S + L_R \mathbf{i}_R{}^S \tag{2}$$

Total stator and rotor inductances L_S, L_R are defined as follows:

$$L_S = \left(L_h + L_{S\sigma}\right) \quad L_R = \left(L_h + L_{R\sigma}\right) \quad L_{S\sigma} = L_S - \frac{L_h{}^2}{L_S L_R} \tag{3}$$

Decrease in stator and rotor resistance R_S, R_R at higher frequencies can be neglected. For a constant angle frequency of the injected signal ω_i, we can write the following equation:

$$0 \cong j\left(\omega_i - \omega_m\right)\left(L_h \mathbf{i}_S{}^S + L_R \mathbf{i}_R{}^S\right) \tag{4}$$

$$\left(\omega_i - \omega_m\right) \neq 0 \quad \Rightarrow \quad \left(L_h \mathbf{i}_S{}^S + L_R \mathbf{i}_R{}^S\right) = 0 \tag{5}$$

$$\mathbf{i}_R{}^S \cong \frac{L_h}{L_R} \mathbf{i}_S{}^S \; \mathbf{u}_S{}^S \cong j\omega_i L_{S\sigma} \mathbf{i}_S{}^S \tag{6}$$

If there is a magnetic asymmetry in induction motor depended on the position of the rotor, stator leakage inductance is not constant and it can be defined as a functional dependence on the position of the rotor. In the stator reference system [α, β] it can be expressed by the equation:

$$\begin{bmatrix} u_{S\alpha} \\ u_{S\beta} \end{bmatrix} = j\omega_i \left\{ \begin{bmatrix} L_{S\sigma} + \Delta L_{S\sigma}\cos\left(h\varepsilon\right) & -\Delta L_{S\sigma}\sin\left(h\varepsilon\right) \\ -\Delta L_{S\sigma}\sin\left(h\varepsilon\right) & L_{S\sigma} + \Delta L_{S\sigma}\cos\left(h\varepsilon\right) \end{bmatrix} \begin{bmatrix} i_{S\alpha} \\ i_{S\beta} \end{bmatrix} \right\} \tag{7}$$

5 High-frequency signal injection

The voltage signal with a higher angular frequency ω_i and constant amplitude U_i can be defined as follow:

$$\begin{bmatrix} u_{S\alpha\_i} \\ u_{S\beta\_i} \end{bmatrix} = U_i \begin{bmatrix} \cos\left(\omega_i t\right) \\ -\sin\left(\omega_i t\right) \end{bmatrix} = U_i \, e^{j\omega_i t} \tag{8}$$

If the induction motor is of asymmetrical leakage inductance, it will be excited with injected voltage signal, the injected current will be through motor stator, and this injected current can be defined as follows:

$$\mathbf{i}_{S\_i}^{S} = \begin{bmatrix} i_{S\alpha\_i} \\ i_{S\beta\_i} \end{bmatrix} = I_{i\_p} \begin{bmatrix} \sin(\omega_i t) \\ \cos(\omega_i t) \end{bmatrix} - I_{i\_n} \begin{bmatrix} \sin(h\varepsilon - \omega_i t) \\ \cos(h\varepsilon - \omega_i t) \end{bmatrix} \tag{9}$$

$$\mathbf{i}_{S\_i}^{S} = -jI_{i\_p} e^{j\omega_i t} + jI_{i\_n} e^{j(h\varepsilon - \omega_i t)} \tag{10}$$

$$I_{i\_p} = L_{S\sigma} \frac{U_i}{\omega_i \left(L_{S\sigma}^2 - \Delta L_{S\sigma}^2\right)} \qquad I_{i\_n} = \Delta L_{S\sigma} \frac{U_i}{\omega_i \left(L_{S\sigma}^2 - \Delta L_{S\sigma}^2\right)} \tag{11}$$

High-frequency injected current contains two basic components. I_{i_p} (positive component which does not contain information of rotor position) and I_{i_n} (negative component which includes information of motor position).

6 Implementation of the injection method

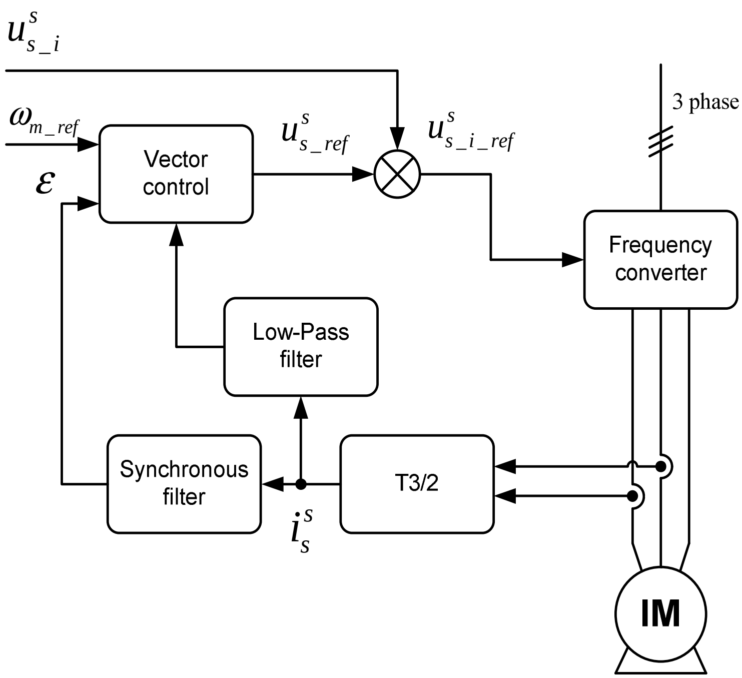


Figure 2: Implementation of the injection method to control structure.

Figure 2 shows the implementation of the injection method to control structure. Auxiliary voltage (injection) is added to the base voltage. The frequency inverter must be able to produce the injected signal with higher frequencies, which results in increased losses in the inverter and also in the induction motor. If we use the injection method to control the entire speed range, there is a reduction in motor speed. Therefore, this method can be used to control low motor speed. To determine the current with higher frequency and to obtain information of rotor position, so-called “synchronous filter” was used. Its activity is based on equation (9) and (11).

7 Conclusion

Main purpose of the article was to present high-frequency injection method for induction motor and its implementation in the control structure. In each chapter I try to describe the basic parts (blocks) that are associated with this proceeding.

8 Acknowledgment

In the paper there are the results of the project SP2011/107 which was supported by Student Grant Competition of VSB-Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

References

1. Brandštetter, P.: “*A.C. Control Drivers – Modern Control Method”,* VŠB Technical University of Ostrava, 1999. ISBN 80-7078-668
2. Neborák, I.: “*Střídavé regulační pohony – Moderní způsoby řízení*”, VŠB Technical University of Ostrava, 2002.
3. Bose,B.K.: “*Modern Power Electronics and AC Drives”,* Knoxville 2002, ISBN 0-13-016743-6.
4. Holtz, J., *“Sensorless Control of Induction Motor Drives”,* Proceedings of the IEEE, Vol. 90, No. 8, pp. 1359-1394, 2002.
5. Ha, J. I., Sul S. K., *“Sensorless Field-Orientation Control of an Induction Machine by High-Frequency Signal Injection”*, IEEE Transactions on Industry Applications, Vol.35, No. 1, pp. 45-51, 1999.

Interconnection of DSC TMS320F28335 and PROFIBUS

Daniel Kalvar

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
daniel.kalvar@vsb.cz

Abstract. This article deals with the PROFIBUS communication and its possibilities of connection to the DSC TMS320F28335. At first is presented the ISO/OSI model, subsequently is analyzed the basic properties of the PROFIBUS including the most often used applications. Then is explained the principle and communication settings for DSC TMS320F28335 through the SCI interface. At the end of this article is outlined possible hardware implementation. It consists mainly in adapting the appropriate voltage levels for connection to the PROFIBUS fieldbus. In practice is most often used for this purpose as a transmitter integrated circuit from one of a number of manufactures.

Keywords: PROFIBUS, ISO/OSI model, DSC, SCI.

1 Introduction

The tendency toward decentralized automation is strongly influencing the technical development of the 90s. Modern microelectronics offers developers of technical information systems nearly unlimited possibilities in the realization of innovative products. The availability of higher performance at relatively low costs enables automation functions to be realized beyond the domain of the central processing unit, for example in machines, converters, aggregates, sensors and actuators.

1.1 The ISO/OSI model

Modern communication systems are designed according to the International Organization for Standardization (ISO)/OSI reference model (ISO 7498) which structures the communication functionality into 7 layers (Figure 1). Each layer extends over the whole communication system and determines a layer-specific protocol that is agreed upon by the systems devices. Services are provided at the interface to the next higher level. With these services the user has a transparent communication system. Which allow simple and efficient data transfer between stations. A transparent communication system means, that the message transmitted over this system are neither disorted or otherwise modified.

The entire communications process is partitioned into a hierarchy of 7 layers to be traversed one by one the message either top down or vice versa during the communication. Every layer provides services to the next higher layer and in turn

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 201–206.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

uses services of the next lower layer to execute its tasks. Thus, within one station, each layer exchanges information only with adjacent layers.

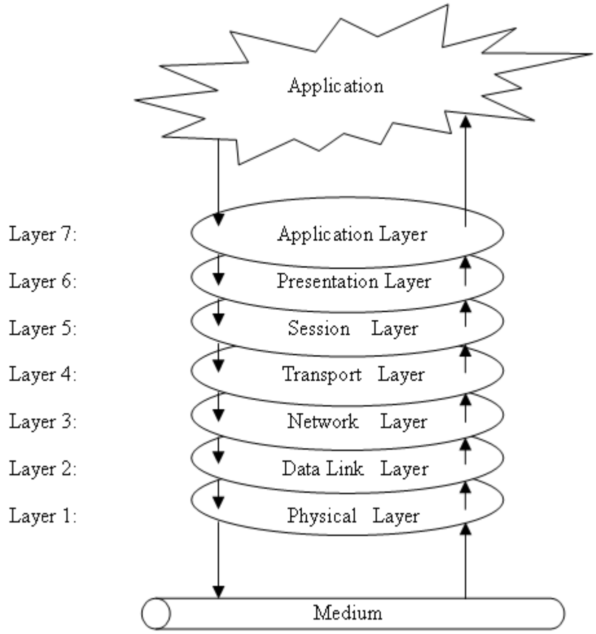


Figure 1 The ISO/OSI basic reference model

2 PROFIBUS

Industrial Profibus (Process Field Bus) is designed for the automation of production lines (automotive, filling lines, storage systems), for home automation (air conditioning, heating), for process automation (chemical and petrochemical industry, paper industry, sewage treatment plants), production management and energy distribution. The history of using this successful bus for industrial began in fact in 1987 when the company Bosch, Klöckner & Möller, Siemens submit a project Profibus of German government. This concept for automation of discrete and continuous production supported the project and member organizations of the Central Association of German Electrical Industry (Zentralverband Elektrotechnik und Elektronikindustrie - ZVEI). Actual association of persons interested in the bus (Profibus Nutzerorganisation) was founded in November 1989 and has since grown into the largest association (Profibus International) in the fieldbus.

2.1 Architecture of PROFIBUS

PROFIBUS standard, as was already mentioned in the introduction, based on the ISO/OSI model. Because of the time optimization of this model defines only the physical layer, line, and application (see Figure 1). Currently there are three variants of industrial Profibus, each of which is designed for specific use.

PROFIBUS DP (Decentralized Periphery)

This is the simplest and most common variant of PROFIBUS. It is used as a sensor bus. It is designed for quick communication master-slave. It is particularly suitable for rapid transmission of signals from the process with decentralized peripherals and remote I/O units. Communication medium is either twisted pair (standard RS-485), or optical fiber at speeds up to 12 Mbits / s.

PROFIBUS FMS (Fieldbus Massage System)

It belongs to a group of fieldbus. It offers a communication standard for communication in a heterogeneous environment with a large set of services for working with data, programs, and alarms. Communication medium is similar to the DP variant Profubus either twisted pair (standard RS-485), or optical fiber, but the speed is already lower.

PROFIBUS PA (Process Automation)

It belongs to a group of Fieldbus. It uses an extended standard PROFIBUS DP is used to control slow process especially in hazardous environments, is responsible for intrinsic safety. In order to use the network in this environment, it is also used for special physical layer - the current loop according to the standard IEC 1158-2 communicating a constant speed of 31,25 kbit/s.

2.2 Transmission technology

RS-485 (high speed – H2)

As physical media is used shielded twisted pair cable. In this variant is used encoding type NRZ (Non Return to Zero). To transfer the character encoding used 11 b-bit type UART (Universal Asynchronous Receiver Transmitter) with one start bit, one stop bit, 8 information bits and one parity bit (even parity) according to DIN 19244. Total 32 stations can be connected in one segment, but a total of 127 stations all together. Maximum length PROFIBUS ranges from 100 m (at a transmission speed of 12 Mbit/s) up to 1200 m (at a transmission speed of 9,6 kbit/s) depending on the choice of transmission speed. Using repeaters can this length extended be up to 10 km.

Transmission rate[kbit/s]	9,6	19,2	45,45	93,75	187,5	500	1500	3000,6000,12000
Length[m]	1200	1200	1200	1200	1000	400	200	100

Table 1 PROFIBUS transmission rate depending on the length of line

Optical fiber - PROFIBUS DP/FMS

Nowadays, used both glass fiber (up to 2 km), and inexpensive plastic optical fiber (length up to 50 m). These optical fibers are now available as special connectors with integrated converter RS-485 to the optical interface and vice versa, which makes it easy to combine copper and fiber distribution in a network.

3 SCI interface on DSC TMS320F28335

Connection between the DSC (Digital Signal Controller) and PROFIBUS is implemented by the DSC via SCI Interface. The devices include three Serial Communications Interface (SCI) modules. The SCI modules support digital communications between the CPU and other asynchronous peripherals that use the standard non-return-to-zero (NRZ) format. The SCI receiver and transmitter are double-buffered, and each has its own separate enable and interrupt bit. Both can be operated independently or simultaneously in the full-duplex mode. To ensure data integrity, the SCI checks received data for break detection, parity, overrun, and framing errors. The bit rate is programmable to over 65000 different speeds through a 16-bit baud-select register.

Features of each SCI module include:

• Two external pins:

– SCITXD: SCI transmit-output pin,

– SCIRXD: SCI receive-input pin.

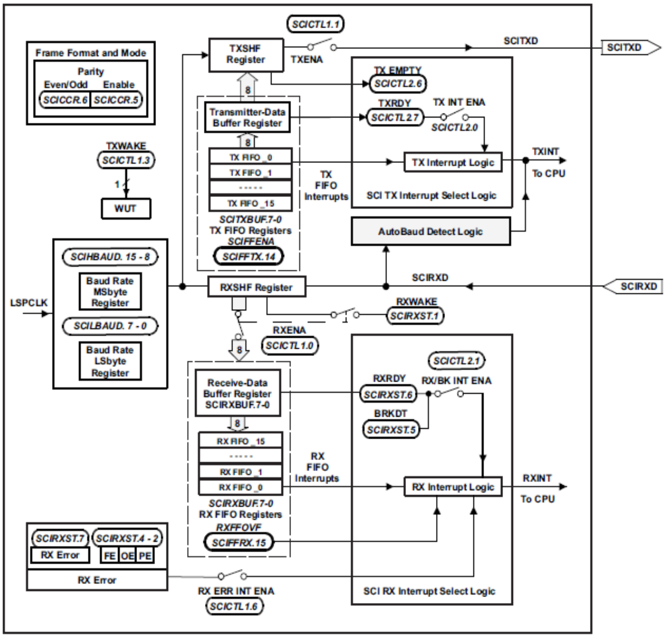


Figure 2 Serial Communication Interface (SCI) block diagram

4 Communication DSC with PROFIBUS

To adjust the voltage levels are in practice usually using transmitters in the form of integrated circuits, which offers many different manufacturers. One of these circuits can be as LTC485. The LTC485 is a low power differential bus/line transceiver designed for multipoint data transmission standard RS-485 applications with extended common-mode range (12V to - 7V). Typical application wiring of this circuit is shown in Figure 3.

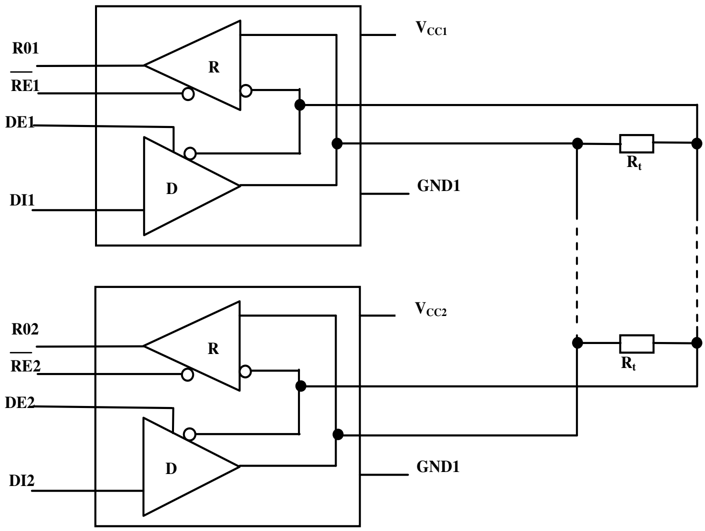


Figure 3 Circuit diagram of typical application

Figure 4 gives a block diagram of the DSC circuit TLC485. This connection is realized by the CPU via the SCI interface UART protocol. TLC485 integrated circuit provides voltage adjustment for standard RS-485.

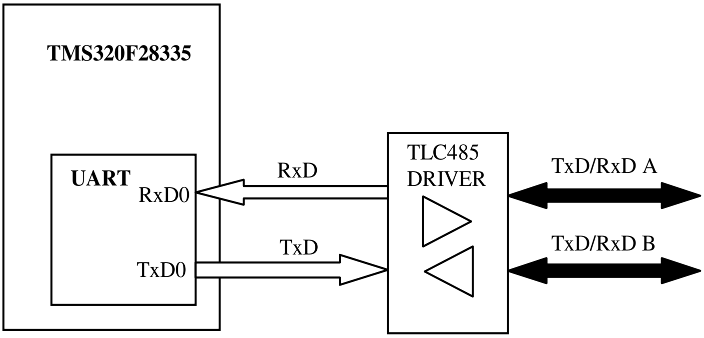


Figure 4 Block diagram of communication DSC and PROFIBUS

5 Conclusion

This article summarizes the basic information on the PROFIBUS including the ISO/OSI basic reference model. As were it provides information on options and settings DSC TMS320F28335 for communication through the SCI interface. Subsequently, was outlined the possibility of interconnection between DSC and PROFIBUS using integrated circuit LTC485. This conception shows possible way of technology implementation for the ENET project. ENET will be solved by the Department of Electronics. The above mentioned matters will be solved by the author of this article. Practical results will be realized and presented later due to timetable of the project.

References

1. Bender K.: PROFIBUS – The Fieldbus for Industrial Automation, Cambridge: The University Press, Cambridge, 1999. ISBN 0-13-012691-8
2. Profibus _system_description_v_2010_English.pdf, <http://www.profibus.com/nc/downloads/downloads/profibus-technology-and-application-system-description/display/>
3. Texas Instruments, Datasheet TMS320F28335, <http://focus.ti.com/lit/ds/sprs439i/sprs439i.pdf>
4. Linear Technology Corporation, Datasheet LTC485, <http://cds.linear.com/docs/Datasheet/485fi.pdf>

Novel Wien Bridge Oscillator Design Using Functional Block Structure with Current Conveyors

Lukáš Klein

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.klein@vsb.cz

Abstract. The purpose of this paper is to generally present possibilities of the current conveyor use in the well-known active electronic circuits. Current conveyors are able to substitute all known active elements. This claim is supported by the fact, that four basic functional block structures can be realized by use of current conveyors - voltage-controlled voltage source, voltage-controlled current source, current-controlled current source and current-controlled voltage source. The paper presents particular example of use of functional block structure with current conveyors in Wien bridge oscillator, where it successfully substitutes operational amplifier on the place of active element. Both theoretical formulae and design description are given. Finally, OrCAD PSpice simulation results are presented.

Keywords: current conveyor, functional block structure, Wien bridge oscillator

1 Introduction

Currently, current conveyor is already well-known active element. The first generation current conveyor (CCI) was introduced in 1968 [1]. However, the electronic industry did not require wider expansion of more current conveyor variations and their wider application in commercially produced electronic circuits and devices.

Second generation current conveyor (CCII) has received the greatest attention of the three existing current conveyor generations ever [2], particularly non-inverting positive second generation current conveyor (CCII+). It also remains the only available current conveyor variation, not counting the universal current conveyor (UCC) [3].

The number of advantages of this active element dominates over disadvantages. Positive properties of current conveyor include low voltage supply, wide frequency range, improved noise immunity, improved circuit dynamics and especially easy integrability of the resulting circuit structures using these elements. The last point comes from the fact that the principle and internal structure of current conveyor are relatively simple. Therefore the design of many modern active elements is based on the current conveyor internal structure and directly or indirectly derived from it. Let us mention for example current feedback amplifier (CFA), whose concept is based right on the current conveyor structure [4].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 207–212.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

The paper tries to theoretically verify the versatility of block structures with current conveyors and their application in electronic circuits on place of standard applied active elements.

2 Principle and Variations of Current Conveyors

General three-port and four-port current conveyor belong to best-known and most used current conveyors (especially in theory). Three-port current conveyor can be found even in real form. Some circuit structures of sophisticated integrated circuits include variation of current conveyor CCII+ as a part of complex circuit. CCII+ input and output terminals are terminated in a limited number of these integrated circuits, so current conveyor block can be practically used. Four-port current conveyor can be currently implemented only using the UCC [3]. However, UCC is still not yet commercially available and it is not sure, if ever UCC will be produced commercially in the future. Schematic symbol of general three-port current conveyor is shown in Fig. 1a, schematic symbol of general four-port current conveyor is shown in Fig. 1b.

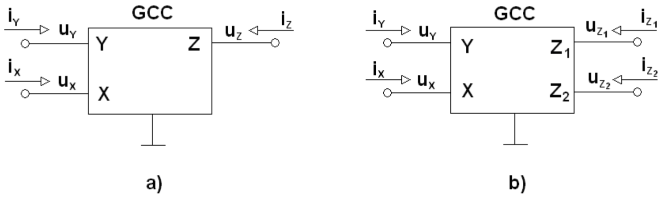


Fig. 1. Schematic symbol of general a) three-port b) four-port current conveyor

Terminals labeled X represent current inputs, terminals Y are voltage inputs. Terminals Z represent current outputs with positive or negative transfer of current from terminal X [5]. In a case of general current conveyor, voltage is also conveyed from terminal Y to terminal X. Selected current conveyor generations (specifically CCI and CCII) moreover convey current from terminal X to terminal Y [4]. These relations can be simply expressed by equations:

$$K_{U_{YX}} = \frac{u_X}{u_Y}\ ,\ \ K_{I_{XY}} = \frac{i_Y}{i_X}\ ,\ \ K_{I_{XZ}} = \frac{i_Z}{i_X}\ . \tag{1a,b,c}$$

These transfers are referred to the coefficients *a, b, c* in the technical literature. Specifically, $K_{U_{YX}} = a$, $K_{I_{XY}} = b$, $K_{I_{XZ}} = c$. The coefficients becomes ideally values of a={-1,1}, b={-1,0,1}, c = {-1,1}, but they may differ from these values in practice. The coefficients *a, b, c* figure in the current conveyor definitional equations. For general three-port current conveyor they have the form:

$$u_X = a \cdot u_Y\ ,\ \ i_Y = b \cdot i_X\ ,\ \ i_Z = c \cdot i_X\ . \tag{2a,b,c}$$

Specific values of the coefficients in the equations stated above directly define the current conveyor variation according to the following rules [6]. If the coefficient *a*=1, it is a non-inverting current conveyor. However, if coefficient *a* will take value -1, it is an inverting current conveyor.

Coefficient *b* sets generation of current conveyor. The first generation current conveyors take coefficient value *b*=1. The second generation current conveyor is defined by coefficient value *b*=0. The third generation current conveyors take coefficient value *b*=-1.

Coefficient *c* determines, if it will be a positive or a negative current conveyor. If the coefficient value *c* equals 1, it is a positive current conveyor. On the other hand, when *c*=-1, it is a negative current conveyor. Overview of three-port current conveyor variations is clearly presented in Tab. 1 [6].

Table 1. Current conveyor variations depending on the coefficients *a*, *b*, *c*

	Coefficients		
Conveyor type	*a*	*b*	*c*
CCI+	1	1	1
CCI-	1	1	-1
CCII+	1	0	1
CCII-	1	0	-1
CCIII+	1	-1	1
CCIII-	1	-1	-1
ICCI+	-1	1	1
ICCI-	-1	1	-1
ICCII+	-1	0	1
ICCII-	-1	0	-1
ICCIII+	-1	-1	1
ICCIII-	-1	-1	-1

3 Functional Block Structures with Current Conveyors

As stated above, current conveyors can be considered as universal elements. This claim stems from the fact, that four basic active functional block structures can be implemented by using current conveyors, as described in [4]. These four basic functional block structures theoretically allow to substitute the basic types of active elements in electronic circuits. This fact is supported by the idea, that all existing active elements can be basically divided into four main groups, namely:

- voltage-controlled voltage source,
- voltage-controlled current source,
- current-controlled current source,
- current-controlled voltage source.

These four groups of active elements are identical with possible functional block structures realizable using current conveyors. For this reason, current conveyors can be considered as universal elements and it is possible to implement all of these types of active elements by using current conveyors [4].

The functional block structure belonging to voltage-controlled voltage source is shown in Fig. 2. This functional block structure is able to substitute operational amplifier, because this active element is also generally classified as voltage-controlled voltage source.

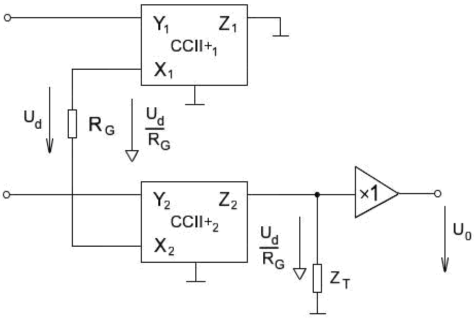


Fig. 2. Functional block structure corresponding to voltage-controlled voltage source

Input terminals of structure Y_1 and Y_2 represent the inverting and non-inverting voltage inputs. Terminal Z_2 represents output and Z_1 have to be grounded [4]. Output signal then flows through voltage follower - a voltage amplifier with unit gain. Resistor R_G defines a total gain of functional block structure. It is evident from formula (3). Z_T is transimpedance, U_d is input differential voltage between inverting and non-inverting terminal and U_0 is output voltage of functional block structure [4].

$$K_U = \frac{U_0}{U_d} = \frac{Z_T}{R_G} \tag{3}$$

The structure is constructed from CCII+, which is the only commercially available variation of current conveyor. It offers the possibility of practical realization and subsequent application in electronic circuits.

4 Proposed Wien Bridge Oscillator using Current Conveyors

There was used well-known circuit of Wien bridge oscillator to verify the theory stated above. There was used functional block structure using two CCII+ described in the previous chapter as the active element in the circuit structure of the oscillator. The structure practically only substitutes an operational amplifier in the circuit. The resulting circuit solution is shown in Fig. 3.

Resistors R and capacitors C shown in schematic diagram make up own circuit of Wien bridge. It is known, that Wien bridge has transfer value 1/3. It is possible to define the frequency of the output signal by suitable choice of passive element values. Circuit connection of resistors R_1, R_2 with an active element make up the non-inverting amplifier. The amplifier has to amplify with gain least 3 to compensate the attenuation of Wien bridge [7]. Calculation of passive element values was performed according to the basic equations:

$$f_0 = \frac{1}{2\pi RC}\,,\quad U_2 = \left(1 + \frac{R_1}{R_2}\right)\cdot U_1\,. \tag{4a,b}$$

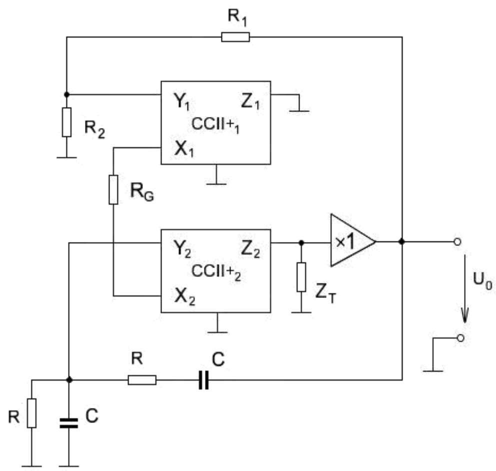


Fig. 3. Wien bridge oscillator using functional block structure with current conveyors

There was chosen the output signal frequency 1MHz for final solution of Wien bridge oscillator using functional block structure with current conveyors. There were given the values of passive elements R=3.3kΩ, C=47pF, R_1=5kΩ and R_2=1kΩ according to the equations stated above.

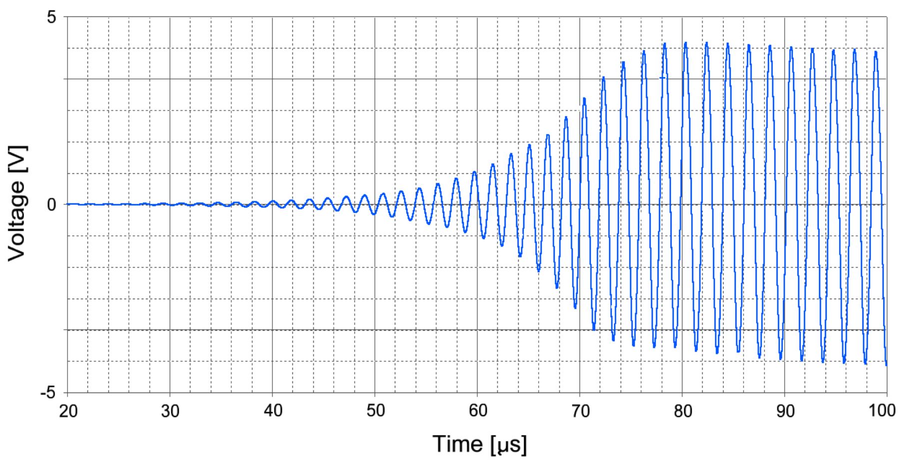


Fig. 4. Output response of Wien bridge oscillator with CCII+

OrCAD PSpice simulation was performed to verify the behavior of the resulting solution of oscillator. The final oscillator output signal response is shown in Fig. 4. There was not added the amplitude stabilizer to the complex circuit solution, therefore output signal amplitude is not stable. However, the output signal response clearly shows the emergence of oscillations that have the frequency value of 1MHz.

5 Conclusion

The modest paper tries to describe basic application possibilities of current conveyors and to demonstrate their possible universal use. There was chosen one of four functional block structures representing the voltage-controlled voltage source as an example. The structure was subsequently used in the circuit of Wien bridge oscillator on the place of active element, where it successfully substituted the operational amplifier. Functional block structures using current conveyors bring many advantages. One of the many is the fact, that the resulting circuits can be used at higher frequencies. In our case, there was demonstratively chosen frequency value of 1MHz. Another advantage is the low voltage supply of current conveyors.

On the other hand, low commercial availability belongs to disadvantages of current conveyors. However, there can be realized a number of interesting circuit solutions using only CCII+, because four basic active functional block structure can be assembled. Therefore, this variation of current conveyor can be considered as universal.

Acknowledgment

In the paper there are the results of the project SP2011/107 which was supported by Student Grant Competition of VSB-Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandstetter, CSc.

References

1. Smith K.C., Sedra A., *The current conveyor: a new circuit building block*, IEEE Proceedings of the CAS, Vol. 56, pp 1368-1369, 1968.
2. Smith K.C., Sedra A., *A second-generation current conveyor and its applications*, IEEE Transactions on Circuit Theory, Vol. 17, pp. 132-134, 1970.
3. Zhang Q. J., Wang C.H., Leng Y., *A Novel Building Block: Universal Current Conveyor*, Conference Proceedings of IEEE 8th International Conference on ASIC, Changsha, China, 2009, Vols 1-2, pp. 1109-1112, ISBN 978-1-4244-3868-6.
4. Punčochář J., *Historie a současnost operačních zesilovačů*, BEN Praha, Praha, 2006, ISBN 80-7300-047-4.
5. Čajka J., Dostál T., *Nové názvosloví a sjednocující pohled na proudové konvejory*, http://www.elektrorevue.cz, Brno, 2001, No. 24. ISSN1213-1539.
6. Kubánek D., Šponar R., Vrba K., *Filtrační obvody s napěťovými konvejory – Tematický okruh G1*, Ústav telekomunikací FEKT VUT v Brně, FEKT VUT v Brně, Brno, 2004.
7. Brandštetter P., *Elektronika*, VŠB-TU Ostrava, Ostrava, 1991, ISBN 80-7078-124-8.

Digital Filter Implementation to DSP Systems

Josef Opluštil

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
josef.oplustil@vsb.cz

Abstract. This paper deals with a comparison of digital and analog filters and digital filter implementation to DSP. In this paper is also solved design of digital filters, their behavior and their applications. Digital filters are investigated in simulations as well as implemented in DSP. For DSP there is a time of execution studied.

Key words: digital filter, FIR, IIR, DSP, TMS320f2812, matlab.

1 Introduction

Today we are more frequently using digital processing to problems, which were solved by analog technology. It is because there is big expansion of digital technologies. In these applications, the analog solution is even simple then the digital one. However, digitally we can solve this on current device without other additional elements.
In these days of miniaturization the producers more like use current hardware (digital) to perform as much as possible functions by one piece of hardware (processor, FPGA ...). Digital filter is one of the elementary applications for these systems.

2 Digital filters

Digital filter is an algorithm or electronic circuit, which converts the frequency spectrum of input. It can be realized by special circuit or by the software. Digital filter must be able to calculate the convolution algorithm in between two input samples.

There are the elementary parts of digital filters:
a)sum, b)gain, c) unit delay.

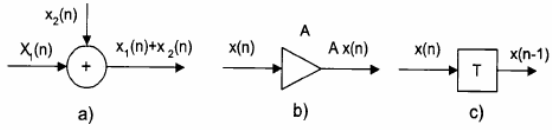


Fig. 1. Digital filter parts

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 213–218.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

There are high precision, easy simulation and design, possibility of adaptive filtration and no drift as advantages of digital filters.
The disadvantages are high requirements to computing performance and it's impossible to use them for high frequencies.

There are two basic digital filters classified by length of the impulse response.

2.1 Finite impulse response filter (FIR)

This type of digital filters does not contains feedback. It is nonrecursive filter. Its behavior is described by equation followed:

$$y(n) = b_0 \cdot x(n) + b_1 \cdot x(n-1) + b_2 \cdot x(n-2) + \ldots + b_M \cdot x(n-M) = \\ = \sum_{i=0}^{M} b_i \cdot x(n-i) \tag{1}$$

The structure of FIR looks like this:

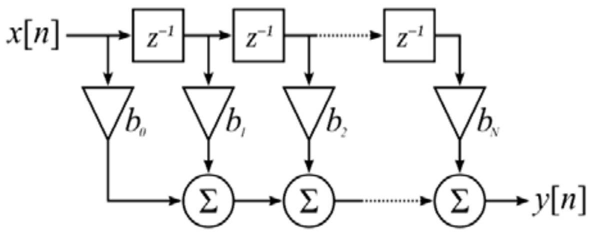


Fig. 2. FIR filter structure

2.2 Infinite impulse response filter (IIR)

This filter needs unless one feedback loop for its function. It is a recursive filter. Filter order for IIR is much smaller then for FIR filters and that is the reason for their faster response.
These filters are described by equation followed:

$$y(n) = \sum_{i=0}^{M} b_i \cdot x(n-i) + \sum_{i=0}^{N} a_i \cdot y(i-N) \tag{2}$$

The structure of IIR can look like this:

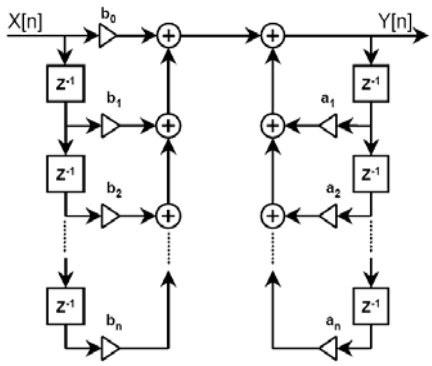


Fig. 3. IIR filter structure

3 Digital filter design

Digital filters are mostly designed by using the computer programs, which are created for this purpose. There is described design in matlab software cursorily.

Generally, digital filter design proceeds in the following steps:

- Filter type choice
- Filter order choice
- Requirements checking

If the proposed filter does not comply, whole procedure is repeated for higher filter order.

Basically the filter design is consists of two steps:

- Definition of required frequency response
- Calculation of the coefficients (calculating the impulse response from last step) it is a calculation of the inverse Fourier transform

We can use the window functions too, it is used to suppress the oscillations. There are many possible window functions (such as Hamming, Hanning, Kaiser and other).

3.1 FIR filter design by Matlab

If we need very quickly calculation of coefficients, we can use Matlab. This software contains function to quick design of any type of FIR filter. It is FIR1 command and it can calculate the coefficients for all of the most common filters. Such as high-pass, low-pass, band-pass a band-stop filters.
It can also directly multiply coefficients by any window function what we need.

The application of this function is below:

bn=fir1(FO,NF,'T',W) **(3)**

The meaning of variables:
bn … coefficients array
FO … filter order
NF … normalized frequency
T … type of the filter (high for high-pass, low for low-pass, stop for band-stop)
W … type of the window function, if it is empty then hamming window will be used

Example:
There is an example for using this function for design the 50th. order of band-pass FIR filter with 3800Hz low cut frequency and 15400Hz high cut frequency. The coefficients are multiply by rectangular window.

bn=fir1(50,[0.1725,0.7],rectwin(51))

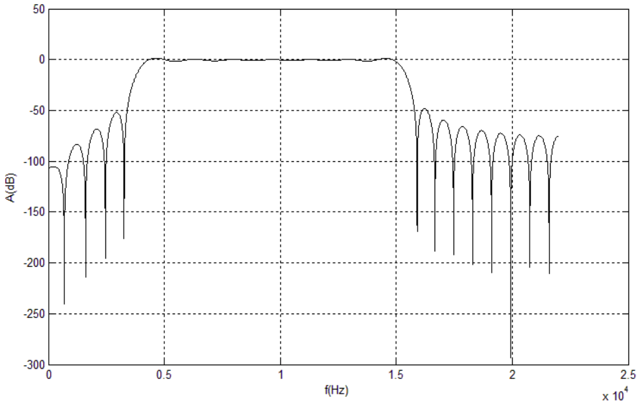


Fig. 4. Designed filter in matlab

4 FIR filter implementation to TMS320F2812

There is the digital signal processor TMS320F2812 used for demonstration of implementation. It is situated on the Spectrum Digital development kit.
To this kit there is connected an additional board with DAC and analog signal conditioning (voltage adjustments to ±10V).
There are practically three possibilities to implement digital filter into the DSP.

- Using C language for whole software
- Using assembly language for whole software
- Combine both

The C language has one big advantage, and it is the speed how we are able to write whole program. But its big disadvantage is that we completely count on compiler. We can never know how it compile our program. Software written in C language always run slower than the assembly one. If we want write whole program in assembly language, we need to know everything about the processor we used and final program will be very difficult to orient in it. The best possibility is to combine C and assembly language. The most of the program we can write in C and for time demanding function, we can use assembly language. The most advantage of DSPs is that it has some instructions designed just for these operations we need. Usually most of instructions consume only one cycle.

4.1 FIR filter realization in C language

FIR filter realization in C language is very simple. It consists only from one FOR cycle.

```
vystup = 0;
 for(i=0;i<(rad_filtru-1);i++)
 {
   vystup = vystup + (int16)(((int32)vzorky[j] * (int32)h_n[i])>>11);
   j++;
    if(j>(rad_filtru-1))
   {
    j = 0;
   }
```

Variables meaning:

vystup	…	final result y(n)
vzorky	…	array of input samples
h_n	…	array of coefficients
i,j	…	other variables

We need 1804 cycles to realize one sample of order 50 FIR filter.

4.2 FIR filter realization in assembler

FIR filter realization in assembly language looks like this:

```
MOVW DP,#Xpointer ;                Load DP with page address of Xpointer
MOVL XAR6,@Xpointer ;              Load XAR6 with current X pointer
MOVL XAR7,#C ;                     Load XAR7 with start address of C array
MOV @AR1,#N ;                      Load AR1 with size of data array N,
SPM −4 ;                           Set product shift mode to ">> 4"
ZAPA ;                             Zero ACC, P, OVC
RPT #N−1 ;                         Repeat next instruction N times
||QMACL P,*AR6%++,*XAR7++ ;        ACC = ACC + P >> 4,
                                   ; P = (*AR6%++ * *XAR7++) >> 32
ADDL ACC,P << PM ;                 Final accumulate
```

MOVL @Xpointer,XAR6 ; *Store XAR6 into current X pointer*
MOVL @Sum,ACC ; *Store result into sum*
This is code for FIR filter by Texas Instruments.

For the same filter we need only 61 cycles. This time is achieved by using the MAC instruction in software together with parallel processing of instructions.

5 Conclusion

As previously been said, each method has pros and cons as we can see, there are big differences in time we need to calculate one sample of output. In C code we need approximately 35 cycles per multiply one sample of input. Other way in assembly language, each instruction (even multiplication) needs only one cycle and that is the reason why it is so quick.

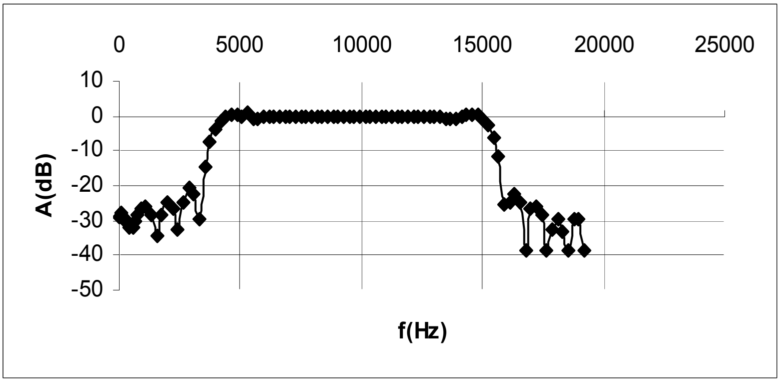


Fig. 5. Designed filter in DSP

Figure 5. shows real results for FIR filter implemented to DSP. It is the same filter as designed by matlab. The shapes of both characteristics are similar, only one difference is in signal attenuation. In real filter we computing with finite resolution which is under the resolution of matlab. This has a significant influence to final characteristic.

Reference

1. The Mathworks Inc.: Matlab R2010b product help
2. Texas Instruments: TMS320f28xx Manuals and User's Guides, www.focus.ti.com
3. SOVKA, Pavel; ČMEJLA, Roman; ŠMEJKAL, Ladislav. Číslicové filtry. *Automatizace*. 2005, 7-8, s. 495-497.

Simulation of Power Dissipation in the Indirect Frequency Converters

Roman Šprlák

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
roman.sprlak@vsb.cz

Abstract. The article deals with the simulation of indirect frequency converter. The frequency converter consists of a rectifier, where is the input current and voltage rectified. The voltage is adjusted to desired values by the inverter. Input and output circuits are isolated dc-link. In the created simulations are used calculations of effective and mean currents. These currents are load for the transistors and diodes three-phase inverter. By applying a special mathematical apparatus can be determined the loss in the inverter which are caused the current conduct and switching losses.

Key words: frequency converters, matlab, inverter, power dissipation.

1 Introduction

As input power for asynchronous and synchronous motors with permanent magnets are most commonly used indirect frequency converters with dc-link voltage. The power part is consists of GTO thyristors, or IGBT transistors, which allows the use of high frequency switching converters. As a result of this, we can achieve the excellent static and dynamic properties. [3]

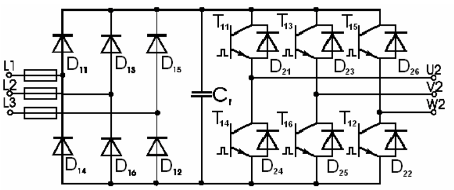


Fig. 1. Indirect frequency converter with dc-link voltage [4]

The aim of this article is:

- Calculation mean and effective values of transistors and zero diode currents . Subsequently define the losses caused by the current line (VA-characteristic is replaced by the "broken curve")
- Calculation of switching losses
- Determination of total losses
- Simulation of power losses in the Matlab Simulink

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 219–224.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Determining power dissipation inverter

The AC drives with sinusoidal pulse width modulation (PWM sine - Pulse Width Modulation) generates three-phase IGBT inverter (Fig. 2) output voltage, which has the character of a sequence of pulses with variable width the duty cycle. In the created simulation is first necessary to define a mathematical apparatus for calculating the individual partial parameters of the system. [4]

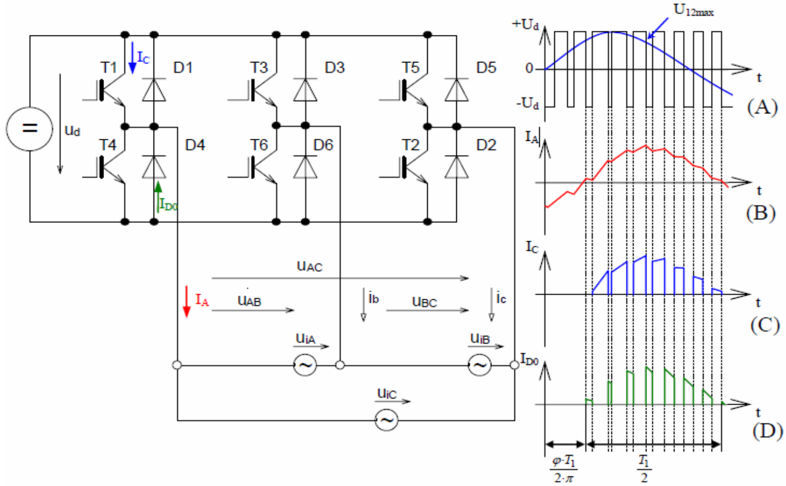


Fig. 2. Three-phase inverter – motor [5]

2.1 Effective value of line voltage inverter

Calculation of effective voltage is described in [1]. It this literature is used so called modulation factor, which is defined in equation (1):

$$M = \frac{U_{AB}}{U_d} \tag{1}$$

Where UAB is effective line voltage and Ud is dc-link voltage. The effective voltage value is calculated from the equation (5 b):

$$U_{ABef} = U_d \sqrt{\frac{2}{\pi} M} = 0{,}7979 \cdot Ud \sqrt{M} \tag{2}$$

2.2 Mean current of the transistor and the diode

Collector current of transistor and zero diode has a shape unipolar pulses as shown in Fig.2. The general calculation of the mean value of this signal is a follows:

$$I_{stř} = \frac{1}{T}\int_{0}^{T_1} i(t)\cdot s(t)dt \tag{3}$$

The function i(t) specifies the height of the envelope function and s(t) instantaneous impulses of duty cycle. In [2] is shown that the mean currents value of the transistor and diode which is in opposite possition can be calculated: [2]

$$I_{T\,stř} = I_{mA}\left(\frac{1}{2\pi} + \frac{M}{4\sqrt{3}}\cdot\cos\varphi\right) \qquad I_{D\,stř} = I_{mA}\left(\frac{1}{2\pi} - \frac{M}{4\sqrt{3}}\cdot\cos\varphi\right) \tag{4 a,b}$$

2.3 The effective current value of a transistor and diodes

For unipolar current pulses can be used general expression:

$$I_{ef} = \sqrt{\frac{1}{T}\int_{0}^{T_1}\int i^2(t)\cdot s(t)dt} \qquad U_{ef} = \sqrt{\frac{1}{T}\int_{0}^{T_1}\int u^2(t)\cdot s(t)dt} \tag{5 a,b}$$

In equation (3) are using only pure sinusoidal functions and the calculation can be done by: [1]

$$I_{Tef} = I_{mA}\sqrt{\frac{1}{8} + \frac{2M}{3\sqrt{3}\pi}\cdot\cos\varphi} \quad I_{Def} = I_{mA}\sqrt{\frac{1}{8} - \frac{2M}{3\sqrt{3}\pi}\cdot\cos\varphi} \tag{6 a,b}$$

2.4 Losses caused by the current conduct

The calculation of current losses can be done by the:

$$P_{ved} = U_p \cdot I_{stř} + R_d \cdot I_{ef}^2 \tag{7}$$

U_p is the threshold voltage and R_d is dynamic resistance. This formula is the same as for the calculation of the transistors and for the calculation of diodes.

2.5 Switching losses

Switching power can be determined from the relation: [2]

$$P_{PZ} = \frac{1}{T}\int_0^{T_1} W_{PZ}\left[i_z(t)\right] f(t)dt \tag{8}$$

In the case, when is the frequency and energy constant equation (8) is adapted to the form:

$$W_{PZ} = \frac{E_{\max}}{I_{Cnom}} \cdot I_z \tag{9}$$

I_Z is the current to load. E_{max}, E_{on} and E_{off} are from catalog data, this data indicate the energy loss. I_{Cnom} is the current of the transistor. We define the current utilization factor of the transistor:

$$K = \frac{I_{Z\max}}{I_{Cnom}} = \frac{I_{mA}}{I_{Cnom}} \tag{10}$$

In equation (8) when is carrier frequency constant, we can modify this equation on form:

$$E_{\max} = W_{PZ}\left[I_{Cnom}\right] = E_{on} + E_{off} \tag{11}$$

$$P_{PZ} = \frac{1}{\pi} \cdot f \cdot E_{\max} \cdot \frac{I_{mA}}{I_{Cnom}} \tag{12}$$

2.6 Total losses in the inverter

The total loss P_C of the transistor is the sum of losses caused by the leadership of current and switching losses.

$$P_C = P_{ved} + P_{PZ} \tag{13}$$

$$P_{D0} = P_{D0\_ved} \tag{14}$$

$$P_{Cztr} = 6 \cdot (P_C + P_{D0}) \tag{15}$$

3 Indirect frequency converter

Simulation of indirect frequency converter is consists from the three parts. Rectifier, which includes a diodes. The diodes are connected to three sinusoidal sources which are mutually shifted by 120 degrees and their amplitude are 315 V. The other part of model is dc-link which is created by capacitor with value 2.35 mF . To the dc-link circuit are connected IGBT modules.

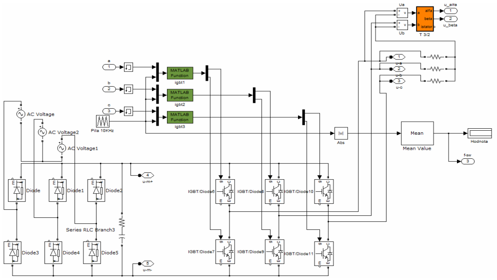


Fig. 3. Indirect frequency converters

4 The simulation results

The simulation was created in Matlab Simulink. The duration of the simulation was set to 2 seconds with the specified step of the 0,0000444. Switching frequency was 8 kHz. Asynchronous motor with a value of 15kW with 1000 rpm. Fig.4 shows the simulation. In the blocks „Výpočet napětí/proud", „Ztráty způsobené vedením proudu", „Spínací ztráty" and „Celkové ztráty" are implemented from above formulas. The value of individual variables are $I_{Tstř}$=27,62A, $I_{Dstř}$=7,412A, I_{Tef}=56,44A, I_{Def}=27,34A, P_{Tved}=52,89W, P_{Dved}=13,1W, P_{PZ}=48,06W, P_{Cztr}=683,7W. These values are the last recorded value after the end of the simulation. Information obtained from the simulations are recorded in a text file for further use, for example, for the simulation of thermal field. Fig.5 shows the individual loss.

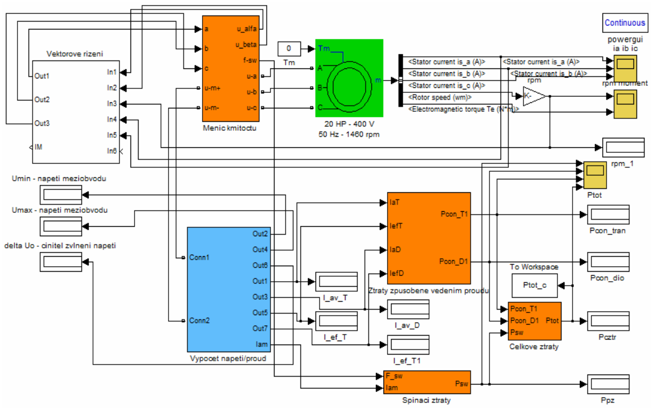


Fig. 4. Simulation of a frequency converter with an asynchronous motor

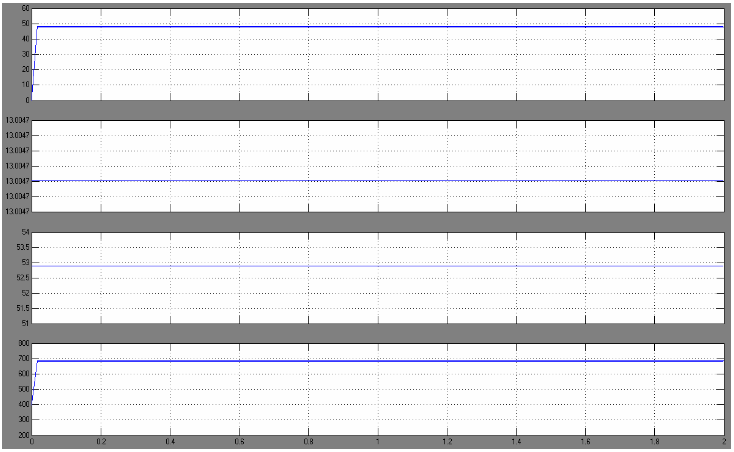


Fig. 5. Power dissipation P_{PZ},P_{D0},P_{vedT},$P_{Cztř}$

5 Conclusion

In this paper, was used different formulas for calculate the individual partial components of the frequency converter. These formulas were applied in Matlab Simulink to model three-phase frequency converter with the Intermediate circuit employs. To the model is connected asynchronous motor with power 15 kW. The aim was to obtained data loss of performance, because when we realize that the experimental measurement of inverter losses is not practically possible because of:

- Output power can not be measured with sufficient accuracy due to interference.
- Calorimetric measurement losses of inverter is for the common practice unusable.

Reference

1. Patočka, M., Matematická analýza signálů sinusové PWM, Sborník XXIV. Celostátní konference o el. pohonech, Plzeň 1995, str. 233-238.
2. Patočka, M., Výkonové poměry v soustavě trojfázový střídač – asynchronní motor,Sborník XXVI. Celostátní konference o el. pohonech, Plzeň 1997.
3. Klumpner Ch., Boldea I., Limited Ride-Through Capabilities for Direkt Frequency Converters, IEEE Transactions on power electronics, vol. 16, no.6, November 2001
4. Brandštetter, P., Střídavé regulační pohony–Moderní způsoby řízení, Ostrava 1999
5. Vaculík, P., Střídač hlavního pohonu elektromobilu s vodním chlazením, diplomová prace 2007, Ostrava 2007

Monitoring the Status of Accumulator Battery Cells

Miroslav Uchoč

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
miroslav.uchoc@vsb.cz

Abstract. This article describes possible ways of monitoring the status of accumulator batteries not only in their operations. Monitoring device with a microprocessor to easily carry out various data collection, processing and subsequent use results to inform, monitor or control other devices. Communications of the measured data from the battery is galvanic separated and sent to the control center.

Keywords: Monitoring, device, cells, battery

1 Principle monitoring battery

In monitoring full battery is necessary to measure each cell separately and detect the states of all cells. When measuring the battery voltage on the cells it is necessary to measure the battery in service. If the battery is disconnected from the load, we have to load or attach the load on each cell separately. Only in this way accurately measure the voltage on each cell and we can then evaluate the status on the cells.

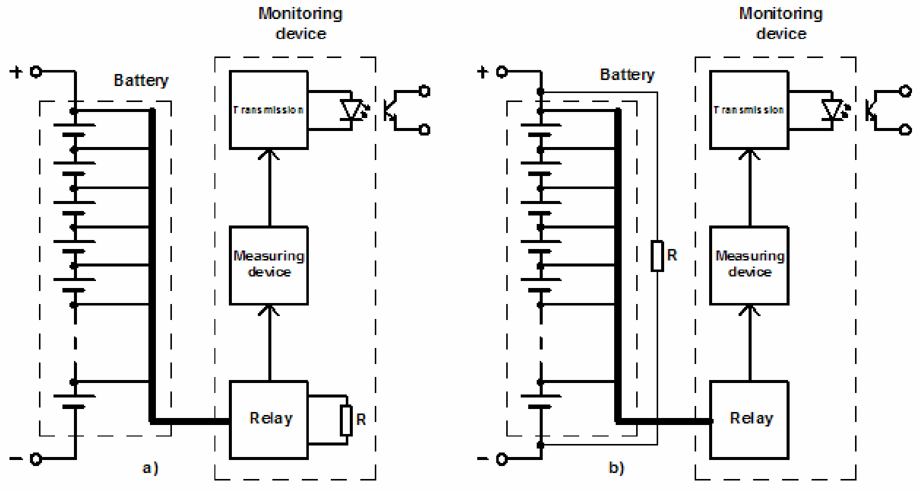


Fig. 1. Block scheme cell monitoring load a) of each cell, b) full battery

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 225–230.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

The best option will be monitoring each cell with a total load of batteries. Using a device that will measure the voltage on the cells without load and the load. In another variant of monitoring cells there will be used one resistor, which will be switched via relay for each battery cell for a short time. Measured data will be galvanic separated from the battery and sent to the control center. Supply all device is done externally.

2 Monitoring of cell battery under load

Measurement of voltage on the cells in the battery during operation, the simplest solution. It does not have to connect the load on the battery or for each cell. Just measure the voltage on each cell and send measured data to the control part.

2.1 Monitoring of full battery in one monitoring device

When monitoring battery full in one device, it must be lead from wire to each cell and measuring section, in which it switches the desired cell and measure appropriately. The disadvantage is the large number of wires that must be switched either electronically or mechanically.

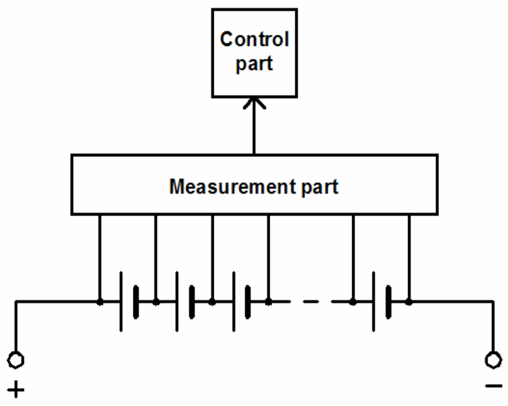


Fig. 2. Monitoring of full battery in one monitoring device

2.2 Monitoring of one cell in one monitoring device

If you want to monitor each cell one monitoring device, we must solve power source of the whole device with just one cell. We use DC / DC converter, which converts us to 2V voltage needed to power the device. We can use the circuit with transistors or circuit use LT1073, which has a everything you need and can work from voltage 1V. Involvement of the transistors is more complicated but substantially cheaper than the LT1073 circuit.

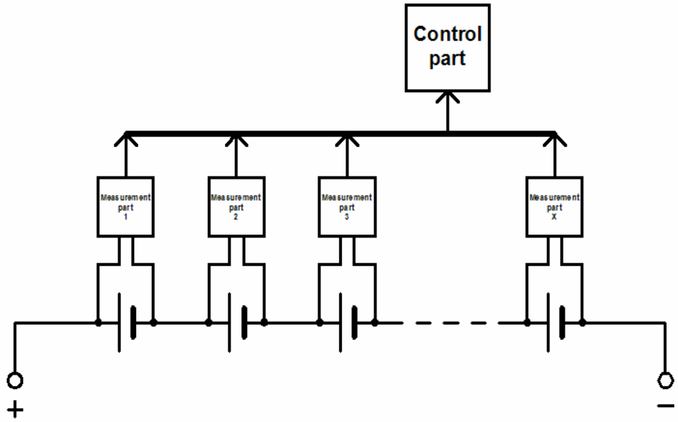


Fig. 3. Monitoring of one cell in one monitoring device

2.3 Monitoring of cell eight one monitoring device

The advantage compared to a single monitoring device is one cell in the number of monitoring devices. Power of measuring part is made through all 8 cells at the same time. The measurement is performed by electronic switching cables to the measuring terminals.

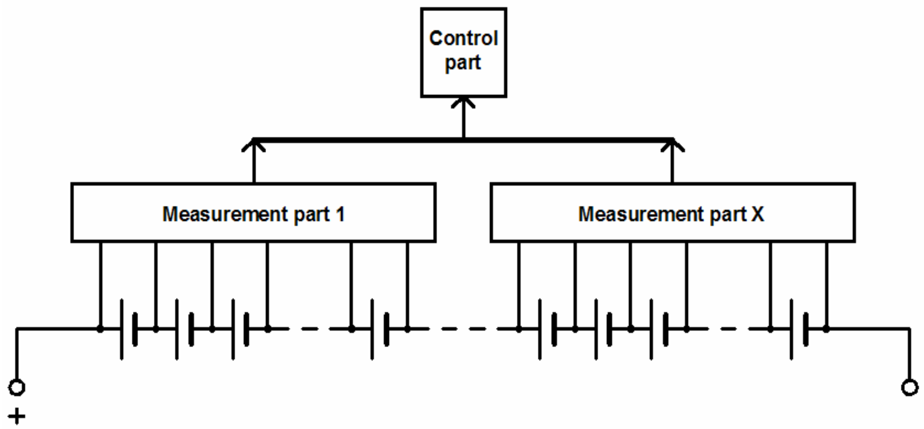


Fig. 4. Monitoring of cell eight one monitoring device

3 The monitoring device

In Figure 5 and 6 are two ways of solving the DC / DC converters that are used to power the monitoring device. Transistor can be used in connection or involvement with the circuit LT1073. The output voltage 5V supply can be powered microprocessor system that will perform measurement cell and subsequent sending of data to the control center.

3.1 DC/DC converter

After connecting the power supply, current flows through resistor R1 to the base of transistor T1 and transistors T1 and T2 are open. On T2 collector voltage is almost zero and transistor T2 and the coil currently flows through L1, which is gradually increasing. At a certain size the current saturated transistor T2 and the coil currently stops increasing. This will increase the voltage at the collector T2 and at a sufficient magnitude of the voltage through resistor R2, both transistors are closed. Interruption of current in inductor L1 to the collector T2 induces a large positive voltage across the diode D1 will charge capacitor C1. Zener diode D2 reduces the voltage on capacitor C1 at 5 V.

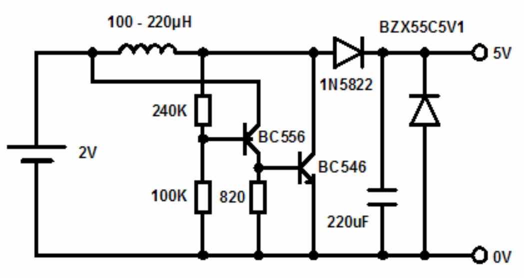


Fig. 5. Converter voltage transistor connecting

LT1073 is universal DC / DC converter. The device requires only three external components to deliver a fixed output of 5V or 12V. The very low minimum supplies voltage of 1V. Average current drain of the LT1073 is only 135µA. The circuit shown can deliver 5V at 40mA from an input as low as 1.25V and 5V at 10mA from a 1V input.

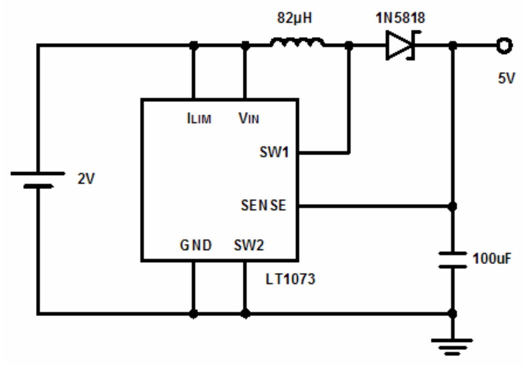


Fig. 6. Converter circuit with LT1073

3.2 Cell eight monitoring one monitoring device

The measured voltage of the first cell will be directly responsible, a second article voltage is calculated from the measured voltage of both cells and subtracting the voltage the first cell. Voltage on other cells are calculated similarly. But it decreases the accuracy of measurement, for the eighth cell will be lower accurate than for the cell first. There is also way to connect up in the monitoring device only four cells, the measuring accuracy was better than the first but it would have to use twice as monitoring devices. Power is realized through all eight cells (four cells).

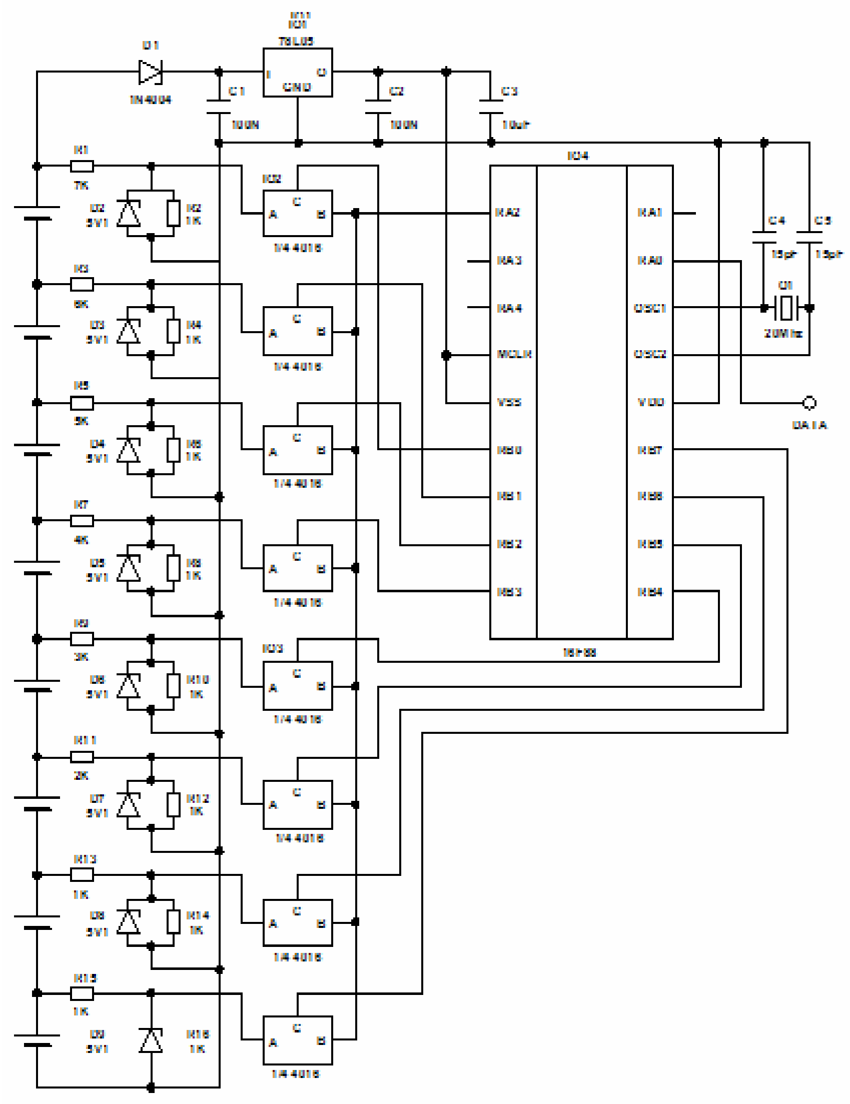


Fig. 7. Cell Eight scheme monitoring one monitoring device

Conclusion

The best way to monitor battery cells is one in which the battery is in operation. We need not to worry about switching complex when we deal with load of every cells separately. During operation, the battery will run a short test to measure the voltage on all the cells and evaluate them.

References

1. Cenek M., Jindra J., Jon M., et al. Accumulators from principle to practice, FCC PUBLIC s.r.o., 2003. ISBN 80-86534-03-0

2. Pospíšil R.,Račanský P. Monitoring devices accumulator batteries in small appliances. Bachelor thesis, FEI VUT, Brno, 1999

3. Hess K., Kocián P., Tmej M., Vaverka D. Universal modular monitoring device. Bachelor thesis, FEI VUT, Brno, 1999

Remote Controlled Instruments Simulators for Educational Purpose

Jan Besina and Petr Bilík

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.besina@vsb.cz, petr.bilik@vsb.cz

Abstract. When working with the remote-controlled device measuring instrument it is beneficial to use the drivers that are created according to the standard SCPI and VXI plug and play. These instrument drivers largely simplify and speed up the work in developing applications for measuring devices. At the end, the work gives a closer look on a simulator that is used for teaching remote control devices. There are described various devices that a simulator contains, and mentioned some of their features.

1 Introduction

Communication with measurement systems is used for two reasons. First of all, it is the possibility to set instrument parameters without actually being physically present or to change the parameters automatically according to the needs during the process. Secondly, it is the possibility of complete automation of the whole process and saving data to a file that enables next data processing.
This article aims to define standards that are used for distant communication with instruments and also to define a created program that serves as a teaching method with instrument drivers and to teaching.

2 Standard IEEE 488

2.1 Standard IEEE 488.1

This standard is defined by standard electrical and mechanical properties for connectors and cables. This norm also defines handshaking, addressing, and general protocol for the transfer of bytes from the devices to the computers. This standard does not address the importance of individual transferred bytes.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 231–236.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2.2 Standard IEEE 488.2

This standard is linked to the previous specification, and is processed so that it remains compliant with IEEE 488.1. It suggests ways in which devices communicate with each other, so that it is completely reliable and highly effective.

To the IEEE 488.2 standard have been implemented commands. These orders were precursors of uniform syntax for communication with devices - SCPI standard.

3 Standard IEEE 488

Standard SCPI (Standard Commands for Programmable Instruments) is a standard of commands for programmable devices. It is a language based on ASCII commands. SCPI commands are based on a hierarchical structure, where the commands are divided by a common root, creating a tree structure of command.

Standard SCPI has extended the existing IEEE 488.2 standard by defining complex commands for all devices.

SCPI standard defines:

- Hierarchically ordered set of SCPI commands for control instrumentation
- A set of general statutory commands designed to identify the state and setting up of the internal registers
- Data formats

4 VXI Plug & Play instruments driver

VXI Plug and Play instrument driver is a set of VI, which is used to control a programmable device. Each VI corresponds to a particular operation that can be carried out with the device, such as configuration, triggering and reading of the measured values from the device. The instrument driver is provided for users to facilitate the initial work with a programmable device and at the same time serves the users to save time in the development and thus leads to a reduction in the cost of the resulting program. The big advantage of the instrument drivers is that the user does not need to know all the programming options for each unit. Due to the modular design it is very easy to change drivers.

Device drivers include high-level VI with intuitive front panel so that end users can quickly test and verify the possibilities of programmable device. End-users therefore do not have to know the low level of the control apparatus itself, such as specific instruments and their syntax. Users create applications to control devices using different building elements inserted into their block diagrams. These building blocks are in the form of instrument driver's subVI.

4.1 The model of instrument driver

A lot of programmable devices have a large amount of features and modes. Due to this complexity it is necessary to provide a simple design of a model that helps both the developers of the instrument drivers and the users themselves. The users are using these to develop instrument drivers to control the application of programmable devices.

4.2 The external structure of the instrument drivers

The instrument control is composed of Vis where the user calls at a higher level. The following diagram shows the interaction of the instrument control with the rest of the system.

The outer structure shows the interactive interface and application programming interfaces. Interface for programming (APIs) is a set of instrument drivers in a form of individual VI that the user can call in their application. An example could be a test application that for its function uses the instrument drivers to communicate with a multimeter and oscilloscope.

Virtual instrument software architecture (VISA - Virtual Instrument Software Architecture) input-output interface in LabVIEW is a set of functions that are used for communication with devices. VISA has established a standard for communication on standard communication interfaces (GPIB, USB, serial).

VISA library functions are able to implement a prescribed exchange of messages regardless of the physical interface that will be used to do this exchange. Interface type to be used for the communication is a parameter of a function for initialization of the communication during which is generated a variable control. Through this control operations of writing and reading are carried out in communication with the device.

4.3 The internal structure of the instrument drivers

Internal structure of the instrument driver defines the organization of the individual instrument drivers in such a way that they are well-organized. For maximum clarity, the instrument drivers are arranged according to their functionality.

- Initialization VI - all files of the instrument drivers should include this VI, which creates the connection between software and device.
- Configuration VIs - is a collection of individual VI that enables to configure devices. Their number depends on the unique capabilities of the device itself.
- Action / Status VIs – are VI that are used to start or end the test installation and measurement.
- Data VIs – these VI enable the transfer of data to or from the device. They include, for example, VI for reading the measured values and curves or for adjusting the digital outputs and digital inputs to read.
- Utility VIs – these VI perform various additional functions such as reset or self-test.

- Closing VI – this VI closes the connection of the instrument and software. Each set of instrument drivers should include this VI.

5 Simulator for teaching remote control

Because it was financially very expensive and also difficult for spatial layout to equip a classroom with just a few measuring devices, it was necessary to create a software simulation of several most commonly used devices in the measuring technique. The instruments were the following: Power supply; Digital multimeter; Signal generator and Oscilloscope.

5.1 Description of the simulator power supply

This simulator of power supply simulates basic functions of a real power supply. Using commands it can set the output voltage, or turn it off. Also, this resource allows adjusting the voltage ramp (sweep mode). The output of the voltage source can be brought to the simulator input of digital multimeter the oscilloscope input.

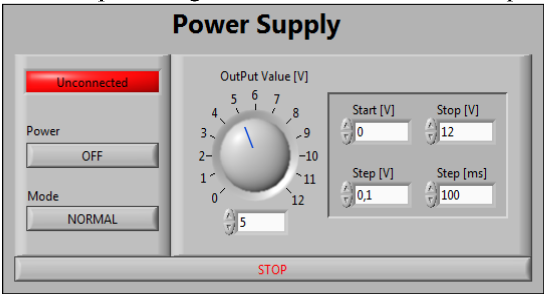


Fig. 1. Front panel of power supply simulator

5.2 Description of the simulator function generator

The second of simulated devices is a signal generator. This generator can generate four types of signals. Amplitude, frequency and DC offset can be set in there. Also, this generator allows sweep mod. The output from the generator can be brought to the simulator of digital multimeter or an oscilloscope where it can be seen its progress.

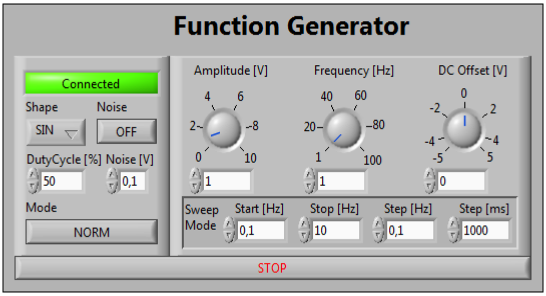


Fig. 2. Front panel of function generator simulator

5.3 Description of the simulator digital multimeter

Next to the last simulated device is a digital multimeter. This multimeter allows to measure several parameters of the input signal, which can be brought from the power source or function generator.

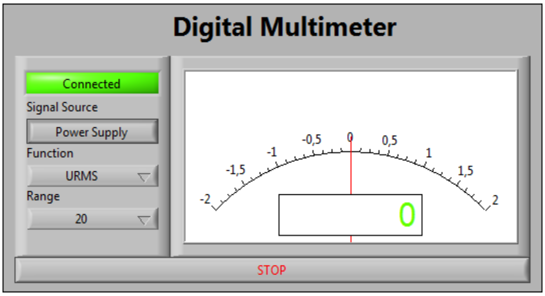


Fig. 3. Front panel of digital multimeter simulator

5.4 Description of the simulator oscilloscope

The last of the simulated device is an oscilloscope. The simulator handles the basic functions of an oscilloscope used in current measurement technology. It allows to set both the time base and voltage range. The measured waveform is displayed next to the device controls. The oscilloscope enables to switch signal source. The first source of signal is the simulator of power supply, the second one is the simulator of the function generator.

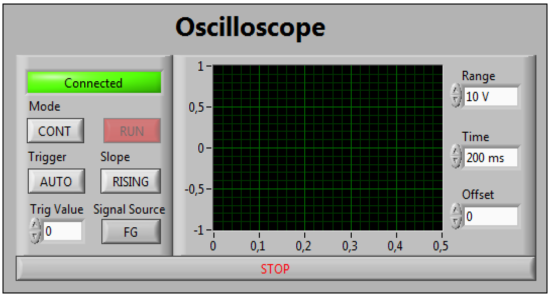


Fig. 4. Front panel of oscilloscope simulator

5.5 Communication between the simulator and control applications

Communication between the simulator and the application being used for remote control proceeds thanks to the TCP / IP protocol on a local IP address. On each of the devices a port is assigned for communication. Ports that are allocated to individual devices are shown in the following table (Table 1). Application Simulator acts as a server to which the clients connect (user application).

Table 1. Communication port

Device	Number of port		Device	Number of port
Power Supply	9999		Digital multimeter	9997
Function Generator	9998		Oscilloscope	9996

6 Conclusion

The described simulator enables to save a considerable amount of money that would otherwise have to be spent on the equipment in the classrooms by these devices. This application of the simulator fully replaces individual devices with regards to the possibility of communication with them.

References

1. Zone.ni.com [online]. 18.11.2010 [cit. 2010-01-10]. GPIB Hardware and Software Specifications. <http://zone.ni.com/devzone/cda/tut/p/id/3388>.
2. Zone.ni.com [online]. 18.11.2010 [cit. 2010-01-10]. Developing LabVIEW Plug and Play Instrument Drivers. <http://zone.ni.com/devzone/cda/tut/p/id/3271>.
3. PIEPER, John. Automatic measurement control : A tutorial on SCPI and IEEE 488.2. Mnichov : Rohde & Schwarz, [199?]. Status and error reporting, s. 158-184. ISBN 978-3-939837-02-2.

Tester for Phasor Measurement Unit Testing

Marek Horinek and Petr Bilík

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{marek.horinek, petr.bilik}@vsb.cz

Abstract. Phasor Measurement Unit (PMU) is used for the evaluation of synchronous phasor. The IEEEC37.118 standard describes features of the PMU. This document is focused on communication description with PMU and PMU measurement accuracy in steady state. Detailed description of the implementation of phasor evaluation in the international standard form doesn't exist.

Key words: PMU, IEEEC37.118, phasor measurement, PXI

1 Introduction

The current state of the electricity network as well as many other industries is result of economic development and political changes of recent decades. As a result, the separation of generation, transmission and distribution of electricity was performed. There are multiple interest groups influencing the planning, development and operation of the transmission system, which operates on a very wide area. Due to the limited construction of new facilities the electricity system is on (sometimes beyond) the limit of its capacity [1].

One of the newly-used means of increasing the reliability of system operation under these circumstances is monitoring and management of large networks (Wide Area Monitoring and Control = WAMC). Basic information for this monitoring and management is provided by a synchronous phasor measurement which allows monitoring system management of large networks to achieve greater accuracy and speed.

Synchronous phasor is described as the angle of measured values in a given time t=0. The importance of measuring this parameter starts to set in, when we analyze the angle between the synchronous phasors detected at the same time in different places of the electricity system and their change over time.

The number of manufacturers now offers for this measurement and monitoring a tool in the form of phasor measurement units (PMU - Phasor Measurement Unit). The framework features of the PMU, which regarding of phasors assessment defines IEEEC37.118 standard [2]. This paper is mainly focused on the description of communication with the PMU and defines the basic accuracy of the PMU in the steady state. A detailed description of the implementation of self-assessment phasors in the form of international standards doesn't exist. The accuracy of phasor evaluated

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 237–242.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

values angle and amplitude depend on many factors resulting from both the measured values and the way of evaluating implementation of the PMU.

This paper deals with the design and development of a test facility for PMU properties testing, which is currently developed at the VSB - Technical University of Ostrava.

2 Phasor Measurement Unit Testing

The need for comparative measurement of PMU is seen abroad, such as the Bonneville Power Administration (BPA) in the U.S. has also prepared a plan for such tests. The document BPA Phasor Measurement Unit Evaluation Test Plan [3] identified a number of performance criteria of these units, which should be verified in the upcoming test.

2.1 PMU Simulation Software

In the first stage of verification of PMU properties the simulator for the test signal simulation has been bought. This simulation software was created by ELCOM, a.s. company. The test signals generation for Phasor Measurement Unit testing can be simulated by this software. The simulator allows you to superimpose a degradation factors on the ideal signal and the different dynamic phenomena. The presence of these factors in a power grid could lead to incorrect measurement of phase and amplitude phasors by the PMU.

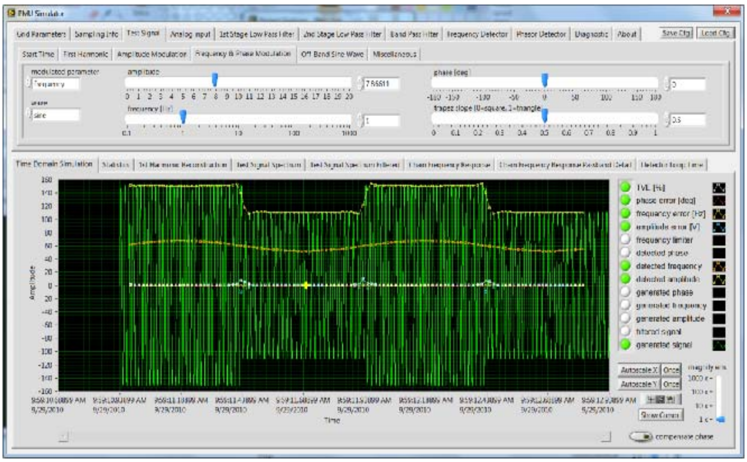


Fig. 1. PMU simulator user interface

The PMU Simulator allows calculating signal, which represents measured data for voltage and current phasor – the user can choose the frequency of this signal (50/60 Hz) and frequency with which the PMU unit generates at its output the measured data and phase in its ideal form.

The user can select the basic parameters of the calculated signal (amplitude, frequency, phase). These parameters represent the test signal in its ideal form.

The PMU simulator allows superimposing any interference on the ideal signal that may occur in real networks: the harmonic components, modulation components from various phenomena associated with running the real measurement system, etc.

The next feature of this simulator is possibility to visualize and store the test signal in its each stage of processing and also from the PMU output. This option allows the user to examine the behavior of the each blocks of simulator structure which process the signal in the presence of various artifacts in the measured signal. It is also possible verify that the results of measurement given in the PMU's protocol by IEEE standard C37.118 for such simulated state of measured system are correct, and errors are within limits which defines the requirements of BPA to test these units.

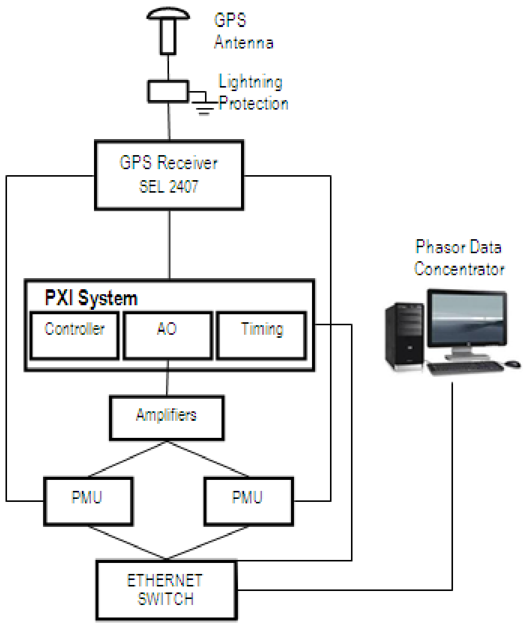


Fig. 2. Basic concept of the PMU test facility

2.2 Phasor Measurement Unit Testing Facility

Hardware concept of the tester, see Figure 2, consists of the following parts:

- GPS receiver SEL2407,
- NI PXI system (NI 8106 controller, NI 6682 timing and synchronization module, NI 6733 high speed analog outputs),
- NI cRIO 9014 + NI 9225 analog input + NI 9227 analog input,
- Power amplifier DEXON DP1500,
- PC equipped with phasor data concentrator software.

The heart of the PMU tester is created by a National Instruments PXI system, which is used to generate test signals.

The generation of test signals is performed by NI PXI-6733 high speed analog outputs module. This module is equipped with 8 analog to digital converters with 16 bits resolution. NI PXI-6733 supports simultaneous generation of each channel up to 1 MS sampling rate. The generated signal from analog output module is further amplified by power amplifier. The tester allows generation waveforms of 3 phase electric system (3 voltages and 3 currents). For the amplification at a higher level (62 Vrms, 3 Arms) the DEXON DAP1500 amplifiers are used. The basic technical characteristic of the amplifier (output power 2 x 800W (20-20000 Hz, THD = 0.05%, 4 Ω) RMS, bandwidth 20 – 20000 Hz, 8Ω) [4]. The precise time synchronization of generated signal is provided by NI PXI-6682 timing and synchronization module.

The waveform of test signals is measured and stored with timestamps by cRIO platform during the test process. For voltage measurement the NI 9225 analog input module is used. This module is equipped with 3 analog inputs which allow direct connection of measured voltage up to 300 Vrms. The resolution of analog to digital (AD) converters is 24 bits. NI 9925 supports simultaneous sampling of each channel up to 50 KS/s. The current measurement is performed by using NI 9227 analog input module. This module allows direct current measurement up to 5 Arms. The sampling parameters of AD converters are the same like at the NI 9225.

The accurate timestamps are derived from a common source of clock signal (SEL 2407), which ensures timing synchronization of all devices participating in the test. The measured values are stored in the form of individual waveforms of generated current and voltage in the TDMS file format. The sampling frequency of measured waveforms is 50 KS/s. Stored data can be used for declaring of test signal accuracy which is used for PMUs testing.

The last part of the test facility is a personal computer (PC) which is equipped with software for collecting measurement data (Phasor Data Concentrator) of the tested PMU units. Evaluation PC communicates with the individual PMUs via Ethernet communication interface. Reading data from the PMUs are stored in a TDMS file format [6].

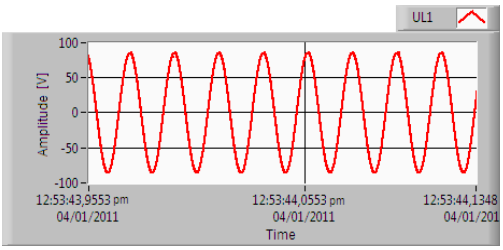


Fig. 2. Basic concept of the PMU test facility

3 Test Signal Accuracy and Results of Basic Tests

Amplitude characteristic of the amplifier was measured using the AGILENT 34401A digital multimeter (6 ½ digit resolution, basic accuracy 0.0035% DC, 0.06% AC, bandwidth 3 Hz to 300 kHz).

Phase characteristic of the amplifier was measured using the NI cDAQ 9188 equipped with NI 9223 module. This module can measure voltages up to 7 Vrms with a sampling rate up to 1 MS/s. The sampling rate used during the phase characteristic measuring of the amplifier was 500 KS/s. The phase error using this sampling rate corresponds to 0.036 ° for the signal frequency of 50 Hz. Methodology of the phase characteristics measurement was performed by measuring the phase shift between the generated signal on the amplifier's output and the PPS signal (Pulse per Second) on the output of the SEL 2407. Figure 4 shows the detail of the measured phase between PPS and the amplifier's output signal for the phase shift of 0 °. The sampling rate of measured waveforms is 500KS/s.

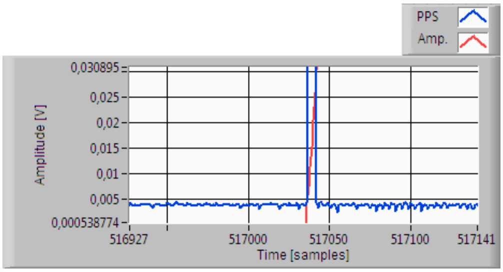


Fig. 4. A detail of the generated signal measured phase (PPS signal and voltage signal)

The testing laboratory was used for verification of the behaviour of the two same professional PMU and two versions of the PMU developed at the university within a diploma thesis written by a student. Due to limited space reserved for this article, only the results of one test are mentioned below: a response to the step change of voltage from 80% to 120% Vn. The X axis is mentioned in the numbers of fundamental frequency periods.

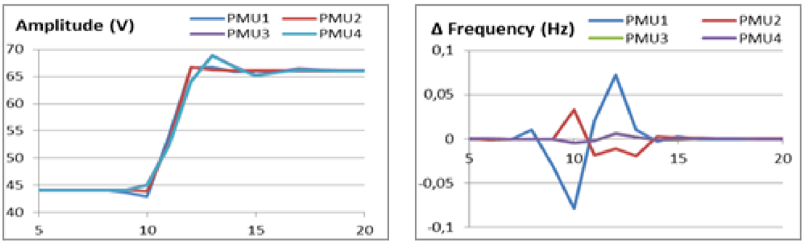


Fig. 5. Amplitude and frequency deviation of tested PMUs

4 Conclusion

Nowadays the development of a tester for PMU testing was completed. A tester conception on the basis of virtual instrumentation has been selected for construction of the testing laboratory. This conception is not limited by the functionality defined by the producer, but it brings the openness and flexibility which is so needed in a university worksite. The prepared tests allow verification of the compliance of the tested PMU units properties with the requirements of the basic standard IEEE C37.118 as well as compliance characteristics of the test unit with the other requirements of end users of the measurement system and also the characteristics of the tested PMU units in the presence of various disturbing factors which may occur in the real network. The behavior of PMUs in occurrence the disturbing factors in the measured signal is a very important aspect of the usability of the result provided by these units for planned and existing systems WAMC, which should increase the reliability of transmission systems.

In the next research, the algorithms for testing of PMU behaviour under more complex conditions in the supply network will be developed. Multiple repetition and subsequent statistic evaluation are inseparable parts of the tests.

References

1. POPELKA, Antonin. Synchronous Phasor of Voltage and Current in Power Grid, Identification of Parameter for Power Traffic Management In . Brno : AIS spol. s r.o., 2006. s. 5. WWW: <http://www.ais-brno.cz/clanky/egu2006.pdf>.
2. IEEE C37.118. *IEEE Standard for synchro phasors for Power Systems*. [s.l.] : IEEE Power Engineering Society, 21.10.2005. 65 s. WWW: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1611105>.
3. BPA Phasor Measurement Unit Evaluation Test Plan. In . [s.l.] : Bonneville Power Administration , 2009.)
4. *DEXON Amplifiers* [online]. 2011 [cit. 2011-04-02]. DEXON DAP1500. WWW: <http://www.dexon.cz/data/popisy/27039.pdf>.
5. *National Instruments* [online]. 2010 [cit. 2011-02-25]. NI PXI-6682. WWW: <http://sine.ni.com/nips/cds/view/p/lang/en/nid/203944>.
6. *National Instruments* [online]. 2010 [cit. 2011-02-25]. TDMS FAQ. WWW: <http://zone.ni.com/devzone/cda/tut/p/id/9995#3>.

Optimization of Takagi-Sugeno Type Fuzzy Regulator Parameters by Genetic Algorithm

Jiri Kocian, Jiri Koziorek, and Miroslav Pokorny

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jiri.kocian1, jiri.koziorek, miroslav.pokorny}@vsb.cz

Abstract. Many papers deals with the usage of fuzzy rules to implement PID type control. Fuzzy models, especially the Takagi-Sugeno-type, have received significant attention from various fields of interest. It is very often very difficult to determine all the parameters of the Takagi-Sugeno-type controller. In this paper we present optimization of Takagi-Sugeno-type fuzzy regulator parameters by genetic algorithm. Implementation of universal fuzzy P/PS/PD function block implemented to the PLC Simatic S7 300/400 is introduced. Mamdani model is used as comparative model. Parameters of Takagi-Sugeno-type fuzzy regulator are determined by genetic algorithm optimization from comparative regulation surface.

1 Introduction

The aim of this work is optimization of Takagi Sugeno-type (TS-type) fuzzy regulator parameters by genetic algorithm. Implementation of universal fuzzy P/PS/PD function block implemented to the PLC Simatic S7 300/400 is introduced.

Mamdani model is used as comparative model. Parameters of TS-type fuzzy regulator are determined by genetic algorithm optimization from comparative regulation surface.

Many papers deal optimization of TS-type fuzzy regulator parameters. [1, 2, 3]

The goal of this work is to design own universal fuzzy control function block for PLC Simatic S7 300/400 with optimized parameters.

Chapter 2 introduces basic knowledge of genetic algorithm.

Chapter 3 contains a description of the implemented fuzzy P/PS/PD function block.

Optimization and results are given in chapter 4.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 243–248.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Genetic Algorithm

The genetic algorithm is a method for solving both constrained and unconstrained optimization problems. The genetic algorithm is based on natural selection, the process that drives biological evolution.

The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population is going toward an optimal solution. The genetic algorithm can be applied to solve a variety of optimization problems that are not well suited for standard optimization algorithms.

The genetic algorithm uses three main types of rules at each step to create the next generation from the current population: selection, crossover and mutation.

Selection rules select the individuals, called parents, that contribute to the population at the next generation.

Crossover rules combine two parents to form children for the next generation.

Mutation rules apply random changes to individual parents to form children.

The genetic algorithm can be described in the form (1).

$$GA = (p, I, P, f, s, c, m, T) \tag{1}$$

Where

$p^0 = (a_1^0, \ldots, a_r^0)$	Initial population of chromosomes,
I	Coding of chromosomes,
$P \in N$	Size of population,
$L \in N$	Length of the chromosome,
$f: I \to R$	The fitness function,
$s: I^P \to I$	Selection of chromosome pairs of parents,
$c: I^2 \to I^2$	Operation of chromosome crossover,
$m: I \to I$	Operation of chromosome mutation,
$T: I^P \to \{0,1\}$	The stop criteria of the genetic algorithm.

3 Fuzzy P/PS/PD regulator

Function block for Siemens Simatic S7 300/400 has been implemented.

Fuzzy controller can be except fuzzy P also fuzzy PS or fuzzy PD according to a logical value of function block input *Sel_PS_PD* (*Sel_PS_PD* in log. 0: PS regulator, *Sel_PS_PD* in log. 1: PD regulator).

For regulation error e the first difference of regulation error Δe and for and manipulated value u (in the case of fuzzy PS regulator the first difference of manipulated value Δu) are implemented weighting factors (K_P, K_S, K_D, K_U). When selecting the type of fuzzy PD regulator the error e is multiplied by K_P factor and the first difference of the error Δe by K_D factor. On the other hand when selecting a fuzzy PS regulator the error e is multiplied by K_S factor and the first difference of the error Δe by K_P factor.

Structure (2) is generally applied for fuzzy PD regulator and structure (3) is generally applied for fuzzy PS regulator.

$$IF(e\ is\ A)\ and\ (\Delta e\ is\ B)\ THEN\ (u\ is\ C) \tag{2}$$

$$IF(e\ is\ A)\ and\ (\Delta e\ is\ B)\ THEN\ (\Delta u\ is\ C) \tag{3}$$

Manipulated value calculation for the fuzzy PD regulator is in (4) and for the fuzzy PS regulator in (5).

$$K_u \cdot u(k) = K_P \cdot e(k) + K_D \cdot \Delta e(k) \tag{4}$$

$$K_u \cdot \Delta u(k) = K_P \cdot \Delta e(k) + K_S \cdot e(k) \tag{5}$$

The range of control error e and the first difference of the error Δe is divided up to seven fuzzy sets (NB - Negative Big, NM - Negative Medium, NS - Negative Small, AZ - Approximately Zero, PS - Positive Small, PM - Positive Medium, PB - Positive Big). Takagi-Sugeno method was used for inferring the manipulated values (6).

$$IF(e\ is\ A) and (\Delta e\ is\ B)\ THEN\ \mathrm{u} = \mathrm{f}(e, \Delta e) \tag{6}$$

The final output is determined as a weighted mean value over all rules according to (7).

$$u = \frac{\sum_{i=1}^{r} \alpha_i \cdot w_i \cdot (u_i + a_i \cdot e + b_i \cdot \Delta e)}{\sum_{i=1}^{r} \alpha_i \cdot w_i} \tag{7}$$

Where

u_i, a_i, b_i Parameters of Takagi Sugeno model,

α_i It is given as a minimum degree of membership of members of the antecedent (corresponding to the logical AND in the antecedent),

w_i Weighing factor of the rule,

u Calculated manipulated variable.

Fuzzy regulator contains an initialization procedure that will get the coordinates of fuzzy sets e, Δe and parameters of Takagi Sugeno model from input structures into static structures of the fuzzy controller.

Initialization procedure also checks the coordinates of fuzzy sets so that any bad assignment of fuzzy sets and no fatal impact to the regulation function should be applied. If errors are detected in the coordinates of fuzzy sets during initialization procedure, the control is stopped and indication of wrong parameter is appeared in function block output.

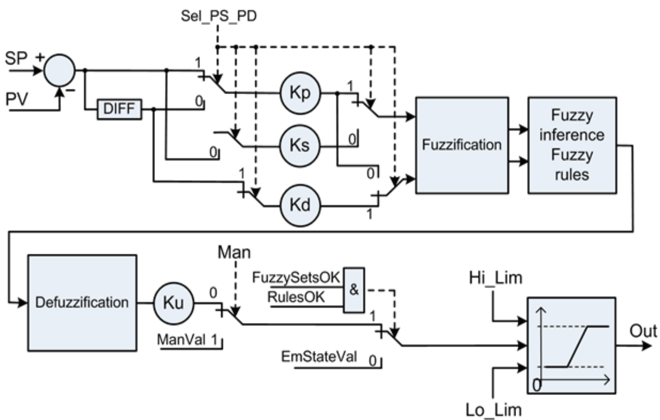


Fig. 1. The block scheme of fuzzy PS/PD regulator.

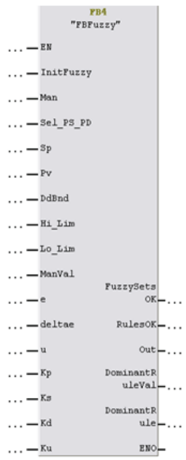


Fig. 2. Implemented fuzzy P/PS/PD regulator.

4 Optimization of parameters by genetic algorithm

With comparison to Mamdani model, TS fuzzy model brings one significant advantage but also one significant disadvantage. Final implementation of Takagi Sugeno type fuzzy model to some control system is often much more convenient.

Deffuzification method for TS-type fuzzy model needs less computing time than Mamdani CG defuzzification method. But on the other hand it is very difficult and even impossible to set all parameters of Takagi Sugeno model correctly without some automatic algorithm. Therefore, many optimization methods are used to solve this problem.

Optimization by genetic algorithm is used in this approach. Mamdani model is used for comparative purposes.

Genetic algorithm Toolbox in Matlab is used.

We have used "ga" function from Matlab:

[X, FVAL, EXITFLAG, OUTPUT] = GA (FITNESSFCN, NVARS, A, b, Aeq, beq, lb, ub, NONLCON, options).

The genetic algorithm minimizes the fitness function (8).

$$y_{val} = \frac{1}{\left(\sum_{i=1}^{r} abs(y_i - y_{GA,i})\right)+1} \tag{8}$$

Where

y_{val} Output value from fitness function,

y_i Value of i-th part of comparative regulation curve or surface (from Mamdani model),

$y_{GA,i}$ Value of i-th part of regulation curve or surface determined by genetic algorithm (TS model).

Fitness function (8) minimizes the sum of the differences across all TS fuzzy model parameters in function values between comparative Mamdani model and the TS model.

4.1 P fuzzy type regulator optimization

Comparative regulation curve of Mamdani fuzzy model is determined for fuzzy P regulator optimization.

Fig. 3 contains optimized regulation curve of TS fuzzy model. Fuzzy model with three fuzzy sets were used.

Optimized parameters of fuzzy P regulator:

$$u_1 = -0{,}142 \cdot e - 0{,}5823$$
$$u_2 = -1{,}2794 \cdot e - 0{,}0005$$
$$u_3 = -0{,}4268 \cdot e + 0{,}3765$$

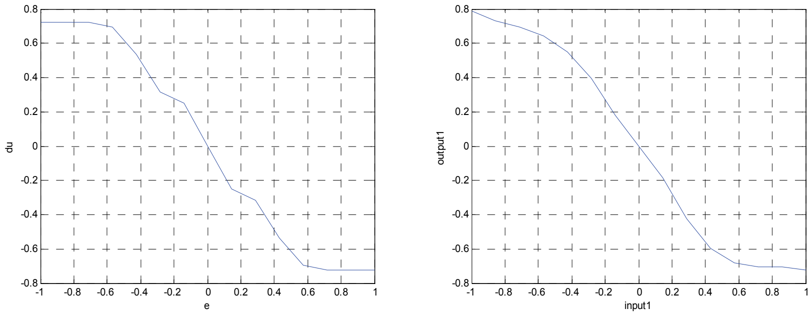


Fig. 3. Comparative regulation curve on the left and optimized regulation curve of TS model on the right.

4.2 PS/PD fuzzy type regulator optimization

Regulation surface of Mamdani fuzzy model is determined for TS fuzzy PS/PD regulator.

Fig. 4 contains optimized regulation surface of TS fuzzy model. Fuzzy models with three fuzzy sets are shown.

Optimized parameters of fuzzy PS regulator:

$$du_1 = 0{,}4284 \cdot e + 0{,}4673 \cdot de - 0{,}0914$$
$$du_2 = -0{,}3780 \cdot e + 1{,}1180 \cdot de - 1{,}0197$$
$$du_3 = 0{,}6289 \cdot e + 0{,}5416 \cdot de - 0{,}3034$$
$$du_4 = 0{,}1429 \cdot e - 0{,}2764 \cdot de - 0{,}9857$$
$$du_5 = 0{,}5452 \cdot e + 0{,}5673 \cdot de + 0{,}0001$$
$$du_6 = 0{,}9321 \cdot e + 0{,}3144 \cdot de + 0{,}4729$$
$$du_7 = 0{,}3666 \cdot e + 0{,}4225 \cdot de + 0{,}1675$$
$$du_8 = -0{,}0577 \cdot e + 1{,}0996 \cdot de + 0{,}1675$$
$$du_9 = 0{,}4808 \cdot e + 0{,}7011 \cdot de - 0{,}1321$$

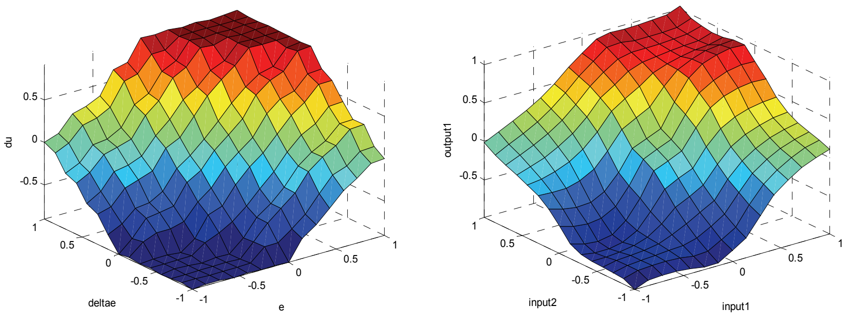


Fig. 4. Comparative regulation surface on the left and optimized regulation surface of TS model on the right.

5 Conclusion

The aim of this work is optimization of Takagi-Sugeno-type fuzzy regulator parameters by genetic algorithm. Implementation of universal fuzzy P/PS/PD function block implemented to the PLC Simatic S7 300/400 is introduced.

Mamdani model is used as comparative model. Parameters of Takagi-Sugeno-type fuzzy regulator are determined by genetic algorithm optimization from comparative regulation curve or surface.

The goal of this work is also design own universal fuzzy control function block for PLC Simatic S7 300/400 with optimized parameters.

Acknowledgements

This work is supported by project SP2011/45, "Data acquisition and processing from large distributed systems" of Student Grant System, VSB-TU Ostrava and also by the company of Ingeteam a.s. Ostrava.

References

[1] Aydogmus, Z.: Implementation of a fuzzy-based level control using SCADA. In Expert Systems with Applications. Volume 36, Issue 3 Part 2, April 2009, Pages 6593-6597. ISSN: 09574174.

[2] Behalek, K.. Analýza vlastností regulačních fuzzy modelů Takagi-Sugeno. Diploma thesis 2004. VSB-TU Ostrava.

[4] Karr, C. L.. Design of an Adaptive Fuzzy Logic Controller Using a Genetic Algorithm. In R. K. Belew & L. B. Booker (Eds.), Proceedings of the Fourth International Conference on Genetic Algorithms (pp. 450-457). 1991. Morgan Kaufmann.

User Perspective Adaptation Enhancement using Autonomous Mobile Devices

Jaromir Konecny, Jiri Kotzian, and Ondrej Krejcar

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jaromir.konecny, jiri.kotzian, ondrej.krejcar}@vsb.cz

Abstract. The need for devices with the ability to detect toxic gases, trapped people and to multifunction has increased. Dangerous places and armed conflicts have increased the demand for remote and autonomous devices. We propose a concept of two such devices with the ability to comfortably and remotely control such devices and even with an autonomous control in remote areas inside the buildings. The localization by WiFi is used to locate a position where the GPS signal is not well presented. The ability to locate a mobile device by a wireless network is a well known possibility. The current problem is precise indoor localization where WiFi signal from the building infrastructure is not strong enough to obtain right position.

1 Introduction

Mobile devices usually use global navigation systems like GPS, Glonass etc. for orientation in open space [1]. Navigation systems are very helpful in our everyday lives. The problem is rising in places with a high density of buildings. The precision of computing positions is too low. Inside buildings the normal navigation is usually not possible at all [2]. The reason is low signal or total signal absence. Different technologies for the navigation of mobile devices have to be used in buildings. For example: the human body uses stereovision for environment detection and orientation in cooperation with "maps" or other information (info panels, labels and indicators). Unfortunately this method is over the computation/power/space possibilities of today's embedded systems in mobile devices. It is also possible to equip rooms of a building with a set of transmitters like GPS. But this method is very expensive and complicated. The best method would use the current data infrastructure of the building – the net of mobile Ethernet access points (WiFi, WiMax,...) [7].This method is described in the following paragraphs. A net of access points is not sufficient on its own. For obtaining the right position of mobile devices, it is necessary to equip the device with other sensors and maps. When needed we can dynamically place additional access points to achieve a higher communication range or a higher precision of position detection.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 249–254.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

1.1 Motivation

Humans are well equipped with set of sensors for environment detection. However, some places are hidden to human vision, are too dangerous or too far away. As a result there is a need to equip the user with some devices which increase human perspective. This mobile device could search for people who have been trapped by an earthquake using infra camera and make audio-visual contact before the rescue come. Another example is searching for dangerous chemicals or to find criminals before authorities arrived. The simplest usage is to send or bring some item to a given position. Remote mobile robots are a common thing now, but this device should fulfill the given task using its own artificial intelligence at the end of the development. A special set of mobile devices was built at the Technical University of Ostrava for this purpose. The set is displayed in Fig. 1.

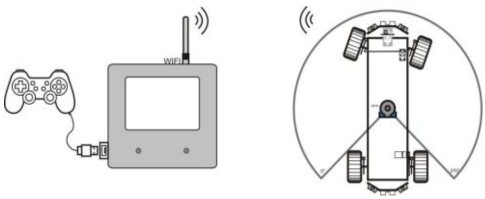


Fig. 1. The set of devices for increased human perspective.

2 Localization

The primary localization is needed to detect a position inside the building. Consequently, the map of the detected location is loaded into a mobile device to activate other sensors to “open the eyes” of our mobile devices. If the mobile device knows the position of the stationary device (transmitter), it also knows its own position within a range of this location provider. The typical range varies from 30 to 100 m where there is WiFi, respectively 50 m where there is BT case or 30 km for GSM. Granularity of the location can be improved by triangulation of two or more visible APs (Access Points). The mobile client currently supports the application in automatically retrieving location information from nearby WiFi, BT and GSM location providers, and in interacting with the PDPT server [3,4].

2.1 Laser scanner measuring

The current work on more precise position is to use distance sensors. The probe is equipped with laser scanner, two pieso-sensors and four corner optical distance sensors. Estimation of the position within a room is done using of fingerprinting localization. For similarity measurement is used correlation between fingerprint and current measurement. Data from laser sensors is array of distances between the sensor and closest subject within given angle. The range is 270° and maximum distance is

20m. The space around the sensor can be displayed by connecting measurement distances/angle. The C# implementation of used formula [1] is shown in Fig. 2.

$$(f * g)_k = \sum_{i=0}^{\infty} f_i \cdot g_{k+i} \qquad [1]$$

```
public Measure XCorr(long[] refer, long[] test)
{
    Measure ret = new Measure();
    for (int n = 0; n < refer.Length; n++)
    {
        long k = 0;
        for (int m = 0; m < refer.Length; m++)
        {
            int i = m + n;
            if (i >= refer.Length)
                i -= refer.Length;
            k += refer[m] * test[i];
        }
        ret.Values[n] = (int)(k / 10000);
    }
    return ret;
}
```

Fig. 2. Implementation of similarity measurement.

The measured data from the laboratory of embedded system are shown on Fig.3. The room is approx. 3m x 4m. The open door can be seen as long rectangle in bottom part. Right picture shows probe rotated (angle 30.75°). The angle is computed using correlation. The correlation curve is displayed on right part of the picture. The measured angle is found as global maximum of the curve.

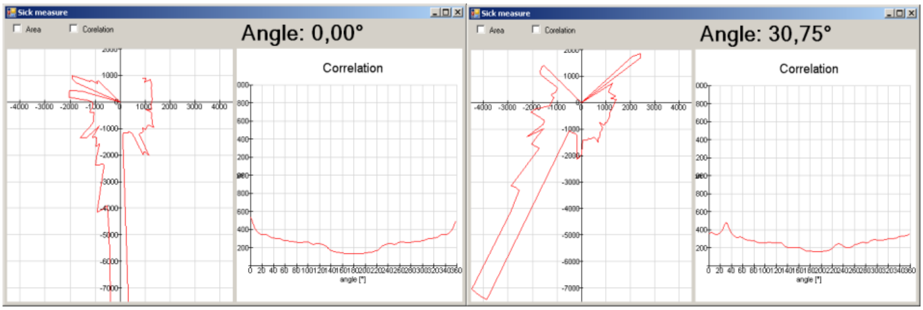


Fig. 3. An angle measurement of the probe.

The measurement is influenced by the difference between fingerprint and current state of the environment. The error in measurement is increasing especially due to placing new furniture or due to movement of people. The given method was tested by movement of a man in different distance from the sensor. The measurement results are shown on Fig. 4. Left picture represent error when human is 30cm from the sensor (minimal due to size of the probe); right picture display situation when human is 1m from the sensor. A human stands in the worst part of the area due to masking view into the corridor. The corridor has high weight in correlation in the case.

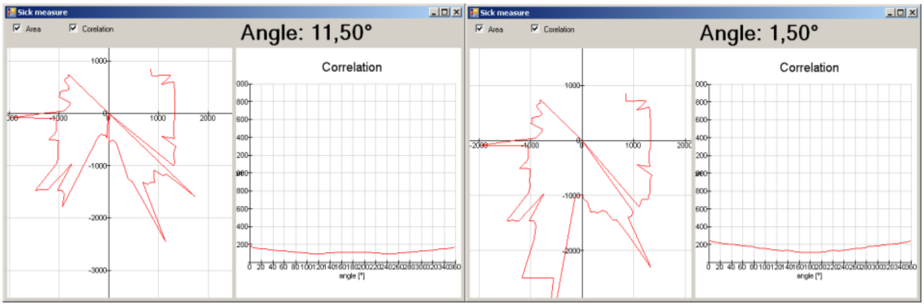


Fig. 4. An angle measurement error of the probe with disturbance 30cm and 1m from the sensor.

As a result we can say that this method is very suitable to obtain more precise position within given room. The method could also be used for determine if estimated position from WiFi infrastructure is right due to correlation of measured data to expected fingerprint.

3 Remote Control

The remote control is the first mobile device from the set. The remote control provides Human-Machine Interface. The remote control is a medium sized handheld device. Remote control provides monitor and control of the probe, display information from sensors, setting the task or the desired position in auto mode, manual control using a gamepad or joystick, maps insertion and actualization, audio-visual interface, diagnostic interface with trends and help. The user has full control of the second mobile device (the probe) or he can give the task to the probe. The user can handle the probe using a color display and touch panel. Manual mode is also available. The probe is controlled by a gamepad in manual mode. Programmed application provides intuitive user interface with a set of buttons, value displays and screens. The remote control uses the iMX31LiteKit embedded controlling board based on the ARM architecture. The core of the board is the Freescale iMX31 microprocessor. The controlling application is stored on external SD card. The remote control communicates with the probe using the WiFi module Owspa311 connected via serial interface. The remote control uses a touch panel placed on the color display with a resolution of 640x480 pixels. It is possible to use a gamepad or joystick in the case of manual mode. The architecture is displayed in Fig. 5.

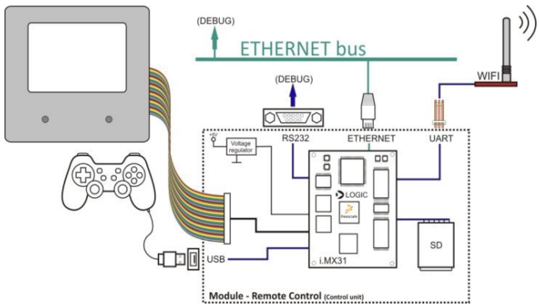


Fig. 5. Architecture of the Remote Control – display, mainboard and WiFi communication.

4 Probe

The second mobile device – the probe – is based on a massive chassis with a distributed control system, motor and high capacity battery. It is approximately 85cm meter long and 60cm high without an antenna. The weight is approximately 10kg including the DC drive, the battery, the CAN based distributed control system and all sensors.

The probe includes a set of sensors for distance measuring, environment measurement and position and movement detection. For example, the Laser scanner can measure an environment up to 20meters in 270 degrees. The probe includes infra-camera, pieso and optical distance sensors, pressure, temperature, GAS sensors, 3-axis accelerometers etc. The audio-visual interface is in the preparation phase. [5,6] The main control unit uses the same HW architecture like the remote control. Fig. 6. Other control boards are based on industrial microcontrollers Freescale HCS12 with the cooperation of the FreeRTOS operating system. These boards are programmed using C programming language using the CodeWarrior IDE. The probe communicates with the remote control using the WiFi module Owspa311.

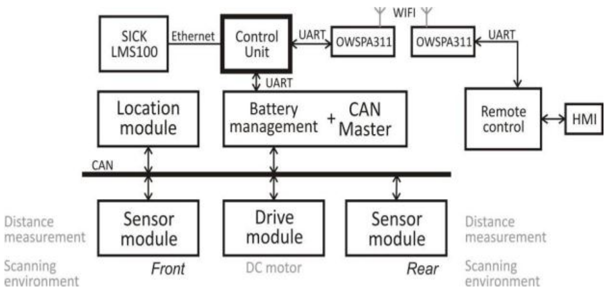


Fig. 6. Architecture of the probe: CAN based distributed control system.

5 Conclusion

The proposed set of mobile devices in the article successfully works at the Technical University of Ostrava. The main purpose of the project is to give the tool to the user which can extend the user perspective. The mobile device is equipped with an algorithm for finding the best way from the current to desired position. The outdoor precision obtain from GPS is sufficient, but the precision of indoor localization using only WiFi is insufficient. The current state of the project due to low precision is to use information from distance and other sensor to precise the localization. As a result the current work is focused on using data from laser scanner. In the future dynamic generation of the map based on information from sensors will be implemented.

Acknowledgements

The work and the contribution were supported by the project FR-TI2/273 “Research and development of progressive methods of long-distance monitoring of physico-mechanic quantity including wire-less transmission of processed data.” and the SGS project SP/2011.

Extract from the article published on: In Intelligent Information and Database Systems. Korea.

References

1. Mohapatra, D., Suma, S.B.: Survey of location based wireless services. In: IEEE International Conference Personal Wireless Communications 2005, ICPWC 2005, pp. 358-362 (2005)
2. Hui, L., Darabi, H., Banarjee, P., Jing, L.: Survey of wireless indoor positioning techniques and systems. In: Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions Volume: 37 , Issue: 6, pp.: 1067 – 1080, ISSN: 1094-6977 (2007)
3. Krejcar, O.: User Localization for Intelligent Crisis Management. In: AIAI 2006, 3rd IFIP Conference on Artificial Intelligence Applications and Innovation, pp. 221-227 (2006)
4. Krejcar, O., Cernohorsky, J.: Database Prebuffering as a Way to Create a Mobile Control and Information System with Better Response Time. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2008, Part I. LNCS, vol. 5101, pp. 489–498. Springer, Heidelberg (2008)
5. Kotzian, J.; Konecny, J.; Prokop, H., Lippa, T.; Kuruc, M. Autonomous explorative mobile robot: Navigation and construction (2010) Proceedings - 9th RoEduNet IEEE International Conference, RoEduNet 2010, art. no. 5541599, pp. 49-54., ISBN: 978-14244-7335-9
6. Srovnal Jr.,V., Machacek Z., Hercik R., Slaby R., Srovnal V.: Intelligent Car Control and Recognition Embedded System, In Proceedings of International Multiconference on Computer Science and Information Technology, IMCSIT 2010 – RTS 2010, 18.-20. October 2010, Wisła, Poland, pp. 831-836, IEEE CFP0964E, ISBN 978-83-60810-22-4, ISSN 1896-7094
7. Evennou, F., Marx, F.: Advanced integration of WiFi and inertial navigation systems for indoor mobile positioning, In Eurasip journal on applied signal processing, Hindawi publishing corp., New York, USA, (2006)

Use of Mobile Phone as Intelligent Sensor for Sound Input Analysis and Sleep State Detection

Ondrej Krejcar, Jakub Jirka, and Dalibor Janckulik

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{ondrej.krejcar, jakub.jirka1, dalibor.janckulik}@vsb.cz

Abstract. A sleep itself is not just a passive process. Throughout the night a specific number of sleep stages take place. Sleeping is a strongly dynamic process that is terminated by waking up. A sleep consists of number of stages that are repeatedly changing in various periods of time. This specific time interval and specific sleep stage are very important in the wake up event. If we wake up during the deep NREM (2-4) stage of a sleep it is far more difficult to wake up because the rest of the body is still sleeping. On the other hand if we wake up during the mild (REM, NREM1) sleep stage it is much more pleasant for us and for our body. This problem led the authors to create this study and as well as a Windows Mobile based devices application called wakeNsmile. The wakeNsmile application records and monitors the sleep stage for specific amount of time before desired alarm time set by users. It uses with a built-in microphone and determines optimal time to wake the user up. Hence, if the user sets an alarm in wakeNsmile to 7:00 and wakeNsmile detects that a more appropriate time to wake up (REM stage) is at 6:50, the alarm will start at 6:50. Current availability and low price of mobile devices is yet another reason to use and develop such an application that will hopefully help someone to wakeNsmile in the morning.

1 Introduction

A sleep is a complex process regulated by our brain and as such is driven by a 24 hour biological rhythm.

As we age, our sleeping habits change rapidly. A new-born baby sleeps in short periods for 18 hours a day. According to studies, [3] hours of sleep is enough for some people but others need 8 or even 10 hours of sleep to feel rested. Recent research 1 shows that differences in biological rhythm of people can explain why some teenagers have such difficulty in waking up. It seems that during puberty this rhythm shifts and adults tend to go to bed much later and to wake up much later as well.

This sleep shift is common and during adolescence it disappears. [2]

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 255–260.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Our biological clocks are controlled by chemical substances that are mostly known to us. One of them is hormone called melatonin which is suspected to make us feel sleepy.

This substance is produced in our brain and some scientists believe that it is also causes our metabolism to slow down before falling asleep. Melatonin secretion leads to a reduction of body temperature, a limitation of blood flow towards brain and a slackness of muscles. [3]

1.1 Sleep Stages

Approximately two hours after we fall asleep our eyes start to move back and forth irregularly. Based on this fact scientists divided sleep stages into two main stages, namely REM sleep with (Rapid Eye Movement) and NREM sleep stage (Non Rapid Eye Movement). NREM sleep is divided into another four sub-stages in which a sleep gets gradually deeper and deeper. [3]

During a healthy sleep, REM and NREM stages interchange several times.

Table 1. describing and outlining the processes in human body during individual sleep stages follows. It shows relations between biological manifestations in different sleep stages [3].
This table and information therein are taken into account in wakeNsmile detection algorithm.

Table 1. Biological manifestations in sleep stages.

Physiological Process	During NREM	During REM
brain activity	decreases from wakefulness	increases in motor and sensory areas, while other areas are similar to NREM
heart rate	slows from wakefulness	increases and varies compared with NREM
blood pressure	decreases from wakefulness	increases (up to 30 percent) and varies from NREM
blood flow to brain	does not change from wakefulness in most regions	increases by 50 to 200 percent from NREM, depending on brain region
respiration	decreases from wakefulness	increases and varies from NREM, but may show brief stoppages (apnea); coughing suppressed
airway resistance	increases from wakefulness	increases and varies from wakefulness
body temperature	is regulated at lower set point than wakefulness; shivering initiated at lower temperature than during wakefulness	is not regulated; no shivering or sweating; temperature drifts toward that of the local environment
muscle tension	decreasing with increase of NREM	increases from NREM

2 Sleep Stage Analysis Algorithm

A detection algorithm is based on an assumption that if a patient is in the REM phase and his sleep is not disturbed or pathological in any way, his muscles start to spasm which leads into movements that can be recorded and detected by the device microphone. If the patient is in NREM stage 2 or any higher sleep stage, all his muscles are relaxed and thus no body movement occurs. [6].

The recording itself is triggered 30 minutes (by default) before the wake up time as set by the user before going to sleep. If a REM stage movement is detected within this interval, the user is awaken at the time when this movement occurs.

The time interval of 30 minutes was estimated on the basis of the facts mentioned in studies [3], [4] and on empirical testing.

This time window represents the REM stage of sleep, when we are closest to being awake.

If no movement sound is detected during this 30 minutes window, the user is awaken at the set up time. According to [3], at least one REM phase occurs during this 30min. window.

2.1 Filterable Pathological Manifestation Sounds – Pathologicals Filter

From the signal analysis point of view, snoring is a periodical signal with similar amplitudes. The snoring sound is periodic with very small deviations and clearly distinguishable from other signals. Snoring is detected by the algorithm as a periodical event with amplitudes with small deviations.

Snoring filter algorithm is based on the following facts:

- Time sample value TX is taken from a peak of a signal.
- Amplitude value AX is a peak value at a time sample stamp.
- If the last two time sample difference TD is within 15% deviation of the total time difference calculated as a moving average of 5 last time samples TA, then the time sample stamp is classified as snoring S and taken out from further processing.
- If the signal at a time sample stamp TX is classified as snoring S, amplitude AX is checked for the difference within 50% deviation of the last maximum signal AM.

2.2 Environmental Noise and other interferences

Every area where the patient is sleeping and where is he being recorded is full noise and other interfering signals. These signals need to be filtered in order to get a necessary quality of a signal for further REM/NREM stage signal analysis and detection. Noise cancelation FIR filter is used in this application. Math.NET Neodym [8] adaptive FIR filter library functions are used to filter environmental noises.

2.3 Physiological Detectable Sounds

This chapter describes the algorithm used to detect the user REM stage movement during a sleep. Data detected by this algorithm are already filtered and are classified to be processed.

The algorithm itself is based on a time-series recorded data analysis. It uses real time sound wave analysis (time, amplitude) that are abnormal and above an isoline. The isoline is computed by a linear moving average algorithm as in noisy areas and thanks to the input devices different noise level the isoline might be above 0. The difference method between the real time data and the moving averaged data filter out both these problems.

3 wakeNsmile Application

The WakeNsmile application (Figure 6) is written in C# programming language and uses .NET Compact Framework version 3.5, which is a special derivative of .NET Framework for mobile devices 10. The application was developed in Visual Studio 2008 Team Edition on Windows Mobile 6.5 emulator.

A number of tests with several types of WM PDA devices have been taken.

3.1 Application design

The WakeNsmile application uses a user control called Alarm that has been created as a part of this project.

The application is using Math.NET [8] neodym library for FIR (Finite Impulse Response) filter design and WaveIn and WaveOut libraries [10] for mobile device sound interface communication.

The application structure is designed with respect to the real time recording and processing. The top level application structure is displayed on (Figure 1).

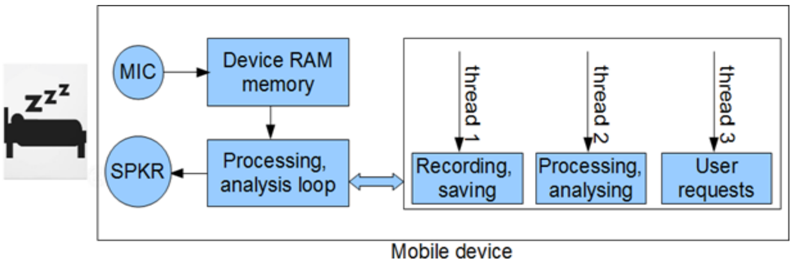


Figure 1. Block diagram of wakeNsmile application

4 Results Discussion

This chapter displays and describes results on tested subjects divided into two main groups as subjects with **pathological sleep (group - P)** and subjects with normal **healthy sleep (group - H)**.

For the purposes of testing, subjects were split into two main groups and subgroups, which, from the sleep sound manifestation point of view, is sufficient to classify and divide tested subjects: **H** – healthy sleep, **P** – pathological sleep (S – snoring, N – no manifestations, O – other issues)

4.2. Healthy Sleeping Subject

Figure 2. Recorded sound data of subject n.1

Result: The subject was correctly woken up by the application 16min., 39sec. before the set up time. **Comment:** The application correctly evaluated situation and woke user up.

5. Project extension and new capabilities

wakeNsmile project and application was developed on Windows mobile 6.5 as an application for REM/NREM sleep stage detection. Both platform and detection methods are being continuously upgraded. New version of an application is being developed on Android platform and REM/NREM stages are detected not only by sound movement detection but by movement detection itself with the application of devices built-in accelerometer sensor. This application extension brings better results then sound only input recording and detection. Application area is also beign extended to epileptical sizure detection in corporation with university hospitals neurological department for small children.

Figure 11. Android application port and extension

Project publications

This project results has been published in Sensors journal 2011, 11(6), 6037-6055; doi:10.3390/s110606037 on 3rd June 2011 and on ICMEE conference 2010 in Kyoto, Japan. [12]

References

1. M. A. Carskadon, C. Acebo, G. S. Richardson, B. A. Tate, and R. Seifer, "An approach to studying circadian rhythms of adolescent humans," *Journal of Biological Rhythms*, vol. 12, pp. 278–289, 1997.
2. J. Michael, M. D. Thorpy, E. Korman, A. J. Spielman, and P. B. Glovinsky, "Delayed sleep phase syndrome in adolescents," *Journal of Adolescent Health Care*, vol. 9, Iss. 1, pp. 22-27, 1988.
3. P. M. Fuller, J. J. Gooley, C. B. Saper, "Neurobiology of the Sleep-Wake Cycle: Sleep Architecture, Circadian Regulation, and Regulatory Feedback," *Journal of Biol Rhythms*, vol. 21, no. 6, pp. 482-493, 2006.
4. Rechtschaffen, "Current perspectives on the function of sleep". *Journal of Perspectives in Biological Medicine*, vol. 41, no. 3, pp. 359–390, 1998.
5. Sleep Cycle mobile application (http://www.mdlabs.se/sleepcycle/). 2010
6. S. Scholle and T. Schäfer, "Atlas of states of sleep and wakefulness in infants and children," *Journal of Somnologie - Schlafforschung und Schlafmedizin*, vol. 3, no. 4, pp. 163-241, 1999.
7. R. Stoohs and Ch. Guilleminault, "Snoring during NREM sleep: respiratory timing, esophageal pressure and EEG arousal," *Journal of Respiration physiology*, vol 85, pp. 151-157, 1991.
8. Math.NET documentation (http://mathnet.opensourcedotnet.info/doc/), OpenSource .NET Projects 2002 - 20010
9. Microsoft developer network MSDN (http://msdn.microsoft.com/), Microsoft 1985-2010
10. Recording and Playing Sound with the Waveform Audio Interface (http://msdn.microsoft.com/en-us/library/aa446573.aspx), Seth Demsey – Microsoft January 2004
11. O. Krejcar, "Problem Solving of Low Data Throughput on Mobile Devices by Artefacts Prebuffering", *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 802523, 8 pages, 2009.
12. O. Krejcar, J. Jirka, D. Janckulik, "Proactive User Adaptive System for Windows Mobile Devices – Processing of Sound Input Signal for Sleep State Detection", in *Proceedings of 2nd International Conference on Mechanical and Electronics Engineering (ICMEE 2010)*, vol. 1, pp. 374-378, Kyoto, Japan, August 2010. DOI 10.1109/ICMEE.2010.5558525

Phasor Measurement Units and Phasor Data Concentrators Data Evaluation on Virtual Instrumentation Concept

Jakub Maňas

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jakub.manas@vsb.cz

Abstract. The article deals with PMU data processing using LabVIEW. PMU provides continuous data stream according IEEE C37.118. For the purpose of PMU testing has been developed application which receives data from several PMUs, provides time correlation of received datagrams and saves synchrophasor data. When precise time synchronized source of voltage and current signal is used as common source for all PMUs, then developed software can be used for PMU test evaluation.

Keywords: Industry applications; Software; Phasor measurement unit; Data visualization

1 Introduction

Energy producers are forced by electricity market development to make serious effort for power system security and reliability. Therefore is more paid attention to power system monitoring increase in the important branching points. Synchronized measurements are made very precisely due to further mentioned reasons. In connection with the stability of power system, there is need for synchronous phasors monitoring. For detailed power system disturbance analysis it is necessary to record synchronously disturbance states happening very often at distant power system places. Developed application (PMU-Monitor) allows monitoring of synchronous phasors (thereinafter synchrophasors) using phasor measurement units and phasor data concentrators.

2 Synchrophasor Monitoring Introduction

Sine value (such as the fundamental harmonic voltage) is defined with (1).

$$x(t) = X\cos(\varpi t + \varphi) \tag{1}$$

Where φ is the angle vector in reference time t=0 s (initial phase) and ϖ is angular frequency. Phasor vector is characterized by amplitude X and initial phase φ.

Due to that transmission and distribution power systems are circuits with spatially distributed parameters therefore voltage and current phases are not constant throughout power system (Fig. 1).

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 261–266.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Synchrophasor data-set is called the phasor frame. This set must be recorded at same time along power system diverse locations. During analysis of frames one phasor is chosen like reference $\varphi_R=0$ rad and others are considered to it. This is called phasor data-set with reference (or relative phasor frame). [1]

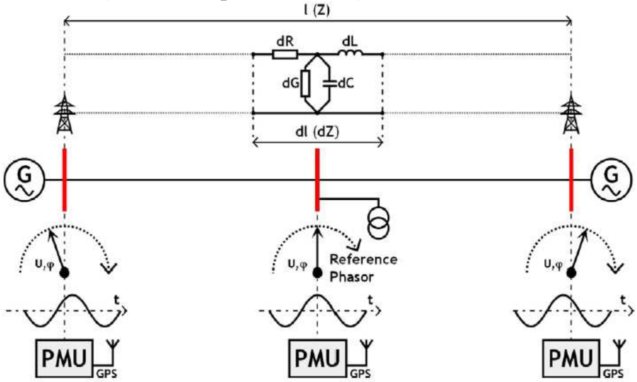


Fig. 1. Relative phasor frame.

2.1 Phasor Monitoring

Phasor measurement is made with Phasor Measurement Unit (thereinafter PMU). Transmission and distribution power system monitoring uses precise time synchronization (GPS technology). PMU evaluates values of voltage and current phasors. Precise timestamp is assigned to each phasor.

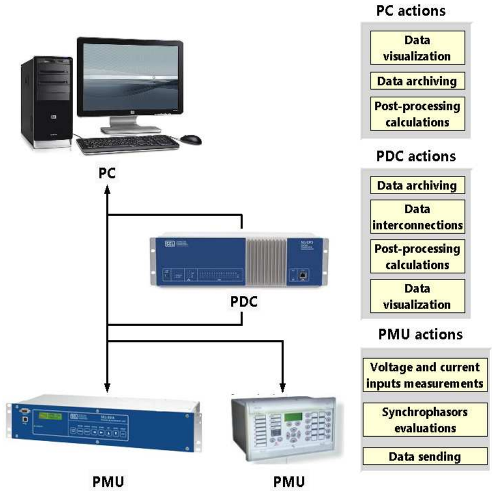


Fig. 2. Typical Synchrophasor System Architecture.

Every PMU sends data-set to Phasor Data Concentrator (thereinafter PDC). PDC provides data receiving from PMUs (Fig. 2), sending data to superior system, control PMUs and data manipulation in real time, off-line analysis (such as time alignment message from PMUs, calculation of derived variables, long-term archiving etc.).

After relative phasor frames evaluation with developed application (running on PC), critical information is displayed and user can find out power system state and then appropriate steps can be taken to ensure power system stability. [1]

2.2 Synchrophasor Monitoring Standardization

IEEE Std C37.118-2005 is basic standard for measurement synchrophasor evaluation in power systems. The standard determinates synchrophasor evaluation, verifying methods and definition of communication protocol which is used by PMU to send data to superior system in real time (for example to PDC).

Communication protocol defines PMU and PDC standard message format and the structure of so-called frames. Frames are divided into data, configuration, header and command. First three types are transmitted by PMUs, last frame type is sent by superior system. In normal operation, each PMU sends only data frames containing information specified in standard (device ID, measured data, timestamp, status, etc.).

The standard message format does not dictate hardware interface. Individual devices can use any serial communication (e.g. RS-232, RS-485), Ethernet, etc. [2]

2.3 Synchrophasor Monitoring Existing Solution

Some similar software solution can be found for synchrophasor monitoring in power systems. For example there are commercially available SEL-5073 developed by Schweitzer Engineering Laboratories and open source “PMU Connection Tester”. [3] [4]

I was forced to develop my application “PMU-Monitor” because I want export data to NI DIAdem™ for subsequent off-line analysis and measurement protocols creating.

3 PMU-Monitor Application

The application purpose is user-friendly synchrophasor visualization from PMUs and PDCs (thereinafter devices). Synchrophasors are displayed graphically and indicator data-set. User can view each device configuration without possibility change it. Application communicates with devices (user can define up to 8 devices) via Ethernet interface. Application holds 10 minutes (user-defined setting) of incoming data frames for each device in buffer. After exceeding this limit the oldest frame is overwritten.

The application is robustness in state that incoming frame is not received at sequence. For example, during network overloading the application is receiving data frames, what it is not time-aligned (e.g. after frame with timestamp “16.11.2010 14:13:15.1”, would be correctly followed “16.11.2010 14:13:15.2”, but the application received frame with timestamp “16.11.2010 14:13:15.3” first.

The application is programmed with the G language NI LabVIEW™ 2010 and NI DIAdem™ 2010 is using for measured data-set analysis.

3.1 Communication Subsystem Description

Simplified communication block diagram between devices and the application describes Fig. 3. The application initiates the communication connects (as a client) to one or more devices (IP addresses, port numbers and device identification numbers must be known). After successful connection the application sends command to devices to send the configuration frame. The device sends the configuration frame that contains needed settings for proper interpretation of data frames (e.g. device name, channel names, data representation and channel types). Then the application sends to devices command “Start sending data frames” and goes into data acquisition loop. When data frame is received it is transferred through FIFO memory to data processing loop running in parallel branch. With adding data processing loop the application robustness is increasing against data frames losses. Data acquisition is running until user does not want to abort communication (variable “Abort”) or error does not occur. Then communication is correctly closed and system memory is unallocated.

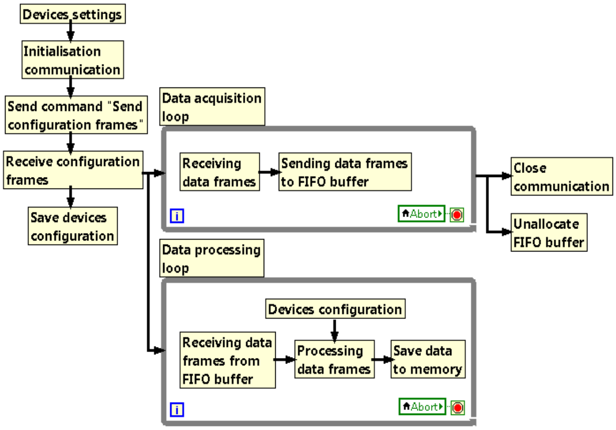


Fig. 3. Communication subsystem description.

3.2 Application Graphical Interface

Application graphical interface was designed with the emphasis on clarity and simplicity. The window size was set up to 1024*768 pixels, because it can be displayed in any monitor. In the upper left corner there are three bookmarks:

- Waveforms (graphical representation of amplitude and phase of selected synchrophasors, Fig. 4)
- Actual data (display last device data frame)
- Device configuration (devices configuration overview)

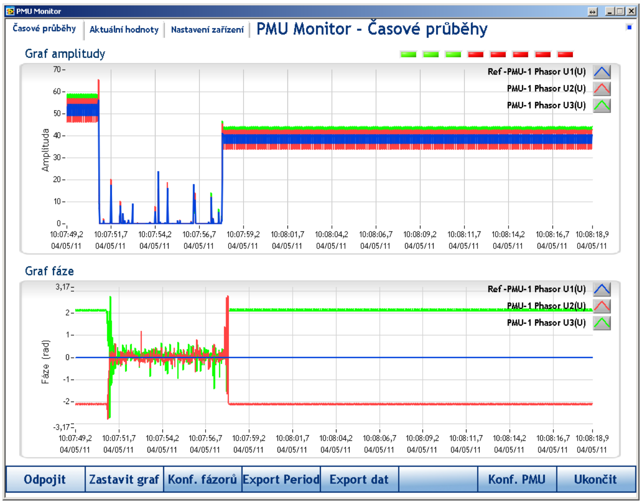


Fig. 4. The application GUI – Waveforms.

3.3 Functionality Verification

To verify whether the application meets the IEEE Std C37.118-2005 standard I used commercially available device - protection system SEL-351A (Fig. 5). The SEL-351A provides phasor measurement unit capabilities when connected to a suitable IRIG-B time source SEL-2407. The device meets the IEEE Std C37.118-2005 standard. [4]

PMU voltage inputs were connected to power system using voltage transformers.

Fig. 5. Application verify functionality using protection system SEL-351A with phasor measurement unit capabilities.

4 Conclusion

The developed application “PMU-Monitor” allows synchrophasors monitoring with display of graphs or indicators. The Software complies with IEEE Std C37.118-2005 standard and works with any commercially available PMUs and PDCs, which correspond with this standard.

Data export (to NI DIAdem™) using opens possibilities for subsequent off-line analysis and measurement protocols creating easily.

The application is as part of a larger project of research and development synchrophasors.

In comparison with similar solution “PMU Connection Tester” developed application “PMU-Monitor” makes possible to monitor more than 1 device (up to 8 devices) and export data to NI DIAdem™.

The solution is based on concept of virtual instrumentation what provides advantage that solution is upgradeable to future requirements without need for greater time and resources spending.

Acknowledgment

This research has been carried out under the financial support of the project, VSB-TU Ostrava FEECS, under Project SGS SP2011/161 and under support of project CZ.3.22/2.3.00/09.01525 of Ministry for Regional Development.

Reference

1. POPELKA, Antonín; JUŘÍK, Daniel. *Synchronní fázory - Měření a možnosti využití*. In Sborník Konference ČK CIRED 2007. Tábor : EGC - ČB, 2007. s. 12. ISBN 978-80-254-0304-4.
2. *IEEE Standard for Synchrophasors for Power Systems*, IEEE Std. C37.118-2005 (Revision of IEEE Std. 1344-1995).
3. *PMU Connection Tester* [online]. 2011 [cit. 2011-06-09]. PMU Connection Tester - Home. Dostupné z WWW: <http://pmuconnectiontester.codeplex.com/>.
4. *Schweitzer Engineering Laboratories, Inc.* [online]. 2011 [cit. 2011-06-09]. - Home. Dostupné z WWW: <http://www.selinc.com/default.aspx>.

A New Method of Precise Fundamental Frequency Estimation from DFT Spectrum

Viktor Michna and Jindřich Černohorský

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{viktor.michna, jindrich.cernohorsky}@vsb.cz

Abstract. This paper explores the possibility how to implement a frequency estimator based on a new principle. It takes advantage of the spectral leakage effect of the discrete Fourier transform. The proposed method is extended to identify the fundamental frequency. The method is then tested on simulated signals and compared with other methods.

1 Introduction

The issue of frequency estimation and pitch detection has been solved for many years and lots of methods have been developed but still there is some space to invent new and more precise methods for certain applications.

The basic problem is how to extract the fundamental frequency from a sound signal, which is usually the lowest frequency harmonic component or partial f_0 [1]. The time-domain methods include the zero-crossing rate method and its improvements, autocorrelation methods, for example the YIN estimator, and phase space methods [2]. The frequency-domain methods include component frequency ratios methods, filter based methods, cepstrum based methods, and multi-resolution (wavelet) methods [3].

This paper presents an idea how to implement a potentially very precise frequency estimator in the frequency domain. The proposed method is intentionally based on the 'blurred' frequency spectrum phenomenon of a sinusoid signal where the frequency of the signal is not an integer multiple of basic frequency.

2 Theory

Let $w(t)$ is the sinusoidal signal before time discretization, where t is time, A_0 is amplitude, k_0 is DFT bin frequency, f_s is sample rate, N is sample count, φ_0 is initial phase. Compute the coefficients of one-sided Fourier series

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 267–272.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

$$W(k)=\frac{2}{T}\int_0^T w(t)\,e^{-j\omega kt}dt=\frac{2f_s}{N}\int_0^{N/f_s} w(t)\,e^{-j2\pi k f_s t/N}dt \tag{1}$$

where $T = N / f_s = 2\pi / \omega$ is DFT window period, k is DFT bin frequency. Then the spectrum of the signal $w(t)$ is

$$W(k)=\frac{jA_0}{2\pi}\left[\frac{e^{j\varphi_0}(e^{-j2\pi(k-k_0)}-1)}{k-k_0}+\frac{e^{-j\varphi_0}(e^{-j2\pi(k+k_0)}-1)}{k+k_0}\right] \tag{2}$$

Now modify the spectrum to be a smooth function fitting in every integer frequency k

$$\tilde{W}(k)=\frac{jA_0}{2\pi}\left[\frac{e^{j\varphi_0}(e^{+j2\pi k_0}-1)}{k-k_0}+\frac{e^{-j\varphi_0}(e^{-j2\pi k_0}-1)}{k+k_0}\right] \tag{3}$$

The equations (2) and (3) imply that if k is integer then $\tilde{W}(k)=W(k)$. For an example of $W(k)$ and $\tilde{W}(k)$ functions, see Fig. 1.

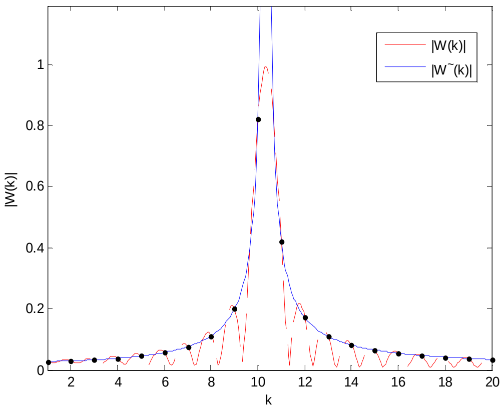


Fig. 1. Comparison of original amplitude spectrum W(k) and modified spectrum W˜(k) for k_0=10+1/3

The value of $\tilde{W}(k)$ is infinite for $k = k_0$. Define a one-sided conditionally inverted spectrum $\tilde{M}(k)$

$$\tilde{M}(k)\overset{\text{def}}{=}\frac{1}{\tilde{W}(k)} \tag{4}$$

Then $\tilde{M}(k_0)=0$. This finding can be used to search the partial DFT bin frequency k_0. Define an auxiliary real function

$$\tilde{M}_{sgn}(k)\overset{\text{def}}{=}\operatorname{sgn}(k-k_0)\left|\tilde{M}(k)\right| \tag{5}$$

This implies that $\tilde{M}_{sgn}(k_0) = 0$. The function $\tilde{M}_{sgn}(k)$ is smooth and near linear around zero crossing and seems more suitable for finding the frequency k_0 than $\tilde{M}(k)$, because it is a real function and its values can be obtained directly from the amplitude spectrum $|W(k)|$ which results from (2-5).

$$k \in \boldsymbol{N} \Rightarrow \tilde{M}_{sgn}(k) = \frac{\operatorname{sgn}(k - k_0)}{|W(k)|} \tag{6}$$

The functions $W(k)$, $\tilde{W}(k)$, $\tilde{M}(k)$ change their phase by π at the bin frequency k_0.

Furthermore, there is a relation between the derivation of one-sided conditionally inverted spectrum $\tilde{M}(k)$ in neighborhood of the critical bin frequency k_0 and the amplitude of the signal A_0

$$A_0 = \frac{\pi}{\left|\frac{\partial \tilde{M}(k_0)}{\partial k}\sin(\pi k_0)\right|} = \frac{\pi}{\left|\frac{\partial \tilde{M}_{sgn}(k_0)}{\partial k}\sin(\pi k_0)\right|} \tag{7}$$

which can be used for the amplitude calculation.

Furthermore, there is a relation between the initial phase of the signal φ_0 and $\tilde{W}(k_0)$

$$\varphi_0 = \lim_{k \to k_0 -} \arg(\tilde{W}(k)) - \frac{3\pi}{2} - \arg\left(e^{j2\pi k_0} - 1\right) = \lim_{k \to k_0 +} \arg(\tilde{W}(k)) - \frac{\pi}{2} - \arg\left(e^{j2\pi k_0} - 1\right) \tag{8}$$

3 Proposed method

The problem of the fundamental frequency estimation can be divided into the following steps – location of harmonic partials, suppression of the localized partials from the spectrum, and finding the fundamental one.

3.1 First step – precise localization of harmonic component

Let $W(k)$ is a normalized one-sided DFT spectrum, where $k = 1, 2, ..., N/2$ and k_e is the DFT bin frequency with the maximum amplitude.

1. Compute

$$\Delta\varphi_a = \pi - \left\|\arg\left(W(k_e)\right) - \arg\left(W(k_e - 1)\right)\right| - \pi\right| \tag{9}$$

$$\Delta\varphi_b = \pi - \left\|\arg\left(W(k_e + 1)\right) - \arg\left(W(k_e)\right)\right| - \pi\right| \tag{10}$$

2. If $\Delta\varphi_a < \Delta\varphi_b$, then $\Delta k_r = 1$, otherwise $\Delta k_r = 0$.
3. Calculate for $k = -2 + k_e + \Delta k_r, \ldots, 1 + k_e + \Delta k_r$

$$\tilde{M}_{sgn}(k) = \begin{cases} k < k_e + \Delta k_r : & -\left|W(k)\right|^{-1} \\ k >= k_e + \Delta k_r : & \left|W(k)\right|^{-1} \end{cases} \tag{11}$$

$$\tilde{\varphi}_{sgn}(k) = \begin{cases} k < k_e + \Delta k_r : & \arg(-W(k)) \\ k >= k_e + \Delta k_r : & \arg(W(k)) \end{cases} \tag{12}$$

   If anyone from the calculated $\left|\tilde{M}_{sgn}(k)\right| = \infty$, then the method SCIPI has failed.
4. Calculate $\tilde{M}_{sgn}(k)$ and $\tilde{\varphi}_{sgn}(k)$ by cubic interpolation in the interval $\left\langle k_e + \Delta k_r - 1; k_e + \Delta k_r \right\rangle$.
5. Find the solution k_0 of the equation $\tilde{M}_{sgn}(k) = 0$. If no solution is found, the method has failed.
6. Calculate the derivation of estimated polynomial $\tilde{M}_{sgn}(k)$ in the point k_0 and substitute in the equation (8) to determine the amplitude A_0 of partial.
7. Determine the initial phase of partial

$$\varphi_0 = \tilde{\varphi}_{sgn}(k_0) - \frac{\pi}{2} - \arg\left(e^{j2\pi k_0} - 1\right) \tag{13}$$

3.2 Second step – selection of the fundamental frequency component in a signal

There have already been many efforts to use harmonic structure models to estimate signal frequency [4],[5]. Our method is using the following principle of exclusion of higher harmonic partials except the fundamental one.

Let $k_0[h]$, $A_0[h]$, $\varphi_0[h]$ is the list of found partial components where $h = 1, 2, \ldots H$, sorted from the one with the biggest amplitude $A_0[h]$.

1. In the cycle for $n = 1,2, \ldots N_{max}$ we are searching for the harmonic number n_{fund} of partial with the biggest amplitude (h=1)
2. For every partial h compute harmonic ratio

$$p_a[n,h] = k_0[h] / (n\, k_0[1]) \tag{14}$$

$$p_b[n,h] = \mathrm{round}(p_a[n,h]) \tag{15}$$

3. Compute the weight of the partial for maximum frequency deviation 50 c = 3%

$$m_a[n,h] = A_0[h]\left[1 - 24\left|\log_2\left(\frac{p_a[n,h]}{p_b[n,h]}\right)\right|\right] \tag{16}$$

If $p_b < 1$ or $p_b > N_{\max}$ or $m_a < 0$, then the partial *h* is out of range, therefore it is skipped ($m_a = 0$) for given *n*.

4. Calculate the sum of weights

$$m_{sum}[n] = \sum_h m_a[n,h] \tag{17}$$

5. Find such n_{fund} that sum of weights $m_{sum}[n_{fund}]$ is minimal.
6. Find such h_{fund} that $p_b[n_{fund}, h_{fund}] = 1$.
7. Then the fundamental frequency and amplitude are

$$f_{fund} = k_{fund} \frac{f_s}{N} = k_0[h_{fund}] \frac{f_s}{N} \tag{18}$$

$$A_{fund} = A_0[h_{fund}] \tag{19}$$

4 Results of testing

Before testing on real musical signals, the precision of the method will be tested on a simulated signal with defined parameters. We will compare our method (SCIPI) with other methods.

UACPI is a time-domain method based on an unbiased autocorrelation. After the calculation of $R(\tau)$, the negative part τ and *N*/4 values are discarded. Then the first maximum is found and frequency is calculated by parabolic interpolation.

FMPI is a simple frequency estimation method based on FFT. The maximum peak in the amplitude spectrum is found and frequency is calculated by parabolic interpolation. There are some advanced modifications of this method [6],[7].

YIN is the well-known fundamental frequency estimator based on a modified autocorrelation method with a number of modifications and a simple algorithm [2].

In the first test we will compare the methods on three types of signal:

a) a sinusoidal signal: $y_1(t) = \sin(2\pi f_0 t)$,
b) a signal composed of three partials with frequency ratios 1:2:3 and equal amplitudes: $y_2(t) = \sin(2\pi f_0 t) + \sin(4\pi f_0 t) + \sin(6\pi f_0 t)$,
c) a square signal: $y_3(t) = \mathrm{sgn}[\sin(2\pi f_0 t)]$.

The frequency of the testing signal is changing from 100 to 5000 Hz with exponential steps. A rectangular window with N=2048 samples and sample rate f_s=44100 Hz is used here. Also the 16-bit quantization is applied. Only one window is sampled and estimated for every tested frequency. Average relative errors of frequency estimation are summarized in Table 1.

Table 1. Comparison of frequency estimation methods

δ [%]	1 harm.	3 harm.	square
uacpi	0.0733	0.0729	0.1780
yin	0.0609	0.0696	0.1872
fmpi	0.9278	0.9109	0.9250
scipi	0.0010	0.0184	0.0066

5 Conclusion

In this paper a new method of frequency estimation of partial signal components, SCIPI, has been proposed. The method draws on the behavior of DFT transform of a periodic signal between the spectral lines caused by the spectral leakage.

The tests on basic simulated signals show that the method is potentially extremely precise. It is about 10^2–10^3 times more precise than a simple parabolic interpolation method in the DFT spectrum and the YIN method is better only at lower frequencies. However, the method is sensitive to the shape of signal and distortion yet.

The method can be used to estimate the fundamental frequency. Also it could be used to analyze polyphonic signal. Another application is the filtration of partial periodic (distortion) components from signal due to the capability of estimating the amplitude and the phase of the partial too.

Future work will focus on more tests to find out the limitation of the method usage. These include testing a degree of interference caused by neighbor partials in a signal and testing on recorded signals from real world.

References

1. Gerhard, D.: Pitch extraction and fundamental frequency: History and current techniques. Tech.Report TR-CS 2003-06. Dept. of Computer Science. University of Regina (2003)
2. de Cheveigné, A., Kawahara, H.: Yin, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, vol.111, no.4 (2002) 1917-1930
3. Zhang, J.X., Christensen, M.G., Jensen, S.H., Moonen, M.: A Robust and Computationally Efficient Subspace-Based Fundamental Frequency Estimator. Audio, Speech, and Language Processing. IEEE Transactions on, vol.18, no.3 (2010) 487-497
4. Shi L., Zhang, J., Han, G.: Multiple Fundamental Frequency Estimation Based on Harmonic Structure Model. Image and Signal Processing. CISP'09. 2nd International Congress on (2009) 1-4
5. Klapuri, A.P.: Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. Speech and Audio Processing. IEEE Transactions on, vol.11, no.6, (2003) 804-816
6. Milivojevic, Z.N., Mirkovic, M.D., Milivojevic, S.M.: Fundamental Frequency Estimation of the Speech Signal Compressed by MP3 Algorithm Using PCC Interpolation. Advances in Electrical and Computer Engineering, vol.10, no.1 (2010)
7. Shi L., Zhang, J., Pang, L.: Piano Fundamental Frequency Estimation Algorithm Based on Weighted Least Square Method. Information Technology and Applications. IFITA '09. International Forum on, vol.2 (2009) 155-158

Automated Testing of Measurement Instruments

Petr Pětvaldský and Petr Bilík

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{petr.petvaldsky, petr.bilik}@vsb.cz

Abstrakt. In these days the term power quality becomes to be significant in industry. Many manufacturers develop equipment for measuring the quality of electrical energy for various applications. However, the qualitative requirements for the operation of these devices are still not precisely defined. So it is not an exception that different manufacturers differently interpret requirements. Then it is more difficult to compare qualitative parameters of these different devices. This paper describes how to test functionality and capability of devices for measuring power quality.

Key words: Power quality, Flicker, Power system measurements, Automatic testing, Instrumentation and measurement;

1 Introduction

In recent years, occurrence of disturbing effects in the electricity system is increasing which results in degradation of electrical energy quality. Therefore, power quality is becoming increasingly important for the electricity distributors. In addition, the interest of European regulatory authorities in the quality of electricity is constantly increasing.

Given these facts, electrical systems for measuring and analyzing power quality are increasingly used in various places of the electrical grid. Although the requirements for evaluation of the power quality are described in IEC 61000-4-30, currently, the testing equipment for testing power quality analyzers is not widespread.

Therefore, we have decided to develop a fully automated system that will be able to test the device for measuring the quality of electric energy. Currently, this device is still under development, but we have already made the first stage of the tester development - the evaluation of flicker. We have also made the first real testing of several analyzers from various manufacturers.

2 Description of test system

Voltage changes were implemented on a programmable voltage source HP6834B Hewlett Packard (now Agilent HP6834B). Power source is capable of carrying up to 3kW, so there is not a problem to supply testing devices. In order to control HP6834B source via GPIB interface LabVIEW application was developed. The power source in

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 273–278.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

combination with PC enables to implement a number of voltage changes. The generated sequences were verified by an independent test application created in LabVIEW using the following hardware: cDAQ chassis that can be connected to a PC via a USB port and NI9225 module, which provides three galvanically isolated voltage inputs with a range of 300V. The entire assembly includes additional control oscilloscope AGILENT. All mentioned devices are shown on block diagram Figure 1.

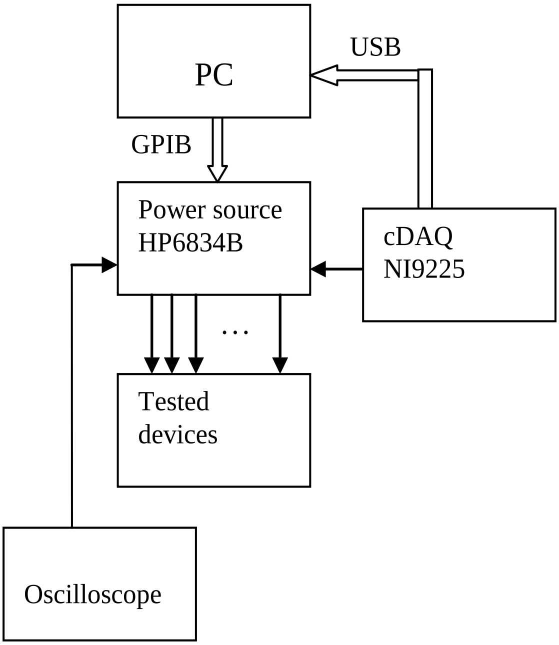


Fig. 1. Test system block diagram

These devices were selected due to the possibility of connection with a computer and ability to control (data acquisition) using LabView programming environment.
Programmable source HP6834B has sufficient capacity to generate almost any testing waveform or to simulate real waveforms.
For a comprehensive comparison of flicker evaluation on different PQ analyzers were first identified the typical voltage changes in distribution network caused by different appliances: sawmill, pellet press and heat pump. Typical changes of voltage caused by such appliances were measured in real distribution network with device TOPAS1000. To capture the rapidly changing rms voltage, data of periodical values of rms voltage were used. Recorded voltage changes were simulated on a programmable voltage power supply HP6834B.

On the output of HP6834B power supply six different types of PQ analyzers were connected:

- Topas 1000
- Memobox 300
- Fluke 1744

- ENA 330
- ENA 450
- PQ Monitor MEG30 NN

Before the test, we verified that all three defined and generated typical courses of voltage changes correspond to voltage changes recorded on a real network, close to the typical real load.
At the beginning of the test, connecting of all devices to a source HP6834B was done. Although the source is three-phase supply, voltage inputs of all devices were connected to phase L1 of the source, to avoid any possibility of connecting different signals to particular input devices. This type of connection makes it impossible to test the evaluation of voltage unbalance, but it was not the subject to the test. [1],[3]

3 Description of test application

The core of the entire test system is a software application created in LabVIEW. This application allows users to control power source HP6834B. The user can define almost any testing waveform with the required parameters, which is then generated on the output of source.
The proposed application is quite extensive, to allow full use of functional possibilities of programmable of source. In this article is presented image of the user interface, which is used for testing flickermeters. This part allows creating any sequence of voltage waveforms, while the number of adjustable parameters is very rich. This way you can create and simulate almost any real waveform of practice.

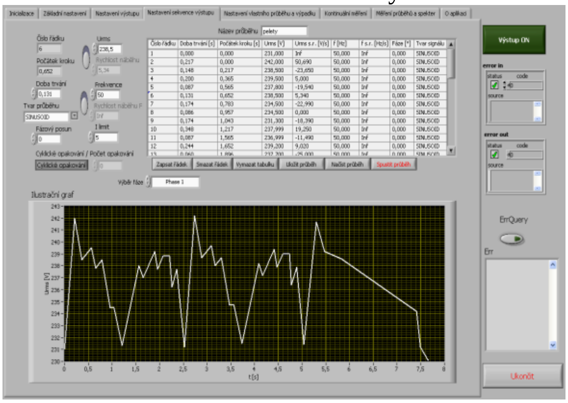


Fig. 2. Test application screenshot

4 Description of test process

As was mentioned in introduction realized tester is testing PQ analyzers in terms of evaluation of flicker parameter.

Flicker is a visible change in brightness of a lamp due to rapid fluctuations in the voltage of the power supply. From the momentary value of flicker is calculated the short term flicker (Pst) accordingly to a statistical process over a ten minutes observation interval. Long term flicker (Plt) is calculated as the cubic average of short term flicker values over a 2 hours observation interval.

4.1 Voltage Changes Caused by Pellet Press

Pellet press being under operation reports varying load and idle operation is mostly absent here. Operation of the press and the resulting consumption of power cause the constant changing of the effective voltage without delay where would be no changes in voltage. It can also be expected to conduct flicker without significant changes.

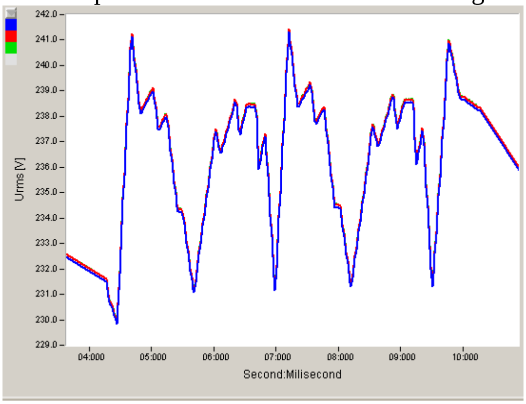


Fig. 3. Voltage changes record – pellet press

The recorded waveforms show that with the exception of the PQ-monitor analyzer (MEG30), which indicated Pst and Plt approximately 2.75, all other analyzers, showed the result of about 2.25. Unlike Pst data recorded by other analyzers, Pst data from analyzer MEG30 show a slight ripple.

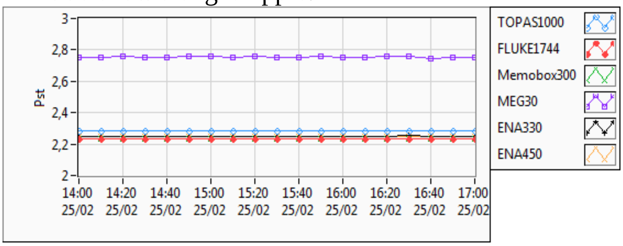


Fig. 4. Pst measured data – pellet press

4.2 Voltage Changes Caused by Heat Pump

Heat pump causes a decrease, which is typical for rapid voltage changes. When the heat pump is turned on a large current peak appear, caused by asynchronous motor. After starting the motor current peak disappears and the engine draws the nominal current. After a time the pump is turned off, the current drops to zero and the voltage returns to its original value. Described action causes an irregular network load specific by the relatively long time when the pump is turned on and off (in the simulation it was 15min and 15 min). Irregular network load causes not too frequent changes in the level of rms voltage so significant changes of the flicker value can be expected.

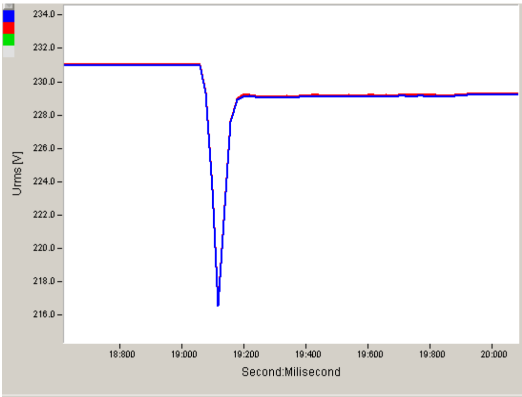


Fig. 5. Voltage changes record – heat pump

The measured Pst data show that with the exception of the analyzer PQ-monitor (MEG30), which indicated peak Pst = 0.22, all other analyzers showed peaks Pst = 0.9. Peaks on Pst from TOPAS1000 analyzer located at different times than other devices. Due to carried out setting of the internal clock of the instruments before the beginning of measurements, it is unlikely that this is caused by inaccurate clock. Probably Topas 1000 has a different algorithm to record a time stamp of the measurement interval. [4],[5],[8]

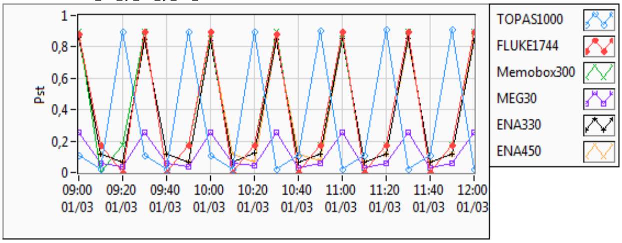


Fig. 6. Pst measured data – heat pump

5 Conclusion

At the beginning of the conclusion has to be noted that the above mentioned tests are not described or required by applicable standards [5] and [8], but they verify the responses of flicker meters on voltage changes which are commonly occurring in practice. So it cannot be said that any device do not evaluate in accordance with applicable standards, but it can be stated that evaluating is different than others, whose results are almost identical. Evaluation of responses to various types of repetitive voltages changes in the form of the measured flicker Pst of the PQ analyzer is shown in Table 1.

Table 1. Pst measured data

Voltage changes	Press	Saw-mill	Heat-pump
PQ analyzer	Pst	Pst	Pst
Topas 1000	2,285	4,044	0,920
Memobox 300	2,235	4,012	0,901
Fluke 1744	2,235	4,013	0,901
MEG30	2,757	3,502	0,256
ENA330	2,252	4,072	0,851
ENA450	2,249	4,089	0,851

From Table 1, clearly indicate the following facts:
In the case of voltage changes from the pellet press responses of the PQ analyzers are almost identical, except MEG30 analyzer which has about 20% higher values of both Pst. For evaluation of voltage changes induced by saw-mill are responses again exactly the same for all analyzers except MEG30 device, which has about 20% lower Pst values.
Flicker measurement for voltage changes caused by operation of heat pump is again exactly the same for all analyzers except MEG30 analyzer, which shows again a different evaluation of Pst, see Table 1.
It can be concluded that only MEG30 device, which is declared by manufacturer as device class "B" according to IEC 61000-4-30 [6], has a different evaluation of the flicker, it may cause in a significant problems for electricity distribution companies in the case of measuring power quality in distribution grid.

Reference

[1] *AC Power Solutions : Agilent 6800 series*. [s.l.] : Agilent, 2000. 187 p.

[2] *National Instruments : Instrument drivers* [online]. 2010 [cit. 2011-03-05]. WWW: <http://zone.ni.com/devzone/cda/tut/p/id/4803>.

[3] BAGGINI, Angelo. *Handbook of Power Quality*. Wiltshire : 2008,. 642 s. ISBN 978-0470065617.

[4] PIEPER, John M. *Automatic Measurement Control*. [s.l.] : Rohde & Schwarz, 2007. 295 p. ISBN 978-3-939837-02-2.

[5] ČSN EN 61000-4-15: Zkušební a měřicí technika – Měřič blikání – Specifikace funkce a dimenzování. Praha : Český normalizační institut, 1999. 28 s.

[6] ČSN EN 61000-4-30 ed.2: Zkušební a měřicí technika – Metody měření kvality energie. Praha : Český normalizační institut, 2009. 80 s.

[7] IEC 61000-4-15:2010: Testing and measurement techniques – Flickermeter – Functional and design specifications. Geneva : International Electrotechnical Commission, 2010. 33 p.

Visual and RFID Checkout Inspection System

Radka Pustková, David Vala, Zdeněk Slanina, and Jindřich Černohorský

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{radka.pustkova, david.vala, zdenek.slanina, jindrich.cernohorsky}@vsb.cz

Abstract. This contribution discusses possibilities of implementation visual inspection and RFID techniques into labeling technology of blood container due demand of increasing process automation level in transfusion center. At the current development of the imaging equipment and computer vision, especially given the reduction of cost can be effectively used for technologies demanding of industrial processes of visual inspection. Another benefit of the application this processes is the possibility of achievement more safety in the handling of biological material in these critical applications. Also in the introduction of the RFID technologies enables automatic matching and verification of the RFID tags and bar codes. This also leads to improved safety especially at the time when all workplace are still not equipped with the RFID readers and therefore are still just using bar codes readers. Currently in cooperation with the Blood center of University Hospital in Ostrava are tested technology possibilities of both the labeling system and their reciprocal compatibility. Detailed description of visual inspection of blood containers label, matching with RFID tags and inspection of the integrity of the container is in the following text.

1 Checkout Inspection System Introduction

Blood transfusion service, blood processing and the concept of the blood centers are forms an integral part of the contemporary medicine. General pressure for reducing the cost in the current growing quality of peoples to health care is necessary to increase the level of automation even in biomedical processes. These main reason leads to the introduction of modern technology and logistics industry, such as the use of RFID technology and visual inspection is, in modern medicine. However, the RFID technology is not covered the all aspects of manipulation with biological materials, such as bag with blood is, because a lot of processes required a straight human intervention. Currently, in Cooperation with Blood Center of in the Faculty hospital and Technical University in Ostrava are in the solution the implementation of the RFID tags in blood transfusion services. This project is also dealt with the issue of safety and reliability of RFID technology for marking blood. In particular then the reliability of RFID tags, marking due to the several weather conditions on different types of bags

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 279–284.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

with blood, which are the need to maintain the readability of both radio and the visual identification information on the labels. The system for automatic visual inspection with and without used of RFID technology are developing in the MATLAB environment 2009. The second solution of visual inspection is evaluating and implementing in the standard automation software, the NI Vision Builder. These two variants of solution are then compared in aspects possibilities and efficiency of implementation in blood centers operating conditions. Using algorithms designed in MATLAB with subsequent industry components using based on the NI Vision Builder is assumed.

1.1 Checkout Technology

Department of the expedition is the place where the bag with blood leaves the blood center. Blood preparations are passing entities in the University Hospital and other organizations. Inspection is carried out on removal transfusion preparation. By work instructions is controlling the appearance and integrity of the bag, blood groups, expiration date, the correspondence from transfusion preparation with particular dates in information system and on the "Order form for before the transfusion examination and blood product "and" Requisition for output transfusion preparation."

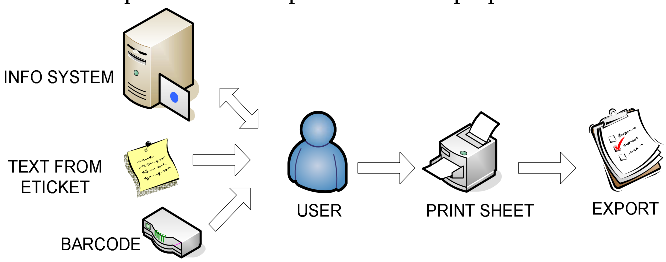


Fig. 1) Blocks of manual operations.

These operations are only performing by an eye visually without any other records about the checks. Bags with blood are identifying by manual barcode reader. This information is transfers into the information system. Information about particular bag with blood is displaying on the monitor. The employee of checkout extention compares the information shown on the label in text form. The single barcode have not be individually visual checked. Movement or delivery note with the required particulars is drawn by the employee. Type sheet depends on the type of bag with blood and target customer. Prescribed particulars of transfusion are also entering into an information system. Subsequently this transfusion product is ready to be transports to the destination.

1.2 RFID Checkout extension

If is use the RFID tags as primary identification is necessary it must by verify consistency of the RFID label with Visually marking bag with blood and accuracy the barcode. Therefore, the expedition department is equipped with a RFID reade zone which is identical with the reading zone of visual identification. RFID tag can help simplify and accelerate searching required blood products in the areas expedition and hand storage. Deposit each of labeled products will be recorded. This information will be transmitted to the information system and validated with the data from the expedition.

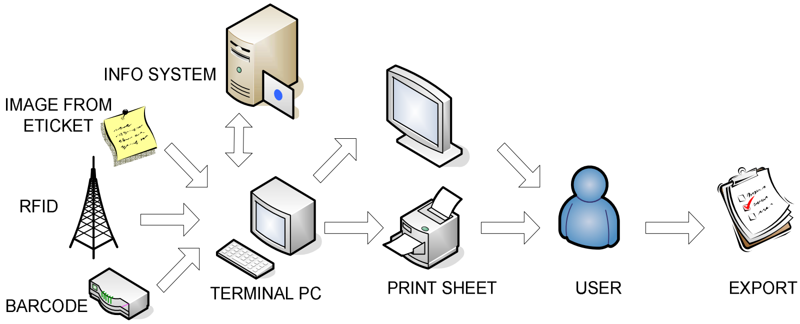


Fig. 2) Block diagram of automation Visual and RFID inspection system**.**

1.3 Visual and RFID system design

In the normal process of the expedition is not enough time to complete inspection. Manual scanning all of the codes and full visually inspection is time demanding to much and therefore the reliance is on an accuracy of recorded datas and identify information in database system. In the proposed control of system is extended with full control all of items on the label of the bag. Barcodes reading, RFID tags reading - if it´s present and visual control the bags with blood are performs in real time.

Package Control

The system verifies visual quality walls of bags. Defect detection was targeted on the recognition damage contention the bag and concentrate on entry into the division. The inspection of the bag size was also performed. The system does not detect only the bags with blood but it also detecting various types of bags and the comparison with the data from the database. It is also performed the visual control of the contents of and the bags identitying with label. The studio model is maintained of due to the possibility of full automation in process. Defect detection of bag with blood is done by computer vision algorithms. Input image is scanned by the camera likes an expertise. The mask of the bag, that has 2 bits of depth, is created during this process. Necessary depth of the object segmentation is set automatically with respect to demands of the process. Defects on bag has been detecting by analyzing performed mask. Founded defects are marks and the operator made removes it.

Label Recognizing

The every label of bag with blood contains of 5 barcode identification. There are recordes number of the product, blood type, Rh factor and also the performance and consumption. The label also contains additional text and graphical information. Color codes help in visual inspection on blood groups. Blood groups identified are used for verification datas comparing textual information with barcode and RFID tag. Individual codes are placed on fixed positions in the label. The identification mades by using the math pattern matrix in the first step. After the identification the location of the label the each elements are controled in segmented blocks on the label. Included automatic barcode readers are reading the informations from each block. The image information reading from the camera along with barcode is also compare from RFID reader. The all datas are also validating with information from the database. If a deviation is identified, such as an expired shelf life or other defect, the operator is also informs about it. The bag with identified defect removes from the process.

1.4 Safety and security management

Increasing the safety and security management with potentially dangerous biological materials and increasing the safety and security of the system are main reason for the implementation of the proposed system.

RFID tags verification
In the case of bags with blood that contains RFID tags are needed to compare the information saved in the RFID tag with information on the label. For this reason, are a place for reading is also provided with an RFID reader. The values are compared with each other. Formal control database dates accuracy are making with visually detected dates from the label. The error is identified in the case of disagreement when comparing data. Operator reading place are removed following identified bag.

Time management
In case of defined place for reading reduces the number of operations required by the personnel department expedition. Automatic label evaluation requires only a one image taken automatically after placing the bag into the reader zone. Manipulation is reduces only to placed into the reading zone.

Communication with the server
Central Blood Center database includes all data resulting in the blood production process. In safety aspects and elimination possibility of unauthorized access to datastore, the critical data are not transmitted from the server. The datas are senting to the server even if they makes evaluation and pairing with database. Consequently, the indication correct or incorrect datas returns serfer to the inspection system. This is elliminate needing remote authorization of an inspection module for database access and disabling demand data stored on central server. This solution requires the installation verification software directly on the server in blood center. Computing power required to perform verification is preserved in terms of other transactions occurring at the server is negligible. In essence it is the execution of a "one database query" to verify the accuracy of all data on the blood.

2 Actual Results

The proposed solution is currently being tested in realistic blood bags and separate labels too. The optimal reading zone location so that it accommodates demands of the expedition workplace ergonomics is currently searching. The Interface design of information system in terms of maximum security of communication between module of visual inspection and blood center information system is currently in the solution. RFID technology is currently only tested and not yet deployed in the direct operation of Blood center of FNS Ostrava. The deployment of this solution in the near future is considered. Introduced system is designed for the needs of the expedition extension of Blood Center; however, its application is possible in other departments. For example, the location of the proposed visual control of the operation of production and labeling bags will speed up the process and increase safety for at the very beginning of the process chain. A doctor and of blood Centre staff will have the maximum possible information about the movement, manipulation and conditions in which the bag found during the process. As mentioned method is now being tested. The conclusion will be specified after a sufficient number of cycles.

Acknowledgements

This application was carried out in collaboration with Blood center of Faculty Hospital of Ostrava. This work is supported by the Ministry of Education of the Czech Republic under Project 1M0567 – Centre of Applied Cybernetics.

References

[1] Cermakova, Z.; Valova, D.; Klement, P.; Kebo, V., RFID TECHNOLOGY IN BLOOD CENTER OSTRAVA - WHY TO IMPLEMENT?, VOX SANGUINIS, 99: 100-100 Suppl. 1 JUL 2010 ISSN: 0042-9007

[2] Van Der Togt, R., Van Lieshout, E.J., Hensbroek, R., Beinat, E., Binnekade, J.M., Bakker, P.J.M., Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment, (2008) JAMA – Journal of the American Medical Association, 299 (24), pp. 2884-2890.

[3] Machacek, Z., Hercik, R., Slaby, R. Smart use adaptive system for intelligent object recognizing (2011) Studies in Computational Intelligence, 351, pp. 197-206.

3 Publications

[1] Pustkova R., Kutalek F., Penhaker M., Novak V.: MEASUREMENT AND CALCULATION OF CEREBROSPINAL FLUID IN PROPORTION TO THE SKULL; 2010 RoEduNet, p. 96-102, ISSN 2068-1038

[2] Pustkova R., Kutalek F., Penhaker M., Novak V.: ALGORITHM CALCULATING THE RATIO OF CEREBROSPINAL FLUID TO THE SKULL; KTTO 2010, 13. December 2010, p. 211-215, ISBN 978-80-248-2330-0
[3] Pustkova R., Vala D., Slanina Z.: VIZUÁLNÍ INSPEKCE V KLINICKÉ PRAXI, Workshop WCRR 2011, Trojanovice 27-29.6. 2011, p.186-191, ISBN 978-80-214-4297-9

Integration of REPACE Central Registry, DASH and ICS 3000 Devices with Hospital Information System

Karel Vlach and Jakub Jirka

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{vla106, jakub.jirka1}@vsb.cz

Abstract. Paper describes the system that is developed by the authors and which allows connection of HIS (Hospital Information System) used at Městská nemocnice Ostrava, REPACE central registry of pacemaker implantations, vital signs monitor DASH and implant control system ICS 3000. This connection allows users to effectively create, store and visualize operation reports. The new system is being developed using C# programming language within .NET environment. The developed application's title is ImplantSys.

1 Introduction

The main goal of this project is to create a system that helps physicians at the cardio stimulation department of the hospital in Ostrava (*Městská nemocnice Ostrava*). After the operation, which is most often the implantation of a pacemaker, there is a need for physicians to make a report. This report contains various information like the patient's identification data or the measured signals from devices. Then it has to be stored in the nationwide central registry named **REPACE**.

There are four sources of data that the new developed system **ImplantSys** has to combine for the most effective workflow. The hospital **HIS** (*Hospital Information System* aka *EHR – Electronic Health Records*) which contains patient's administration data, **DASH** vital signs monitor used mainly for measuring ECG *(Electro Cardio Graph*), implant control system **ICS 3000** mainly used for measuring IECG (*Internal Electro Cardio Graph*) and programming the pacemaker, and the **REPACE** central registry.

From this the **ImplantSys** allows the physicians to work faster and more effectively with the patient's records in a user-friendly way. It will be the first system that allows this.

2 The New System

The ImplantSys application is running on the PC in the block of processing (see Fig. 1). There is a HUB that connects HIS, DASH and REPACE via an Ethernet interface.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 285–290.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

The ICS3000 does not support online connection in real-time. Therefore the data from the ICS3000 is transferred via one of the supported media (USB flash drive or CD) in XML format. Communication with the REPACE registry is via the internet.

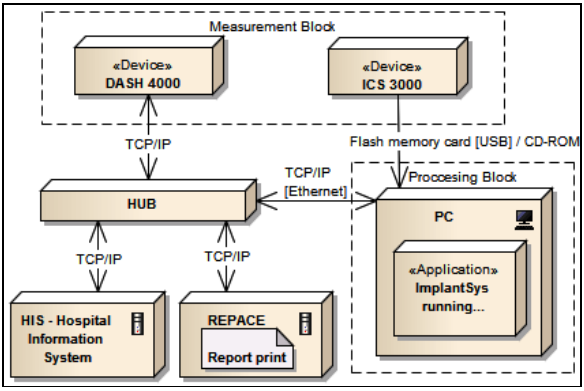


Fig. 1. Physical interconnection architecture

The data can then be imported into the application safely, reliably and above all automatically. This is much more efficient compared to the current method of manual rewriting.

3 Hospital Information System

There is important data in **HIS** which needs to be imported to the **ImplantSys**. Such data includes patient's name, birthdate, gender, weight, height, diagnosis, address, insurance and other relevant information. This will help the physician to automatically fill in the form of the report and prevent them from making a potential mistake, which can happen while rewriting the data manually.

The name of the in service HIS is **CLINICOM**. **CLINICOM** is a robust hospital information system with a modular design and an object-relational database *Caché* from the *InterSystems* company [1]. It is used by several hospitals in Czech Republic [2].

Communication between **ImplantSys** and the **CLINICOM** database is solved by the interface implemented as a dynamic-link library in C#. This interface contains several methods for getting the relevant data from the database to the ImplantSys application. The form for the report creation is filled in with the information and can then be edited. Relevant data is picked from the database by a method in which the personal identification number serves as the primary key for selection. If there is more than one record associated with the patient another key for selection is used. Date/time is the secondary key.

4 Vital Signs Monitor DASH

Various vital and electric parameters are measured during the pacemaker implantation. External and internal ECG are the primary parameters of those vital signs. Other parameters such as respiration, O2 saturation, blood pressure and others are observed only in more severe cases of a patient's condition. The vital signs monitor **DASH 4000** is used for the measurement.

4.1 Communication Protocol

The **DASH** device uses a closed protocol of the "Unity Network" [3]. There are three known types of UDP messages: *RWHAT*, *BEDSIDE* and *WAVEFORM*.

1) RWHAT message

This message serves to identify all devices on the Marquette network. Each device sends this message typically every 10 seconds. Each device involved maintains a list of these devices.

2) BEDSIDE message

This message serves to call the functions of the devices and to change the parameters. There are two types of the message - a request ("request" packet) and answer ("response" packet). The requests are used for reading and writing of the parameters, or to request a subscription ("subscribe") of the parameters and waveforms in real-time. The answer then contains the specific parameters or results of operations.

3) WAVEFORM message

This message serves to transmit the monitored waveforms in real-time. The specification of the message is not complete yet. However, there is a functional way to handle such messages and process them.
The vital signs monitor **DASH** filters out these waveforms and computes another three (1) from them automatically. But the **DASH** only displays them and does not send them to the PC so the **ImplantSys** must compute these as well. These three computed waveforms are: aVR, aVL, and aVF, which are augmented limb leads [4].

$$aVR = -\frac{I + II}{2},\ aVL = I - \frac{II}{2},$$
$$aVF = II - \frac{I}{2}\ . \tag{1}$$

For long-term remote monitoring by the application, it was necessary to create a driver that allows a reliable connection to the **DASH** device in real-time. The **ImplantSys** uses advanced design technology with regard to the real-time data processing [5]. Based on the described specification, the driver for patient monitor **DASH** was developed in the .NET framework [6].

Developed ***TMqExplorer*** component is used to detect devices on the network. The main component is ***TMqDevice*** which contains information about supported services ***TMqService*** and provides an interface for the collection of the waveforms of the measured signals ***TMqWaveform*** and for update of the parameters ***TMqParameters***. ***TMqWaveform*** component organizes the measured signals into the individual channels ***TMqChannel*** and each channel is then formed by samples ***TMqSample***.

The direct usage of these classes is too complex so that is why another component was created **-** ***TSimpleDASHReader***. This component is specifically designed to collect the waveforms from the **DASH** monitor at the constant sampling rate.

Class allows it to:

1. Connect to the target monitor based on its IP address.
2. Get a list of all currently measured channels from the monitor.
3. Continuously measure the selected channels. Measured waveforms are returned as a table of values (channel, time). Possible error conditions and failures in the measurement are replaced by the constant value NaN.

5 Implant Control System ICS 3000

Biotronik **ICS3000** serves as a mobile implant control system that helps surgeons to fully control an implanted pacemaker. Also it is used for the measurement of electrical parameters and as the external pacemaker. As for the measured electrical variables, there are, for example IECG, voltage threshold, impedance of the electrodes and others.

In order to integrate the **ICS3000** device and the patient data into the **ImplantSys** application, an interface component needs to be implemented. Once this component is used the software is able to extract, transform and translate the internationally standardized *Biotronik IEEE 11073-10103 XML* data and programmer data into the **ImplantSys** software or hospital's **EHR** system.

The interface component is programmed in .NET 3.5, .NET 4.0 in a mixed assembly manner and compiled as a dynamic library with a public API (*Application Programming Interface*) that can be used by any .NET language.

Data export process and integration into the **ImplantSys** application is done with respect to the specification provided by *Biotronik* [7]. In order to read **ICS3000** exported data, transformation of this data into the internationally standardized format needs to be performed. This transformation from *Biotronik* programmer XML file format to internationally standardized *Biotronik IEEE 11073-10103 XML* is done via a command line application called Programmer Adapter which is provided freely from the *Biotronik* website [8].

5.1 Customized Adapter Implementation

The customized adapter is implemented in C# .NET 3.5 and .NET 4.0 mixed assemblies and compiled as a dynamic library included into **ImplantSys** application.

The customized adapter's XML parser expects either *IEEE 11073-10103 XML* format or ICS programmer XML format. In the case of ICS programmer XML format, XML is transformed into IEEE format by a *Biotronik* Programmer Adapter command line utility [8]. Encoded backup zip file reading is not implemented yet as *chilkat* component is not freely available and the password protection of these backups is therefore unknown.

Customized adapter's XML parser is using a language-integrated query to parse and process XML stored data. All data stored in the XML format is described in the document provided by *Biotronik* [9]. Parsed XML data is mapped into object structures. Customized adapter is working as a *Biotronik* XML to object mapper.

6 Central Registry REPACE

REPACE is the name of the nationwide central registry developed by the *Institute of Biostatistics and Analyses at the Faculty of Medicine and the Faculty of Science of the Masaryk University, Brno, Czech Republic*. This registry contains reports from the implantations of pacemakers and other relevant operations such as re-implantations.

Reports contain information about the operation such as patient's data, implanted devices, electrodes, complications, measured data and the names of the operational personnel. All this data is inserted via the web page interface.

Most of this data **ImplantSys** can obtain automatically – from the **CLINICOM**, **DASH** and **ICS3000**. An option to export the data to the **REPACE** is another one of the developed main features. There is an ongoing implementation on the interconnection. A .NET interface will serve this purpose. **REPACE** is built on the **ORACLE** database.

7 Conclusion

ImplantSys (please see Fig. 2) is very helpful as it connects four sources of information and combines it into one framework where the user – the physician – can effectively work with it in the modern GUI.

Every part of the system can be used alone for developing other applications. For example there can be a simple application used in ambulances car to store important vital signs during patient transfer where the **DASH** is used. This data can later be used for analysis. Another software application can evaluate the length of the interval between two pulses from the pacemaker. Longer intervals mean that the battery is discharging. This software can be used to determine the status of the battery. This information is very important not only in the case that the patient is pacemaker dependent.

Other applications can be developed with the **ICS300** dynamic library or the **REPACE** interface anywhere it is needed. From this point of view this is the most valuable contribution of this whole project and its main system the **ImplantSys**.

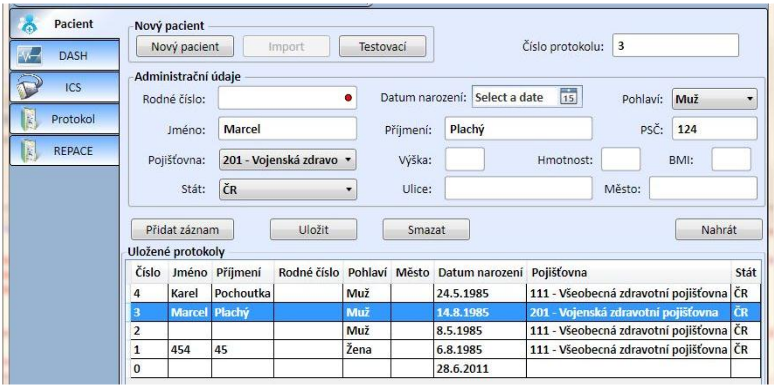


Fig. 2. ImplantSys screen with patient's data

References

1. CompuGroup. *CareCenter G3 NIS CLINICOM* [online]. c2010, [cit. 2011-30-03]. <http://www.compugroup.cz/underwood/download/files/cc_g3_cz_cg_newdesign_v01.pdf>
2. SMS spol s.r.o.. *Reference SMS system implementation* [online]. c2005, [cit. 2011-30-03]. <http://test.nlogy.cz/smsbrno/public/reference.html>
3. MICHNA, Viktor – VLACH, Karel. System for monitoring pacemaker implantation. In *Proceedings of 8th International PhD Student's Workshop on Control and Information Technology – IWCIT 2009*. Brno: Brno University of Technology, 2009. pp. 97-102. (ISBN 978-80-214-3949-8).
4. Wikipedia. *Electrocardiography* [online]. c2007, last revision 4th of April 2011 [cit. 2011-30-03]. <https://secure.wikimedia.org/wikipedia/en/wiki/Electrocardiography>
5. CERNOHORSKY, Jindrich – KOZIOREK, Jiri – HRUDKA, Gustav. The examples of implementation of two specialized case tools as extensions of a general framework. In *IECON 2006 – 32nd Annual Conference on IEEE Industrial Electronics 2006*. New York: IEEE, 2006. Vols 1-11., pp. 4909-4914. (ISSN: 1553-572X, ISBN: 978-1-4244-0135-2).
6. MICHNA, Viktor – BABUSIAK, Branko. Zobrazování a analýza EKG z pacientského monitoru DASH. *Mezinárodní konference Technical Computing Prague 2009*. HUMUSOFT s.r.o., 2009. pp. 72. (ISBN 978-80-7080-733-0). <http://www.humusoft.cz/akce/matlab09/sbornik/>
7. Biotronik SE & Co. KG. *IT-Specialist Guide for integrating BIOTRONIK Home Monitoring® and programmer data into hospital Electronic Health Records system* [online]. c2009, [cit. 2011-30-03]. <http://91.204.11.72/sixcms/media.php/211/EHR%20Tech%20Guide.pdf>
8. Biotronik SE & Co. KG. *Technical documentation for EHR Integration of BIOTRONIK programmer data* [online]. c2009, [cit. 2011-30-03]. <http://www.biotronik.com/biohm/technical-documentation-for-ehr-integration-of-biotronik-programmer-data/28769>
9. Biotronik SE & Co. KG. *Documentation of the BIOTRONIK IEEE 11073-10103 XML Structure* [online]. c2009, last revision 7th of February 2010 [cit. 2011-30-03]. <http://www.biotronik.com/sixcms/media.php/211/BIOTRONIK_IEEE_11073-10103_XML_Structure.pdf>

Using Mathematical Modelling of Temperature Conditions inside Geothermal Resources

Petr Vojčinák

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.vojcinak@vsb.cz

Abstract. This paper deals with some usage of mathematical apparatus for modelling of temperature conditions inside geothermal resources – in this case, we are interested in low-temperature ones in the forms of research polygons and rocks, whereas it is required to consider lots of factors, describing this problem correctly by using relevant mathematical formulas and equations, based on famous Fourier partial differential equation and other empirically developed equations. Using some sophisticated software applications, it is possible to evaluate the acquired data, whereas there is mentioned software application of EED only. Its borehole design process, including practical example, focused on approximate calculations of borehole length, depth, and number.

Keywords: Boreholes, Collector, Conductivity, EED software application, Fourier partial differential equation, Heat pumps, Mathematical modelling, Resistance, Thermodynamic temperature, TRT test.

1 Introduction

Modelling of temperature conditions in the surrounding of boreholes need not be prerequisite for operation of heat pumps (HP). It is required to review their necessity by somebody, who is erudite (i. e. experience and knowledge) to analyze some geological conditions in the place of projected borehole, thus:

- kinds of rocks,
- underground water affluents,
- other specifics.

From economic viewpoint, it is required to study locality conditions (a borehole location for a heat pump) very carefully, because it is possible that the planned borehole could be overdesigned, so it is needlessly expensive. On the other hand, this borehole could be underdesigned – in this case, effectiveness of this system is lower. From pieces of knowledge, observed via modelling, it is possible to quesstimate these (optimal) parameters:

- length of boreholes,
- number of boreholes,
- location of boreholes.

In practice, some user-developed application for modelling of temperature field in the neighbourhood of boreholes can be used. The next stage of some borehole design (e.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 291–297.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

g. COP parameter calculation, borehole length, or number of boreholes) has to be solved via either some other software tool (or application) or traditional “paper” calculation. In this case, it is very advantageous to use some sophisticated software application (e. g. EED 2.0 or higher). In case of a polygon of boreholes or some special conditions, it is required and useful to know, how this temperature field is conceived and how it will act at steady state or long-time operation of one heat pump or more ones. Some user-friendly software application , which could realize both data modelling and data generation used for a borehole design, has not been developed yet or its development is in progress now. [1]
Reservoir engineers and structural dynamicists, for example, routinely use advanced finite difference and finite methods. But drillers have traditionally relied upon simpler handbook formulas and tables, that are convenient at the rigsite. Simulation methods are powerful, to be sure, but they also have their limitations. [3]

1.1 Borehole Dimensioning Approach

There are these subprocesses during a borehole dimensioning for systems with heat pumps:

- determining borehole number,
- determining depth of each borehole,
- determining distance between two boreholes and among more ones.

There are several fundamental formulas used for this mathematical modelling. Four-variable function of rock temperature in the borehole neighbourhood *T(r, φ, z, t)* conforms to Fourier partial diffential equation of heat transfer in cylindrical coordinates, see (1) and (2):

$$\frac{1}{a}\cdot\frac{\partial T}{\partial t}=\frac{\partial^2 T}{\partial r^2}+\frac{1}{r^2}\cdot\frac{\partial^2 T}{\partial \varphi^2}+\frac{1}{r}\cdot\frac{\partial T}{\partial r}+\frac{\partial^2 T}{\partial z^2} \qquad \textbf{(1)}$$

And

$$T=T(r,\varphi,z,t)=f(r,\varphi,z,t) \qquad \textbf{(2)}$$

And

$$a=\frac{\lambda}{c\cdot\rho} \qquad \textbf{(3)}$$

Where
r radial distance from borehole symmetric axis (m),
φ angle of radius vector (rad),
z borehole vertical coordinate in z-axis (m),
t continuous time (s),
T rock thermodynamic temperature (K),
a thermal diffusivity ($m^2 \cdot s^{-1}$),
λ thermal conductivity of rocks ($W \cdot m^{-1} \cdot K^{-1}$).
c thermal capacity of rocks ($J \cdot kg^{-1} \cdot K^{-1}$) at constant pressure (Pa).

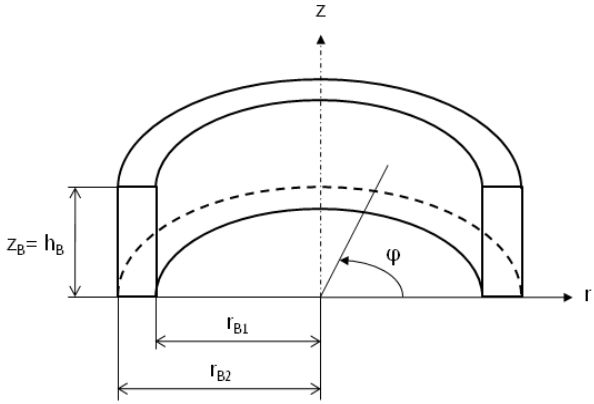


Fig. 1. Geometrical configuration of borehole in cylindrical coordinates – $r_{B1} = r_B$ is an inner borehole radius, r_{B2} is an outer borehole radius, and $z_B = h_B$ is a borehole depth.

If T is linearly dependent on φ (in this case, the first-order partial derivative of T on φ is constant), then:

$$\frac{1}{a} \cdot \frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \cdot \frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial z^2} \tag{4}$$

And

$$T = T(r, z, t) = f(r, z, t) \tag{5}$$

A numerical simulation, focused on this problem, showed that some simplifications can be used to determinate a borehole depth very accurately, namely at real geological conditions. During mathematical modelling, only one average value of rock thermal conductivity can be used, if a rock massif is homogeneous from thermal conductivity value viewpoint.

Due ot seasons, air/surface thermal resistance, snow cover and frost have not effect on borehole with connected heat pump. In this case, annual average air temperature of specific locality is very important parameter. Simplified and boundary conditions – based on a form of (5) and Figure 1 – are as follows [2]:

Simplified initial condition of rock temperature:

$$T(r, z, t = 0) = T_g \tag{6}$$

Simplified boundary condition of surface:

$$T(r, z = 0, t) = T_o(t) \tag{7}$$

Basic boundary conditions inside borehole:

$$T(r = r_B, z, t) = T_B(t) \tag{8}$$

And

$$q(t) = \frac{2 \cdot \pi \cdot \lambda}{h_B} \cdot \int_0^{h_B} r_B \cdot \frac{\partial T}{\partial r}\bigg|_{r=r_B} \cdot dz \tag{9}$$

Where

r_B borehole radius (m),

h_B borehole depth (m),

$T_B(t)$ thermodynamic sidewall temperature (K),

T_g average rock temperature in the neighbourhood of uninfluenced borehole (K),

$q(t)$ average heat consumption ($W \cdot m^{-1}$).

During the borehole design process, some influence of underground water affluent is not impeached. During EED borehole design process, a rock thermal response (borehole neighbourhood and time interval) is converted into a file of thermal response non-dimensional coefficients or G-functions. Then, total heat consumption is converted into a file of finite thermal pulses, whereas temperature of borehole sidewall is set by this G-function (for $T_B(t)$ calculation), thus:

$$T_g + \sum_{i=1}^{N} \frac{q_i - q_{i-1}}{2 \cdot \pi \cdot \lambda} \cdot g\left(\frac{t_i - t_{i-1}}{t_s}, \frac{r_B}{h_B} \right) \tag{10}$$

And

$$t_s = \frac{1}{a} \cdot \left(\frac{h_B}{3} \right)^2 = \frac{h_B^2}{9 \cdot a} \tag{11}$$

Where

t_s stationary time of heat consumption (s),

$g(.)$G-function of rock thermal response (-),

q_i borehole heat consumption for a period of ith thermal pulse ($W \cdot m^{-1}$),

t_i time period of ith thermal pulse (s).

Size of G-function depends upon number of boreholes and their placement.

Parameter of total thermal resistance R, i. e. borehole resistance versus rock anti-freeze mixture heat transfer, is very important to borehole design and it is also dependent upon borehole filling, position of collector pipes, and mode of collector anti-freeze mixture flow. It is desired to minimize the R value, namely from viewpoint of effective rock-collector heat transfer. It was simulationally verified, that this condition is qualified, when:

- thermal conductivity value of the borehole filling material is very high,
- equidistance between collector pipes is available along the borehole depth,
- minimal distance between collector pipes and sidewall is available,
- mode of anti-freeze mixture is turbulent.

Equation (10) mathematically denotes some calculating thermal resistance value between sidewall and anti-freeze mixture inside collector, thus:

$$R = R_B + R_g = \frac{\Delta T}{q} \Rightarrow \frac{T_B(t) - T_F(t)}{q(t)} = \frac{T_B(t)}{q(t)} - \frac{T_{IN} - T_{OUT}}{2 \cdot q(t)} \tag{12}$$

Where

$T_F(t)$ average input temperature value of anti-freeze mixture (K),
$T_B(t)$ average output temperature value of anti-freeze mixture (K),
T_{IN} input temperature value of anti-freeze mixture (K),
T_{OUT} output temperature value of anti-freeze mixture (K),
R_B internal thermal resistance of borehole ($K \cdot m \cdot W^{-1}$),
R_g thermal resistance between sidewall and a border, where borehole thermal influence has no effect ($K \cdot m \cdot W^{-1}$).

Above-mentioned formulas and most quantities are used for EED borehole design process.

1.2 EED Evaluation Software Features

EED was developed by Swedish and German researchers, who are the specialists for the fields of building object physics and mathematical physics. It is only useful for ground/water system borehole design, whereas finite depth equals to 200 meters. EED enables to calculate these parameters:

- average temperature values of anti-freeze mixture inside collectors (at specific borehole depth),
- borehole depth (at specific temperature range of anti-freeze mixture inside collectors).

In practical use, the second approach is preferred, because the first one often leads to borehole depth undersizing and undesirable rock chilling inside borehole. Specific input conditions naturally determine the corresponding (maybe optimal) borehole depth to ensure:

- minimal rock temperature increasing in the neighbourhood of boreholes,
- constant value of HP heating power.

User of this application has to specify these input parameters:

- value of drilled rock thermal conductivity λ – one specific value for each borehole,
- value of earth heat flux ψ for specific locality,
- value of annual average surface temperature,
- value of litre volume of circulating pumps inside HP primary circuit,
- type of borehole concrete mixture,
- type of anti-freeze mixture inside HP collectors,
- monthly energy balance of specific building,
- monthly estimation of HP operation at its maximum value of heating power.

2 Example of Approximate Calculation of Optimal Borehole Length via EED 2.0

In Ostrava – Poruba, a TRT (Thermal Response Test) test was practiced to determine some physical properties of rock massif inside two boreholes, equipped with PE collector and grout. The parameters in the form of input data used for EED calculation process are as follows:

- *thermal conductivity of rock massif* $\lambda = 2.15\ \mathrm{W \cdot m^{-1} \cdot K^{-1}}$,
- *thermal resistance ("classical" grout)* $R_B = 0.16\ \mathrm{K \cdot m \cdot W^{-1}}$,
- *thermal resistance (grout with sand)* $R_B = 0.12\ \mathrm{K \cdot m \cdot W^{-1}}$.

On the basis of energy balance, some other parameters were also used for the EED calculation of borehole length:

- *planned heat power of heat pumps* 650 kW,
- *annual average air temperature* 8.45 °C.

Table 1 summarizes the results of these tests.

Table 1. Summary of the TRT test results – values of borehole depth, collector length, and total number of boreholes.

Parameter	Classical type	Type with sand ingredient
Depth of boreholes	130 m	130 m
Approximate collector length	cca 15 km	cca 14 km
Total number of boreholes	116 pieces	108 pieces

Figure 2 shows some curves of collector anti-freeze mixture temperature (in Celsius degree) versus long-time operation mode of heat pump (in years).

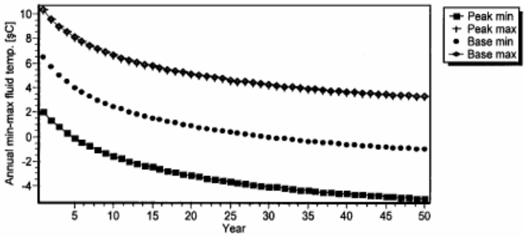


Fig. 2. Three curves of collector anti-freeze mixture temperature (y-axis in Celsius degrees) versus long-time operation mode of heat pump (x-axis in years).

A curve of "Base min" is fundamental for some dimensioning of borehole depth. Figure 2 also shows, that long-time decrease of minimal temperature of heat carrier fluid (inside collector and rock) settles on temperature level of 0 °C due to HP operation. This statement does not take account of "rock massif charging" by air-conditioning heat in summer.. Though, borehole depth of 130 meters should ensure reliable building heating for a long time. A curve of “Peak min” shows some short-time fluctuations due to bitter cold and requirement of heat energy supply.

3 Conclusion

The first part of this paper is focused on some theoretical description of heat, internal and external geothermal resources, physical processes inside rocks, and distinguished thermal field parameters at specific conditions and mathematical formulas, based on Fourier partial differential equation in cylidrical coordinates.
The second part is very important, because it focuses on mathematical modelling of temperature conditions in the surrounding of boreholes, when relevant formulas are mentioned and discussed in connection with practical use. There is also some borehole dimensioning approach, based on P. Eskilson’s work, knowledge, and parameters useful for modelling via some sophisticated evaluation software application of EED. Features of this are mentioned. Only EED borehole design process, including practical example of approximate calculation of borehole length, and its results are described and discussed. A TRT test approach is not described.

Aknowledgement

This work is supported by the project SP2011/45, named “Data Acquisition and Processing from Large Distributed Systems.”, of Student Grant System (VŠB – Technical University of Ostrava) and by the project TA01020932, named “Using Geothermal Energy for Renewable Energy Sources Systems Including Verification of Energy Accumulation.”, of Technology Agency of the Czech Republic.
This research was supervised by doc. Ing. Jiří Koziorek, Ph.D.

References

1. P. Lukes, “Use of Computer Modelling in Geothermal Energy Utilization”, The Czech Geological Survey, Prague (CZ), 2008, pp. 225 – 228, ISSN 0514-8057.
2. P. Bujok, M. Vrtek, B. Horak, R. Hajovsky, G. Hellstrom, *Study of Thermal Response of Rock Massif for Installations of Heat Pumps*, study for the Czech Energy Agency, Prague (CZ), 2005, 152 pages.
3. W. C. Chin, *Borehole Flow Modeling in Horizontal, Deviated, and Vertical Wells*, Gulf Publishing Company, Houston, TX (USA), 1992, ISBN 0-88415-034-8.

Method of Robot Motion Control using Cubic Hermite Splines

Petr Wagner and Jan Kordas

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{petr.wagner, jan.kordas}@vsb.cz

Abstract. This paper presents the results of motion control method for wheeled robots that utilizes cubic Hermite splines for path planning. The method is designed to allow computation and recomputation of robot trajectory in real-time using the current computer technology. In the first part, the main components of the system are described including the Kalman filter facilitating the filtering and prediction of the robot position, the trajectory planner based on Hermite splines and the feedback and feedforward controller controlling the movement along the calculated trajectory. In the experimental part of the paper, the calculated trajectories and its velocity profiles are compared with the measured ones on the test bed. The experiment was carried on with sampling rate 100 cycles per second.

1 Introduction

The problem of robot control consist of several parts that alone have many various practical applications: the capturing of images with the digital camera, the adaptation of the camera setting to current lighting conditions, the digital image processing, the strategic control (it can utilize the elements of artificial intelligence, expert systems, neural networks or/and multiagent systems), the reliable wireless transmission systems, and/or the design and the realization of embedded control systems e.g. [7].

This paper tries to cover the topic of robot trajectory planning and presents some experimental results achieved using the described method. It does not solve the problem of optimal trajectory planning. This topic is covered e.g. in [2], [4], [6]. The goal of our work and this paper was the realization of real-time trajectory planning with respect to acceleration and physical constraints of robots. The term real-time signifies the capability of the system to compute at least five trajectories during one cycle of the control system which was for fast reaction set to 10ms.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 298–303.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Motion Control

The Motion control performs the robot movement control on the low level. The input information of this module consists of robot positions, angles and robot velocities that are acquired from the vision system. The required positions within the environment and the target velocities that robots are supposed to have in the target positions are the other input of the module. The higher strategies provide this information. The aim of Motion control module is to calculate a trajectory from current position of robot to the target position using the input data and to control the move of robot on the calculated trajectory.

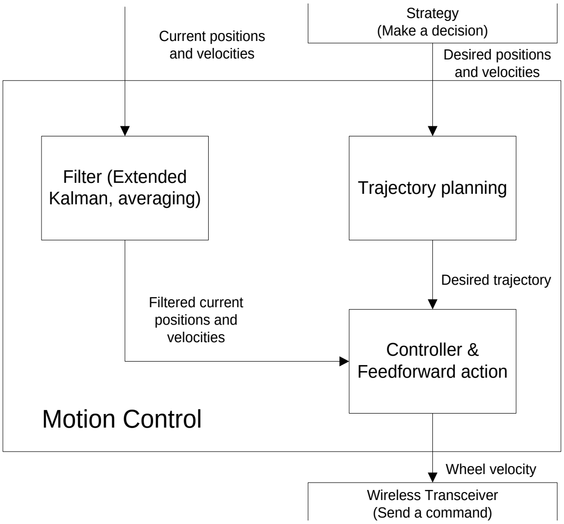


Fig. 1. The structure of Motion Control module

The module Motion control consists of three parts that perform the control. They are: the filtering module, the trajectory planning module and the controller module. The structure of the Motion control module is shown in Fig. 1. More detailed description can by found in [1], [3].

2.1 Filtering Module

The aim of the filtering module is to filter the incoming data from the vision system. In regard of quite short sampling period, the data obtained from the vision system are badly influenced by noise.

The extended Kalman filter is used to filter this data e.g. [5]. The filter uses the kinematical model of mobile robots. Because the model of robot is nonlinear and nonholonomic it is necessary to linearize the model in a neighborhood of the working point and the new matrix A (describes the dynamic of the system) has to be calculated in every operating point.

2.2 Trajectory Planning

The trajectory planning module calculates and generates the desired trajectory of a robot from the current position of the robot to its target position. The current and target positions, angle of robot and velocities in the current and target position are the parameters for this module. The module computes the trajectory and its velocity profile, which consist of the maximum velocities that respect the motion constraints for every point of the trajectory.

2.3 Controller and Feedforward System

The controller and the feedforward system have to keep the robot on the calculated trajectory in every moment of motion. They have to keep not only the trajectory but also the time schedule. The control consists of two parts. The feedforward system controls the motion on the calculated trajectory without any feedback information. This type of control is sufficient if there are no disturbance signals that influence the behavior of robot. This condition cannot be fulfilled in the real world system hence the feedback controller is used. The feedback controller compensates the disturbance signals that affect the mobile robot. See Fig. 2.

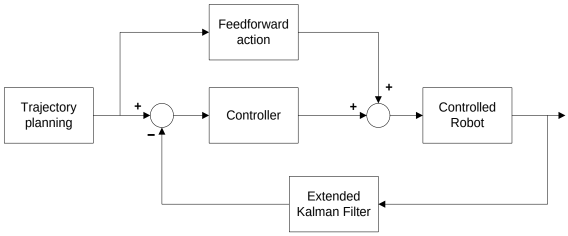


Fig. 2. Feedforward and feedback controllers

The movement of robots is planned with using of the third degree polynomials. To be more specific, the Hermite splines are used. These splines are constructed by using of 4 control points. Two of them are the origin and target (in the parametric form with the parameter u and the other two specify directions of the spline in the origin and target points. The curve constructed of the Hermite splines can be used for trajectories with at most two arches. If the more complex curves are required, they can be put together from several Hermite splines.

3 Experimental Results

The experiment with the real robot has been carried out. The initial trajectory was being recomputed in every control cycle (100 times per second). The resulting trajectory of the robot that was captured by a camera is compared with the desired trajectory (see Fig. 3). The correspondence between the trajectories is evident. In the origin point, the robot does not have the simulated initial orientation. Therefore, one can easily notice a difference between the measured and computed trajectory at the beginning. But the feedback controller was able to reduce this error and keep the robot on the trajectory.

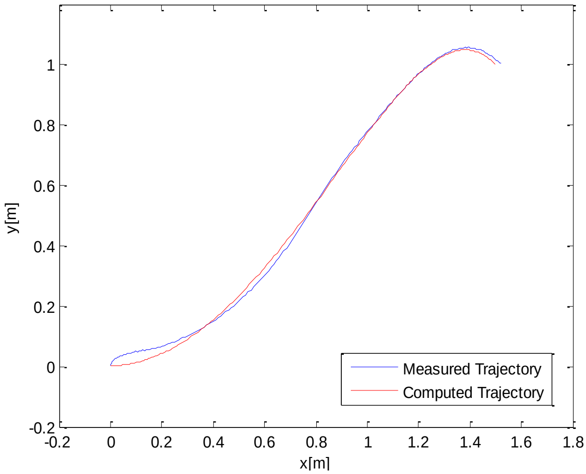


Fig. 3. Comparison of trajectories

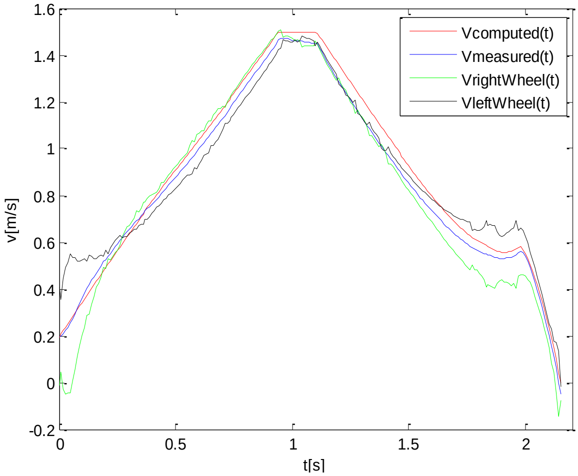


Fig. 4. Comparison of velocity

In Fig. 4, the comparison of the computed velocity profile and the measured one is depicted. In the experiment, the robot performed the movement from a resting state. And again, there is the correspondence between the two charts. The differences are caused by corrections of the feedback part of the controller.

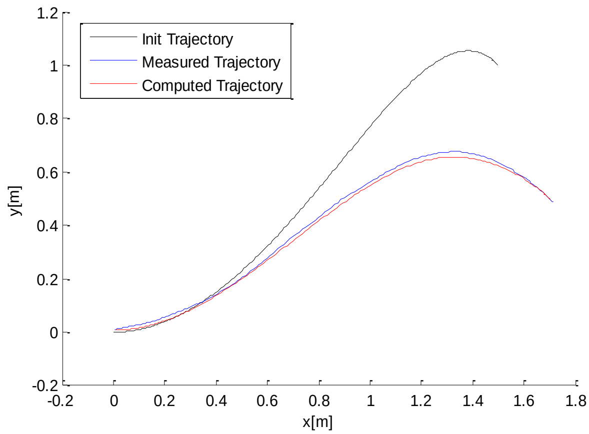


Fig. 5. Comparison of trajectories (adjustment of the target point)

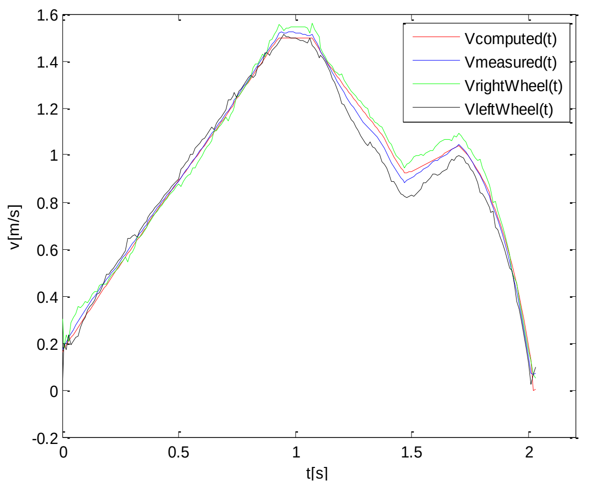


Fig. 6. Comparison of velocity profiles (adjustment of the target)

The similar experiment was executed with the adjustment of the target point. See Fig. 5 and 6. The target point was being changed during the course of the movement. The change took place during 150 consecutive system cycles and the target point was changed from initial position (x;y)=(1.5;1) to final position (x;y)=(1.7;0.5). In Fig. 5, the initial and the resulting trajectory and target points are depicted along with the measured trajectory of the robot. It is shown that the recomputing of the trajectory by the presented method is working. The resulting movement is smooth and the controller is able to keep the robot on the changing trajectory.

4 Conclusions

This paper presented the results of trajectory computation using the proposed method that facilitates the cubic Hermite splines in its core with a view to the computation of the trajectory in the real-time. The implemented discrete method of computation is capable to provide the required trajectory with a sufficient accuracy. If the trajectory is computed in 100-200 points in dependence on the trajectory length it is possible to achieve the accuracy to perform the real movement. The computing time is in this case shorter than 1ms and that allows using this control method in the real-time and it exceeds the required time.

The presented experimental results show that the proposed solution of the trajectory planning, profile computation and trajectory adjustment can be used on a real robot. The trajectory adjustment solution allows applying the trajectory planning algorithm in the environment where the target point changes in the course of movement.

References

1. Wagner, P.; Kotzian, J.; Kordas, J.; Michna, V.: Path Planning and Tracking for Robots Based on Cubic Hermite Splines in Real-time. In Proceedings of 2010 IEEE Conference on Emerging Technologies and Factory Automation, Bilbao (2010) 1-8
2. Gajdusek, M., et al.: Time Optimal Trajectory Generation Applied to Control of Differentially Driven Robot. AT&P journal, Vol. 2, (2006) 86-89
3. Michna, V.; Wagner, P.; Cernohorsky, J Constrained optimization of robot trajectory and obstacle avoidance. In Proceedings of 2010 IEEE Conference on Emerging Technologies and Factory Automation, Bilbao (2010) 1-4
4. Grepl, R., Byoungsoo, L., et al.: Exact Linearization of Nonholonomic System Dynamics Applied to Control of Differentially Driven Soccer Robot. Engineering Mechanics, Vol. 15, (2008) 311-318
5. Jetto, L., et al.: Development and Experimental Validation of an Adaptive Extended Kalman Filter for the Localization of Mobile Robots. IEEE Transactions on Robotics and Automation, Vol. 15, (1999) 219-229
6. Lepetic, M., Klancar, G., et al.: Path Planning and Path Tracking for Nonholonomic Robots. Mobile Robots: New Research, (2005) 345-368
7. Tucnik, P., et al.: Multicriterial Decision-Making in Multi-agent Systems. Lecture Notes in Computer Science, Springer-Verlag, Berlin (2006) 711-718

Using Concept Lattices in Publications' Classification Analysis

Saleh Alwahaishi

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
salehw@vsb.cz

Abstract. The definitive classification of scientific journals depends on their aims and scopes details. In this paper, we present an approach to facilitate the journals classification of the DBLP datasets. For the analysis, the DBLP data sets were pre-processed by assigning each journal attributes defined by its topics and then the theory of formal concept analysis is introduced. It is subsequently shown how this theory can be applied to analyze the relations between journals and the extracted topics from their aims and scopes. The result is a concept lattice that contains information on journal-topic relational context depending on how they are associated.

1 Introduction

Formal Concept Analysis (FCA) was invented in the early 1980s by Rudolf Wille as a mathematical theory [1]. FCA is concerned with the formalization of concepts and conceptual thinking and has been applied in many disciplines such as software engineering, knowledge discovery and information retrieved during the last two decades. The mathematical foundation of FCA is described in [2].

FCA considers the extensional and the intensional part equally important, whereas the other formalisms are biased towards the intensional aspect of concepts. FCA is thus complementing other conceptual knowledge representations; and the combination of FCA with other representations has been the topic of many publications. For instance, several approaches combined FCA with description logics [3, 4, 5, 6] and with conceptual graphs [7, 8, 9].

In this paper, we describe how we used FCA to create a visual overview of the DBLP scientific journals classification based on their aims and scopes. As a case study, we zoom in on the top journals based on their impact factors. The remainder of this paper is composed as follows: In section 2 we introduce an overview of the Digital bibliography and Library project (DBLP). Section 3 visualizes, with an example, the literature using FCA lattice. In section 4 we explained the classification criterion of journals and applied the concept lattice on the selected journals. Section 5 concludes the paper.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 304–309.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Digital Bibliography & Library Project (DBLP)

Digital libraries are collections of resources and services stored in digital formats and accessed by computers. Studying them offers an interesting case study for researches for the following reasons: Firstly, they grow quickly, Secondly, they represent a multidisciplinary domain which has attracted researchers from a wide area of expertise. DBLP (Digital Bibliography & Library Project) is a computer science bibliography database hosted at University of Trier, in Germany.

It was started at the end of 1993 and listed more than one million articles on computer science in January 2010. These articles were published in Journals such as VLDB, the IEEE and the ACM Transactions and Conference proceedings [10, 11]. Besides DBLP has been a credible resource for finding publications, its dataset has been widely investigated in a number of studies related to data mining and social networks to solve different tasks such as recommender systems, experts finding, name ambiguity, etc. Even though, DBLP dataset provides abundant information about author relationships, conferences, and scientific communities it has a major limitation that is its records provide only the paper title without the abstract and index terms.

2.1 Using DBLP Dataset in Recommender Systems

In addition to using the DBLP dataset for finding academic experts, it has been used extensively in academic recommender systems. A number of studies were conducted to recommend academic events and collaborators for researchers using different methods and techniques. For example, a recommender system for academic collaboration called DBconnect was presented in [12], the authors of this paper used DBLP data to generate bipartite (author-conference) and tripartite (authorconference-topics) graph models, and designed a random walk algorithm for these models to calculate the relevance score between authors. And in another study [13] a recommender system for events and scientific communities for researchers was proposed based on social network analysis.

3 Concept Analysis Of Journals Classification

This section describes how formal concept analysis is employed to analyze the DBLP's journals classification. A group of interested similar journals, which covered the scope of computer science, were selected. The list of selected journals was obtained from well-known DBLP database that contains information about the published articles and their authors as well. The selected list of links to journals has the size of 115 items. The next step was to identify main topics, which each of the journals covers. From the journal web sites we have found the aim and scope of each journal, and have manually extracted the main topics, such as Pattern Recognition, Image Processing, etc.

3.1 Journals Selection and Categorization

Each journal has been identified by an existing classifier by company due to the problem with using their own names or similar names of topics. The used classifier that contains about 1224 sub disciplines classified to disciplines and those classified to discipline field, e.g. sub discipline Pattern Recognition is in disciplines Artificial Intelligence and Image Processing and that is in Information and computing sciences [14]. We selected only sub disciplines in the field Technology and Information and computing sciences. Our manually extracted topic from journals in many cases correspond the classified disciplines, but in some cases it was necessary to assign the extracted topic to sub discipline, which was almost similar. Therefore, journals were classified into a list of topics based in their relation to the topic

3.2 Lattice Construction for Journal-Topic Matrix

A journal is represented as a list of topics. The topics are the disciplines that being covered by all journals, based on the extracted data from their aims and scopes. Each topic is assigned a weight of 0 or 1. A topic's weight for a journal expresses the coverage possibility of the topic by the related journal. A value of 1 denotes that the journal covers the column's topic and 0 denotes the lack of coverage.

Formally, these data can be represented as a matrix of journals by topics whose m rows and n columns correspond to m journals and n topics, respectively.

The formal concept analysis of the data starts with the creation of a formal context. The formal objects of the formal context are the journals J_i that were retrieved from DBLP database. The set of these journals is denoted by J. Using the information that was extracted from the aim and scope of the journals in J. The coverage possibility T_j that shows the topic coverage by the journals in J, constitute the formal attributes of the formal context. The set containing these attributes is denoted by T.

The cross table of the resulting formal context has a row for each journals in J, a column for each topic in T and a cross in the row of J_i and the column of T_j if the corresponding weight y_{ij} is 1. To minimize the cross table size, journals impact factors will be considered to decrease the number of tested journals. The journals with an impact factor of 3.0 and above will be enlisted in the matrix, dropping the number of selected journals to be 18. After the formal context is constructed, formal concept analysis is applied to produce the concept lattice.

The formal concepts of the contexts are shown as extent-intent pairs, where the extent is a subset of the journals and the intent is a subset of the topics.

Abbreviation	Main Topic	Subfields
AIIP	Artificial Intelligence and Image Processing	Adaptive Agents and Intelligent Robotics
		Neural, Evolutionary and Fuzzy Computation
		Simulation and Modeling
		Computer Vision
		Pattern Recognition and Data Mining
		Signal processing
		Image Processing
CTM	Computation Theory and Mathematics	Computer Graphics
		Other Computation Theory and Mathematics
		Numerical Computation
		Applied Discrete Mathematics
		Computational Logic and Formal Languages
		Analysis of Algorithms and Complexity
CS	Computer Software	Software Engineering
		Operating Systems
		Computer System Security
		Bioinformatics Software
ISLIS	Information Systems and Library and Information Studies	Database and Database Management
		Information Retrieval and Web Search
		Inter-organizational Information Systems and Web Services
		Information Systems Management
		Information Systems Development Methodologies
DF	Data Format	Data Encryption
		Data Structures
DC	Distributed Computing	Mobile Technologies
		Distributed Computing
CT	Communications Technologies	Computer Communications Networks (computer network)
		Wireless Communications
		Other Communications Technologies (telecommunications)
CA	Computer Architecture	
DIP	Data and Information Processing	
STVV	Software Testing and Verification & Validation	

Table 1. Journals' main topics and subfields

({F,G,J,Q,R}, {AIIP,CTM})
({F,Q,R}, {AIIP,CTM,CS})
({A,D,F,H,L,Q,R}, {CTM,CS})
({C,G,J,Q}, {AIIP,CTM,CS,ISLIS})
({G,N,R}, {CS,CT})
({G,O,R}, {AIIP,DF})
({B,F,Q,R}, {AIIP,CA})
({G}, {AIIP,CTM,ISLIS,DF,CT})
({H}, {CTM,CS,STVV})
({Q}, {AIIP,CTM,CS,ISLIS,CA,DIP})
({R}, {AIIP,CTM,CS,DF,CT,CA})

The intent of each formal concept contains precisely those topics covered by all journals in the extent. Conversely, the extent contains precisely those journals sharing all topics in the intent.

The line diagram of the concept lattice, showing the partially ordered set of concepts is shown in Fig 1, has the minimal set of edges necessary; all other edges can be derived by using reflexivity and transitivity. Journals and topics label the node that represents the formal concept they generate. All concept nodes above a node labeled by a journal have the journal in their extent. All concept nodes below a node labeled by a topic have the topic in their intent. The extent of the concept node labeled by the topic "STVV" for example is easily found by collecting the journal H labeling this concept node on a path going downward.

The concept generated by the topic "DF" is a sub concept of the concept generated by the topic "AIIP", for the extent of the former concept is contained in the extent of the latter concept. All journals classified by the topic "DF" were also classified by the topic "AIIP", suggesting that within the given formal context "DF" is a more specific topic than "AIIP".

Another multi constructed example is found in the extent of the concept node labeled by the topic "DIP", which is found by collecting the journal Q labeling this concept node on a path going downward. The intent of this concept is found by collecting the topics "CTM", "CA", and "ISLIS" labeling the three concepts found on paths going upward. The latter two topics, however, are sub concepts of the concept generated by the topic "AIIP". The resulting extent-intent pair of this concept is ({Q}, {AIIP,CTM,CS,ISLIS,CA,DIP}).

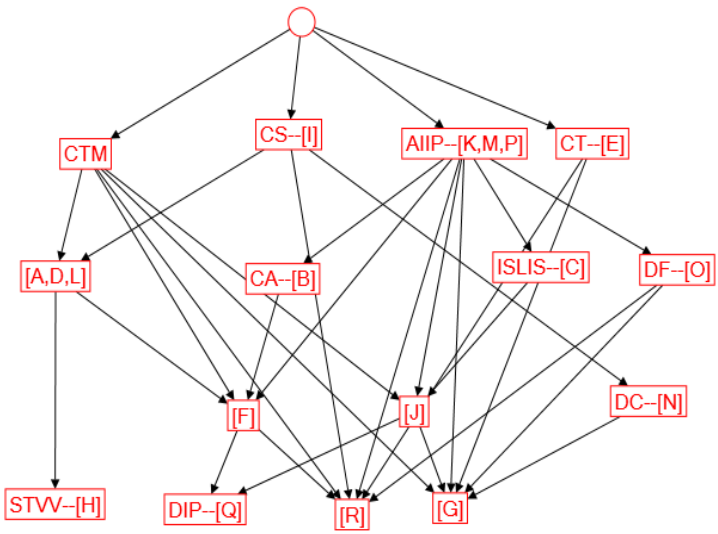


Fig.1. Concept lattice for journals classification

4 Conclusion

The concept lattice uncovers relational and contextual information. Journals' topic categorizations are put into relational context depending on how they are associated by the journals' aims and scopes. The topics "Computer Theory and Mathematics – CTM", and "Data and Information Processing –DIP" for example are shown as related because these topics share a similar classification context. The implicit structures revealed help researchers to classify journals more efficiently. This approach has the potential to support the emergence of new knowledge by identifying concept relations, making these explicit and enabling researchers to inspect these concept relations.

5 References

1. Wille. R. (1982). Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival (Ed.). Ordered sets. Reidel. Dordrecht-Boston. 445-470.
2. Ganter, B., Wille, R. (1999) Formal Concept Analysis: Mathematical foundations. Springer
3. Beydoun, G. (2009) Using Formal Concept Analysis towards Cooperative E-Learning. D. Richards and B.H. Kang (Eds.): PKAW, LNAI 5465, 109-117. Springer
4. Beydoun, G., Kultchitsky, R., Manasseh, G. (2007) Evolving semantic web with social navigation. Expert Systems with Applications 32, 265-276. Elsevier
5. Maddouri, M. (2005) A Formal Concept Analysis Approach to Discover Association Rules from Data. R. Belohlavek. V. Snasel (Eds.): CLA 10-21
6. Priss, U. (2006), Formal Concept Analysis in Information Science. Cronin. Blaise (ed.). Annual Review of Information Science and Technology, ASIST, Vol. 40.
7. Ganter, B., Kuznetsov, S.O. (2008) Scale Coarsening as Feature Selection. Medina and S. Obiedkov (Eds.) : ICFCA, LNAI 4933, 217-228. Springer.
8. Stumme, G., Bestride, Y., Taouil, R., Lakhal, L, (2002) computing Iceberg Concept lattices with TITANIC. Data and knowledge engineering 42 (2), 189-222.
9. Stumme, G., Wille, R., Wille, U. (1998) Conceptual knowledge discovery in databases using Formal Concept Analysis Methods. PKDD, 450-458.
10. Ley, M. (2002) The dblp computer science bibliography: Evolution, research issues, perspectives. SPIRE 2002: Proceedings of the 9th International Symposium on String Processing and Information Retrieval. London, UK: Springer-Verlag, pp. 1–10.
11. URL, http://en.wikipedia.org/wiki/DBLP.
12. Zaiane ,O. R. Chen, J. and Goebel, R. (2007) Dbconnect: mining research community on dblp data. WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. New York, NY, USA: ACM, pp. 74–81.
13. R. Klamma, P. M. Cuong, and Y. Cao (2009) You never walk alone: Recommending academic events based on social network analysis. Complex (1), pp. 657–670.
14. Obadi G., Drazdilova P., Hlavacek L., Martinovic J., and Snasel V. (2010) A Tolerance Rough Set Based Overlapping Clustering for the DBLP Data, Web Intelligence and Intelligent Agent Technology, International Conference on, pp. 57-60.

Use of Neural Networks in Recognition of Weather Impact on Radio Signal

Petr Blaha[1], Marek Dvorský[2], Libor Michálek[2], and Roman Šebesta[2]

[1]Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.blaha@vsb.cz
[2]Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{marek.dvorsky, libor.michalek, roman.sebesta}@vsb.cz

Abstract. This paper deals with using a neural network in field of recognition weather impacts on a telecommunication network. In this work was gathered data from a meteorological service from VŠB-Technical University of Ostrava, GSM signal. Analysis with strong focus on intercorrelations was done for the data. For this purpose was studied and used theory of multilayer perceptron network with back propagation algorithm for clustering.

Keywords. data mining, neural networks, GSM, weather

1 Introduction

Gsm (Global System for Mobile Communication) network is part of our outer environment, what people use every day while talking through mobile phones, browsing internet or receiving data from distant devices, like flood control sensors. GSM network use an air interface as a transfer. To be able to describe reliability of GSM service, there is a need to distinguish attributes of weather affecting it most. With such knowledge, it is possible to forecast when weather conditions change from standard conditions.
Neural networks are widely used for applications, such as weather condition forecasting [1].

2 Testing Data

Testing data consists of two subsets. First subset is weather related data. Second subset is GSM parameters related data.

VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Data related to weather were acquired by meteorological service of VŠB-Technical University of Ostrava. These data was measured during five weeks (May – June 2011) with the measurement period of five minutes. Weather data consist of temperature and humidity.

Data related to GSM was measured in idle mode in real GSM network. For measuring, it was developed the application which synchronous automatically saving current GSM data and saving current weather data. GSM data consists of signal strength from one base transceiver station. Totally were analyzed 4831 values.

3 Data Preparation

Before the data could be used for neural network as input, clean and synchronize procedures had to be done.

GSM logger is logging data from one service and up to 6 neighbor base transceiver stations. The order of stations changes due to changing of signal strength from respective station.

There was need to have for the same time the data sequence which can be be given as an input for neural network. In the next step, the median of GSM signal strength for 5 minutes period was calculated in order to have the most balanced input for neural network.

4 Computing and Results

At first, eight categories were analyzed, see Table 1.

Table 1. Division of data into categories based on signal, humidity and temperature

Category Name	Row Count
1	922
2	725
3	800
4	718
5	666
6	225
7	449
8	326

Neural network discovered rules, which can provide compact view about categories and associations between attributes.

– At first, the most influence has humidity category. There can be said that temperature was less than 11,7 degrees Celsius where humidity grew above 85 %.
– Second and third category is rather based on temperature, while about other attribute sets a mute is ongoing.

- Fourth category can be assigned to forenoon. Humidity is leading descriptor where the temperature is rising in the morning.
- Fifth, sixth and eight categories are also one-attribute related, other attributes play insignificant role.
- Seventh category is most significant and from its data structure a strong observation can be made. Signal is above -54.5 dBm, when humidity is very low, less than 51 %.

Detailed categorization is shown in Table 2. Graphical representation of significance in categories is shown in Figure 2.

Table 2. Details Of Attribute Significance In Categories Based On Signal, humidity and temperature

Category	Attribute	Rule	Favour
1	Humidity	>= 85	100
1	Temperature	< 11,73	25
1	Temperature	11,73 - 15,70	7
1	Signal Strength	-57,7 - -54,53	2
2	Temperature	15,71 - 19,02	100
2	Humidity	62 - 75	7
3	Temperature	19,02 - 23,02	100
3	Humidity	62 - 75	7
3	Humidity	51 - 62	5
3	Signal Strength	-63,99 - -61,5	1
4	Humidity	75 - 85	100
4	Temperature	11,74 - 15,71	10
4	Temperature	< 11,74	3
4	Signal Strength	-57,7 - -54,53	1
5	Temperature	>= 23,02	100
5	Humidity	< 51	18
5	Signal Strength	-63,99 - -61,5	4
5	Signal Strength	< -63,99	4
5	Humidity	51 - 62	3
6	Humidity	51 - 62	100
6	Signal Strength	< -63,99	11
6	Temperature	19,02 - 23,02	1
7	Humidity	< 51	100
7	Signal Strength	>= -54,53	88
7	Temperature	>= 23,02	12
7	Temperature	19,02 - 23,02	9
8	Humidity	62 - 75	100
8	Temperature	11,74 - 15,70	33
8	Signal Strength	>= -54,53	4
8	Signal Strength	-57,7 - -54,53	1

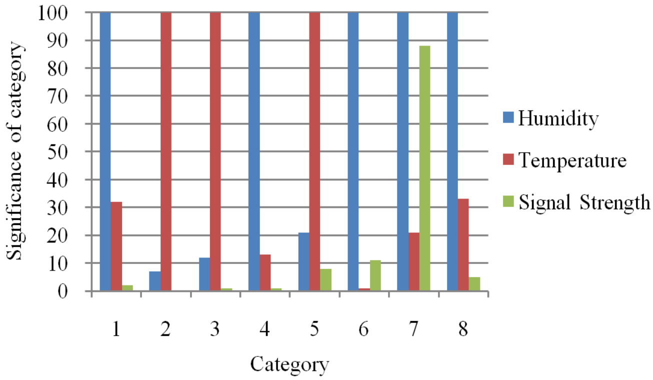


Fig. 2. Graphical representation of significance in categories.

5 Conclusion

Measurement and calculation neural networks methods have not found significant correlation between the monitored atmospheric parameters and GSM signal strength yet. In the next phase of the project we want to focus in detail on the monitoring of related parameters measured in shorter periods, which will include a sudden weather change.

References

1. Jef Hooyberghs, Clemens Mensink, Gerwin Dumont, Frans Fierens, Olivier Brasseur, A neural network forecast for daily average PM10 concentrations in Belgium, ISSN 1352-2310, DOI: 10.1016/j.atmosenv.2005.01.050.

Classification Methods for Brain-Computer Interface

Pavel Bobrov[1,2], Alexander A. Frolov[2], and Dušan Húsek[3]

[1]Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
p-bobrov@yandex.ru
[2]Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Moscow, Russia
aafrolov@mail.ru
[3]Institute of Computer Science,Academy of Sciences of the Czech Republic, Praha, Czech Republic
dusan@cs.cas.cz

Abstract. The performance of four classifiers for Brain Computer Interface (BCI) systems based on multichannel EEG recordings is tested in this work. The classifiers are designed to distinguish EEG patterns corresponding to performance of several mental tasks. It is shown that relatively simple classifiers based on the Bayesian approach are comparable in classification accuracy with more sophisticated classifiers based on Common Spatial Patterns and Common Tensor Discriminant Analysis.

1 Introduction

Brain-Computer Interface in general is a direct communication pathway between the human or animal brain and an external device. The EEG-based BCI considered in the paper operates in the following way. The set of mental task is chosen and each task is attributed to a particular command to an external device. Thus to operate the device via the interface one has to perform these mental tasks. In case the device movement is controlled, it seems natural for human to perform motor imagery (MI) tasks, i.e. imagine moving his extremities. Moreover, it has been shown that such mental states could be classified rather accurately by means of the EEG pattern analysis, as demonstrated by such successful BCI projects as Graz ([1], [2]) and Berlin [3] BCI. But applications of BCI extend beyond motion control, including controlling home appliances, selecting contacts in a phone address book or web search engine manipulation. Such tasks are more naturally accomplished by controlling the BCI with voluntary generation of corresponding visual images.
Since EEG pattern classifier is the core of the system we conducted experiments to test four classification methods on different data acquired during execution of motor and visual imagery tasks.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 314–319.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Data Sets

The classifiers were tested on three data sets. The first one is the data from BCI Competition IV (BCIC data set, [4]), the second data set contains the records from our own experiments involving four mental tasks: left and right hand movement imagining, foot movement imagining, and relaxation (MTR data set, [5]), the third data set contains the records from our experiments involving such mental tasks as relaxation and imagining one of two pictures (VIS data set, [6]). All data sets contain recordings of two sessions for each subject.

3 Classifiers

The accuracy and speed of four classifiers were tested. Let us denote the number of different mental tasks by L and covariance matrix of the signal corresponding to performing of the i^{th} task by $\mathbf{C}_i$.

3.1 Bayesian Approach

The Bayesian approach to classification can be directly applied to the EEG data. In the assumption that EEG signal distribution is Gaussian with zero mean and covariance matrix varying when different mental tasks are performed the probability of signal $\boldsymbol{x}$ (column vector) to correspond to the i^{th} mental task performing is

$$P(\boldsymbol{x}|i) \propto \exp(-1/2\ (\mathrm{trace}(\boldsymbol{C}_i^{-1}\boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}) + \ln(\det(\boldsymbol{C}_i)))) = \exp(-1/2\ \mathrm{V}_i)\ , \quad \textbf{(1)}$$

where $\boldsymbol{C}_i$ is covariance matrix of the signal corresponding to performing of the i^{th} task, $i = 1, .., L$. The signal $\boldsymbol{x}$ can be attributed to the task with number $k = \mathrm{argmax}_i(P(\boldsymbol{x}|i)) = \mathrm{argmin}_i(\mathrm{V}_i)$. In order to classify an EEG signal epoch we computed mean values of V_i and used them to determine the class number. Bayesian approach will be called BC throughout the paper. See [5] for more detailed description.

3.2 Multi-band Bayesian Approach

A natural way to take EEG frequency structure into consideration is to filter EEG within several non-overlapping pass bands and to use a different Bayesian classifier for each band.

Following this approach covariance matrices of the signals corresponding to different tasks are estimated for each band separately. The values of V_i are also computed separately for each band and then averaged for all I over all bands. The number of the smallest averaged value yields the class to which the signal should be attributed.

We will refer to this approach as MBBC. See [5] for more detailed description.

3.3 Multi-class Common Spatial Patterns

The method of Common Spatial Patterns is considered to be one of the most effective classification methods for BCI. Originally, it was proposed for distinguishing between two classes but later it was generalized onto multi-class problems [7].
Classification method based on Multi-class Common Spatial Patterns (MCSP) can be described as follows. First, covariance matrices $\mathbf{C}_i$ are estimated, then matrices $\mathbf{W}_i$ are sought to meet the following requirements

$$\mathbf{W}_i\mathbf{C}_i\mathbf{W}^{\mathrm{T}} = \mathbf{D}_i \qquad (2)$$

$$\mathbf{W}_i\mathbf{C}\mathbf{W}^{\mathrm{T}} = \mathbf{I}\,,$$

where $\mathbf{C} = \mathbf{C}_1 + \ldots + \mathbf{C}_L$, $\mathbf{I}$ is the identity matrix, and $\mathbf{D}_i$ are diagonal matrices.
The problem (**2**) can be solved explicitly using SVD technique.
After the matrices $\mathbf{W}_i$ are found, signal corresponding to a certain state is segmented into epochs and for each epoch $\mathbf{X}$ vectors $\boldsymbol{v}_i$, $i = 1, .., L$, are computed by estimating variances of all components of vectors $\mathbf{W}_i\mathbf{X}$, Then $\mathbf{X}$ is mapped onto a feature vector $\ln(\boldsymbol{v})$, where $\boldsymbol{v}$ is concatenation of all vectors $\boldsymbol{v}_i$ and $\ln(\cdot)$ means component-wise log transform.
Classification of the feature vectors can be accomplished by any multiclass classifier.

3.4 Common Tensor Discriminant Analysis

A multivariate MSCP generalization, namely Common Tensor Discriminant Analysis (CTDA), was proposed in [8] to take EEG time-frequency structure into account. The main idea of Common Tensor Discriminant Analysis (CTDA) for BCI is to extract features from EEG represented as a tensor by removing both inter-channel and inter-frequency correlations.
The detailed description of CTDA can be found in [8].

4 Results

EEG recordings corresponding to execution of mental tasks were then split into epochs of 1 second length. Then 70% of epochs corresponding to each state were chosen randomly for classifier training and the remaining 30% of epochs were used to test classifier. 100 such classification trials were made.
EEG pattern recognition accuracy was evaluated using three measures: average probability of correct recognition (p), mutual information between recognized states and presented instructions (g), and Cohen`s kappa index (k).
Values of the classification accuracy indices (p, g, k) resulting from BCIC, MTR, and VIS data set processing are represented in Fig. 1, Fig. 2, and Fig. 3 respectively. In each figure panels A, B, C demonstrate the results of the first session data classification and panels D, E, F demonstrate the results for the second session.
Averaged values of p, g, and k indices for all data sets are given in Table 1, Table 2, and Table 3 respectively.

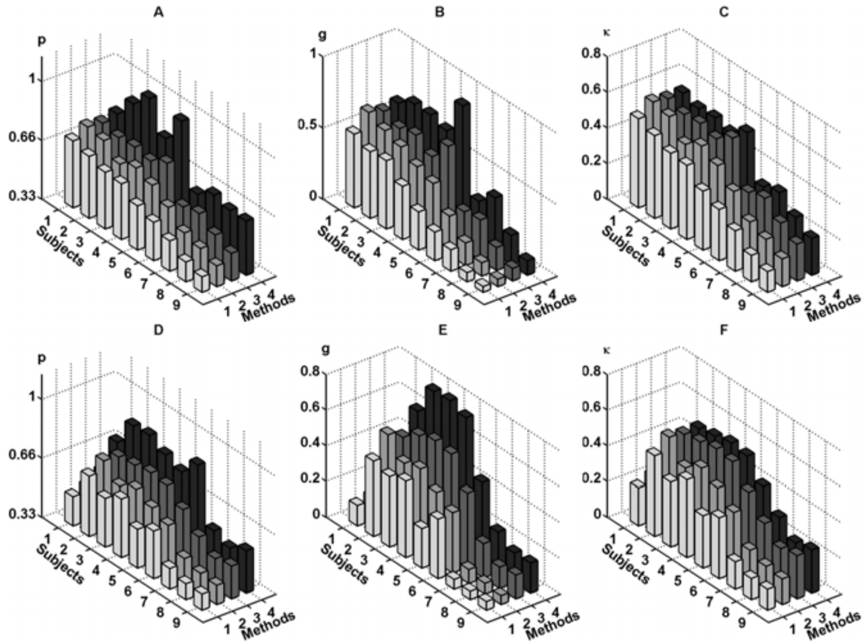


Fig. 1. Indices *p*, *g* and *k* obtained as a result of BCIC data classification for all subjects, using all methods: BC(1), MCSP(2), MBBC(3), CTDA(4). Note that graphs A and D represent exceeding of p over level of random classification.

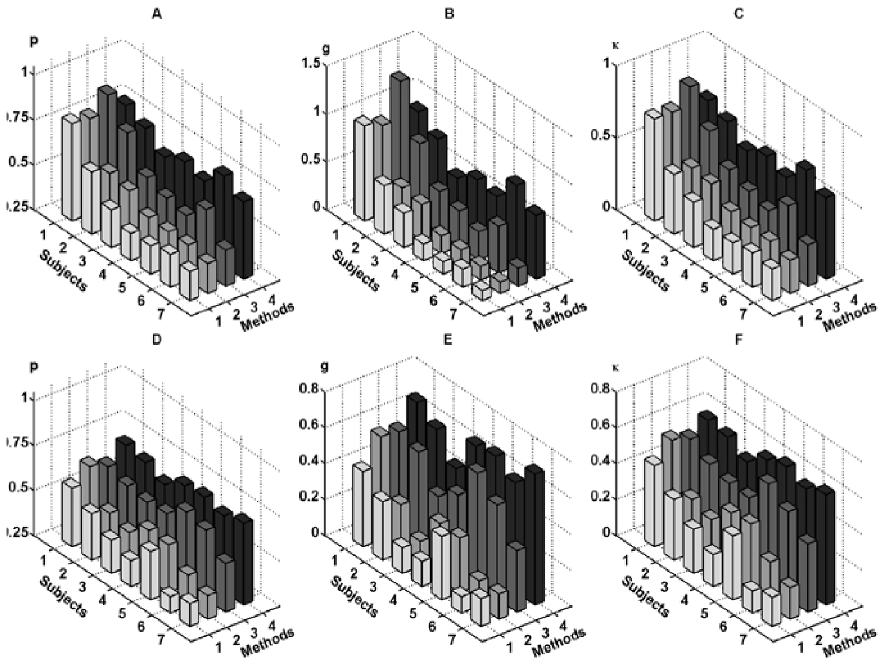


Fig. 2. Indices *p*, *g* and *k* obtained as a result of MTR data classification for all subjects, using all methods BC(1), MCSP(2), MBBC(3), CTDA(4).

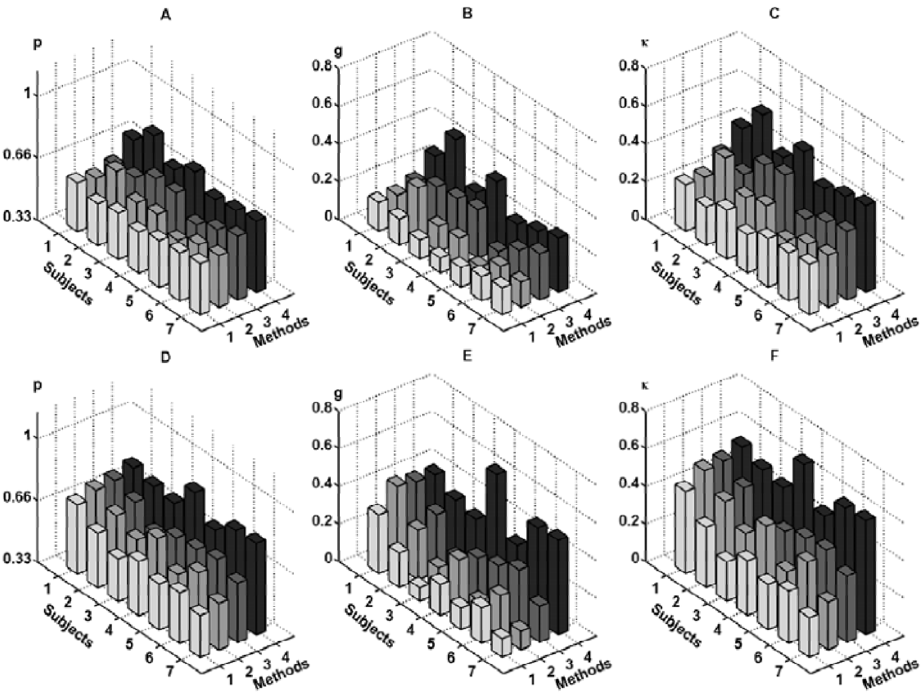


Fig. 3. Indices *p*, *g* and *k* obtained as a result of VIS data classification for all subjects, using all methods BC(1), MCSP(2), MBBC(3), CTDA(4).

Table 1. Average values of *p* index computed after processing the different datasets

Data Set	BC	MCSP	MBBC	CTDA
BCIC, 1st session	0.50 ± 0.05	0.52 ± 0.05	0.61 ± 0.06	0.70 ± 0.02
BCIC, 2nd session	0.45 ± 0.03	0.49 ± 0.03	0.59 ± 0.02	0.68 ± 0.01
MTR, 1st session	0.50 ± 0.01	0.52 ± 0.02	0.56 ± 0.02	0.64 ± 0.02
MTR, 2nd session	0.52 ± 0.02	0.57 ± 0.02	0.60 ± 0.02	0.71 ± 0.02
VIS, 1st session	0.47 ± 0.04	0.51 ± 0.04	0.54 ± 0.04	0.66 ± 0.05
VIS, 2nd session	0.46 ± 0.04	0.49 ± 0.04	0.55 ± 0.04	0.64 ± 0.06

Table 2. Average values of *g* index computed after processing the different datasets

Data Set	BC	MCSP	MBBC	CTDA
BCIC, 1st session	0.35 ± 0.12	0.33 ± 0.11	0.55 ± 0.14	0.70 ± 0.06
BCIC, 2nd session	0.23 ± 0.05	0.27 ± 0.06	0.48 ± 0.04	0.63 ± 0.03
MTR, 1st session	0.12 ± 0.01	0.15 ± 0.02	0.21 ± 0.02	0.29 ± 0.03
MTR, 2nd session	0.16 ± 0.03	0.21 ± 0.04	0.26 ± 0.03	0.43 ± 0.04
VIS, 1st session	0.30 ± 0.05	0.35 ± 0.06	0.38 ± 0.05	0.44 ± 0.06
VIS, 2nd session	0.28 ± 0.06	0.32 ± 0.06	0.40 ± 0.05	0.46 ± 0.06

Table 3. Average values of *k* index computed after processing the different datasets

Data Set	BC	MCSP	MBBC	CTDA
BCIC, 1st session	0.33 ± 0.07	0.37 ± 0.06	0.48 ± 0.07	0.60 ± 0.02
BCIC, 2nd session	0.26 ± 0.05	0.33 ± 0.05	0.46 ± 0.02	0.58 ± 0.02
MTR, 1st session	0.24 ± 0.01	0.29 ± 0.03	0.35 ± 0.02	0.46 ± 0.02
MTR, 2nd session	0.28 ± 0.03	0.37 ± 0.04	0.40 ± 0.03	0.56 ± 0.03
VIS, 1st session	0.25 ± 0.07	0.31 ± 0.08	0.35 ± 0.07	0.43 ± 0.09
VIS, 2nd session	0.25 ± 0.07	0.28 ± 0.08	0.36 ± 0.08	0.45 ± 0.10

5 Conclusion

Classifier comparison yielded quite similar results for all data sets. BC and MCSP classifiers based solely on inter-channel covariance are shown to be comparable in performance while losing to MBBC and CDTA classifiers which additionally account for EEG frequency structure. Although for MTR data set maximal quality of performance is provided by MBBC, CTDA based classifier is shown to provide the best classification accuracy on average over subjects for all data set. Surprising is that computationally inexpensive basic Bayesian classifier showed accuracy comparable with that of much more sophisticated classifiers.

References

1. Pfurtscheller G., Flotzinger D., Kalcher J.: Brain-computer Interface–a new communication device for handicapped persons. Journal of Microcomputer Applications 16 (1993).293-299
2. Neuper C., Scherer R., Wriessnegger S, Pfurtscheller G.: Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a braincomputer interface. Clinical neurophysiology 120 (2009) 239-247
3. Blankertz B., Dornhege G., Krauledat M., Muller K., Curio G.: The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. NeuroImage 37 (2007) 539-550
4. BCI competition IV. http://www.bbci.de/competition/
5. Frolov A,, Husek D., Bobrov P.: Comparison of four classification methods for brain computer interface Neural Network World 21 (2011) 101-192
6. Bobrov P., Frolov A., Cantor C., Fedulova I., Bakhnyan M. et al.: Brain-Computer Interface Based on Generation of Visual Images. PLoS ONE 6(6) (2011) e20674. doi:10.1371/journal.pone.0020674.
7. Dornhege G., Blankertz B., Curio G., Müller K.: Increase information transfer rates in BCI by CSP extension to multi-class. Advances in Neural Information Processing Systems 16 (2004) 733–740
7. Zhao Q., Zhang L., Cichocki A.: Multilinear generalization of Common Spatial Pattern; In. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2009, 19-24 April 2009 (2009) 525-528

Generating Parallel Applications from Models Based on Petri Nets

Stanislav Böhm

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
stanislav.bohm@vsb.cz

Abstract. This paper briefly describes the tool intended for modelling, simulation and generation of parallel applications. A developer is able to model parallel programs and different aspects of communication using Kaira. The used model is based on the variant of Coloured Petri Nets. Our tool can automatically generate standalone parallel applications from models. The parallelism of a final application is currently based on threads or MPI.

1 Introduction

In the world of scientific and technical computations parallel computers are natural and common tools. It gives us possibility to decrease computational time or run applications too large for a single machine. But these profits are paid by a more complex development of applications for such systems compared to their sequential counterparts. The problems with parallel applications arise in each part of the development process from designing parallel algorithm, through implementation to testing and debugging.

As a solution for easier creation of parallel programs we start to work on a programming tool Kaira. It allows to create a visual model of communication and parallelism of a program and a programmer can insert sequential codes into such model. A behaviour of a model can be observed in simulations and at the end a programmer can generate a final stand-alone application by "one click". In the current version sequential codes inside a model are written in C/C++ and resulting applications can use threads or Message Passing Interface (MPI)[1] as parallel backends. Visual models are also used for debugging. The debug information can be shown to a user in a more high-level way compared to classic debugging solutions so a user gains a global overview of a state of computation more easily. The tool is an open source project and can be obtain at http://verif.cs.vsb.cz/kaira.

The variant of Coloured Petri Nets (CPN)[2] are used in Kaira as the graphical language for modelling. From all tools (that we are familiar with) we found *CPN Tools* [3, 2] and *Renew* [4, 5] as the most inspiring and closest to Kaira.

CPN Tools is one of the most famous tools for designing CPN. CPN ML (the variant of Standard ML) is used as the network inscription language. The strong points of this tool are simulations, the state space analysis and the performance

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 320–324.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

analysis. The *CPN Tools* is (for example) very good at modelling protocols. The main difference from Kaira is the absence of a code generation. So a user can model applications but when he wants to get executable codes he needs to write them by hand from scratch.

Renew is the tool for designing reference nets. These nets try to combine the object oriented programming with Petri Nets. Java is used as the inscription language and any Java library can be used in models. Renew is also good at modelling parallelism. A user can have more parallel running instances of a network. Communication is ensured by synchronization of firing transitions, instead of flows of tokens like in Kaira. The simulation in Renew can run simultaneously on more processors but it is restricted to symmetric multiprocessing machines. Any program generated by Kaira can also run on computers without shared memory due to MPI. Moreover Renew cannot generate stand-alone applications. The typical usage of Renew is designing multi agent systems. It is not designed for the high performance computing.

2 Introduction to Kaira

In this text we provide only basic intuition how the tool can be used. More details can be found in papers [6] and [7].

The tool itself provides a standalone development environment so a whole application can be written in it. The most basic function of the tool is a creation of graphical models. The tool also assists a developer with inserting custom codes into models. A user is able to simulate these models and generate stand-alone parallel applications.

Kaira also provides other possibilities such as recording of generated application's runs. Created logs can be visualized in the original model. The tool also offers an automatic agglomeration so a user can create a high-level model with a large number of parallel activities and mapping to bigger tasks is done automatically.

Our own variant of Coloured Petri Nets is used as a graphical language for Kaira models. In CPN the tokens are not only *black dots* like in the ordinary (Place/Transition) Petri Nets but they have values like number *6* or string *Hello*. In the figure 1 you can see an example of our CPN. In the example, each place has assigned a data type and only tokens with values of a corresponding type can be stored in a place. An assigned data type is displayed at the right bottom side of a place. A place's initial content is written at the upper right side.

Let us consider the following problem. We want to perform some computations for an interval of numbers. We can divide our task to separate subtasks but a computation time of each subinterval is notably different and we cannot guess it in advance. It is ineffective to simply divide all subtasks to working nodes at the beginning. As a result we introduce a master node that divides parts of the work to other nodes. When a working node finishes computation of an assigned subtask, it sends results to the master node and waits for a new job. For the sake

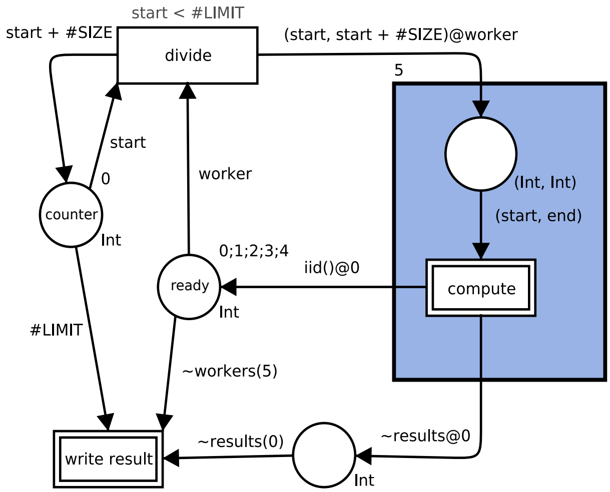


Fig. 1. The example of a model

of simplicity we fix the number of working nodes to five. The network solving this problem is shown in the figure 1.

The place *ready* in our example represents idling workers. At the beginning it contains an id (a number) of each worker. The place *counter* stores an integer representing the start of the next assigned interval. When the transition *divide* is fired then it takes an id of an idling worker from the place *ready* and it assigns a new subinterval and increases the number in the place *counter*.

The significant feature of our modification of CPN are blue areas. It allows to express replication of a part of a network where each replication has own copy of a subnetwork enclosed by a blue rectangle. In our example there are five replications of the right part of the network containing the transition *compute* and one place. Each replication represents a working node in our algorithm. The expression after "@" in the inscription of the edge from the transition *divide* specifies an identification number of replications where we want to send an interval. The double border of the transition *compute* means that there is a C++ function inside the transition. This function performs a computation on an interval of numbers. When this transition finishes a computation then it returns a token into the place *ready* and it also sends results to the master node. When we reach the limit in the place *counter* and all workers finish their computations then all results are written at once using the transition *write results*.

The integration with C++ code into the model is done by two ways: The code can be inserted into transitions and places or C++ types and functions can be integrated into the inscription language. The latter allows for example integrate some C++ library for matrix operations. Tokens can represent matrices and on the edges we can have expressions for operations with them. Only inserting a code into a transition will be shown here. The similar idea with auto-generated templates works also for other cases.

Let us assume that we want to insert a code into the transition *divide*. The tool opens a new editor tab with the following code:

```
struct Vars {
    int start;
    int worker;
};

void function(CaContext &ctx, Vars &var)
{

}
```

This code is generated from the model and it will be updated when the model is changed. A user cannot modify this code but he can write what he wants as a function's body. The second parameter gives us an access to variables used on edges around the transition. The first parameter gives an accesses to some internal informations. Every time when the transition is fired then this function is called. The same idea works also for places but the codes inside places serve for initialization purposes, for the example loading data from files.

3 Conclusion

We propose a tool for modelling, simulating and generating parallel applications. We have been motivated by real-life scientific problems and our work is focused on parallel applications with nontrivial data flows and communication. We are interested in scientific computations where to get an application (and its results) quickly is more important than a handmade solution with a slightly better performance. To summarize, we propose the extensions of syntax and semantics of CPN that are useful in modelling of parallel algorithms. Also we offer the tool that helps to create and simulate parallel algorithms. Important advantage is the ability to generate stand-alone applications from these models.

This workshop should also summarizes the results of students. The author's work is focused on two topics. The first is the tool described here. The second is a work on one-counter machines. The latest result is the following: *The language equivalence of two real-time one-counter machines is NL-complete* [8]. A real-time one counter machine can be seen as a deterministic push down automaton without epsilon transitions and only one stack symbol (+ one special symbol on the bottom of the stack). So it can store a one arbitrary large number and tests a zero.

The author has been partially supported by the Czech Ministry of Education, project No. 1M0567, GACR P202/11/0340 and students research project SP2011/25.

References

1. Snir, M., Otto, S., Huss-Lederman, S., Walker, D., Dongarra, J.: MPI-The Complete Reference, Volume 1: The MPI Core. 2nd. (revised) edn. MIT Press, Cambridge, MA, USA (1998)
2. Jensen, K., Kristensen, L.M.: Coloured Petri Nets: Modelling and Validation of Concurrent Systems. 1st edn. Springer Publishing Company, Incorporated (2009)
3. (CPN Tools) http://cpntools.org/.
4. Kummer, O., Wienberg, F., Duvigneau, M., Schumacher, J., Khler, M., Moldt, D., Rlke, H., Valk, R.: An extensible editor and simulation engine for petri nets: Renew. In: Applications and Theory of Petri Nets 2004. Volume 3099 of LNCS., Springer Berlin (2004) 484–493
5. (Renew) http://renew.de/.
6. Bohm, S., Behalek, M.: Kaira: Modelling and generation tool based on petri nets for parallel applications. In: 13th International Conference on Computer Modelling and Simulation (UKSim2011), Cambridge, United Kingdom (2011)
7. Böhm, S., Běhálek, M., Garncarz, O.: Developing parallel applications using kaira. In Snasel, V., Platos, J., El-Qawasmeh, E., eds.: Digital Information Processing and Communications. Volume 188 of Communications in Computer and Information Science., Springer (2011) 237–251
8. Böhm, S., Göller, S.: Language equivalence of deterministic real-time one-counter automata is NL-complete. In: 36th International Symposium on Mathematical Foundations of Computer Science 2011, ARCoSS subline of LNCS. (2011) to appear in August 2011.

Using LLVM for a Functional Programming Language Implementation*

Nikola Ciprich

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
nikola.ciprich@linuxbox.cz

Abstract. This paper outlines the possibilities which LLVM (Low Level Virtual Machine) framework offers to implementors of functional programming languages. First, the LLVM is briefly introduced. Then its use is demonstrated on implementation of a simple functional programming language named Tiny. Our primary goal is to implement the $\mathcal{TIL}$-`Script` functional programming language that is a computational variant of Transparent Intensional Logic (TIL). We outline the main principles of its implementation using the LLVM framework.

Keywords: LLVM, Lambda calculus, Transparent Intensional Logic, $\mathcal{TIL}$-`Script`

Introduction

The implementation of programming languages, either declarative or imperative, procedural or object-oriented, compiled or interpreted is a complex task consisting of many partial tasks starting with lexical analysis and ending with optimisations. Rather then doing all those tasks 'from scratch', it is usually much more plausible to use existing resources at least for some of them. The LLVM can very much be such a resource. The name stands for the Low Level Virtual Machine that is a fully fledged compiler framework with a strong emphasis on supporting optimisations in all stages of a program lifecycle.

1 The LLVM framework

LLVM provides a middle layer for compilers, as well as a backend. Initialy it has been created as a research project for C and C++ languages, but a language independent design as well as the support of large commercial companies such as Google Inc. and Apple Inc increased its popularity and thus a variety of related

* This research has been supported by the Grant Agency of the Czech Republic, projects 401/09/H007 'Logical Foundations of Semantics' and 401/10/0792 'Temporal aspects of knowledge and information', and also by the internal grant agency of VSB-TU of Ostrava - SP/2010214 Modeling, simulation and verification of software processes II

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 325–330.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

projects emerged. Support for various languages ranging from C through Python to ones such as Haskell is now available, albeit not always fully complete yet.

In this paper we will use LLVM Python bindings to create the interpreter for a very simple functional language called Tiny. This interpreter will accept Tiny source as its input and translate it into the LLVM intermediate representation (IR). LLVM will then optimise and compile the code and evaluate the result.

IR code LLVM uses is a low level assembly-like language of a register virtual machine. There is a noteworthy feature of IR, namely its variables handling; instructions are in the so-called *Static Single Assignment* form (SSA), which means that every variable can be written only once. Then its content is frozen and cannot be changed. While this may seem to be a complication for programmers, it offers some interesting optimisation possibilities (for example, variable dependency analysis is almost trivial for SSA languages). For more information on SSA, see [5].

Another important feature of LLVM is a language independent type system supporting both basic (integer and floating point numeric types) and more advanced (pointers, arrays, vectors, structures and functions) types. This gives programmer a strong tool to implement programming languages with advanced features like support for *first-class objects*, *object oriented programming* and others. Support for easy implementation of exception handling is also provided.

1.1 LLVM instruction set

LLVM instruction set consists of only 31 instructions. This is due to the fact that unlike common assembly languages, LLVM does not use multiple instructions for one and the same operation and some instructions are overloaded by supporting multiple datatypes operations. Most of the instructions are in the three-address form, meaning that they accept two arguments of the same type and return the result of this type. For example the instruction for adding two numbers has the following form:

```
<result> = fadd <type> <op1>, <op2>
```

meaning that it will yield the result of `<type>` type. Thus the following code:

```
%tmp2 = fadd float %tmp, 1.20000e+00
```

will add the number 1.2 to the variable *tmp* of the float type and store the resulting value into a new float variable *tmp*2. Full list of LLVM assembly instructions is beyond the scope of this document and is available in [4].

2 Tiny language

Tiny is a very simple functional language used to demonstrate LLVM ease of use. The syntax of the language is trivial (inspired by lambda calculus) and it supports only one type which is the type of floating numbers (thus all variables are implicitly floats). It is possible to define (recursive) functions and use a branching construct by applying the `if`-then-else operator. Basic numeric operations like addition, subtraction, multiplication and division are supported.

Here is an example code to define and use the well-known factorial function:

```
$ def \x fact if x < 2 then 1 else (* x fact (- x 1))
$ (fact 6)
 result = 720.0
```

2.1 Tiny implementation

Example implementation is done using Python programming languge and LLVM bindings. First two components of Tiny language are syntactic and lexical analyser. Their aim is to parse the source code of a target language, check it for syntactic errors and produce its interpretation in the form of *abstract syntax tree* (AST). For simple languages, those can be implemented directly, for more complex ones, it is more suitable to use some kind of a parser generator. Parser generator is a tool whose input is a grammar (often in BNF) of a programming language and whose generated output is the source code of a parser that can be afterwards used as a component of an interpreter or compiler. Here is Tiny parser grammar in Backus-Naur (for more information on BNF, see for example [7]):

```
EXPR ::= NUMBER
       | VARIABLE
       | ( OP EXPR* )
       | "def" LVAR* NAME EXPR
       | "if" EXPR "then" EXPR "else" EXPR
OP ::= NAME | "+" | "-" | "*" | "/" | "<" | ">" | "=="
LVAR ::= "\"NAME
NAME ::= [a-z]+
NUMBER ::= [0-9]+"."[0-9]+
```

The main rule describing expressions is divided into five subrules: the first subrule describes numeric values, the second subrule describes variables, the third one describes application of operators or functions on their arguments, the fourth one describes definition of functions (with zero or more variables) and the last, most complex one describes the if-then-else conditional. Other rules are just helpers for the main rules with obvious purposes. The parser then produces the AST composed from instances of the following respective classes: `NumberExpr`, `VarExpr`, `FunctionExpr`, `FunctionDefExpr`, `FunctionBodyExpr`, `IfExpr`.

Names of all defined variables are stored in the global list (`symbolTable`), function names including their arity are stored in the global hashlist (`funcTable`). Every AST node implements a `genCode()` method which will generate the respective LLVM code. For numerical expressions, the `genCode()` method is simple, it just returns LLVM floating point numbers constant:

```
def genCode(self):
    return Constant.real(Type.double(), self.value)
```

The variable expressions code generating method is also simple. It checks the symbol table for the presence of a variable, and returns it or raises exception if the variable was not defined:

```
def genCode(self}:
    if self.varName in symbolTable:
        return symbolTable[self.varName]
    else:
        raise RuntimeError, 'Undefined variable: %s' % self.varName)
```

The limitation of the code is evident — for the sake of simplicity, it does not distinguish global variables from local ones.
So far, the code has been simple; no LLVM assembly was needed at all. However, the implementation of functional application is a bit more complicated. Tiny supports the application of named (user defined) functions, as well as implicitly defined mathematical operators. The implementation of `FunctionExpr` is demonstrated on the plus (+) operator and on named functions:

```
def CodeGen(self):
    if self.funcName in ['+', '-', ...]:
        argCount=2
    else:
        if self.funcName in funcTable:
            argCount=len(funcTable[self.funcName])
        else:
            raise RuntimeError, 'Undefined function: %s' % self.funcName
    if len(self.args)<>argCount:
        raise RuntimeError, 'Invalid number of args to %s' % self.funcName
    args = [ a.genCode() for a in self.args ]
if self.funcName == '+':
    return builder.fadd(args[0], args[1], 'tmpRes')
elif self.funcName == '-':
.
. # code for other operators
.
else:
    func = jit.get_function_named(self.funcName)
    return builder.call(func, args, 'tmpCall')
```

First, arguments count is compared to the number of arguments that the function (or operator) accepts. If they match, then all arguments are recursively evaluated, their results inserted into the args array. Then the corresponding LLVM assembly is emitted using `llvm.core.Builder` instance. To adduce an example, the resulting LLVM assembly for the expression `(+ x 2)` comes down to this code:

```
define double @0() {
entry:
    %tmpRes = fadd double %x, 2.000000e+00
    ret double %tmpRes
}
```

while calling example defined function `(inc x)` comes down to:

```
define double @1() {
entry:
    %tmpCall = call double @inc (double %x)
    ret double %tmpCall
}
```

The most complex task is the definition of functions. Since we want to support recursive functions, it is useful to split function definitions into two steps: function prototype and function body. The function prototype `genCode()` comes down to:

```
def genCode(self):
    ft = Type.function(Type.double(), [Type.double()] * len(self.args), False)
    func = Function.new(jit, ft, self.name)
    for arg, name in zip(func.args, self.args):
        arg.name = name
        symbolTable[name] = arg
    return func
```

The code defines a new function prototype and adds variables to global symbol table. For the sake of simplicity, it lacks even the elementary checks for function redefinition. A more proper code can be found in [6]. After the function prototype is prepared, we then need to populate it with the real code. To this end a new (global) `Builder` instance is created, variable scope cleared and then the code is emitted:

```
def genCode(self):
    symbolTable.clear()
    func = self.prototype.genCode()
    block = func.append_basic_block('entry')
    global builder
    builder = Builder.new(block)
    try:
        retVal = self.body.genCode()
        builder.ret(retVal)
        function.verify()
    except:
        func.delete()
        raise RuntimeError, 'function emit failed!'
    return func
```

What makes even such a simple language as Tiny rather powerful is the support for control flow: together with recursion it is then possible to implement almost anything. Conditional statements are represented using `if-then-else` expression. Thus we need to implement `genCode()` method for three last AST expression nodes: `IfExpr`. It contains pointers to all three subexpressions: the condition itself, the code for the branch when the condition is true and the code for the branch when the condition is false.

```
def genCode(self):
    cond = self.cond.genCode()
    cond_bool = builder.fcmp(FCMP_ONE, cond, Constant.real(Type.double(), 0), 'ifcond')
    func = builder.basic_block.function
    then_block = func.append_basic_block('then')
    else_block = func.append_basic_block('else')
    merge_block = func.append_basic_block('ifcond')
    builder.cbranch(cond_bool, then_block, else_block)
    builder.position_at_end(then_block)
    then_value = self.then_branch.genCode()
    builder.branch(merge_block)
    then_block = builder.basic_block
    builder.position_at_end(else_block)
    else_value = self.else_branch.CodeGen()
    builder.branch(merge_block)
    else_block = builder.basic_block
    builder.position_at_end(merge_block)
    phi = builder.phi(Type.double(), 'iftmp')
    phi.add_incoming(then_value, then_block)
    phi.add_incoming(else_value, else_block)
    return phi
```

The code for the flow control support introduces one special function which LLVM provides: the Φ function. While LLVM almost shields SSA variables handling from a programmer by automatically creating variables versions, when implementing conditional expressions, SSA support needs to be taken into account. Let us have a code, which assigns either value 1 or 2 to a variable, depending on a condition evaluation. The compiler automatically creates alternate variable versions, but it needs to be hinted which will be the recent version of the variable; this is the situation when the Φ function is useful:

$$
\begin{aligned}
&\texttt{if(cond)}:\\
&\quad a_1 = 1;\\
&\texttt{else}:\\
&\quad a_2 = 2;\\
&a_3 = \Phi(a_1, a_2)
\end{aligned}
$$

It "merges" the variables a_1 and a_2 and assigns a value to the variable a_3 accordingly to the branch of the code that has been executed. In the Tiny language, the whole `if-then-else` expression returns value of one of its subexpressions depending on the evaluation of the condition. Thus Φ is used to return the evaluation of a proper code block.

3 The $\mathcal{TIL}$-`Script` language

The $\mathcal{TIL}$-`Script` language is a computational implementation of *Transparent Intensional Logic* (TIL) constructions. The language of TIL constructions, invented by the significant Czech logician Pavel Tichý([9]) is a hyperintensional, typed, partial lambda calculus; hyperintensional, because TIL lambda terms denote *constructions* of functions (intensions) rather than the functions conceived as set-theoretical mappings; typed, because all the entities TIL works with, including constructions, receive a type; and partial, because TIL constructions operate over partial rather than total functions. Constructions are algorithmically structured (higher-order) instructions on how to obtain a (lower-order) output product given some input entities.

4 Conclusion

The expressive power of TIL has been demonstrated in a number of papers, especially in the area of *language processing*, specification, conceptual modelling and communication. For more rigorous description, the [10] is the most complete and up to date source of information. This paper outlined the possibilities of its implementation. Its further use in practice is then mainly in the area of natural language processing, comunication and deduction.

References

1. Pierce, B. C. Types and Programming Languages. 2002 MIT Press. ISBN: 978-0262162098.
2. Rémy D. Using, Understanding, and Unraveling The OCaml Language Available at: `http://caml.inria.fr/pub/docs/u3-ocaml/index.html`
3. Adve V., Lattner C. LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation. *In Proceedings of the 2004 International Symposium on Code Generation and Optimization (CGO'04), Palo Alto, California*
4. LLVM Language Reference Manual. `http://llvm.org/docs/LangRef.html`
5. Cytron R., Ferrante J., Rosen B. K. , Wegman M. N., Zadeck F. K.: Efficiently computing static single assignment form and the control dependence graph. *Trans. Prog. Lang. and Sys.*, pp 451–490, 1991.
6. Lattner C., Shawabkeh M. The LLVM Kaleidoscope tutorial for `llvm-py` project. `http://mdevan.org/llvm-py/kaleidoscope/PythonLangImpl1.html`
7. Wikipedia article on Backus-Naur form `http://en.wikipedia.org/wiki/Backus-Naur_Form`
8. Ciprich N., Duží M., Košinár M. The $\mathcal{TIL}$-`Script` language. *In Information Modelling and Knowledge Bases XX, Y. Kiyoki, T. Tokuda, H. Jaakola, X. Chen, N. Yoshida (eds.).* Amsterdam: IOS Press, 166-179.
9. Tichý P. *The Foundations of Frege's Logic.* 1988. De Gruyter. ISBN: 3-11-011668-5.
10. Duží M., Jespersen B., Materna P. *Procedural Semantics for Hyperintensional Logic; Foundations and applications of TIL.* 2010. Berlin: Springer. ISBN: 978-90-481-8811-6

Using Kohonen Maps and Singular Value Decomposition for Plagiarism Detection

Asim M. El Tahir Ali, Hussam M. Dahwa Abdulla, and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
asim070@yahoo.com, hussamdahwa@hotmail.com, vaclav.snasel@vsb.cz

Abstract. Plagiarism has become one area of interest for researchers due to its importance, and its fast growing rates. Effective clustering methods and faster search tools for matching and discovering the similarities between documents were the main two areas for the researchers. Many tools and techniques have been developed for plagiarism detection. In this paper we use singular value decomposition for its effective clustering of the documents in-order to reduce search time by creating a new matrix with fewer dimensions used for clustering the original (source) documents, and we use Neural Networks for local matching and comparison between a suspicious document and a source document, Kohonen maps used to visualized and comparison of the result, in which represent the result as picture that easier to be analyzed.

Keywords: Plagiarism detection, Kohonen Maps, VSD.

1 Introduction

With the hug of the information on WWW and digital libraries, plagiarism became one of the most important issues for universities, schools and researcher's fields. It is so easy through the internet and due to using advanced search engine to find documents or journals by students [1]. Some of the researchers are just copying and pasting others works without reference to the owner of the document. A good survey of ideas about how to define plagiarism is done by Baeza-Yates, R. & Ribeiro-Neto, and B.1999 [2]. In plagiarism detection document clustering of the documents and visualization of comparison result are the most important issues in plagiarism detection, looking for fast, quick, effective and efficient tools for plagiarism detection is our objective. Matrix decomposition used by singular value decomposition, as effective methods for information retrieval and clustering of the documents, is known as Latent Semantic Indexing (LSI) or Latent Semantic Analysis which is used for discovering latent between documents [3], [4]. We are going to apply this approach to formal concept analysis (FCA) [5]. Measuring similarities by using quickly tool is useful to determine whether the document has been plagiarized or not. For document

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 331–336.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

comparison, there are some useful techniques like inter-document similarity which is used to measure the distance and to match similarities between documents that can be easily to be clustered by matrix and to visualize the result by using Kohonen maps (SOM).

2 Definition of Plagiarism

Many definitions have been introduced by some researcher: “Plagiarism, the act of taking the writings of another person and passing them off as one’s own. The fraudulence is closely related to forgery and piracy practices generally in violation of copyright laws” Encyclopedia Britannica [6]. According to the Merriam-Webster Online Dictionary, to "plagiarize" means:

- To steal and pass off (the ideas or words of another) as one's own.
- To use (another's production) without crediting the source.
- To commit literary theft.
- To present as new and original an idea or product derived from an existing source.

2.1 Singular Value decomposition

Singular value decomposition (SVD) is one of the most important matrix decompositions of it used in information retrieval as Latent Semantic indexing (LSI). SVD is especially suitable in its variant for sparse matrices [4].

Theorem 1: Let A is an $m \;\; n$ rank-r matrix. Be $\sigma_1 \geq \cdots \geq \sigma_r$ Eigen values of a matrix $\sqrt{AA^T}$.Then there exist orthogonal matrices $U = (u_1, \ldots, u_r)$ and $V = (v_1, \ldots, v_r)$, whose column vectors are orthonormal, and a diagonal matrix $\Sigma = diag(\sigma_1, \ldots, \sigma_r)$. The decomposition $A = U \Sigma V^T$ is called singular value decomposition of matrix A and numbers $\sigma_1, \ldots, \sigma_r$ are singular values of the matrix A. Columns of U (or V) are called left (or right) singular vectors of matrix A. [17].

Now we have a decomposition of the original matrix A. It is not needed to say, that the left and right singular vectors are not sparse. We have at most r nonzero singular numbers, where rank r is the smaller of the two matrix dimensions.

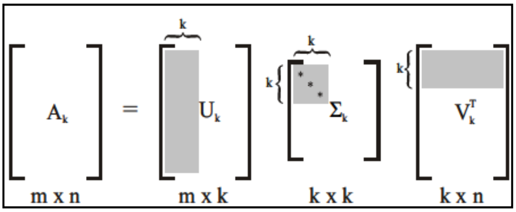


Figure 1: k- reduced singular value decomposition.

Let us have k, $0 < k < r$ and singular value decomposition of A

$$A = U \sum V^T = \left(U_k U_0\right) \begin{pmatrix} \sum_k & 0 \\ 0 & \sum_0 \end{pmatrix} \begin{pmatrix} V_k^T \\ V_0^T \end{pmatrix}$$

We call $A_k = U_k \Sigma_k V_k^T$ a k-reduced *singular value decomposition* (*rank-k* SVD)

In information retrieval, if every document is relevant to only one topic, we obtain a latent semantics – semantically related words and documents will have similar vectors in the reduced space. For an illustration of rank-k SVD see figure 1, the grey areas determine first k coordinates from singular vectors, which are being used. [17].

Theorem 2: (Eckart-Young) among all m × n matrices C of rank at most k A_k is the one, that minimizes $\left\| A_K - A \right\|_F^2 = \sum_{i,j} \left(A_{i,j} - C_{w,j} \right)^2$

Because rank-k SVD is the best rank-*k* approximation of original matrix A, any other decomposition will increase the sum of squares of matrix $A - A_k$. . [17].

2.2 Self-organization maps:

A Kohonen map also known as Self-Organizing Map (SOM) [7] has been developed on 1989 by Teuvo Kohonenis, it's a competitive artificial neural network tool. SOM considered as one of the efficient tool used for visualization of multidimensional data also an efficient mechanism in signal processing and data mining applications. SOM is special types of Neural Networks that can map or simplify a complex high dimensional input signal into a simpler low dimensional. It's used for classification and clustering data set according to their similarities [8]. SOM is an artificial network that is structured in two layers of neurons. The first layer represents the input data, the second layers is a neuron's grid, usually bi-dimensional, full connected to each other. All input nodes are connected to all output neurons (nodes). The output neurons are usually arranged into two or three low dimensional grids. Attached to every neuron there is a weight vector with the same dimensionality as the input vectors. The number of output grid dimension is usually lower than the input dimensions. SOMs are mainly used for dimensionality reduction rather than expansion.

2.3 Unified distance matrix

The unified distance matrix or U-matrix [9] is a representation of the Self-Organizing Map that visualizes the distances between the network neurons or units'-matrices are useful tools for visualizing clusters in input data without having any priori information about the clusters. [10].The multidimensional visualization of the Self-Organizing Map (SOM) data can be possible represented by using by the Unified distance matrix [11]. This is achieved by using topological relations property among neurons after the learning process. U-matrix contains the distances from each unit center to all of its neighbors'. The neurons of the SOM network are represented here by hexagonal cells, by using U-matrix we can detect topological relations among neurons and infer about the input data structure. Color schema used to represent the similarities, the dark coloring (High values in the U-matrix) indicate or represents that there is a gap between the values in the input space this is mean that their no similarities, and the A light coloring (low values in the U-Matrix) represent or indicate the vectors are so close to each other in the input space and this is mean that there are a high similarities between the inputs. This representation can be used to visualize the structure of the input space and to get an impression of otherwise invisible structures in a multidimensional data space [12].

3 Related Work

- C.M.X. Benjamin, W.L. Woon and K.S.D. Wong: developed software for document comparison and visualization. Allow the user to visualize document similarities in an intuitive way via Kohonen Maps. Used in plagiarism detection. [14]
- José Alfredo F. Costa and Márcio L. de Andrade Netto, proposed two algorithms used in automatic partitioning and labeling Kohonen maps (SOM) networks in clusters of neurons and for generating a hierarchy of maps based on the detected data clusters[15].
- Tommy W. S. Chow and M. K. M. Rahman: proposed approach for document retrieval and plagiarism detection they used multi layer self organization map (MLSOM) and a tree structure to improve the document retrieval and plagiarism detection, which can represent document, pages and paragraph. MLSOM used for matching the similarities between document, and they usde MLSOM algorithms for clustering documents that have been developed by the author [16].

4 Our Approach

Our main objectives is to reduce the search time in plagiarism detection tools and to simplify the result analysis, mainly for the manual plagiarism detection like code source plagiarism detection, later we can apply in to automatic plagiarism detection, so in-order to reduce the search time we focus on the clustering mechanism that can reduce or summarized documents in a better way , we used SVD for effective clustering of the documents by creating a new matrix with fewer dimensions used for clustering the original (source) documents and a suspicious document, this is stage one , The second stage in our approach is how to use Neural Networks for local matching and comparison between a suspicious document and a source document, and to use SOM for visualization and comparison of the result, in which the result is represented as picture that is easier to be analyzed .

4.1 Stage one Experiment:

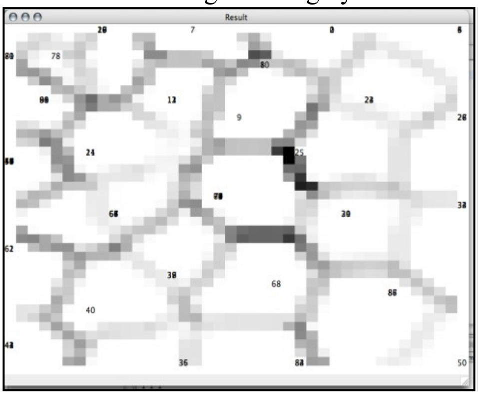
Data source:

Our experiment data was asset of documents (ten of student's assignments about one topic) which are considered as an excellent example to show the similarities between documents, which can be easily representing data in graph. We assume that row represents a document and columns represent documents attributes. Assumptions: let **1** represent an attributes (similar word or text in documents) in the appropriate row if its exit, and column crossing, otherwise **0** (zero**)** if it's not exit. This data can easily used by Singular Value decomposition (SVD).

Experiment steps:

Our experiment divided into three steps as follow:

- First step: reading and transforming data:
  As we mentioned before that the data have been collected from some student's assignments, after we apply the stage by using SVD we obtain a new matrix (92x10) that had 92 rows and 10 columns (row represent documents, and column represent their attributes) If the document has the similar words and texts this means that its attributes exit in this case represented by 1 otherwise 0. This incidence matrix was used in SVD, formal concept analysis and NMF computations.
- Second step: Using Singular Value Decomposition :
  We use SVDLIBC-fix software in our SVD computations, this software is written by C language and it's free of charge, we use incidence matrix as input, where this software computer the three matrixes U, $\sum$ and V^T . k-rank chosen k=4, where matrixes were U(92x10), $\sum$(10x10) and V^T(10x10).
- Third Step: Result Visualization:
  In this stage we use SOMs for representation of the original data and data after applying SVD to the original data, and we obtain two SOMs networks this according to type of input data, for visualization of the network we applied Unified distance matrix (U-matrix) algorithm for the computation of the data, we obtain two figures in grayscale. This is shown down in figure2 and figure3:



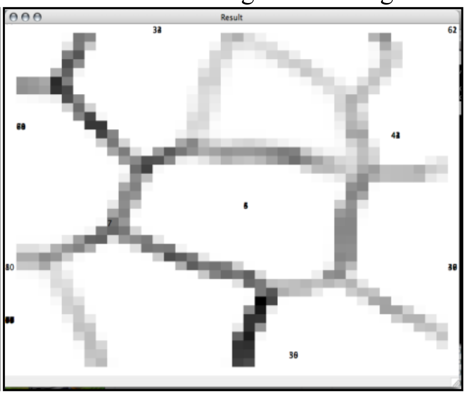


Figure2: The original data Figure3: Data after using SVD

4.2 Stage two: Plagiarism detection:

Clustering techniques are not enough by itself to judge plagiarism but can be used in the candidate retrieval stage to group similar documents that discuss the same subject, it should be followed by another level of plagiarism analysis and detection methods this what happened in the first stage where the above two SOMs figures show that the number of the differences that has been reduced by using SVD computation, the figures are easy now to be analyzed. The time needed to locate the similarities between documents if it clustered in figure3 is shorter and the effort is less than less that the time and effort that we will spend by using figure2. SOMs provide easier form for detecting similarities and it's suitable for code source plagiarism detection which is requiring manual detection. but the plagiarism detection system used clustering to find similar documents, then documents in the same cluster were compared until two similar paragraph are found, paragraph were compared in details.

5 Conclusion and further work

Kohonen Maps and Singular value decomposition were used in our paper to improve clustering and plagiarism detection, we used Singular value decomposition (SVD) for clustering our data for its effective clustering, and we used Kohonen Maps (MOS) for representation of the results, after we applied our above approach (SVD) we had obtained two u-matrix visualization, we had improved the clustering and reduced the number of the concept, also it shows the easiest way for result visualization in graph which is easier to be analyzed and used in comparison between documents that may help the earlier code source plagiarism detection stage. In our future work we are going to use multilayer SOM for automatic plagiarism detection.

References.

1. Plagiarism. http://www.plagiarism.org.
2. Baeza-Yates, R. & Ribeiro-Neto, B. Modern Information Retrieval. Addison-Wesley, 1999.
3. M. Berry and M. Browne. Understanding search engines. Mathematical Modeling and Text Retrieval, 1999.
4. T.Letsche M. Berry, S. Dumais. Computation methods for intelligent information access. Proceedings of the 1995 ACM/IEEE Supercomputing Conference,1995.
5. B. Ganter and R. Wille. Formal Concept Analysis. Springer-Verlag Berlin Heidelberg, 1999.
6. EncyclopediaBritannica, http://www.britannica.com/EBchecked/topic/462640/plagiarism
7. T. Kohonen. Self-organizing maps. Springer, 2001
8. C. Lee C. Kiu. Discovering ontological semantics using fca and som. M2USIC 2004, 2004.
9. Ultsch, A.; Siemon, H. (1990). Kohonen.s self-organizing feature maps for exploratory data analysis. Proceedings INNC.90P. 305 .308, Dordrecht, The Netherlands).
10. J. Hollmén, "Process Modelling using the Self-Organizing Map," M.Sc.thesis, Dept. Computer Science, Helsinki Univ. of Technology, Finland,1996.
11. SOM-based data visualization methods, 1999
12. Using SOM for information visualization and knowledge discovery in complex geospatial datasets. Koura, E.L. (ITC).
13. CLEF (Notebook Papers/LABs/Workshops) 2010.
14. C.M.X. Benjamin, W.L. Woon, K.S.D. Wong: A Graphical and Convenient Tool for Document Comparison and Visualization, ICCE- 13-15, 2008 Kuala Lumpur, Malaysia.
15. José Alfredo F. Costa and Márcio L. de Andrade Netto, Clustering of complex shaped data sets via Kohonen maps and mathematical morphology
16. Tommy W. S. Chow, M. K. M. Rahman: Multilayer SOM With Tree-Structured Data for Efficient Document Retrieval and Plagiarism Detection.
17. Václav Snásel, Martin Polovincak, Hussam M. Dahwa Abdulla: Concept Lattice Reduction by Singular Value Decomposition.SYRCoDIS 2007.

Iris Segmentation in Eye Image

Tomáš Fabián[1], Jan Gaura[1], and Petr Kotas[2]

[1]Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{tomas.fabian, jan.gaura}@vsb.cz
[2]Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.kotas@vsb.cz

Abstract. In this paper we present our previously published work on iris segmentation. The iris extraction algorithm described here is based on our observation of local brightness properties of close-up iris images. Our algorithm consists of three simple steps. We detect the bright point inside the pupil, outer limbic boundary is found via statistical measurements of outer boundary points and inner boundary points are searched by means of defined cost function maximization. Performance of the presented method is evaluated on series of iris images and compared with the traditional Hough method as well.

Key words: iris extraction, iris segmentation, iris localization, biometrics

1 Introduction

Iris segmentation has two major real life applications. The primary motivation for iris segmentation is to get biometric data unique for each person. The secondary usage is in medicine, where some eye diseases can be diagnosed directly by separating some distinct features like different color patches or completely different color tone of the iris. Iris localization is currently solved by many different approaches. The most profound methods are that of [1] and [8]. Daugman uses integrodifferential operator (IDO) for precise pupil and iris localization. Wildes uses two stage passes; the first pass detects edge map; the second pass use Hough transform to detect circles in the image. Localization of inner and outer iris boundary together with eyelids (top and bottom) is presented in [5]. Our method is simple yet effective. It is based on statistical point of view when searching for limbic boundary and rather analytical approach when detecting pupillary boundary. In this way, it is well suitable for parallel implementation with subsequent real-time usage.

The rest of the paper is organized as follows. In the Section 2, we introduce our algorithm for iris segmentation and discuss all parts, Section 3 presents experimental results, comparison to Hough transform and to manually created ground truth data. Section 4 bears final thoughts and conclusion.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 337–342.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Algorithm

Our algorithm consists of three main stages. In the first step, we detect bright spot inside the pupil. The second stage locates limbic boundary and starts from detected bright spot position that is taken as an approximated location of the pupil's center. We are looking for points that compose limbic boundary. Localization of the boundary points is based on statistical evaluation of inverse gradient of the source image. In the third stage, inner edge called pupillary boundary is detected by means of cost function optimization.

2.1 Inner Point Detection

Nowadays, iris databases use a bright light in a close-up distance from the photographed eye. This technique creates a bright spot positioned in the center of the eye. This point represents our first approximation of limbic boundary center $\mathbf{c}_l$. To do so, we need to preprocess original image I to remove surrounding bright artifacts caused by random specular highlights on the eye. This is done by weighting original image by two–dimensional Gaussian function.

2.2 Limbic Boundary Detection

With the starting point $\mathbf{c}_b$ being set, we are able to proceed with localization of the limbic boundary. At first, we need to convert the color eye image I to so-called inverse gradient $I_g = 1/(1 + |\nabla I(x,y)|)$ where $|\nabla I(x,y)|$ is the size of gradient of the image I. Our experiments show that this type of grayscale image provides more information than classic gradient.

The eye image is segmented by a technique similar to polar coordinates samples that means that window $\Omega_{i,j}$ is moving from detected point $\mathbf{c}_b$ under angle α_i with some sample rate (see Fig. 1). Result of such process is depicted in Fig. 2. To produce such plot we used moving window defined as follows

$$d_{i,j} = \frac{1}{A(\Omega_{i,j})} \iint\limits_{\Omega_{i,j}} I_g(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{1}$$

where $\Omega_{i,j}$ is the circular region, I is an input image and A returns the area of $\Omega_{i,j}$ in pixels. We experimented with different types and shapes of window, but our experiment shows that circular shape with constant radius is the best choice since it is invariant to change of angle. In Fig. 2, it can be clearly seen that we have used twenty ρ values (i.e. $n = 20$) ranging from 0 (i.e. image center) to $\min\{I_w, I_h\}/2$ (i.e. image border) per detection line. Figure 2 shows a typical data from plotting all samples acquired by described procedure.
It is evident that local minima located at both sides of the plot correspond, in majority of cases, to our limbic boundary points. After we find minima for every angle, we construct a vector of points as the starting estimate for next phase of circle fitting. We use the Levenberg–Marquardt algorithm [6] for finding center and radius of the circle.

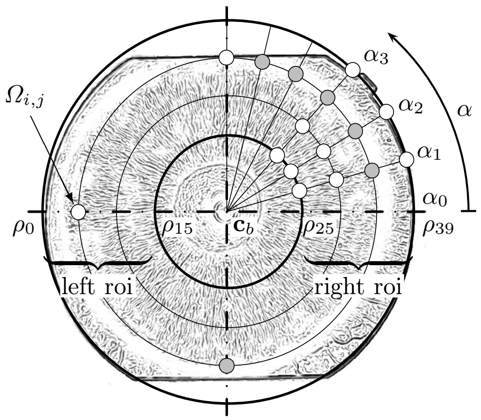


Fig. 1. Scheme of the boundary points extraction. Figure shows centric circles whose radii represent each position of sampling area $\Omega_{i,j}$. Samples with gray color represent points where function $d_{i,j}$ reaches its local minima and so marks the most probable presence of the limbic boundary. The image in the background is result of inverse gradient application to the input image

2.3 Pupillary Boundary Detection

From the previous step of our algorithm, we have the center and radius of the limbic boundary. We use these two information as rough estimate of pupillary boundary center (x_e, y_e) and radius r_e (which is set to be 40 % of the radius of limbic boundary). At this point, we extract square region around (x_e, y_e) with both sides equal to $2r_e$ and convert this area from RGB color space to the 1D brightness space defined by Equation (2)

$$I_s(x,y) = \frac{\sqrt{3}}{3d}\left[I_r(x,y) + I_g(x,y) + I_b(x,y)\right], \tag{2}$$

where suffixes r, g and b specify components of standard RGB color space and d is the length of diagonal of this space.

After analysis of brightness function in the area of the pupil we have found out following observation. The crossing area between pupil and iris shows out remarkable drop of the brightness. Idea utilizing such observation can be formally described as follows

$$J(x,y,r) = 2\pi r\left[\oint\limits_{C(x,y,r)} I_s(l)\,\mathrm{d}l\right]^{-1}, \tag{3}$$

where $C(x,y,r)$ is the circle with center in (x,y) and radius r. I_s is previously computed grayscale image defined in Equation (2). Informal observation about brightness is represented as the second term in Equation (3) that, in essence,

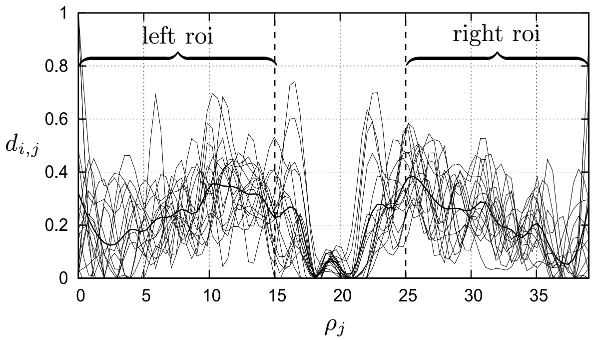


Fig. 2. This graph represents the course of values $d_{i,j}$. Each of the thin lines was created using interpolation of piece–wise linear curve through points $(d_{i,j}, \rho_j)$ for a given angle

specifies the darkness of the circle C. Since the problem is defined as maximization task, the first term embodies our requirements on circle radius maximization. Finally, by multiplying these two terms, we get resulting formula presented in Equation (3). It is obvious that searched circle C can be found by maximization of Equation (3). The function course for fixed y-value is depicted in Fig. (3). Initial estimate of local extreme is marked by cross labeled *init*. It is easy to see that we are looking for local maximum solution of J within basin of convergence from *init* point.

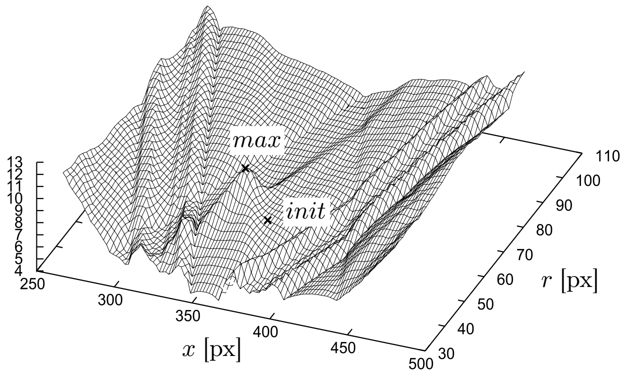


Fig. 3. Example of the cost function maximization

We compute maximal solution of J by Nelder–Mead minimization algorithm [7] simply by setting our cost function to $-J(x, y, r)$. By this computation, we obtain circle parameters that describe pupillary boundary. Solving this problem in two separate steps allow us to check result by intuition in 3D space (see Fig. 3).

3 Experiments

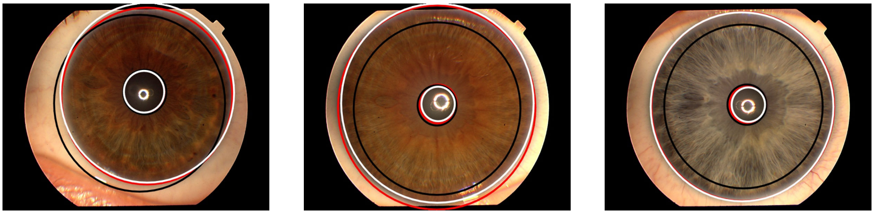


Fig. 4. Examples of evaluation results. Red circles represent manually generated ground truth, white circles show output of our method and black circles represent output from the Hough transform

To evaluate the effectiveness of the presented algorithm we have used the iris database of [2]. We have defined three metrics to precisely evaluate and compare our algorithm with the reference implementation that uses Hough transform. Ground truth data of the pupillary and limbic boundary have been created by precise manual identification. The first normalized quality criterion is described as follows

$$e_a = \frac{1}{|\Gamma|} \sum_{\gamma} \frac{A\left(r_\gamma, r_\gamma^R, ||\mathbf{c}_\gamma^R - \mathbf{c}_\gamma||\right)}{\pi \left(\max\left\{r_\gamma, r_\gamma^R\right\}\right)^2}\,, \tag{4}$$

where function A returns common area of both circles. Index $\gamma \in \Gamma = \{$l - limbic b., p - pupillary b.$\}$ indicates whether the parameter is representing a circle of the outer (i.e. limbic) or inner (i.e. pupillary) boundary. The upper index R specifies the ground truth values. It is obvious that this criterion does not take into account the rate of axle offset between compared circles in certain mutual position of each compared circle. Therefore, to acquire the most objective results of our measurements, we have introduced two new indicators

$$e_c = \sum_{\gamma} ||\mathbf{c}_\gamma^R - \mathbf{c}_\gamma||\,, \qquad e_r = \sum_{\gamma} |r_\gamma^R - r_\gamma|\,. \tag{5}$$

The last two mentioned criterion are not relative and represent absolute distance in pixels between evaluated circle parameters and the corresponding parameters of reference circle. To acquire the relative values, we have divided both values of e_c and e_r by the width of the input image in pixels. Results obtained using described error functions are summarized in Table 1.

It is evident that from our measurements, the proposed method reaches better results than the traditional method using Hough transform. This is obvious from the values e_c and e_r that are approximately three times better in comparison with the implementation of the iris localization based on Hough transform. Moreover, the false detection ratio has also been improved in our method.

Table 1. Results comparing the described method with approach using Hough transform

Selected method	e_a [%]	e_c [px]	e_r [px]	e_c [%]	e_r [%]
Our method	90.9	7.8 ±3.8	8.7 ±2.7	1.02	1.13
Hough transform	86.3	25.6 ±21.8	24.2 ±11.0	3.33	3.15

4 Conclusion

We have introduced a new algorithm for localizing iris that allows its extraction from the close-up eye images. It is possible to use this localization method in numerous applications, including personal identification. The effectiveness of the algorithm has been evaluated on a set of real life images that have been taken by standard equipment without infra-red light. We have compared implementation of the presented algorithm with the traditional method that is using the Hough transform. Achieved results are shown in Table 1 and are based on indicators defined in Equations (4) and (5). Measured error rate is in the case of our method approximately three times lower in comparison to the method that is using Hough transform.

5 Publications

Presented work was published in proceedings of the 2nd International Conference on Image Processing Theory in Paris (to appear on ISI WoS) [3] and extended version is accepted for publication in International Journal of Signal and Imaging Systems Engineering (to appear on Scopus) [4].

References

1. J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:21–30, 2002.
2. M. Dobeš and L. Machala. Iris database, 2003.
3. T. Fabián, J. Gaura, and P. Kotas. An algorithm for iris extraction. In *2nd International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 464–468, Paris, France, July 2010.
4. T. Fabián, J. Gaura, and P. Kotas. Iris localization and extraction algorithm from an eye image. *International Journal of Signal and Imaging Systems Engineering (IJSISE)*, 2011.
5. H. B. R. Ghassan J. Mohammed and A. A. Al-Kazzaz. A new localization algorithm for iris recognition. *Information Technology Journal*, 8(2):226–230, 2009.
6. J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, August 2000.
7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. New York, NY, USA, 2007.
8. R. Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, Sept. 1997.

Understanding the Context of Statements in Transparent Intensional Logic

Adam Farnik

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
adam.farnik@vsb.cz

Abstract. This paper concerns with intensional and extensional interpretation of statements (de dicto and de re) as taken in Transparent Intensional Logic (TIL). It provides a brief introduction to the language of TIL, its concept of typing and the way of formalizing statements. The whole described concept is illustrated on the example. It provides a basic description of TIL's anti-contextualism.

1 Introduction

Some statements in natural language can be understood in more than one way. This had obviously been causing problems in formulations of such statements in first order logics. Languages of intensional logics allow to distinct between these interpretational differences. This paper briefly describes the way, in which Transparent Intensional Logic (TIL) – a branch of intensional logic, formulates different interpretations.

2 Interpretation differences

In first order predicate logic, all mentioned objects are understood "extensionally", i.e. as individuals of universe. First order logic understands the statement

"John supports the president of USA." **(1)**

as if John supports the person (particular individual), who is currently holding the presidential function in USA. In natural language however, one could understand this statement as if John would support any person holding the presidential function in USA during their term of office. In other words, language of first order logic can not express a statement referring to a role – it always refers to the individual occupying the role.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 343–346.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2.1 De dicto and de re

In order to distinguish between the two aforementioned interpretations of natural language statements, phrases *de dicto* (“of the word”) and *de re* (“of the thing”) are used.

Saying that a statement is meant *de re* signifies, that it should be interpreted extensionally. In case of the example (1), the *de re* interpretation says that John supports the individual, who is currently occupying the presidential function in USA. On the contrary, the *de dicto* interpretation refers to the presidential function (Whoever would he be, if he’s a president of USA, John will support him.).

3 Language of Transparent Intensional Logic

The language of TIL is based on the one of lambda calculus developed by Alonzo Church and as such it is based on functions. In TIL, however, the functions are partial, meaning that function can fail during computation i.e. not provide any output value.

Following sub-chapters briefly describe the specific language of TIL.

3.1 Type basis

Type basis is a set of atomic types used in the language. The standard basis of TIL language consists of four types:

o a set of truth-values (true, false)

ι a set of individuals

τ a set of real numbers, also used as a set of instant points in time

ω a set of possible worlds

The types of the basis can be used to define types of the first order above the basis i.e. types of partial functions

$$(\alpha\beta_1 ... \beta_n) \tag{2}$$

where α and β are types of basis and while α defines the function’s return value type, $\beta_1 ... \beta_n$ are the types of function’s arguments. This typing convention can be used recursively, creating types of second and higher orders.

3.2 Constructions

Constructions are the abstract procedures defining the meaning of statements. TIL language inherits two basic constructions from lambda calculus – abstraction (which creates function and is marked as λ) and application (or composition; it applies the arguments to the function, computing the outcome).

TIL language also defines such construction as variable, which type is defined above the basis as described in previous sub-chapter. Variables can be *trivialized* (i.e. simply mentioned; marked as ^{0}), *valuated* (i.e. executed, so that the valuation of the variable mentions the object constructed by the variable; this execution is called *v*-

constructing, where v is a parameter of valuation) and *double executed* (i.e. the resultant object of valuation is valuated again under the same v parameter, marked as ^{2}).

3.3 Object typing

When formalizing a statement, the first step is to define the objects and variables used, and their types. Following the language convention, we write

$$X/\alpha \tag{3}$$

when the object X is of type α and

$$Y \to \beta \tag{4}$$

when the valuation of Y constructs the object of type β. Saying that object is of type $\alpha_{\tau\omega}$ is equivalent notation to $(\alpha\tau)\omega$ and variables $t \to \tau$ (constructing times) and $w \to \omega$ (constructing possible worlds) are commonly used implicitly.

4 Formalization

On the basis of the brief description of the TIL language in the previous chapter, it is now possible to understand difference in formalization of the example statement (1) reading it *de dicto* and *de re*.

4.1 Interpretation de re

The typing of the objects used is as follows:

$$\begin{aligned} &John/\iota \\ &PresidentOfUSA/\iota_{\tau\omega} \\ &Supports/(o\iota\iota)_{\tau\omega} \end{aligned} \tag{5}$$

Having the objects typed, one can now proceed to defining the construction of the statement:

$$\lambda w \lambda t\ [{}^0Supports_{wt}\ {}^0John\ {}^0PresidentOfUSA_{wt}] \tag{6}$$

Reading this construction one can see, that it abstracts from the specific state of the world and time, so that it can be used in any situation. Following, there is a composition of arguments ${}^0John$ and ${}^0PresidentOfUSA_{wt}$ on the function ${}^0Supports_{wt}$. Both arguments and function are only mentioned (trivialized) – the execution happens on the level of the composition.

To both *PresidentOfUSA* and *Supports* objects, there is the same time and state of the world applied, so that on the level of composition $PresidentOfUSA_{wt}$ we know who the president is (we can point on the particular individual), just as we know whether John supports him in the same time and state of the world on the level of the outer composition (marked by square brackets).

4.2 Interpretation de dicto

The object typing differs in this case:

$$John/\iota \qquad PresidentOfUSA/\iota_{\tau\omega} \qquad Supports/(o\iota\iota_{\tau\omega})_{\tau\omega} \tag{7}$$

And so does the construction:

$$\lambda w\lambda t\ [{}^0Supports_{wt}\ {}^0John\ {}^0PresidentOfUSA] \tag{8}$$

Note, that the object *PresidentOfUSA* did not change in type, but it is not a part of composition applying specific time and state of the world to it. On the contrary, the object *Supports* have changed its type, and while in the *de re* interpretation it was person-to-person relation, in this case it is person-to-post relation.

5 Conclusion

The concept of *de dicto* and *de re* interpretations shown on the example in this paper is illustrating the anti-contextualism typical for TIL. The paper shows how by the use of typed objects, TIL language can distinguish between the meaning of statements, which in natural language can be understood in different ways depending on the context.

References

1. Duží, M.: Propoziční a pojmové postoje. In knize. Ed. Svoboda, V., Praha:Filosofia (2010) 149-191
2. Duží, M., Jespersen, B., Materna, P.: Procedural Semantics for Hyperintensional Logic (Foundations and Applications of TIL). Springer (2010)

Watershed Inspired Heat Flooding Segmentation

Radovan Fusek and Karel Mozdřeň

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{radovan.fusek.st, karel.mozdren}@vsb.cz

Abstract. In this paper, we introduce a new image segmentation approach based on the diffusion (heat flooding) equation inspired by watershed transformation. Flooding by water from the watershed segmentation is replaced by flooding of the heat. The main improvement of this method is that we are using a local minima as a starting point for the heat distribution through the image. It also takes the size of segments in account and that leads to further reduction of over-segmentation which is a main problem of watershed segmentation. In this paper we present our method, and show some segmentation results.

1 Introduction

Image segmentation is an important task of computer vision. Many algorithms were developed and one of them is watershed transformation [11]. The water is flooding the image from the local minima and a segmentation lines are created at places where two floods meet. Many modification exist [7, 1, 2, 4, 9, 6, 5, 3, 10]. The main problem for this method is over-segmentation. This happens when the segmentation has too many segments. These are separate even when they could form a bigger one. There are many methods that solve this problem. Some are based on postprocessing and some use preprocessing. In our method we use a diffusion [8] as an inspiration of the watershed transformation for the segmentation. Our paper has follow structure: First we describe our method in a detail. Then we perform some experiments and discuss the results.

2 Algorithm

First of all we need to define a set of markers in the image. Those markers will indicate the location of initial seeds for the heat flooding. Markers can be set manually or automatically. In the first case, the seed is manually inserted into each object of interest. This is not very comfortable because it requires a manual control. And this approach can be difficult for the images with many objects of interest. We are looked for inspiration in watershed transformation as a method for automatic marking. In our case those markers (local minima from watershed) are used as the kernel of heat flooding and it also defines the energy of the segment. During the process of heat transfer (flooding) smaller segment

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 347–352.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

(with less energy) are absorbed by the larger ones. This causes reduction of over-segmentation. The diffusion is used on each marker separately. This creates diffusion map for each marker which defines the get area of segments. After this we get final segmentation of the image. In the following subsections we describe all presented steps in detail.

2.1 Watershed and markers

The Watershed Transformation (WT) is well known image segmentation method. Basically the principle of segments creation is based on geographical grounds. We can imagine the image as a topological region which is flooded by water from a local minima. When two flooded basins are connected the dam is build (segment border). This means that we start flooding the image from the darkest parts of the image to the most bright ones. But all WT methods have the same property. When we want WT to respect the edges in the image, we must apply it on the gradient image instead of on the original image. As it has been said before finding the local minima is crucial part of the image segmentation, because it determines how many segments there would be in the final segmentation. For automatic markers detection the WT is used. Every time new basin emerges, it is marked as marker. It is formed from all neighbour pixels of the same value as the first minimal pixel found. Example of the local minima in the image could be seen in Figure (Fig. 1).

30	28	28	28	28	28	28	28	28	29
30	20	20	20	20	20	20	20	20	30
30	12	18	16	20	22	22	22	22	30
30	10	18	16	22	15	9	8	9	30
29	9	5	5	15	13	10	11	15	29
20	9	5	5	15	14	11	14	17	20
23	9	8	10	15	14	15	15	18	23
30	9	9	10	15	14	15	15	18	30
30	22	15	15	15	15	15	15	18	18
29	28	28	28	28	28	28	28	28	30

Fig. 1. Example of local minima in the image. Isolated single pixel with the value 8 and a coupled pixels with the value 5

2.2 Heat flooding

Heat transfer is based on diffusion equation which is already used in some segmentation and filtration algorithms with good results. In our case we have developed a method which is inspired by WS, but instead of flooding the regions with water, we flood regions with heat. The main idea is based on the thought that heat can spread over the image better than water, because it counts in the whole image for computation. We do not have to build dams to separate segments, so the only thing that can slowdown the heat spread is gradient of the image, which is alternative to the thermal insulator in real-live. The example of heat transfer in the image with and without the heat insulators is show in Figures (Fig. 2).

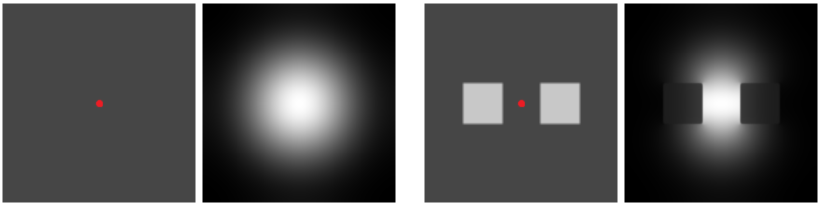


Fig. 2. Transfer of the heat in the image with and without heat insulators

As it has been mentioned before, the size of marker determines how much energy will the corresponding segment have. Each marker is processed separately and output for flooding of each marker is then combined to get the final segmentation. When we have all heat sources identified, it is easy to apply the flooding on them and say which part of the image belong to which segment (heat source). The pixel then belongs to that heat source which changed its temperature the most (the most heat transfered from according heat source). This is done as follows (Alg. 1).

Algorithm 1 Heat Flooding

1: $M \leftarrow Markers\ from\ WS$
2: $i_n \leftarrow number\ of\ iterations$
3: $indexes \leftarrow matrix\ for\ indexes$
4: **for** $m \in M$ **do**
5:     create $matrix$
6:     copy $m \rightarrow matrix$
7:     $i \leftarrow 0$
8:     **repeat**
9:         one diffusion iteration on $matrix$
10:     **until** $i < i_n$
11:     z-buffer on $matrix \rightarrow indexes$
12: **end for**

As it can be seen in step 11 of our algorithm we use z-buffer. It helps us determine the segment pixel belongs to. A new value is written to the z-buffer only if the new value is bigger than a old one. And also on the corresponding place in the *indexes* matrix the index of the identified heat source from which the most heat was transfered is set. With this approach it is certain that in the end of this algorithm there would be index numbers in *indexes* matrix from those heat sources which transfered to the corresponding element the most heat.

3 Experiments

The basic principle of our method is based on the fact that heat flooding process decreases the number of segments. In other words, bigger segments with higher energy are consuming smaller segments with a little energy. Time (numbers of iterations) is an important factor of the calculation because it affects the number of final segments. This effect is also visible on real images (Fig. 3) and (Fig. 4). Over-segmentation is decreasing with is increasing the number of iterations.

orig. 50 500 3500

Fig. 3. Example segmentations. The number of iterations is under the images.

4 Conclusion

We have proposed a method for efficient image segmentation inspired by watershed and heat flooding. In the first step of our segmentation method we find local minima in the image. Resulting minima are used as markers and then as seed point for heat flooding in the image. We have found out that the longer we let the heat flood the image the better the segmentation is. There are less segments because the bigger and more important segments consume smaller less important segments so it reduces over-segmentation. We experimented on real images and showed our results.

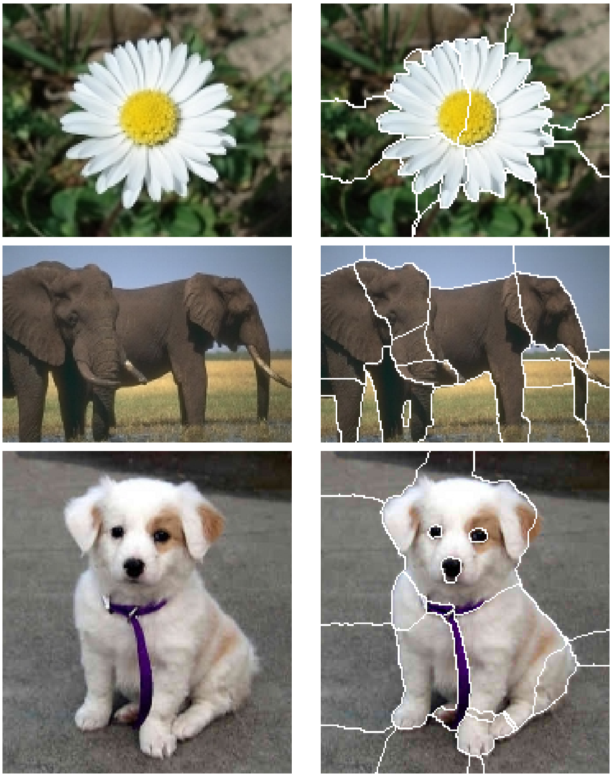


Fig. 4. Example segmentations of real images

5 Acknowledgments

This paper was carried out under SGS 2011 - SP2011/163.

References

1. S. Beucher and C. Lantuejoul. Use of Watersheds in Contour Detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, Rennes, France.*, September 1979.
2. S. Beucher and Centre De Morphologie Mathmatique. The watershed transformation applied to image segmentation. In *Scanning Microscopy International*, pages 299–314, 1991.
3. Andreas Bieniek and Alina N. Moga. An efficient watershed algorithm based on connected components. *Pattern Recognition*, pages 907–916, 2000.
4. André Bleau and L. Joshua Leon. Watershed-based segmentation and region merging. *Comput. Vis. Image Underst.*, 77:317–370, January 2000.
5. Shao-Yi Chien, Yu-Wen Huang, and Liang-Gee Chen. Predictive watershed: a fast watershed algorithm for video segmentation. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(5):453 – 461, may 2003.

6. Shao-Yi Chien, Yu-Wen Huang, Shyh-Yih Ma, and Liang-Gee Chen. Predictive watershed for image sequences segmentation. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 3, pages III–3196 –III–3199, may 2002.
7. Víctor Osma-Ruiz, Juan I. Godino-Llorente, Nicolás Sáenz-Lechón, and Pedro Gómez-Vilda. An improved watershed algorithm based on efficient computation of shortest paths. *Pattern Recogn.*, 40:1078–1090, March 2007.
8. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12:629–639, July 1990.
9. C. Rambabu and I. Chakrabarti. An efficient immersion-based watershed transform method and its prototype architecture. *J. Syst. Archit.*, 53:210–226, April 2007.
10. C. Rambabu, T.S. Rathore, and I. Chakrabarti. A new watershed algorithm based on hillclimbing technique for image segmentation. In *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*, volume 4, pages 1404 – 1408 Vol.4, oct. 2003.
11. L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, June 1991.

Diffusion Spectral Clustering for Image Segmentation

Jan Gaura and Eduard Sojka

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jan.gaura, eduard.sojka}@vsb.cz

Abstract. In this paper, we present a new image segmentation method coming from the idea of spectral clustering. Contrary to the known algorithms in this area, it uses the process that we call diffusion spectral clustering in which the standard spectral clustering based on the spectral decomposition of the Laplacian matrix is improved by incorporating the diffusion process. For clustering itself, the mean shift algorithm is used instead of the k-means algorithm that is usually mentioned in this context. Both the diffusion process as well the as the mean shift algorithm contribute to the good properties of the new method. In the paper, the needed theory and the results of experiments are presented.

1 Introduction

Image segmentation is a process of finding the points (pixels) of image that are lying close to each other and have a similar brightness or colour. Many image segmentation algorithms exist that are based on various principles. The work by Ng *et al.* [1] may be taken as an example of standard spectral clustering. For each image pixel, a vector of new coordinates is computed using the spectral decomposition of the Laplacian matrix of image. According to their new coordinates, the pixels are then clustered using the k-means algorithm. The problem is that the images often contain unnecessary data, e.g., noise that have to be removed before the image segmentation step can be applied. The needed image preprocessing is usually carried out by Gaussian filtering. Gaussian filtering, however, has a disadvantage that it does not preserve the edges in images. This drawback can be overcome by making use of diffusion filtration [6] that should be generally done before the decomposition.

In this paper, we propose a new method carrying out image segmentation using spectral clustering and also incorporating diffusion filtering in the same step. Moreover, in contrast to the known methods, we do not use the k-means algorithm for the clustering itself. The main drawback seems to be that the number of segments must be known *a priori* in this method, which can be regarded as problematic in the given context. Instead, we use a different clustering algorithm called mean-shift [5] that is based on kernel density estimation and does not require the *a priori* knowledge of the number of clusters.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 353–358.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

The paper is organized as follows. In Sections 2 and 3, the fundamentals of spectral clustering and diffusion filtration are summarised for convenience of the reader. Our original contribution, diffusion spectral clustering with mean-shift, is described in Section 4. The experimental results are presented in Section 5.

2 Spectral Clustering in Image Segmentation

In the discrete case, the image is represented as a mesh of vertices, which is a graph. The vertices in the graph correspond to the pixels in the image. Each inner vertex (pixel) is connected to its neighbours by edges. The weight of each edge is chosen so that it decreases with increasing local contrast (difference between the brightness or colour of pixels). Let $\mathbf{L}$ be the Laplacian matrix of image (i.e., the mesh representing the image).

The value of the expression $\boldsymbol{x}^\top \mathbf{L} \boldsymbol{x}$, where $\boldsymbol{x}$ is a vector, is interesting. Considering the shape of the Laplacian matrix, the following equation can be obtained

$$\boldsymbol{x}^\top \mathbf{L} \boldsymbol{x} = \sum_{e_{ij}} w_{ij} (x_i - x_j)^2, \tag{1}$$

where the sum on the right side is over all edges e_{ij}; the value w_{ij} is the weight of the edge e_{ij}. The interesting case is if the value of $\boldsymbol{x}^\top \mathbf{L} \boldsymbol{x}$ is minimal, which can be understood that we are searching for a vector $\boldsymbol{x}$ of new coordinates such that the points connected by an edge with a high value of weight (low contrast) will have similar coordinates. Uninteresting trivial solution $\boldsymbol{x} = \mathbf{0}$ leading to $\boldsymbol{x}^\top \mathbf{L} \boldsymbol{x} = 0$ should be excluded, which can be done by introducing the additional constraint $\boldsymbol{x}^\top \boldsymbol{x} = 1$. The problem of minimization with a constraint can be solved using the Lagrange multipliers, which requires to minimize the expression

$$\boldsymbol{x}^\top \mathbf{L} \boldsymbol{x} + \lambda (\boldsymbol{x}^\top \boldsymbol{x} - 1). \tag{2}$$

Using the usual evaluation of the gradient and taking into account the symmetry of $\mathbf{L}$, we can simply obtain the following condition

$$\nabla \boldsymbol{x} = 2(\mathbf{L} \boldsymbol{x} - \lambda \boldsymbol{x}) = \mathbf{0}. \tag{3}$$

As can be seen, the minimum is achieved if the sought vector $\boldsymbol{x}$ is an eigenvector of the matrix $\mathbf{L}$. Denote by $\boldsymbol{u}_k$ and λ_k the k-th eigenvector and, respectively, the k-th eigenvalue. Considering the condition of ortonormality of eigenvectors, it can be easily seen that $\boldsymbol{u}_k^\top \mathbf{L} \boldsymbol{u}_k = \lambda_k$. It also follows that the solution is $\boldsymbol{x} = \boldsymbol{u}_2$. We point out that the smallest eigenvalue is $\lambda_1 = 0$, which gives $\boldsymbol{u}_1^\top \mathbf{L} \boldsymbol{u}_1 = 0$, but the corresponding eigenvector is $\boldsymbol{u}_1 = (1, 1, \dots, 1)$, which is uninteresting. It can be concluded that by this procedure, a new coordinate has been assigned to each node (pixel) such that the edges that are connected by edges with a high value of weight have this coordinate similar to each other, which follows from the right side of Eq. (1).

Similarly, we can also proceed in the case if more that one new coordinate should be assigned. In this case, we are searching for a matrix $\mathbf{X} =$

$[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_q]$ such that $\mathbf{X}^\top \mathbf{L} \mathbf{X}$ is minimal if $\mathbf{X}^\top \mathbf{X} = \mathbf{I}$ holds. Here, the property is required that the nodes that are connected by the edges with a high value of weight will have similar values of the newly introduced coordinates.

The idea of using spectral clustering for image segmentation is based on what was already mentioned. The segments in the images are considered to be the areas with approximately the same brightness or colour. The values of weights of the edges connecting the pixels in one area are thus high. The segmentation can be done as follows: (1) Construct the Laplacian matrix of image and compute its eigenvalues and eigenvectors. (2) Take the chosen number of eigenvectors corresponding to a chosen number of the lowest eigenvalues. This eigenvectors define new coordinates of nodes (pixels) in image. (3) Carry out the clustering on the new coordinates introduced in the previous step using the standard methods, for example, k-means. Each of the clusters is then declared to be a segment of image. The above described method of spectral clustering is known [1].

3 Diffusion Equation in Image Filtering

The drawback of the method mentioned in the previous paragraph is that, for a successful segmentation, the pre-filtering of image is usually necessary to get rid of unwanted details that interfere with the results of segmentation. One of the methods that can be used for filtering is a method based on the analogy to diffusion. Diffusion is the process during which a substance diffuses from the places of its greater concentration to the places where the concentration is lower. In images, the concentration is a value that corresponds to brightness. The brightness of original image describes the initial concentration. During the diffusion process, the concentration is changing, which can be calculated. The concentration at any time can, in turn, be considered as brightness.

Given a compact Riemannian manifold M. Let $f(\boldsymbol{x}, t)$ be the amount of the substance that undergoes the process of diffusion, e.g., heat at a point $\boldsymbol{x}$ of M and at a time t. Moreover, some initial distribution $f(\boldsymbol{x}, t = 0)$ is known. The $f(\boldsymbol{x}, t)$ satisfies the diffusion equation $\partial f(\boldsymbol{x}, t)/\partial t = -\Delta_M f(\boldsymbol{x}, t)$, where Δ_M stands for the Laplace-Beltrami operator of M. For the discrete case (the manifold is approximated by a mesh), the equation can be rewritten as

$$\left(\frac{\partial}{\partial t} + \mathbf{L} \right) \boldsymbol{f}(t) = 0, \tag{4}$$

where $\mathbf{L}$ is a Laplacian matrix of the mesh and the heat-amount vector $\boldsymbol{f}(t) = (f_1(t), \ldots, f_n(t))^\top$ is indexed by the nodes of the mesh. The solution can be written in the form of

$$\boldsymbol{f}(t) = \mathbf{H}(t) \boldsymbol{f}(0), \tag{5}$$

where $\mathbf{H}(t)$ is a diffusion operator (a matrix). The entry $h_t(p, q)$ of $\mathbf{H}(t)$ expresses the amount of heat that is transported, by the diffusion process, from the q-th node into the p-th node (or vice versa since $h_t(p, q) = h_t(q, p)$) during the time interval $[0, t]$. It can be easily shown that the following is required for $h_t(p, q)$ to satisfy Eq. (5)

$$\mathbf{H}(t) = e^{-t\mathbf{L}}, \tag{6}$$

where

$$e^{-t\mathbf{L}} = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} \mathbf{L}^k. \tag{7}$$

4 Diffusion Spectral Clustering in Image Segmentation

In this section, the main result is presented, which is a diffusion spectral clustering. For clustering itself, the mean-shift algorithm is used, which is a method based on kernel density estimation.

We can think of spectral clustering as a combination of filtration and clustering. Assume that we have found the eigenvectors of $\mathbf{L}$. The matrix then can be rewritten as follows

$$\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \sum_{k=1}^{n} \lambda_k \boldsymbol{u}_k \boldsymbol{u}_k^\top. \tag{8}$$

Exploiting Eqs. (5), (8), we obtain

$$\mathbf{H}(t) = e^{-t\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top} = \mathbf{U}e^{-t\boldsymbol{\Lambda}}\mathbf{U}^\top. \tag{9}$$

The elements of $\mathbf{H}(t)$ are

$$h_t(p, q) = \sum_{k=1}^{n} e^{-t\lambda_k} u_{pk} u_{qk} = \sum_{k=1}^{n} \left(e^{-\frac{1}{2}t\lambda_k} u_{pk}\right) \left(e^{-\frac{1}{2}t\lambda_k} u_{qk}\right), \tag{10}$$

where u_{pk} stands for the p-th entry of the k-th eigenvector of $\mathbf{L}$ and λ_k is the k-th eigenvalue (the smallest eigenvalue of $\mathbf{L}$ is $\lambda_1 = 0$). For the i-th mesh node, the following vector can be created

$$\Phi_t(i) = \left(e^{-\frac{1}{2}t\lambda_1} u_{i1}, e^{-\frac{1}{2}t\lambda_2} u_{i2}, \ldots, e^{-\frac{1}{2}t\lambda_n} u_{in}\right)^\top, \tag{11}$$

which may be regarded as coordinates of the node in a new space. It can be easily seen that h_t can be written as a dot product $h_t(p, q) = \langle \Phi_t(p), \Phi_t(q) \rangle$. We note that t is a parameter that cannot be excluded; it determines the amount of filtration.

Diffusion spectral clustering is a process of clustering in the space that was newly introduced in Eq. (11). In contrast with the ordinary spectral clustering, the coordinates contain the term $e^{-\frac{1}{2}t\lambda_k}$, which is important. As was verified by the experiments, the parameter t influences the properties of the final result substantially.

For carrying out the clustering itself, we use mean-shift algorithm, which is a density based clustering algorithm. The mean-shift procedure is based on finding the local maxima of kernel density estimate. Usually, a certain class of radially symmetric kernels is used, such as, Epanechnikov, unit step, or Gaussian kernel. The main advantage over the k-means algorithm is that mean-shift does not require *a priori* knowledge of number of clusters (number of image segments).

5 Experiments

In this section, we present an evaluation of diffusion spectral decomposition used in conjunction with the mean-shift clustering algorithm using the Gaussian kernel. (i.e., diffusion spectral decomposition is responsible for computing the new coordinates of points that are clustered afterwards; clustering itself is carried out by making use of the mean-shift algorithm.)

Real-life images from The Berkeley Segmentation Dataset and Benchmark [4] have been used. We present a comparison with the hand made segmentation presented in [4] as well as k-means spectral clustering segmentation.

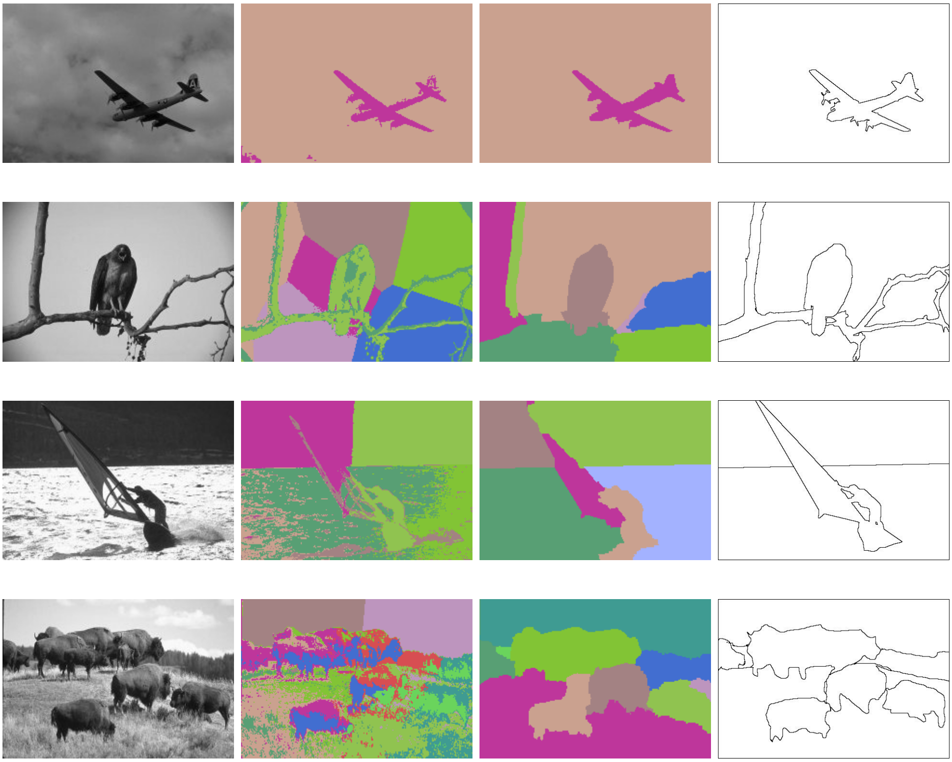


Fig. 1: Various images segmented using mean-shift algorithm. From left to right: source image, result of k-means spectral clustering segmentation, result of diffusion spectral clustering using the mean-shift algorithm, segmentation provided from the dataset

6 Conclusion

We have proposed a new method for segmenting images using diffusion spectral clustering. Our method combines image filtering and spectral clustering into one unified approach. Moreover, for clustering itself, the mean shift algorithm is used instead of the k-means algorithm that is usually mentioned in this context. Diffusion spectral clustering takes the pixels from the image and assigns them new coordinates that are based on their properties and on the degree of filtering that is required. In this way, the filtration step is entirely embedded into the spectral clustering by adding additional information to the new coordinates. The new coordinates are then used in a clustering method, mean-shift in our case that has been chosen for its good properties. We also present experimental results confirming a good performance of the new method.

Acknowledgements. This work was partially supported by the grant SP 2011/163 of VŠB - Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science.

Related papers. This paper is a new original work loosely based on [2], [3].

References

1. Ng, A., Jordan, M., and Weiss, Y.: On Spectral Clustering: Analysis and an Algorithm. Advances in Neural Information Processing Systems 14, 2002, pp. 849-856
2. Gaura, J., Sojka, E., Krumnikl, M.: Image Segmentation Based on Electrical Proximity in a Resistor-Capacitor Network. In Proc. of Advanced Concepts for Intelligent Vision Systems, 2011, pp. 216-227
3. Gaura, J., Sojka, E., Krumnikl, M.: Image Segmentation Based on k-Means Clustering and Energy-Transfer Proximity. To appear in 7th Internatinal Symposium on Visual Computing, 2011
4. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics, In Proc. of 8th International Conference of Computer Vision, 2001, pp. 416-423
5. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 2002, pp. 1-18
6. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (7), 1990, pp. 629-639

Extension of Method for Fuzzy Rules Extraction by Means of Artificial Neural Networks

Jan Górecki

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.gorecki@vsb.cz

Abstract. Knowledge extraction from data in the form of rules is a widespread direction in data mining area which allows to obtain interesting relationships in data from large databases in for a human easily understandable form. In this article I deal with one of the methods for extraction of rules from data which extract rules in form of a formula in considered fuzzy logic by means of artificial neural networks with very special architecture. Using artificial neural networks in extraction process, above mentioned methods gain good approximation of the analyzed data and also thanks to special architecture allows to extract well human-understandable knowledge. In the method described in this paper was, however, missing any module that is a standard part of the most of methods used for rules extraction from data, that would allow to the user subjective selection of the best ratio between accuracy and comprehensibility of the model, which is especially important feature for solving data mining tasks called *searching of concepts descriptions*, which methods allow the user to get a good insight and understanding of the analyzed data. Thus, the main purpose of this article is a design of such a module inspired by a similar module in methods for extraction of the so-called *decision trees* and subsequent illustrations and evaluations of the results obtained by the original method, extended by a new module on sample dataset.

Keywords: knowledge extraction from data, comprehensibility and accuracy of data models, artificial neural network, fuzzy logic, disjunctive normal form

1 Introduction

Extraction of logical formulas (so-called *rules*) from data is in the last twenty years one of the most important direction in data mining, i.e. in the IT field which deals with extraction of understandable, valid and potentially interesting non-trivial dependences in the data, also called as *knowledge*. For knowledge discovery from data in the form of rules there were developed many techniques based on quite different approaches and theoretical principles - there are methods using different types of logic (Boolean, fuzzy logic or observational), methods

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 359–364.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

are based on various mathematical and statistical approaches or based on different methods from field of machine learning (an overview and comparison of several methods can be found e.g. in [9]). One very important group of methods are methods for extracting rules using framework of Artificial Neural Networks (ANN's). However, methods based on commonly used types of artificial neural networks (e.g. feed-forward Multi-Layer Perceptrons (MLP's)) have one big disadvantage - though models obtained by these methods have very good approximation properties, they are not easily human-comprehensible (in terms of [1] models provide a high *data fit* but a low *mental fit*). It is the difficult comprehensibility of such models that motivated research into the problem of extracting from them more easily comprehensible sets of logical rules [7]. One way of solving the problem is transformation of the ANN architecture and of parameters determining the computed mapping into a set of logical rules of a prescribed kind, e.g., into a set of appropriate implications or equivalences [3, 4]. Another way to solve the problem is to use ANN's with special architectures which correspond to the form of a rules. These special types of ANN's don't have as good approximation properties as more general types, but for this price gain good comprehensibility of extracted model, which is in form of rules (see e.g. ANFIS [10] or NEFCLASS [11] methods).

However, these latter mentioned ANN-based fuzzy rule extraction methods share important feature, which is that most of them rely primarily on heuristics, and their underlying theoretical principles are not very deep. Most of them allow to interpret different logical connectives by means of algebraic operations that do not belong to the same algebra. Typically, the conjunction is interpreted in a G-algebra, but the negation in a MV-algebra, due to the fact that the interpretation of the negation in a MV-algebra is involutive, whereas in a G-algebra, it is not. Consequently, the extracted rules then can not be correctly evaluated in any single fuzzy logic model. The solution of this problem is outlined in [8], where is showed, that such a mixing of different interpretations is not needed and a method for the extraction of fuzzy rules in a disjunctive normal form (DNF) which can be evaluated in a single fuzzy logic model is proposed.

The implementation of this method created at the Institute of Computer Science in Prague I obtained from the author for testing purposes. This implementation has already been used in several real applications (e.g. data from EEG spectral analysis in neurophysiology [8, 5]) and also was used for comparison with representatives of other types of methods for extracting rules from the data based on several measures of ruleset quality (see [9]). When using this implementation, however, I missed a module that would allowed me to set an subjectively appropriate balance between accuracy and comprehensibility of the final rule (more about this in chapter 2), which such a module is a standard part of other methods used for rules extraction from the data (e.g. [2]). Design of such a module, illustration of its functionality and evaluation of its results on sample data is a major purpose of this article. In the next section there will be very briefly outlined basics of original method (for more details see [8, 7]) and due to space limitation also brief description of module for rule simplification which allows to the user of original method select appropriate balance between accuracy and comprehensibility of results given by original method.

2 Original method outline and its extension

Original method extract DNF rules in form $\Psi \equiv \bigvee_{i=1}^{d} \bigwedge_{j \in C_i} \varphi_{i,j}$, where $C_1, ..., C_d$ are non-empty sets, $\Psi,\ \varphi_{i,j},\ i = 1, ..., d,\ j \in C_i$ are atomic formulas and $\vee$,$\wedge$ are connected via *De Morgan law*: $\alpha \vee \beta = \neg(\neg\alpha \wedge \neg\beta)$.

To compute truth value of formula all connectives have to be interpreted in same algebra (e.g. MV-algebra, G-algebra, PŁ-algebra). All atomic formulas $\varphi_{i,j}$ are interpreted by the same kind of finitely-parameterizable fuzzy sets on crisp domain $\mathbb{R}$ (i.e. $\tilde{\varphi}_{i,j} \in \mathcal{F}(\mathbb{R})$, where $\mathcal{F}(\mathbb{R})$ denotes the set of fuzzy sets on $\mathbb{R}$ and $\tilde{\varphi}_{i,j}$ denotes interpretation of formula $\varphi_{i,j}$ in some parametrization model). In implementation of original method user choose from nearly dozen of the parameterizations, e.g. $\mathcal{F}(\mathbb{R}) = \Gamma_{(a,b)}$, where $a \in \mathbb{R}, b > 0$ and $\Gamma_{(a,b)}(x) = e^{-\frac{(x-a)^2}{2b}}$ for $\forall x \in \mathbb{R}$. Extracted rule than can have the following form:

Class is Virginica $\equiv$
(SepalLength is $\Gamma_{(5.8101,491.5838)}$ $\wedge$ SepalWidth is $\Gamma_{(2.4312,9.0511)}$ $\wedge$
PetalLength is $\Gamma_{(6.3667,2.0397)}$ $\wedge$ PetalWidth is $\Gamma_{(2.1893,0.97068)}$)
$\vee$ (SepalLength is $\Gamma_{(6.2354,0.75672)}$ $\wedge$ SepalWidth is $\Gamma_{(2.8615,0.57805)}$ $\wedge$
PetalLength is $\Gamma_{(5.1952,1.6948)}$ $\wedge$ PetalWidth is $\Gamma_{(2.1865,0.49648)}$)
$\vee$ (SepalLength is $\Gamma_{(5.7062,0.77931)}$ $\wedge$ SepalWidth is $\Gamma_{(2.2046,0.59679)}$ $\wedge$
PetalLength is $\Gamma_{(5.7177,1.1318)}$ $\wedge$ PetalWidth is $\Gamma_{(1.2866,0.9082)}$)

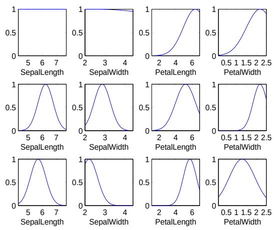


Fig. 1. Visualization of rules extracted from data Fisher Iris (4 input real-valued attributes and one classification attribute (with values $\in$ {Setosa, Versicolor, Virginica}) for Ψ = (Class is Virginica), and for parametrization $\tilde{\varphi}_{i,j}$ of type $\Gamma_{(a,b)}$ and for $d = 3$.

Visualization of this rule can be seen in Fig. 1. Individual graphs corresponds to interpretation of atomic formulas $\varphi_{i,j}$. In i-th row there are always depicted interpretations of formulas $\varphi_{i,1}, \varphi_{i,2}, \varphi_{i,3}, \varphi_{i,4}$, which are for each i connected via conjunction, individual rows are then connected via disjunction.

However, when examining the extracted rules I realized that most of the atomic formulas in the final rules are redundant, i.e. the inaccuracy of the model while omitting these formulas does not change significantly. Thus arose the question how to remove these redundant formulas.

To solve this problem I got inspired from other method for knowledge extraction from data, namely the method for Classification And Regression Trees induction (CART) [2]. In this method redundant nodes in induced tree are removed in a way, that in every step of algorithm there is removed node, which does affect the tree accuracy the least. I used this approach to form (as I call it) *greedy* algorithm, where in every step (similarly to CART approach) there is removed atomic formula $\varphi_{i,j}$ which has the smallest impact on rule accuracy (i.e. the accuracy measured with Mean Absolute Error (MAE) on training data increases the least possible). To formalize this process, I defined binary matrix $\mathbb{C} \in \{0,1\}^{d,n}$ which expresses presence of atomic formula $\varphi_{i,j}$ in extracted rule as $(\forall i = 1, ..., d)$ $(\forall j = 1, ..., n)$ $(j \in C_i \equiv \mathbb{C}_{i,j} = 1)$.

So, with this definition the problem is transformed to find in every step of algorithm *the best* matrix $\tilde{\mathbb{C}}$ (measured by MAE)[1]. The output of this greedy algorithm was sequence of $dn+1$ matrices $\tilde{\mathbb{C}}_1, ..., \tilde{\mathbb{C}}_{dn+1}$ with decreasing number of ones, where matrix $\tilde{\mathbb{C}}_1$ contained dn ones and matrix $\tilde{\mathbb{C}}_{dn+1}$ contained only zeros. The evolution of MAE in this sequence on sample data can be seen on the left part in Fig. 2.

After design and testing of this algorithm there arose the question, if I can not get smaller increase of MAE in case, that there will be removed in every step of the algorithm more then one atomic formula. To get answer for this question I designed second algorithm (which I called *localsearch* algorithm), where in every step there was removed one *or more* atomic formulas at once. Evolution of MAE for sequence of the best matrices $\tilde{\mathbb{C}}_1, ..., \tilde{\mathbb{C}}_k$, where $k \leq dn+1$ founded for the same sample data as for the previous algorithm can be seen also in Fig. 2.

3 Experiments

Algorithms described in previous section have been tested on traditionally used data set Fisher Iris [6]. Dataset contain 4 real-valued attributes and one nominal attribute with values $\in$ {Setosa, Versicolor, Virginica}. Using original method I extracted rule for $\Psi =$ (Class is Virginica). I used parametrization type $\Gamma_{(a,b)}$ and Łukasiewicz logic, which was giving the best results on the dataset (measured with 5-fold crossvalidation) Parameter d was set to value of 3 also according to results of crossvalidation. Text form of extracted rule can be seen in Section 2 and its visualization can be seen in Fig. 1.

I used this rule as input for algorithms presented in previous chapter. Evolution of MAE for sequence $\tilde{\mathbb{C}}^1, ..., \tilde{\mathbb{C}}^{dn+1}$ given as output of *greedy* algorithm is depicted on Fig. 2.

Based on evolution of MAE on this sequence it can be chosen the most convenient member from sequence of matrices in accordance with user criteria. For testing and comparison purposes I wanted as much comprehensible (short)

[1] Due to space limitation I omit more precise description of algorithm.

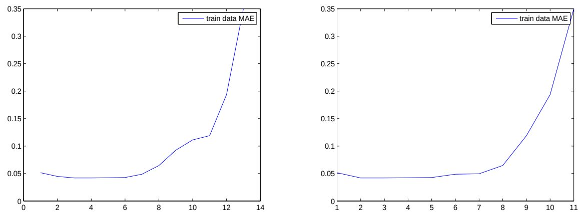


Fig. 2. Evolution of MAE for sequence $\tilde{\mathbb{C}}^1, ..., \tilde{\mathbb{C}}^{13}(\tilde{\mathbb{C}}^{11})$ given by greedy algorithm on the left side, given by localsearch algorithm on the right side respectively. On x axis is showed index k for corresponding MAE measured for $\tilde{\mathbb{C}}^k$.

rule as possible without significant decrease in accuracy, so I chose 10-th member of sequence (I denote as $\tilde{\mathbb{C}}^{10}_{greedy}$), which has the form

$$\tilde{\mathbb{C}}^{10}_{greedy} = \begin{pmatrix} 0\,0\,1\,1 \\ 0\,0\,0\,0 \\ 0\,0\,1\,0 \end{pmatrix}. \tag{1}$$

and corresponding rule contained only 3 atomic formulas $\varphi_{i,j}$ (as number of ones in matrix). The corresponding rule has the form

Class is Virginica $\equiv$ (PetalLength is $\Gamma_{(6.3667,2.0397)}$ $\wedge$ PetalWidth is $\Gamma_{(2.1893,0.97068)}$) $\vee$ (PetalLength is $\Gamma_{(5.7177,1.1318)}$)

MAE measured on the sample data with this rule is 0.1113.

Evolution of MAE while searching for sequence $\tilde{\mathbb{C}}^1, ..., \tilde{\mathbb{C}}^k$ with localsearch algorithm is illustrated in Fig. 2. To compare rule simplification given by this algorithm with rule sinmplification given by greedy algorithm I chose also rule containing 3 atomic formula $\varphi_{i,j}$. The corresponding matrix from the sequence founded by localsearch algorithm is 8-th member, which I denote $\tilde{\mathbb{C}}^{8}_{localsearch}$), which has the form

$$\tilde{\mathbb{C}}^{8}_{localsearch} = \begin{pmatrix} 0\,0\,1\,1 \\ 0\,0\,0\,1 \\ 0\,0\,0\,0 \end{pmatrix}. \tag{2}$$

The corresponding rule has the form

Class is Virginica $\equiv$ (PetalLength is $\Gamma_{(6.3667,2.0397)}$ $\wedge$ PetalWidth is $\Gamma_{(2.1893,0.97068)}$) $\vee$ (PetalWidth is $\Gamma_{(2.1865,0.49648)}$)

MAE corresponding to this rule was 0.0646, what is roughly half of MAE measured for the rule of *the same length* founded by greedy algorithm.

All results obtained with both greedy and localsearch algorithm are in accordance with results obtained on the same dataset with method CART.

4 Conclusions

The article describes the previously proposed method for fuzzy rules extraction using artificial neural networks and then design of the new module for simplification of rules obtained by this method. Two algorithms which are solving the task of simplification were proposed. The first of these two algorithms is similar to the algorithm for pruning of extracted tree in CART method, which was adapted to the needs of the methods described in this article. The second algorithm is an improvement on the first algorithm that searches the space of possible models to a user-specified depth, and which can find a better solution than the first algorithm, what is illustrated in chapter 3. So, the original method was extended with the new module for simplification of extracted rules, and hence the method is now suitable for solving the tasks of searching concepts in the data, where the main aim is a good insight into the data and a good understanding of data that can be obtained thanks to a model with a properly balanced ratio of precision and comprehensibility.

References

1. M. Berthold and D. Hand. *Intelligent data analysis: an introduction.* Springer-Verlag, Berlin, 1999.
2. L. Breiman, J. Freidman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Wadsworth, 1984.
3. A. S. d'Avila Garcez, K. Broda, and D. M. Gabbay. Symbolic knowledge extraction from trained neural networks: A sound approach. *Artificial Intelligence*, 125(1-2):155 – 207, 2001.
4. W. Duch, R. Adamczak, and K. Grabczewski. Extraction of logical rules from neural networks. *Neural Processing Letters*, 7:211–219, 1998.
5. J. Faber, M. Novák, P. Svoboda, and V. Tatarinov. Electrical brain wave analysis during hypnagogium. *Neural Network World*, 13:41–54, 2003.
6. R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2):179–188, 1936.
7. M. Holeňa. Extraction of fuzzy logic rules form data by means of artificial neural networks. *Kybernetika*, 41:297–314, 2005.
8. M. Holeňa. Neural networks for extraction of fuzzy logic rules with application to eeg data. In B. Ribeiro, R. F. Albrecht, A. Dobnikar, D. W. Pearson, and N. C. Steele, editors, *Adaptive and Natural Computing Algorithms*, pages 369–372. Springer Vienna, 2005.
9. M. Holeňa. Measures of ruleset quality for general rules extraction methods. *International Journal of Approximate Reasoning*, 50(6):867 – 879, 2009. Ninth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007).
10. J. Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:665–685, 1993.
11. D. Nauck and R. Kruse. Nefclass-x — a soft computing tool to build readable fuzzy classifiers. *BT Technology Journal*, 16:180–190, 1998.

Using of SVD and NMF in Social Networks

Lukáš Hlaváček and Michal Výmola

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.hlavacek@vsb.cz

Abstract. The aim of this work is to make use of the singular value decomposition method (SVD) for reduction of social network graphs. This method is plentifully used for issues where is the dimension reduction requirement such as an image compression or searching in document collections. The other aim is to use a method of nonnegative matrix factorization (NMF) which is used for dimension reduction as well. We want to compare these methods SVD and NMF from point of view of graph reduction and better identification of significant authors in large graphs.

Key words: SVD,NMF,DBLP,Social Networks

1 Introduction

This days, social networks are quite a big phenomenon and we can hear about them in all kind of media or scientific quarters every day. Social networks show us how people comunicate, how do they behave in internet environment, which friends do they have and many other interesting information. These networks grow very fast and keep a load of information. When the network reach some limit border it is becoming really hard to analyze it. It might be also difficult to analyze properities of this large network with mathematical methods. All calculations made on the whole network take a long time and they are not very effective. That is why we want to reduce this large networks. Also we want to keep the most important node of networks.

Social networks are represented by so called 'sociograms' whose are graphs actually so we can generalize our problem to graph reduction, subgraphs searching and searching of paterns in graphs. We can reach this aim by some different methods, for instance Markov + Laplacian, which can identify groups of object with some common property in matrix representation of network of these objects.

We have chosen slightly different approach which is matrix partitioning by SVD. This method used for dimensions reduction might be also very suitable method for reduction of complexity of social network. SVD makes it possible to recognize important components in network. To compare our SVD implementation with another method we have chosen nonnegative matrix factorization (NMF).

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 365–370.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

We will try to compare the fruitfulness of our approach with both methods together to only SVD approach. Our data source will be the network of authors in DBLP database where is possible to specify not only the relation between authors but also a strength of this relation. This could also bring us higher precision in identification of particular subgroups.

2 Graph reduction

Given a graph, partition it into subgraphs which are sparsely connected with one another, with each subgraph being possibly densely connected internally. For example the problem of partitioning a circuit amongst circuit boards connected by as few wires as possible can be directly seen as a graph partitioning problem. More generally, graph partitioning is the key to employing divide and conquer techniques on graphs, and arises in many areas [1–3]. Many social network data can be seen as object-attribute data or simply matrix (binary and fuzzy)[9]. Therefore they can be processed using matrix factorization methods which have been proven to be useful in many data mining applications dealing with large scale problems. V. Snášel et al. shows graph reduction using matrix factorization methods in [7, 10].

3 Singular Value Decomposition (SVD)

Let A be an $m \times n$ rank-r matrix. Be $\sigma_1, \ldots, \sigma_y$ eigenvalues of a matrix $\sqrt{AA^T}$. There exist orthogonal matrices $U = (u_1, \ldots, u_r)$ and $V = (v_1, \ldots, v_r)$, whose column vectors are orthonormal, and diagonal matrix $\Sigma = diag(\sigma_1, \ldots, \sigma_r)$. The decomposition $A = U\Sigma V^T$ is referred to as a singular value decomposition of matrix A and numbers $\sigma_1, \ldots, \sigma_r$ are singular values of the matrix A. Columns of U (or V) are referred to as left (or right) singular vectors of matrix A. Singular value decomposition (SVD) is well-known because of its application in information retrieval as LSI. SVD is especially suitable in its variant for sparse matrices [4, 5].

Among all $m \times n$ matrices C of rank at most k, A_k is the one that minimizes

$$||Ak - A||_F^2 = SUM(A_{i,j} - C_{w,j})^2 \tag{1}$$

Because rank-k SVD is the best rank-k approximation of original matrix A, any other decomposition will increase the sum of squares of matrix $A - A_k$.

4 Nonnegative Matrix Factorization (NMF)

Nonnegative matrix factorization (NMF) differs from other rank reduction methods for vector space models in text mining by using of constraints that produce nonnegative basis vectors. There vectors make the concept of a parts-based representation possible. First introduced the notion of parts-based representations

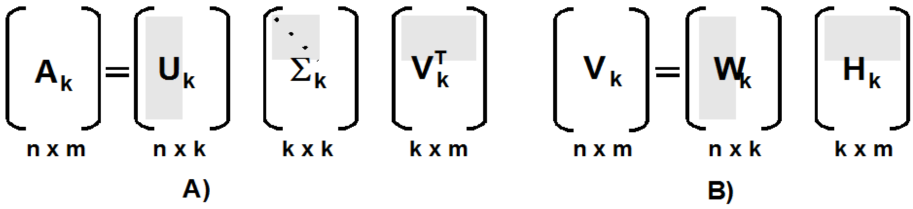


Fig. 1. A) K-reduced SVD, B) K-reduced NMF

for problems in image analysis or text mining that occupy nonnegative subspaces in a vector-space model [6].

Basis vectors contain non-negative entries. This allows only additive combinations of the vectors to reproduce the original. The perception of the whole becomes a combination of it is parts represented by these basis vectors. In text mining, the vectors represent or identify some semantic features. If a document is viewed as a combination of basis vectors, then it can be categorized as belonging to the topic represented by it is principal vector. Thus, NMF can be used to organize text collections into partitioned structures or clusters directly derived from the nonnegative factors.

Common approaches to NMF obtain an approximation of V by computing a (W,H) pair to minimize the Frobenius norm of the difference $V \bullet WH$. Let $V \in R^{m \times n}$ be a non-negative matrix and $W \in R^{k \times n}$ and $H \in R^{k \times n}$ for $0 < k << min(m, n)$. Then, the objective function or minimization problem can be expressed as $min||V - WH||_F^2$ with $W_{ij} > 0$ and $H_{ij} > 0$ for each i and j.

The matrices W and H are not unique. Usually H is initialized to zero and W to a randomly generated matrix where each $W_{ij} > 0$ and these initial estimates are improved or updated with alternating iterations of the algorithm [6].

5 Modular scheme

In the past we have already worked with analysis of social networks so in this paper we continue in our previous research. Previously we dealed with analysis of various features of social network where we used DBLP social network. That is why we decided to use this tested database again in this work. In the Fig 2 is the basic scheme showing our approach. In the first step we parse original data into our own format which is easier to work with.

6 Results

The process begins by choosing of particular author. Then we can create author's social network (sociogram) up to set level. For our experiment we choose few different size networks with various number of edges and nodes as well. We choose 4 social networks from 24 to 286 nodes and from 34 to 807 edges. Each

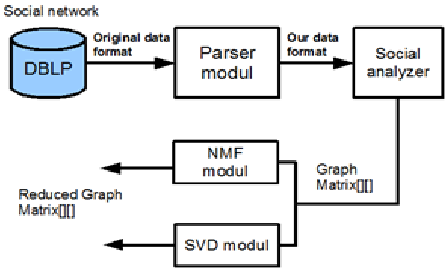


Fig. 2. Modular scheme for reduction of DBLP social networks

graph is represented by matrix, which would have a size for instance for 286 nodes 286 rows × 286 columns. Then we apply reduction dimension methods to this matrix.

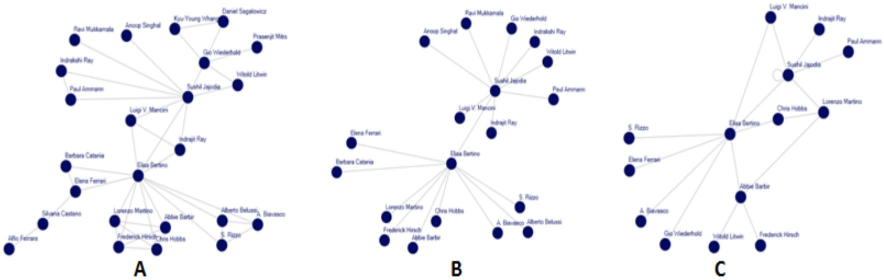


Fig. 3. A) Original graph, B) SVD reduction with k-rank=3, C) NMF reduction with k-rank=3

After using these methods on original matrix A and with the appropriate setting of k-rank, we obtain a new reduced matrix A_k. For the right setting of k-rank with SVD we need to first analyze singular values matrix Σ. If singular value is less then 1, we considet this value as not important (noise), so we can set it to zero. This is actually, how we reduce the matrix. After elimination all singular values less then 1, matrix A_k still haven not lost any data and that is why number of authors in matrix does not change.

According to our tests, we can consider on the average 65 percent of singular values as a noise. This way we can decrease matrix size up to 65 percent. With NMF we set k-rank differently. There is no rule for the right choose of k-rank so we had to choose this value experimetally. We found out that the best value for k-rank is close about 1/2 of dimension size. or instance if we have a matrix of 6×6, we set k-rank to 3. By setting of higher value of k-rank the matrix A_k will not reduce. We set maximum number of cycles of algorithm to 200. Calculation may finish either earlier in case when the diffrence between matrix A and A_k is under the set error value.

Lower k-rank more unimportant people will be took out from the graph by reduction. This corespond to authors with the highest degree centrality. In Fig 3 we can see graphs after reduction when k-rank was set to 3. We can see that NMF reduce the graph better with the same k-rank then SVD. As the undesirable effect of NMF we consider the fact that there may appear new edges between two authors which were not in the original graph.

We made experiments with graphs of sizes of 24, 62, 67, 123 and 286 authors. We applied SVD and NMF reduction on all these graphs. Each graph was tested with k-rank values from maximum (original dimension) up to zero with the step of 1. For each step we noted down the number of nodes/edges in currently reduced graph. Table 1 shows us graphs for picked authors with various size of their network. Also we can see 5 different levels of reduction for each author. The first row for each author represents the original unreduced graph. The second row represents the point when graph began to reduce. Columns '% nodes' and '%edges' show percentual difference of reduction rate of NMF to SVD (positive numbers say that NMF is more efficient).

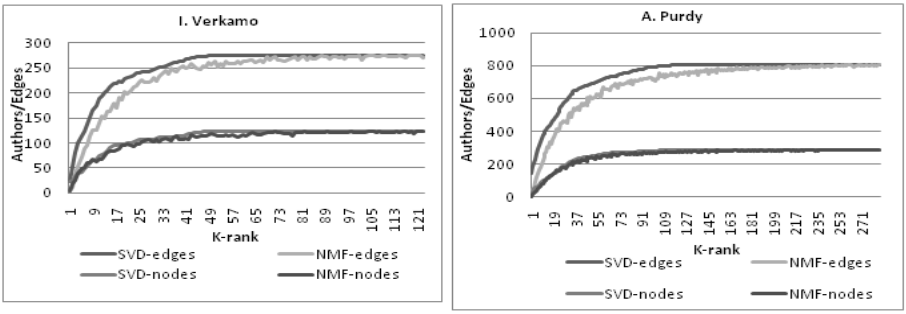


Fig. 4. I.Verkamo and A.Purdy

7 Conclusion

We made experiments with SVD and NMF methods and we found out following facts. With SVD method it is easier to choose k-rank value from which is a graph reduced. This value can be identified from singular values in matrix Σ. With NMF we found the rihgt k-rank based on executed experiments. Mostly this value is about 50% of original dimension size. In SVD, this vaue was about 35%. Based on result values appears that NMF method reduces the graph better with the same k-rank value. However NMF is more efficient in reduction it may not be more quality at the same time (NMF may add new edges). Both methods keep significant authors which fulfilled our presumption.

Table 1. Reduction dependent on k-rank

Author	SVD edges	SVD nodes	NMF edges	NMF nodes	K-rank	% edges	% nodes
Ke-li	131	62	131	62	62	0,0	0,0
	129	61	119	57	17	7,8	6,6
	119	56	106	57	10	10,9	-1,8
	91	43	79	44	5	13,2	-2,3
	65	29	48	43	3	26,2	-48,3
I.Verkamo	275	123	275	123	123	0,0	0,0
	272	121	247	115	46	9,2	5,0
	228	97	202	96	20	11,4	1,0
	175	66	125	62	10	28,6	6,1
	82	31	40	26	3	51,2	16,1
A.Purdy	807	286	807	286	286	0,0	0,0
	806	286	760	276	133	5,7	3,5
	689	248	605	243	50	12,2	2,0
	546	177	450	165	25	17,6	6,8
	376	99	221	80	10	41,2	19,2
	208	50	54	23	3	74,0	54,0

References

1. Abhiram G. Ranade, Some uses of spectral methods ,Indian Institute of Technology, Mumbai, *India, 2008*
2. P. DRINEAS, Clustering Large Graphs via the Singular Value Decomposition, 2004
3. Christos Faloutsos,Hanghang Tong, Large Graph Mining:Patterns, Tools and Case Studies, *ICDE,2009*
4. Petr Praus,SVD-based principal component analysis of geochemical data,2005
5. Hussam Dahwa Abdulla, Martin Polovincak, Vaclav Snasel, Using a Matrix Decomposition for Clustering Data,2009
6. Václav Snášel, Zdeněk Horák, Jana Kočibová,Reducing Social Network Dimensions Using Matrix Factorization Methods,2009
7. Václav Snášel, Hussam Dahwa Abdulla, Solution for search result clustering by Using Singular Value Decomposition,
8. Soo-Chang Pei, Ja-Han Chang, Jian-Jiun Ding,Quaternion matrix singular value decomposition and its applications for color image processing, 2003
9. D.B. Skillicorn, Social Network Analysis via Matrix Decompositions: al Qaeda,2004
10. Hussam Dahwa Abdulla, Václav Snášel, Search results clustering using Nonnegative Matrix Factorization(NMF), *IEEE Xplore 2009*

Computers and the Handicaped

Jaroslav Hořejší

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jaroslav.horejsi@vsb.cz

Abstract. In the context of this article is to meet the skilled public with the work of the handicapped on the computers. The handicapped use computers to work They use it for the fun. They need to adapt a mouse or keyboard to the comfortable control. They need to adjust websites for simplier use, too.

Keywords: the rules of Web Acessibility, the handicapped, computer, devices

1 Introduction

I am very interested in the work of the handicapped with the computers. I am the handicapped. I have the child cerebral polio. I am the kvadruplegig. I move on the mechanic wheelchair. The work of the computer I predominantly control with my right hand.

2 The handicapped and computers

In the contemporary period the grande widespread of computers to school and to the household has come up. Internet is widespread too. The handicapped especially need the Internet their life. Internet is for them the only connection with the world. Hereafter we meet different kinds of the disabled. We also meet the tools for the control for the computer and the rules of Web Accessibility.

I would like to introduce to you a preparatory conclusion of my survey among the handicapped. My survey investigates a facilitation of the work of the handicapped on the computer.

2.1 The handicapped users

The handicapped have mainly the disabled hands. These handicapped can control the computer only with difficulty. If they can control partially their hands and they can control a bit a mouse or a keyboard. If they cannot control the mouse, they need to control the computer with the keyboard. All the functions must be accessible for the keyboard.

The forms of motoric troubles is a whole row. Some users cannot use their hands after the following disease.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 371–374.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

– Injuries of a backbone with a wound of spinal cord. The lesion of the mobility and the sensitivity of the limb come up with these injuries.
– Child cerebral polio:
  • Hemiparethc forms – disability of upper extremities
  • Kvadruparehic forms – disability of all four extremities
  • Muscular dystrophy – progressive disability of all four extremities
  • Parkinson disease – a man loses the ability to control his movement
– Different diseases that cause the loss of mobility of limbs

It is valid for all the users that they cannot predominantly use the mouse. A lot of the handicapped have also the troubles with the use of the keyboard That is why it is urgent to adjust the mousses or the keyboard for the handicapped.

2.2 Tools for the control of the computer by the handicapped

The handicapped users have a fundamental problem with the control of the computer. That is why different controllable mousses have been designed. If they cannot use to the control of the adapted mouse, they have to control all with the keyboard.

– Among the most used mousses belong *trackballs*, see Fig. 1(a). The control is absolutely fluent and after a couple of minutes of the exercise is very exact. For the work with pictures, video and other multimedia it is completely sufficient. Of course we cannot want impossible.
– Among the most used keyboards belong *ergonomic keyboard*. Ergonomic keyboard, see Fig. 1(b), allows the comfortable writing thanks to the option of the setting of the different settings. You can adapt the keyboard to your natural position of the arm and the hand. This keyboard can serve the people with the damaged motor activity of the hands. This keyboard can be divided in the middle into two parts.
– In case that they cannot control the computer with the hand, so they can control the computer with eyes (I4Control) or with Head pointer.
  I4Control – System I4Control, see Fig. 1(c), introduces a new computer periphery. This periphery allows to control the computer with the movement of eyes or head. Thanks to this, it allows the user to communicate with the installed application by means of the simple way via the eye.
  HeadPointer – It deals with the HeadPointer, see Fig. 1(d), for the control of the keyboard with he movement of the head. The pointed stick is situated under the japone. It does not hinder from the view of eyes of the user. The construction is light (aluminium) and over the different setting elements allow to adapt the stick to each user. The length of the stick is 25cm.

3 Web accessibility

The handicapped users are not an insignificant percent. A lot of people among them use the computer and internet commonly. For some people it is even a tool to be able to remain in the contact with everyday reality. From the side of the owners of the website the opinions can be heard, the opinions such as that their websites are not given to people with some handicap. To adapt the inaccessible website for the handicapped is not easy. It requires more financial resources and more demanding creating of websites. It allows the handicapped users better and quicker orientation on the websites.

3.1 The rules for the creation of the Web Accessibility for the handicapped

Under the conception of accessibility we understand the state, when a given thing does not lay the users any fundamental barrier. It means that the conception of the Web Accessibility is limited to the functionality of websites. The accessible website is used for each user of internet. It is independent of his disability, abilities, knowledge, experience or visualization options.

4 Survey – How to facilitate the work on the computer for the handicapped?

For the discovery of the survey, how to facilitate the work on the computer for the handicapped. I elaborated the questionnaires for the tutors and the handicappeds. I sent these questionnaires to about seventy organizations. These organization interest in the training and the increase of the employment of people with disabilities. It especially deal with the following organizations:

- the centers for the handicapped in the various regions,
- the branches of the union of the handicapped in the various regions,
- the institutions for the handicapped,
- the secondary schools for the handicapped.

This survey has not been finished, but from the received answers is seen the following:

- The knowledge of the handicapped about the rules of web accessibility is minimal
- Next it is shown, that the hard handicapped (use for example Headpointer) want to be able to control websites over the keyboard and key abbreviations.

5 Accessibility

The concept of the accessibility means: The websites are accessible to all users without regard to their disabilities. The web should have the accessible tables, the accessible navigation, and also the accessible pictures. The web has to contain unobtrusive style sheet, unobtrusive JavaScript. In some cases it is necessary to create the alternative website. The standards for the accessibility were accepted in some countries. In the USA for example American Disabilities Act(ADA), Individuals with Disabilities Education Act(IDEA), Section 508 of the Rehabilitation Act of 1973 and Web Content Accessibility Guidelines 1.0 (WCAG 1.0) A WCAG 2.0.

(a) Mouse

(b) Ergonomic keyboard

(c) I4Control

(d) Head Pointer

Fig. 1. Tools for the control of the computer by the handicapped

6 Conclusion

This article sums up the options the handicapped have with the work at the computer and internet It does not describe all the options. It only describes fundamental options of work of the handicapped with computers.

References

1. David Špinar: Tvoříme přístupné webové stránky. Zoner Press, 2004, 360 pp., ISBN 80-86815-11-0, in Czech
2. Pawan Vora : Web Application Design Patterns, Morgan Kaufmann, 448 pp., 2009
3. Hana Řezánková: Analýza dat z dotazníkových šetření, Professional Publishing, 220 pp., 2010
4. http://www.petit-os.cz/

Lossless R-tree Compression using Variable-Length Codes

Peter Chovanec and Michal Krátký

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{peter.chovanec,michal.kratky}@vsb.cz

Abstract. The R-tree is one of the most popular multidimensional data structure. This data structure bounds spatially near points in multidimensional rectangles and supports various types of queries, e.g. point and range queries. When a compression of the data structure is considered, we follow two objectives. The first objective is a smaller index file and the second one is a reduction of the query processing time. In this paper, we introduce a lossless R-tree compression using variable-length codes. Although variable-length codes are well known in the area of data compression, they have not been yet successfully applied in the case of the data structure compression. The main reasons of this fact are inefficient decoding/encoding algorithms. In this paper, we apply recently introduced fast decoding algorithms and we show that these codes save up-to 84% of the index file's size compared to the uncompressed R-tree and provide more efficient query processing time.

Key words: R-tree and its variants, variable-length codes, fast decompression algorithm, lossless compression method

1 Introduction

Multimedia databases have become increasingly important in many application areas such as medicine, CAD, geography, and molecular biology. Processing of multi-dimensional data is requested in almost all fields. There are a lot of applications of *multi-dimensional data structures* [15], e.g., data mining [12], XML documents [14], text documents and images [7]. Query processing in high-dimensional spaces has therefore been a very prominent research area over the last few years.

Since Guttman proposed his method [11] in 1984, R-trees have become the most cited and most used reference data structure in this field. The R-trees can be thought of as an extension of B-trees in a multi-dimensional space. It corresponds to a hierarchy of nested n-dimensional *minimum bounding rectangles* (MBR).

Since the R-tree clusters spatially near tuples on pages using MBRs, redundancy appears, and various compression methods may be utilized. As far as the data structures are concerned, the query processing efficiency is measured by the *Disk Access Cost* (*DAC*) and the query processing time [15]. Obviously, the smaller index leads to lower DAC during the query processing. Consequently, the

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 375–380.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

lower DAC leads to the faster query processing time. Therefore, using the R-tree compression, we try to decrease the index size and reduce the query processing time.

In this paper, we introduce a lossless R-tree compression using variable-length codes. Although the variable-length codes like the Elias-delta code [8] and the Fibonacci family codes [9, 6] are well known in the area of data compression [16], they have not been yet successfully applied in the case of data structure compression. The main reasons of this fact are inefficient decoding/encoding algorithms. In this paper, we apply the recently introduced fast decoding algorithms [13, 18] and we show that these codes provide significantly more efficient query processing time than the uncompressed R-tree.

The outline of the paper is as follows. Section 2 introduces the lossless R-tree compression using variable-length codes. In Section 3, we put forward the experimental results. In the last section, we conclude this paper and outline possibilities of our future works.

2 R-tree Compression using Variable-Length Codes

We present a lossless R-tree compression using variable-length codes. Fast decoding algorithms recently introduced in [13, 18] have been applied.

A common principle of R-tree compression methods is as follows (see [3, 2] for detail): tuples are decompressed always when they are accessed during a tree's operation. It means that the compression methods must be as fast as possible. Another parameter is the compression ratio. Variable-length codes encode the smaller number by a shorter code. When an R-tree's page is split, a split technique is applied. Since tuples are clustered using MBRs, tuples' coordinates in an MBR are close to each other. We can use two types of compression techniques:

1. Tuple's coordinates are directly encoded by variable-length codes
2. A node includes reference items (or reference tuples) resulting in smaller numbers encoded by variable-length codes. This technique is often called difference coding [16].

A structure of the compressed node is shown in Figure 1. Each node includes compressed items (MBR's in inner nodes or tuples in leaf nodes) and uncompressed reference tuples (if they are used). Let us note that leaf nodes can also contain MBRs in leaf node. Since compressed tuples have variable-lengths and operations of the R-tree require random accesses to the tuples, nodes should be compressed by tuples, not by coordinates of a dimension. Therefore, the node includes a table with pointers (or links) to each tuple. In the next subsections, we describe some issues related to the compression using variable-length codes.

Although the compression using the reference tuple is more efficient from the compression ratio point view, reference items create an extra overhead for insert and update operations.

If an inserted tuple includes coordinates lower than reference tuple's coordinates, the reference tuple must be changed and all encoded tuples must be decoded and encoded again using the new reference tuple.

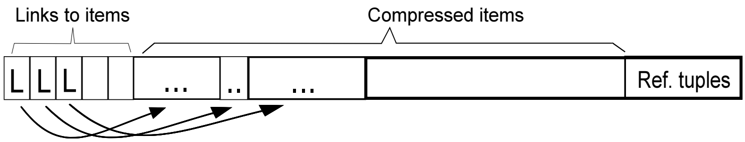


Fig. 1. A structure of the compressed node

Therefore, we introduce the usage of more reference tuples in one page. More reference tuples mean lower encoded values and the higher compression ratio. From the insert and update operations point of view, more reference tuples mean lower overhead because fewer tuples are recompressed during the operation when a reference item must be changed.

Some problems are related to the utilization of reference tuples: we must decide which reference item is appropriate for a new inserted tuple or we must decide whether the new reference tuple should by created. Since we need to cluster tuples with similar values in all dimensions, we utilize the Taxi distance between two tuples T_1 and T_2 of the n dimension:

$$d(T_1, T_2) = \sum_{i=1}^{n} |T_{2i} - T_{1i}| \tag{1}$$

When a tuple is inserted into a node, the Taxi distance between this tuple and the reference tuple is computed. If the distance is greater than a threshold, the new reference tuple is created and the new inserted tuple is related to the new reference item. Evidently, the new reference tuple is the same as the new inserted tuple. If a node contains more reference tuples, we choose the reference item with the lowest Taxi distance to the inserted tuple.

3 Experimental Results

In our experiments[1], we compare the R*-tree utilizing the proposed lossless compression with the R*-tree and the R*-tree utilizing RLE and lossy quantization compressions. All data structures have been implemented in C++. Two real collections have been chosen for these tests. The first collection, titled as XML, represents a set of paths in the XMark collection [17] including 1,031,080 10-dimensional tuples [14]. The second collection, titled as Energy, is a relational collection including records of failures in power networks [10]. This collection includes 239,117 tuples of dimension 43. In all experiments, we turn off the OS's disk read cache to prevent the OS from file caching.

Efficiency of the range query processing was measured by DAC, node read time from the secondary storage, and complete query processing time. We used

[1] The experiments were executed on an AMD Opteron 865 1.8Ghz, 2.0 MB L2 cache; 2GB of DDR333; Windows 2008 Server.

20 various range queries, each query was processed 10× and the results were averaged. Since the range query is processed by random accesses, DAC is measured by the number of disk accesses. In the case of the variable-length compression, we have tested codes Elias-Delta, Elias-Fibonacci and Fibonacci of orders 2 and 3. We use the abbreviation 'RI' instead of 'reference item' (or 'reference tuple'), so 'Fibonacci 2 with RI' means 'Fibonacci 2 with the reference item'.

3.1 Tests of XML Collection

Index sizes and build times are shown in Table 1. Since the collection is rather sparse, we used maximal 3 reference tuples. As we see, the Fibonacci 3 code produces the smallest index file. The compression without the reference item saves approximately 32% of the index size and the compression with the reference item saves up to 50% of the index size. From the build time point of view, the compression using variable-length codes is approximatelly 2× slower. The essential part of the build time is the tuple decompression time. This fact is caused by multiple decompressions during the tuple insertion.

Table 1. XML Collection: Build Statistics and Query Processing Statistics

Type	**Build**		**Query Processing**		
	Index Size [MB]	**Time [s]**	**DAC**	**Time [s]**	**Read Time [s]**
R^*-tree	72.43	678	3,432	1.86	1.72
Elias-Delta	59.14	1,112	2,898	1.74	1.24
Elias-Fibonacci	57.76	1,152	2,691	1.51	1.03
Fibonacci 2	62.35	1,213	3,114	1.99	1.53
Fibonacci 3	**56.61**	1,187	2,352	1.74	1.31
Elias-delta with RI	38.36	994	1,665	**0.81**	0.56
Elias-Fibonacci with RI	39.87	1,086	1,951	1.15	0.69
Fibonacci 2 with RI	36.81	1,150	1,945	1.10	0.67
Fibonacci 3 with RI	**36.23**	1,107	2,302	1.22	0.73

The compression with the reference item increases the node capacity up-to 2× compared to the uncompressed R^*-tree. In the case of the compression without reference items, the compression is not so efficient.

The results of the range query processing are shown in Table 1 as well. Evidently, the read time is an essential part of the query processing time. The most efficient query processing time was achieved by the Elias-delta code. The compression using variable-length codes improves disk accesses up-to 2× and saves 60% of the query processing time.

3.2 Tests of Energy Collection

The Energy collection mainly contains values of a small domain. Therefore, we may efficiently compress them without the reference item; we use only 1 reference item. The statistics of build and query processing are shown in Table 2. The Elias-delta code produces the smallest index in this case; the compression with the reference item saves up to 84% of the index size. The most efficient query processing time is achieved by the Elias-delta code. The query processing time is 4× lower than in the case of the uncompressed R*-tree and we achieved up-to 10× lower read time.

Table 2. Energy Collection: Build Statistics and Query Processing Statistics

Type	**Build**		**Query Processing**		
	Index Size [MB]	**Time [s]**	**DAC**	**Time [s]**	**Read Time [s]**
R*-tree	82.39	827	6,449	6.20	5.64
Elias-delta	**18.65**	1,342	1,829	**1.31**	0.68
Elias-Fibonacci	19.97	1,237	1,935	1.46	0.71
Fibonacci 2	20.11	1,384	1,969	1.48	0.70
Fibonacci 3	20.99	1,255	2,004	1.42	0.65
Elias-delta with RI	**13.17**	1,513	1,417	**1.45**	0.53
Elias-Fibonacci with RI	15.75	1,252	1,614	1.51	0.60
Fibonacci 2 with RI	15.73	1,376	1,615	1.47	0.58
Fibonacci 3 with RI	18.09	1,276	1,818	1.61	0.67

4 Conclusion

In this article, we introduced a new lossless R-tree compression using variable-length codes. We performed a series of experiments with various compression techniques. Variable-length codes with a reference item are mostly more efficient than methods not using the reference item. Moreover, we see that compression is a relevant way to improve the query processing time; the query processing time was improved $2-4\times$. Although, we see that our method is not so dependent on the size of the space domain and the query processing time is always more efficient, the build time is often inefficient. In our future work, we should solve this issue.

5 Projects and Publications

Peter Chovanec is a member of team solving the grant of GAČR, Czech Republic, No. P202/10/0573, Handling XML Data in Heterogeneous and Dynamic Environments, the grant of SGS, Czech Republic, No. SP/2010138, Datové struktury podporující techniky indexování XML dat and the grant of SGS SP2011/172, Detekce plagiátů v textových dokumentech (řešitel: Jan Martinovič). He is co-author of five publications in conference proceedings [1, 3, 4, 2, 5]. One publication is indexed in Web of Science, another two publications are indexed in SCOPUS.

Own References

1. P. Chovanec and M. Krátký. Efficiency Improvement of Narrow Range Query Processing in R-tree. In *Proceedings of DATESO 2009, Czech Republic*, 2009.

2. P. Chovanec and M. Krátký. Benchmarking of Lossless R-tree Compression for Data of Failures in Electrical Power Networks. In *Proceedings of the 7th Workshop of ELNET, Czech Republic*, 2010.
3. P. Chovanec, M. Krátký, and R. Bača. Optimization of Disk Accesses for Multidimensional Range Queries. In *Proceedings of DEXA 2010, Spain*, LNCS. Springer-Verlag, 2010.
4. P. Chovanec, M. Krátký, and J. Walder. Lossless R-tree Compression using Variable-Length Codes. In *Proceedings of the 5th International Conference for Internet Technology and Secured Transactions (ICITST-2010), IEEE*, 2010.
5. F. Křižka, M. Krátký, R. Bača, and P. Chovanec. Multidimensional Implementation of Stream ADT. In *Proceedings of The Third International Workshop on XML Data Management (XMLDM 2011).* Springer–Verlag, 2011, accepted.

Other References

6. A. Apostolico and A. Fraenkel. Robust transmission of unbounded strings using Fibonacci representations. *IEEE trans. inform.*, 33(2):238–245, 1987.
7. P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In *Proceedings of VLDB 1997*, pages 426–435, 1997.
8. P. Elias. Universal Codeword Sets and Representations of the Integers. *IEEE Transactions on Information Theory*, IT-21(2):194–203, 1975.
9. A. Fraenkel and S. Klein. Robust Universal Complete Codes as Alternatives to Huffman Codes. Technical Report Tech. Report CS85-16, Dept. of Appl. Math., The Weizmann Institute of Science, Rehovot, 1985.
10. R. Goňo, S. Rusek, and M. Krátký. Reliability analysis of distribution networks. In *Proceedings of EPQU 2007. Barcelona, Spain.* IEEE Press, 2007.
11. A. Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In *Proceedings of SIGMOD 1984*, pages 47–57. ACM Press, June 1984.
12. N. Karayannidis, A. Tsois, T. Sellis, R. Pieringer, V. Markl, F. Ramsak, R. Fenk, K. Elhardt, and R. Bayer. Processing Star Queries on Hierarchically-Clustered Fact Tables. In *Proceedings of 28th International Conference on Very Large Data Bases (VLDB'02), Hongkong, China.* Morgan Kaufmann, 2002.
13. S. T. Klein and M. K. Ben-Nissan. Using Fibonacci Compression Codes as Alternatives to Dense Codes. In *Proceedings of DCC'08*, pages 472–481, USA, 2008. IEEE CS.
14. M. Krátký, J. Pokorný, and V. Snášel. Implementation of XPath Axes in the Multidimensional Approach to Indexing XML Data. In *Current Trends in Database Technology, EDBT 2004*, volume 3268 of *LNCS*, pages 219–229. Springer–Verlag, 2004.
15. Y. Manolopoulos, Y. Theodoridis, and V. J. Tsotras. *Advanced Database Indexing.* Kluwer Academic Publisher, 2001.
16. D. Salomon. *Data Compression The Complete Reference.* Third Edition, Springer–Verlag, New York, 2004.
17. A. R. Schmidt and at al. The XML Benchmark. Technical Report INS-R0103, CWI, The Netherlands, April, 2001, `http://monetdb.cwi.nl/xml/`.
18. J. Walder, M. Krátký, R. Bača, J. Platoš, and V. Snášel. Fast decoding algorithms for variable-lengths codes. *To Appear in Information Sciences, Accepted Manuscript, ISSN 0020-0255.*, Submitted: February 16 2010, Revised: December 13 2010, Accepted: June 15 2011.

Vienna Development Method

L'uboš Chupáč

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lubos.chupac@vsb.cz

Abstract. This article deals with Vienna Development Method. The primary goal of this article is to give a first introduction on VDM to the reader. First, we will introduce the VDM, further we will describe the origins of the VDM. In the third chapter is described structure of VDM and modules. Using of the VDM is in the last section.

1 Introduction

The Vienna Development Method (VDM) is a formal method for the description and development of software systems. VDM consists of two components:

1. A mathematically based specification language in which specifications can be written in a very abstract and formal form.
2. A method describing a process for the development of an implementation of a software system using an arbitrary programming language from a formal specification. The method consists of a series of development steps by which a VDM specification is transformed into an implementation (an executable program) and of a number of proof obligations that must be discharged to justify each development step. Each step generates a new specifications which is less abstract (more concrete) and closer to the implementation than the previous one. However, no guidelines as to how the development steps should be selected are present in the VDM method. It is only suggested that the level of abstraction should be reduced. The process of moving between the development steps is known as refinement.

The original specification language, Meta-IV, evolved into several different variants, These variants reflect two different styles in the use of VDM: the Danish school and English school. The difference between these schools is primarily caused by different application areas: respectively, 'systems software' and 'algorithm and data refinement'. The main difference between the two schools is that the Danish school focused on modeling large software systems using an explicit style, and thus used the meta-language as an abstraction mechanism to obtain a better understanding of the complexity of the systems, whereas the English school has focused more on the correctness aspects of software systems. The English school used a more implicit style (using pre- and post-conditions) in order the ease the verification of the systems [6].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 381–386.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Origins of the Vienna Development Method

'VDM' is a generic term: the development of VDM has given rise to several VDM 'dialects'. The development of VDM started back in 1970, when in the IBM laboratory in Vienna a group formed by Heinz Zemanek worked on formal language definition and compiler design [1]. They built on ideas of Elgot, Landin and McCarthy to create an operational semantics approach capable of defining the whole of PL/I, including the parallel features of the language. For this purpose they used a meta- language which was called *Vienna Definition Language (VDL)*. The approach taken was successful, but it also showed that operational semantics could complicate formal reasoning in an unnecessary way. A new approach was taken, called *denotational semantics,* in late 1972. A PL/I compiler was designed using a meta-language called *Meta- IV*. Due to external reasons the compiler was never finished, but the formal definition of PL/I in a denotational style is generally seen as the birth of VDM [2].

3 Structure of a VDM

VDM is a model-oriented formal method based on a denotational semantic setting, intended to support stepwise refinement of abstract models into concrete implementations. The method includes a formal specification language, VDM-SL, which supports various forms of abstraction.

Representational abstraction is supported through data modeling facilities. These facilities are based on six mathematical data structuring mechanisms: sets, sequences, maps, composite objects, cartesian products and unions. At a lower level the language provides various numeric types, booleans, tokens and enumeration types. By using the data structuring mechanism and the basic data types, compound data types can be formed, in VDM denoted by the term *domains*. Domains form in general infinite classes of objects. These classes have a specific mathematical structure. Subtyping is supported by attaching *domain invariants* to domain definitions.

Operational abstraction is supported by both functional abstraction and relational abstraction, the first by means of (fully referentially transparent) function specification, the second by operation specification. Both functions and operations may be specified *implicitly* using *pre and post conditions,* or *explicitly* using *applicative* constructions to specify functions, and *imperative* constructions to specify operations[2].

A VDM specification is structured into various blocks, where each block is identified by a distinct keyword [3]:

types
<type definitions>
values
<value definitions>
functions
<function definitions>
operations
<operation definitions>

state <state name> **of**
<state definition>
end

ISO standard for VDM-SL contains different principles of structuring. Modules play several roles in specification.

- They help structure complex systems when specifications scale up. Here, they simply transpose programming language constructs.
- In the context of model-based languages, modules can by seen as a concrete way to encapsulate the elements of an abstract machine (state variables and operations) or abstract data type (types and functions).
- Specification languages often include a notion of refinement. Modules can then be used do distinguish between several layers of abstraction in a refinement lattice[5].

These all follow traditional information hiding principles with modules and they can be explained as:

Module naming: Each module is syntactically started with the keyword module followed by the name of the module. At the end of a module the keyword end is written followed again by the name of the module.
Importing: It is possible to import definitions that has been exported from other modules. This is done in an *imports section* that is started off with the keyword dimports and followed by a sequence of imports from different modules. Each of these module imports are started with the keyword from followed by the name of the module and a module signature. The *module signature* can either simply be the keyword all indicating the import of all definitions exported from that module, or it can be a sequence of import signatures. The import signatures are specific for types, values, functions and operations and each of these are started with the corresponding keyword. In addition these import signatures name the constructs that there is a desire to get access to. In addition optional type information can be present and finally it is possible to *rename* each of the constructs upon import. For types one needs also to use the keyword struct if one wish to get access to the *internal structure* of a particular type.
Exporting: The definitions from a module that one wish other modules to have access to are exported using the keyword exports followed by an exports module signature. The *exports module signature* can either simply consist of the keyword all or as a sequence of export signatures. Such *export signatures* are specific for types, values, functions and operations and each of these are started with the corresponding keyword. In case one wish to export the internal structure of a type the keyword struct must be used.
More exotic features: In earlier versions of the VDM-SL tools there was also support for parameterized modules and instantiations of such modules. However, these features was taken out of VDMTools around 2000 because they was hardly ever used in industrial applications and there was a substantial number of tool challenges with these features[4][5].

4 Support and industrial experience

A number of different tools support VDM:

- VDMTools are the leading commercial tools for VDM and VDM++, owned, marketed, maintained and developed by CSK Systems, building on earlier versions developed by the Danish Company IFAD. The manuals) and a practical tutorial are available. All licenses are available, free of charge, for the full version of the tool. The full version includes automatic code generation for Java and C++, dynamic link library and CORBA support.
- Overture is a community-based open source initiative aimed at providing freely available tool support for VDM++ on top of the Eclipse platform. Its aim is to develop a framework for interoperable tools that will be useful for industrial application, research and education.
- SpecBox: from Adelard provides syntax checking, some simple semantic checking, and generation of a LaTeX file enabling specifications to be printed in mathematical notation. This tool is freely available but it is not being further maintained.
- LaTeX and LaTeX2e macros are available to support the presentation of VDM models in the ISO Standard Language's mathematical syntax. They have been developed and are maintained by the National Physical Laboratory in the UK.

VDM has been applied widely in a variety of application domains. The most well-known of these applications are:

- Ada and CHILL compilers: The first European validated Ada compiler was developed by DDC-International using VDM. Likewise the semantics of CHILL and Modula-2 were described in their standards using VDM.
- ConForm: An experiment at British Aerospace comparing the conventional development of a trusted gateway with a development using VDM.
- Dust-Expert: A project carried out by Adelard in the UK for a safety related application determining that the safety is appropriate in the layout of industrial plants.
- The development of VDMTools: Most components of the VDMTools tool suite are themselves developed using VDM. This development has been made at IFAD in Denmark and CSK in Japan.
- TradeOne: Certain key components of the TradeOne back-office system developed by CSK systems for the Japanese stock exchange were developed using VDM. Comparative measurements exist for developer productivity and

defect density of the VDM-developed components versus the conventionally developed code.

- FeliCa Networks have reported the development of an operating system for an integrated circuit for cellular telephone applications [5].

5 Example

In this chapter I will try to describe how to use VDM on a very simple example. This is an example of an implicit function definition. The function returns the element from a set of positive integers:

```
max(s:set of nat)r: nat
pre card s > 0
post r in set s and
      forall r' in set s & r' <= r
```

The postcondition characterizes the result rather than defining an algorithm for obtaining it. The precondition is needed because no function could return an r in set s when the set is empty.

6 Conclusion

In this article we briefly presented Vienna Development Method. It is one of the formal methods. This method is very attractive for academic and industrial world. It was used in many domains for develop software systems for aircrafts, automobile industry and space program. VDM-SL is an ISO standardized language, which has formal definition.

References

1. Muller, A., VDM - The Vienna Development Method. 2009.
2. Plat, N. and P.G. Larsen, An Overview of the Iso/Vdm-Sl Standard. Sigplan Notices, 1992. **27**(8): p. 76-82.
3. V.S. Alagar, K. Periyasamy, Specification of Software Systems, 405 Texts in Computer Science, DOI 10.1007/978-0-85729-277-3_16, © Springer-Verlag London Limited 2011
4. Fitzgerald, A., Larsen, P. G., Modeling systems: Practical Tools and Techniques in Software Development. Second ed2009, Cambridge. 288.
5. Yves Ledru, Potet, M.L., Sanlaville, R., VDM Modules.
6. Andrews, D.J., *A Formal Definition of VDM-SL*, 1998.

Unsupervised Algorithm for Segmenting Categorical Time Series

Tomáš Kocyan*

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
tomas.kocyan@vsb.cz

Abstract. This article discusses the possible use of Voting Experts algorithm as one of the tools for "Retrieve" part of Case-Based Reasoning (CBR) methodology. The Voting Expert Algorithm is a domain-independent unsupervised algorithm for segmenting categorical time series into the meaningful episodes.

Keywords: Voting Experts, Segmenting, Information Retrieval, Episode-Based Reasoning, Case-Based Reasoning

1 Introduction

Case-based reasoning (CBR) methodology, whose application is the main topic of my thesis, is the process of solving new problems based on the solutions of similar past problems. Each of these cases consists of a specific problem, its solution, and the way it was achieved. For purposes of computer reasoning the CBR has been formalized as a four-step abstract process [1]:

- **Retrieve**: Retrieve the most similar cases from case database that are relevant to solving it.
- **Reuse**: Map the found solutions from the previous cases to the new problem.
- **Revise**: Test the derived solution in the real conditions and, if necessary, revise it.
- **Retain**: After the solution has been successfully adapted to the target problem, the resulting experience is stored as a new case in memory, where can be used for prediction in next cycle.

For achieving the best results using Case-Based Reasoning methodology, it is necessary to successfully manage the first step. There are many of supervised and unsupervised methods for looking for patterns and similar situations. But the most of them have a common problem: they cannot handle searching for patterns of different lengths. One of the algorithms, which are able to handle this, is the Voting Experts algorithm.

* This paper is based on article named 'Text Segmentation Using Voting Experts Algorithm'. Therefore I would like to thank co-authors, namely Jan Martinovic and Jiri Dvorsky, for cooperation with this article and agreement to its publication.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 387–392.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Voting experts

The *Voting Expert Algorithm* is a domain-independent unsupervised algorithm for segmenting categorical time series into the meaningful episodes. It was first presented by Cohen and Adams in 2001 [6]. Since this introduction, the algorithm has been extended and improved in many ways, but the main idea is always the same. The basic Voting Experts algorithm is based on the simple hypothesis that natural breaks in a sequence are usually accompanied by two statistical indicators [7]: low internal entropy of episode and high boundary entropy between episodes.

The basic Voting Experts algorithm consists of following three main steps[1]:

– Build an nGram tree from the input, calculate statistics for each node of this tree (internal and boundary entropy) and standardize these values at the same level of length.
– Pass a sliding window of length n over the input and let experts vote. Each of the experts has its own point of view on current context (current content of the sliding window) and votes for the best location for a split. The first expert votes for locations with the highest boundary entropy, the second expert votes for locations with a minimal sum of internal spit entropy. By this way, the votes are counted for each location in the input.
– Look for local maximums which overcome selected threshold. These points are adepts for a split of sequence.

Tests showed that the algorithm is able to segment selected input into meaningful episodes successfully. It was tested in many domains of interest, such as looking for words in a text [6] or segmenting of speech record [8].

2.1 Two-way voting experts

There are several ways how to improve the basic Voting Experts algorithm. Simply we can divide these improvements into the two main groups: On the one hand, you can add your own “expert” to voting process (for example Markov Expert in [11]) and receive additional point of view on your input. On the other hand, there are methods based on repeated or hierarchical segmenting the input [10, 9].

One of the simplest ways how to slightly improve performance of segmenting is two-way passing of the sliding window. It means using classic voting algorithm supplemented by segmenting of reversed input.

This idea was outlined in [10] where the way how to make high-precision cut points by selection of higher values of a threshold was showed. Additionally, reversing the corpus, and segmenting the reversed input with Voting Experts, generates a different set of backward cut points. The subsequent intersection of sets of cut points offers high precision segmenting. However, on the other hand, this high precision is redeemed by loss of recall.

[1] For detailed explanation of each of mentioned steps see [7].

For this reason, we implemented "two-way voting" idea in another way. After running classic voting experts, we don't make cuts, but we keep the votes for each location. Then we run the backward voting on the reversed input and keep the votes again. Just after that we check the total votes against the selected threshold. Moreover, these received votes can be further balanced by FORWARD and BACKWARD multipliers which can highlight or suppress selected direction of voting.

3 Experiments

As it was mentioned in introduction Voting Experts Algorithm as segmentation algorithm was tested on texts written in natural language. The segmentation of text written in language known to human reader is very easy task. It is also easy task for computer reader, due to the existence of spaces in text. But is these spaces is deleted from the text the task is much harder for human reader and difficult to realize for computer reader. Therefore, the reconstruction of deleted spaces was chosen as test task for Voting Experts Algorithm. To compare performance on different languages same text written in English and Czech language was chosen[2].

In order to simplify evaluation of segmentation quality, precision and recall were measured and Van Risjbergen's [12]:

$$F_\beta = \frac{1+\beta^2}{\frac{\beta^2}{Recall} + \frac{1}{Precision}} = \frac{\left(1+\beta^2\right) Recall\ Precison}{\beta^2 Precison + Recall} \tag{1}$$

where β indicates the ratio of significance between precision and recall. For example, when β is an even 0.5, it means that the user is twice as interested in precision than in recall and when β is an 2, the users interest is vice versa. β was to 1 in our experiment.

In the first experiment, the segmentation of English text was performed. The best result were provided by following methods:

- Forward segmenting (sliding window size = 6, threshold = 3),
- Backward segmenting (sliding window size = 6, threshold = 3),
- Two-way balanced segmenting (sl. window size = 5 and 7, threshold = 7).

Results of segmentation using methods mentioned above are given in Fig. 1. The two-way segmentation repaired some single way segmentation errors e.g. space was between the words `THE-GOOD`, and conversely, the incorrect space was removed from `SOLDI-ER`. Evaluation of the best obtained results is described in Table 1.

The second experiment was aimed on segmentation of Czech text. The Czech version segmentation achieves worse results in both single and two way voting, see Table 2. It is caused by high number of various suffixes in the Czech language.

[2] We choose novel "Good Soldier Svejk" by Jaroslav Hašek.

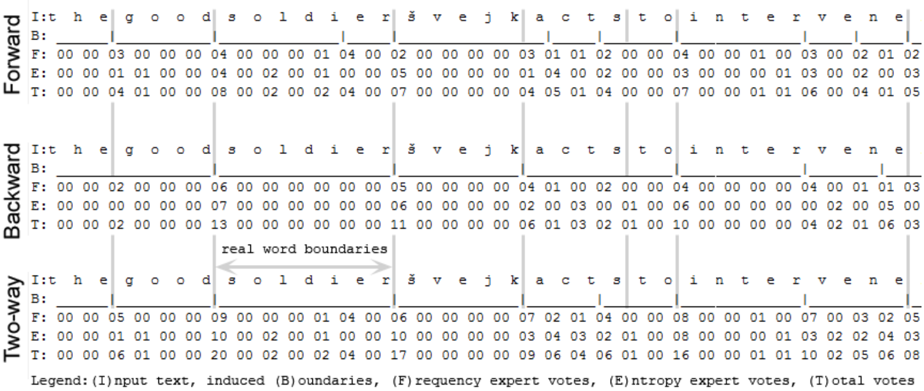


Fig. 1. Segmentation of English Text.

Table 1. English text results

	F-measure	Precision	Recall	Window size	Threshold
Forward	0.75	0.77	0.72	6	3
Backward	0.71	0.74	0.68	6	3
Two way Multipliers: 1.0 – 0.9	0.77	0.73	0.82	5 forward 7 backward	7

The word form is very variable in Slavic language, e.g. Czech. There are number of suffixes which change their form according the gender and their position in a tense. Therefore there is a very high boundary entropy between these suffixes and word ends. For this reason it is difficult to find out true word boundaries. Regardless of this, in the two-way segmentation successfully correct some of wrong forward or backward segmentation, see Fig. 2.

Table 2. Czech text results

	F-measure	Precision	Recall	Window size	Threshold
Forward	0.67	0.69	0.66	7	3
Backward	0.69	0.70	0.67	6	3
Two way Multipliers: 1.0 – 0.9	0.71	0.72	0.71	7 forward 6 backward	7

Values in Table 2 may seem to be not so satisfactory. However, we have to realize that these results are not directly comparable with the English ones. A better way is to compare changes for the better of the two-way approach against the single way algorithm. In the Table 3 you can see the percentage improvements of both versions of text. The translated version has improvement

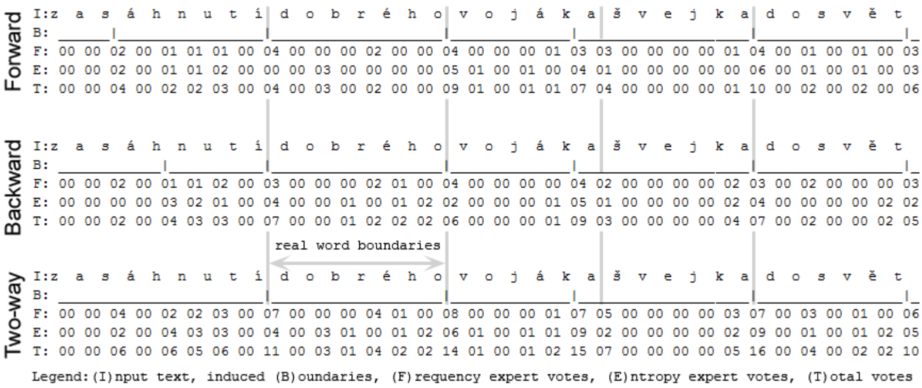


Fig. 2. Segmentation of Czech text version

only about 3.5% whereas the original Czech version reached the improvement about 6.3%.

Table 3. Percentage improvement

	Voting Expert Algorithm			Percentage
Language	Forward	Backward	Two-way	Improvement
English	0.75	0.71	0.77	3.47%
Czech	0.67	0.69	0.71	6.28%

4 Conclusion

Test showed that the *Voting Experts* algorithm is able to segment input stream of categorical data into the meaningful episodes. Even though the algorithm was tested on segmentation of an unbroken text, it can be generalized for segmenting of any categorical time series.

One of the biggest advantages of this algorithm is the fact, that *Voting Experts* is an unsupervised algorithm and it is able to find out episodes of various lengths. These found episodes can be further post-processed (e.g. they can be clustered or balanced) or directly used in the first step (Retrieve) of *Case-Based Reasoning* methodology.

5 List of publications

Floreon-DIP: Vyuziti metody CBR v systemu pro prevenci skod
Kocyan T., Martinovic, J., Janckulik D., Unucka J., Vondrak, I., In ZNALOSTI 2009.

FLOREON$^+$: Using CBR in System for Flood Predictions.
Kocyan, T., Martinovic, J., Unucka, J., Vondrak, I., Disaster Management and Human Health Risk: Reducing Risk, Improving Outcomes, ISBN: 978-1-84564-202-0.

Hydrometeorologic Social Network With CBR Prediction
Kocyan T., Martinovic J., Valickova A., Sir B., Horinkova M., Rihova V.
ECMS 2010, Kuala Lumpur 2010-07-31

Case-Based Reasoning in a System for Disaster Prediction
Kocyan T., In WOFEX. 2010. ISBN 978-80-248-2276-1.

References

1. Aamodt: *Case-Based Reasoning: Foundational Issues.* Methodological Variations & System Approaches, AI Communications, Vol. 7 Nr. 1, March 1994.
2. Bjarne, K. Hanse and Riordan, D.: *Fuzzy Case-Based Prediction of Cloud Ceiling and Visibility.* Dalhousie University, Halifax, NS, Canada, 2003.
3. Watson, I.: *Applying Case-Based Reasoning: Techniques for Enterprise Systems.* Morgan Kaufman, 1997.
4. Watson, I.: *Case-based reasoning is a methodology not a technology.* AI-CBR, University of Salford, Salford M5 4WT, UK
5. Sanchez-Marre, M., Cortes, U., Martinez, M., Comas, J. and Rodriguez-Roda, I.: *An Approach for Temporal Case-Based Reasoning: Episode-Based Reasoning.* Springer Berlin / Heidelberg, ISSN 0302-9743.
6. Cohen, P. R., Adams, N.: *An Algorithm for Segmenting Categorical Time Series into Meaningful Episodes.* Proceedings of the Fourth Symposium on Intelligent Data Analysis, Lecture Notes in Computer Science. 2001.
7. Cohen, P. R., Adams, N., Heeringa, B.: *Voting Experts: An Unsupervised Algorithm for Segmenting Sequences.* To appear in Journal of Intelligent Data Analysis. 2007.
8. Miller, M., Wong, P., Stoytchev A.: *Unsupervised Segmentation of Audio Speech Using the Voting Experts Algorithm.* Proceedings of the Second Conference on Artificial General Intelligence (AGI). 2009.
9. Miller, M., Stoytchev, A.: *Hierarchical Voting Experts: An Unsupervised Algorithm for Hierarchical Sequence Segmentation.* Proceedings of the 7th IEEE International Conference on Development and Learning (ICDL). (Best Paper Award, ICDL 2008)
10. Hewlett, D., Cohen, P.: *Bootstrap Voting Experts.* Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI). 2009.
11. Cheng, J., Mitzenmacher M.: *Markov Experts.* Proceedings of the Data Compression Conference (DCC). 2005.
12. Van Rijsbergen, C. J.: *Information Retrieval, Second Edition.* Department of Computer Science, University of Glasgow, 1979.

Formal Methods in Software Process Modeling

Michal Košinár

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.kosinar@gmail.com

Abstract. Documented software processes and their assessments are the basics of modern software development. Currently the semantic web, knowledge bases and knowledge management have many applications. Yet, applications to support software processes (and business processes in general) are surprisingly being ignored. In this paper we focus on conclusions done in research that focus on applying a knowledge layer into software processes and on the design of such a knowledge base. We will recap proposal of set of methods for building semantic annotations for software processes and an improvement based on the application of a machine readable knowledge base based on such ontology.

Keywords: knowledge base, knowledge management, knowledge modeling, process modeling, software processes

1 Introduction

The main goal of every software company is to develop high-quality software with a minimal cost of development. One way to assure this is to follow good practices that are described in software processes standards. Since such standards should improve software process development, they became an integrated part of every software company. Every software company uses some type of software process. Even if this process is undocumented and/or unknown, it is still there.
Describing and maturing the software process is a key element of a company's business strategy, because more mature software process means higher quality and lower expenses. Maturity of the software process is recognized through the assessment and its evaluation. According to the reference standards, a company's software process maturity is rated at a level from 0 to 5. If the company wants to have more mature process, the process must follow appropriate good practices for a higher level (see [6], [7]).
Building a knowledge base that describes the practices in a company is an essential practice that assures that every employee in company knows what to do. Nowadays, almost every company has some type of human readable knowledge base (HRKB) that describes a variety of practices in the company. There even exist human readable knowledge bases that describe reference software processes and/or good practices. What is still missing in this area is the systematic usage of the machine readable knowledge bases (MRKB) and appropriate automated knowledge management.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 393–398.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Formal approach in software processes

The idea of knowledge support for development, the assessment and evaluation of software processes is based on the fact, that in our opinion, development, the assessment, enhancement and monitoring can be supported by the creation and usage of machine readable knowledge bases. A lot of manual procedures can be automated.
One of the many tasks of this domain area is the comparison between the reference software process and real software process that is used in the company. The issue is to find the similarity between the real software process and the reference software process and provide an evaluation of the current state. Typically, the real process is assessed and human readable knowledge bases are searched for similarities. Everything is performed manually. Our proposed approach shows the possibility of using a machine readable knowledge base for the automatic evaluation of the similarity between the real and reference processes [4].
To do this, we have to transform human readable documentation to machine readable knowledge base. We will build semantic annotations of process items, thus we will create formal content of RUP® software process domain.

2.1 Research background

Our approach can be basically described as follows:

1. The first step is the creation of the particular *reference knowledge base* - Knowledge base transformation. Documented software processes that are described in the human readable knowledge bases are analyzed and the particular ontology for each software process is created. Next, the ontology and knowledge obtained from the HRKB is transformed into the machine readable knowledge base. A new MRKB is created for every type of the reference software process. See [4] for detailed information.
2. The second step is the study and creation of the *real knowledge base* and the comparison of the reference and real knowledge base - enhancing software processes with knowledge management. A real software process is modeled and the ontology for this process is created. The created ontology must then be mapped into the reference software process ontology. The mapped ontology set and the knowledge obtained from the model is then transformed into the machine readable knowledge base.

Both knowledge bases should be then automatically compared and the result is a number that shows the similarity of real and reference software processes.
It is obvious that the first step can be performed only once for every reference software process. The second step is then applied for every real software process that we want to evaluate. Our approach is graphically interpreted on figure 1 (top-right block of scheme marked with pinned note is work in progress and middle-left block with light bulb is the goal of our research).

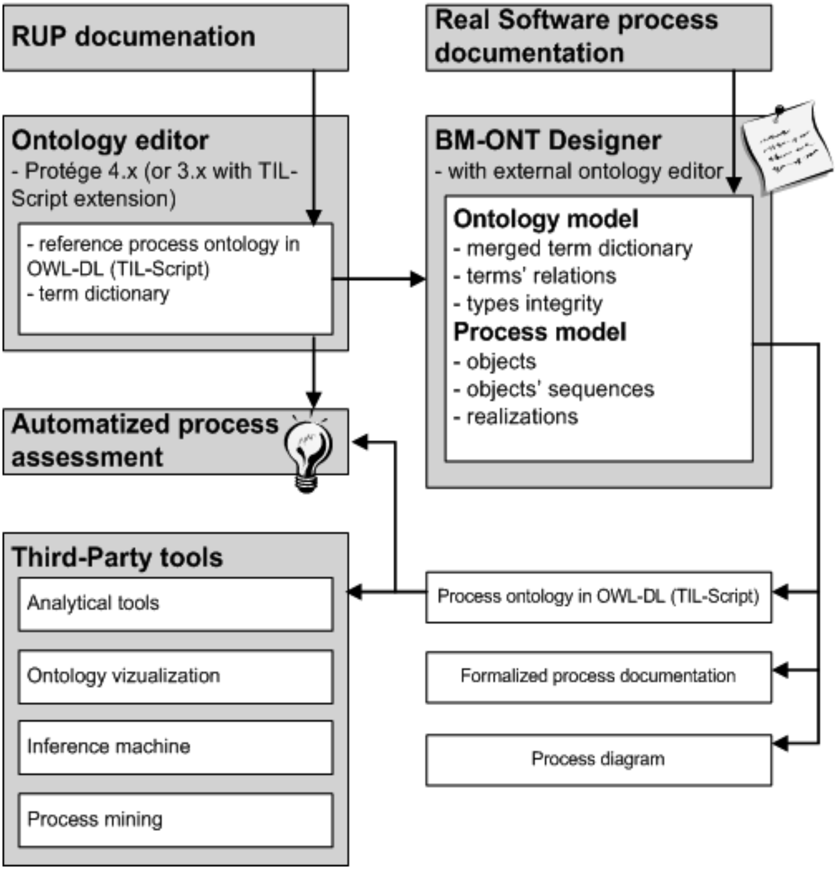


Fig. 1. Knowledge support for software processes

2.2 Process knowledge transformation

The transformation can be done by building ontology for a software process using the human readable reference documentation. Then, the knowledge base is defined by such ontology's content. We can see particular transformation on figure 2. Greater details of the transformation are in [4].

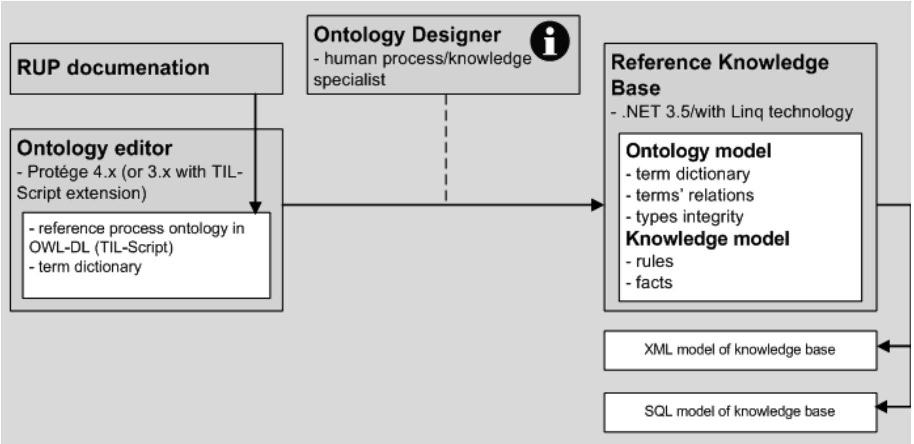


Fig. 2. Knowledge base transformation

An *Ontology Designer* has to describe the software process domain to the ontology data file (using tools like Protégé etc.) and then fill the reference knowledge base (using some content language like OWL-DL [5], TIL-Script or PROLOG) describing the *reference domain* (using terms and relations from domain ontology). For more detailed information on knowledge and ontologies see [1], [2], [3] and [5].

2.3 Mapping software process elements

To create semantic annotations for process we have to find right and exact terms for ontology interpretations. We have used the general business process metamodel to achieve this [6].

Software process ontology design proceeds on three levels of abstraction – Level 0 – *Process* itself, Level 1 – *Activities* of process and Level 3 – *Steps of activities* (see table 1).

Table 1. Process abstraction levels

Level	Description	Example
0	Process (one process of Software process) consisting of activities.	Change Request Management.
1	Activities. Processes from Level 0 are composed of these Activities.	Review CR, Submit CR, …
2	Steps. Activities of Level 1 are composed of atomic Steps.	Think about CR, Send e-mail, …

3 Conclusion

This presented phase is one part of our comprehensive approach to the assessment, evaluation and improvement of software processes. The first step is the extension of the presented process to the automated comparison of specific parts of software processes. The next steps are then e.g. automated comparisons with more than one specific process at once, involvement of process modeling to the approach that will be used for the automated search, evaluation and improvement suggestion using the template software process models etc… A lot of future work is needed to solve all the problems that arise during the development of this new approach. Our work is also supported by the experience that is gained through the practical experiments of this approach in real software companies.

Although, according to our preliminary use cases studies, this approach seems to be very promising, the further use case studies are needed to continuously develop and enhance the approach and support its inclusion into the software process assessment models and improvement techniques.

4 Recent publications

Košinár, M., Kohut, O.: Applying Statistical Methods in Knowledge Management of a Multiagent System. *Organon F*, XVII, n.2, SAV Bratislava, 2010.

Košinár, M., Kohut, O.: *Methods in Knowledge Management of a Knowledge Based Multiagent System*, In GIS Ostrava 2010, Faculty of electrical engineering and computer science, VŠB - Technical University of Ostrava, 2010, Editors: Lucie Hrubá, Michal Podhoranyi, Adrian Kapias, ISBN 978-80-248-2171-9

Košinár, M., Štolfa, S., Kožusznik, J.: *Využití znalostí a ontologií v softwarových procesech*, In Pavel Smrž (Ed.) Znalosti 2010, Vysoká škola ekonomická v Praze, Fakulta managementu, 2010, ISBN 978-80-245-1636-3

Košinár, M., Štolfa, S., Kožusznik, J.: *Knowledge Support for Software Processes*, To appear in ENASE 2010, Athens, Greece

5 Awards

Author was *Grand aided student of Municipality of Ostrava, Czech Republic* in academic year 2010/2011.

References

1. Ciprich, N., Duží, M. and Košinár, M.: *The TIL-Script language*. In the Proceedings of the 18th European Japanese Conference on Information Modelling and Knowledge Bases (EJC 2008), Tsukuba, Japan 2008.
2. Číhalová, M., Duží, M., Menšík, M.: Tvorba ontologií z pohledu logiky. In *DATAKON 09*, 2009, 171-181, ISBN 978-80-245-1568-7
3. Duží, M., Materna P.: Concepts and Ontologies. In *Information Modelling and Knowledge Bases XX* . Ed. Y. Kiyoki, T. Tokuda, H. Jaakola, X. Chen, N. Yoshida, Amsterdam :IOS Press , 2009, 45-64, ISBN 978-1-58603-957-8
4. Košinár, M., Štolfa, S., Kožusznik, J.: *Knowledge Support for Software Processes*, To appear in ENASE 2010, Athens, Greece
5. *Ontology Web Language guide* [online]. 2010 [cit. 2010-03-12]. Available from WWW: < www.w3.org/TR/owl-guide >
6. Vondrák, I.: Methods of Business Modeling. VŠB-TUO, Czech Republic, Ostrava, 2004.
7. Vondrák, I.: Software Engineering. VŠB-TUO, Czech Republic, Ostrava, 2002.

Annotation:

This research has been supported by the internal grant agency of VSB-TU of Ostrava - SP2011/56 Knowledge approach to the modeling, simulation and visualization of software processes. Author is also supported as a *Grand aided student of Municipality of Ostrava, Czech Republic*.

Quaternions and Molecular Dynamics of Rigid Bodies

Martin Kramář

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba

Abstract. In this paper we are going to introduce the system of quaternions which extends the complex numbers. In the first part we will show basic properties of quaternions with a main focus on their product in relation to three dimensional rotation. In the second part we will show their using in molecular simulation of rigid bodies. This approach leads to less memory demands and gets rid of singularities compared to the Euler angles.

1 Introduction

If we want to describe rotation about the origin in the geometric plane, we can use the properties of complex number and we can express the rotation as the multiplication of complex numbers. We consider a point $A = (a_x, a_y)$, where a_x and a_y are the real numbers which express x and y coordinate respectively. The point A can be associated with the complex number $a = a_x + a_y\imath$, where $\imath$ is imaginary unit and the rotation about the angle α with the complex number $r = \exp(\alpha\imath)$, where r is in the polar form. Using the Euler's formula, we obtain $r = \cos\alpha + \imath\sin\alpha$. The product $a_r = a \cdot r$ is complex number where the real part of a_r is the x coordinate of the point A rotated by the angle α and the imaginary part of a_r is the y coordinate of the rotated point A, respectively.Now, there is the question, does a similar number system which express a rotation in three dimensional space as multiplication exist? The answer sounds yes and it is the quaternion algebra introduced by the Sir William Rowan Hamilton in 1843.

The paper is organised as follows. At first we are going to introduce quaternions briefly after that we will discuss the rotational quaternion and in the third section we will show how to use quaternions in the molecular dynamic simulations.

2 Quaternions

As the set, the quaternions $\mathbb{H}$ are equal to the four-dimensional vector space $\mathbb{R}^4$ over the real numbers. Quaternions are usually written

$$q = q_0 + q_1 i + q_2 j + q_3 k,$$

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 399–404.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

where q_0, q_1, q_2, q_3 are the real numbers and $1, i, j, k$ are elements of the basis $\mathbb{H}$ which satisfy equation

$$i^2 = j^2 = k^2 = ijk = -1.$$

The system $\mathbb{H}$ id equipped with addition, scalar multiplication and quaternion multiplication. We consider the quaternions $p = p_0 + p_1 i + p_2 j + p_3 k$ and $q = q_0 + q_1 i + q_2 j + q_3 k$ and the scalar α. Now we can define the addition of p and q

$$(p + q) = (q_0 + p_0) + (q_1 + p_1)i + (q_2 + p_2)j + (q_3 + p_3)k.$$

The scalar multiplication of the scalar α and the quaternion q is defined

$$(\alpha q) = \alpha q_0 + \alpha q_1 i + \alpha q_2 j + \alpha q_3 k.$$

The quaternion multiplication of p and q is defined

$$\begin{aligned}(pq) &= (p_0 + p_1 i + p_2 j + p_3 k)(q_0 + q_1 i + q_2 j + q_3 k) = \cdots = \\ &= (p_0 q_0 - p_1 q_1 - p_2 q_2 - p_3 q_3) \\ &+ (p_0 q_1 + p_1 q_0 + p_2 q_3 - p_3 q_2)i \\ &+ (p_0 q_2 - p_1 q_3 + p_2 q_0 + p_3 q_1)j \\ &+ (p_0 q_3 + p_1 q_2 - p_2 q_1 + p_3 q_0)k.\end{aligned}$$

If we write quaternion $q = q_s + \mathbf{q}_v$ as the sum of the scalar part $q_s = q_0$ and the vector part $\mathbf{q}_v = (q_1, q_2, q_3)$ then we can define the quaternion multiplication

$$(pq) = (p_s + \mathbf{p}_v)(q_s + \mathbf{q}_v) = (p_s q_s - \mathbf{p}_v \cdot \mathbf{q}_v) + (p_s \mathbf{q}_v + q_s \mathbf{p}_v + \mathbf{p}_v \times \mathbf{q}_v), \quad (1)$$

where $\mathbf{p}_v \cdot \mathbf{q}_v$ and $\mathbf{p}_v \times \mathbf{q}_v$ denote scalar and cross vector product respectively. The conjugation of quaternion q we denote q^* and define

$$q^* = q_0 - q_1 i - q_2 j - q_3 k.$$

Now we can define the norm of quaternion

$$\|q\| = \sqrt{qq^*} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$$

and using the norm and the conjugation makes it possible to define the reciprocal quaternion

$$q^{-1} = \frac{q^*}{\|q\|^2},$$

which satisfies $qq^{-1} = 1$.

2.1 Rotational Quaternions

In the previous part we defined all necessary properties of quaternions to describe rotations in three dimensions. Now we consider a vector $\mathbf{a} = (x, y, z) \in \mathbb{R}^3$. This vector can be easily represented by the quaternion $q = q_s + \mathbf{q}_v$, where $q_s = 0$

and $\mathbf{q}_v = \mathbf{a}$. We consider that $\mathbf{k}$ is a unit vector describing an axis of rotation about which we want to rotate $\mathbf{a}$ by an angle α then a rotational quaternion is defined as

$$r = \cos\frac{\alpha}{2} + \mathbf{k}\sin\frac{\alpha}{2}. \tag{2}$$

Now we can compute the rotation as the quaternion multiplication

$$\begin{aligned} q_r = rqr^{-1} &= (\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\mathbf{k})\mathbf{a}(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\mathbf{k}) \\ &= \mathbf{a}\cos^2\frac{\alpha}{2} + (\mathbf{ka} - \mathbf{ak})\sin\frac{\alpha}{2}\cos\frac{\alpha}{2} - \mathbf{kak}\sin\frac{\alpha}{2}, \end{aligned} \tag{3}$$

where q_r is the quaternion represented the rotated vector $\mathbf{a}$. Using (1) and identity $a \times b \times c = b(a \cdot c) - c(a \cdot b)$ we obtain

$$\begin{aligned} \mathbf{ka} - \mathbf{ak} &= (-\mathbf{k}\cdot\mathbf{a} + \mathbf{k}\times\mathbf{a}) - (-\mathbf{a}\cdot\mathbf{k} + \mathbf{a}\times\mathbf{k}) = 2(\mathbf{k}\times\mathbf{a}) \\ \mathbf{kak} &= -(\mathbf{k}\times\mathbf{a})\cdot\mathbf{k} + (-\mathbf{k}\cdot\mathbf{a})\mathbf{k} + (\mathbf{k}\times\mathbf{a})\times\mathbf{k} \\ &= -2\mathbf{k}(\mathbf{k}\cdot\mathbf{a}) + \mathbf{a}(\mathbf{k}\cdot\mathbf{k}), \end{aligned}$$

because $(\mathbf{k}\times\mathbf{a})$ and $\mathbf{k}$ are perpendicular hence $-(\mathbf{k}\times\mathbf{a})\cdot\mathbf{k} = 0$. Now we put $\mathbf{ka} - \mathbf{ak}$ and $\mathbf{kak}$ in (3) and we obtain

$$\begin{aligned} q_r &= \mathbf{a}\cos^2\frac{\alpha}{2} + 2(\mathbf{k}\times\mathbf{a})\sin\frac{\alpha}{2}\cos\frac{\alpha}{2} - (\mathbf{a}(\mathbf{k}\cdot\mathbf{k}) - 2\mathbf{k}(\mathbf{k}\cdot\mathbf{a}))\sin^2\frac{\alpha}{2} \\ &= \mathbf{a}(\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}) + (\mathbf{k}\times\mathbf{a})(2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}) + \mathbf{k}(\mathbf{k}\cdot\mathbf{a})(2\sin^2\frac{\alpha}{2}) \\ &= \mathbf{a}\cos\alpha + (\mathbf{k}\times\mathbf{a})\sin\alpha + \mathbf{k}(\mathbf{k}\cdot\mathbf{a})(1-\cos\alpha). \end{aligned}$$

Now we can express the rotated vector $\mathbf{a}_r$ as

$$\mathbf{a}_r = (\mathbf{a} - \mathbf{k}(\mathbf{k}\cdot\mathbf{a}))\cos\alpha + (\mathbf{k}\times\mathbf{a})\sin\alpha + \mathbf{k}(\mathbf{k}\cdot\mathbf{a}). \tag{4}$$

In the next section we will clarify meanings of the all terms in (4).

2.2 Three Dimensional Rotation

Let's have a look at rotation from a different point of view. We consider again a vector $\mathbf{a}$ which we want to rotate around an axis z represented by a unit vector $\mathbf{k}$ by an angle α which is shown in Fig. 1. How can we express a rotated vector $\mathbf{a}_r$ in terms of $\mathbf{a}, \mathbf{k}, \alpha$? Using the vector algebra, we obtain $\mathbf{a}_x = \mathbf{a} - \mathbf{a}_\mathbf{z}$, where

$$\mathbf{a}_z = (\mathbf{k}\cdot\mathbf{a})\mathbf{k}. \tag{5}$$

Now we define a vector $\mathbf{w} = \mathbf{k}\times\mathbf{a}$ and we can see that $\mathbf{a}_x$ and $\mathbf{w}$ are in the same xy plane and $\|\mathbf{a}_x\| = \|\mathbf{w}\|$. Hence

$$\mathbf{a}_{xrot} = \mathbf{a}_x\cos\alpha + \mathbf{w}\sin\alpha, \tag{6}$$

where $\mathbf{a}_{xrot}$ is the $\mathbf{a}_x$ rotated by the α in xy plane. Using equations (5), (6), $\mathbf{a}_x = \mathbf{a} - \mathbf{a}_\mathbf{z}$ and $\mathbf{w} = \mathbf{k}\times\mathbf{a}$, we obtain

$$\mathbf{a}_{rot} = \mathbf{a}_{xrot} + \mathbf{a}_z = (\mathbf{a} - (\mathbf{k}\cdot\mathbf{a})\mathbf{k})\cos\alpha + (\mathbf{k}\times\mathbf{a})\sin\alpha + (\mathbf{k}\cdot\mathbf{a})\mathbf{k}. \tag{7}$$

We can see now, that vectors $\mathbf{a}_r(4)$ and $\mathbf{a}_{rot}(7)$ are the same and the rotational quaternions (2) and (3) describe rotation in three dimensional space.

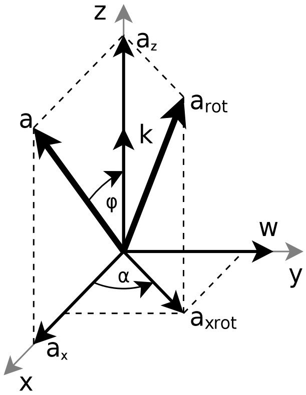


Fig. 1. Rotation of the vector $\mathbf{a}$ around the axis z in 3D.

3 Rigid Body Motion

Motion of rigid bodies can be decomposed into translation motion of the center of mass and rotation around the center. Translation motion is governed by the Newtonian equation

$$\mathbf{F}(t) = m\ddot{\mathbf{r}}(t),$$

where $\mathbf{r}(t)$, $\mathbf{F}(t)$ and m are the position vector of the particle, forces acting on the particle and the particle mass respectively. A conventional integrators such as Verlet or Runge-Kutta can be used to integrate the Newtonian equation. In this paper we focus on the rotation around the center of mass described by quaternions and proposed by Evans [1], [2] originally.

3.1 Coordinate Systems

If we work with rotations we usually need to work in different coordinate systems and we need to know the relations between them. We consider three different types of coordinate system: a laboratory $\mathcal{L}$, a reference $\mathcal{R}$ and system $\mathcal{F}$ fixed to the body. The relation among them can be seen easily in Fig. 2.

Each coordinate system is defined by the origin and three coordinate axes. We define the laboratory coordinate system $\mathcal{L} = (0, x, y, z)$, where 0 is the origin and x, y, z are coordinate axes. We consider a point $T_{\mathcal{L}}$ which is expressed in the $\mathcal{L}$. Now we can define the reference coordinate system $\mathcal{R} = (T_{\mathcal{L}}, x_r, y_r, z_r)$ and the body fixed system $\mathcal{F} = (T_{\mathcal{L}}, x_f, y_f, z_f)$ with the same origin at point $T_{\mathcal{L}}$.The axes x, y, z are parallel with axes x_r, y_r, z_r and axes x_f, y_f, z_f of the body fixed system are determined by given rotational quaternion r.

We consider a particle $P_{\mathcal{F}} = [p_{xf}, p_{yf}, p_{zf}]$ expressed in $\mathcal{F}$ and the same particle $P_{\mathcal{R}} = [p_{xr}, p_{yr}, p_{zr}]$ expressed in $\mathcal{R}$. $P_{\mathcal{F}}$ and $P_{\mathcal{R}}$ represent a vector parts

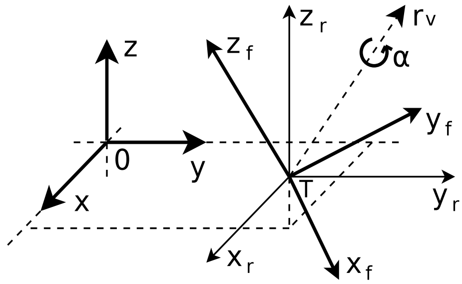


Fig. 2. Illustration of coordinate systems $\mathcal{L} = (0, x, y, z)$, $\mathcal{R} = (T, x_r, y_r, z_r)$,$\mathcal{F} = (T, x_f, y_f, z_f)$ and a relation between $\mathcal{R}$ and $\mathcal{F}$ described by the quaternion $r = \cos(\alpha/2) + \mathbf{r}_v \sin(\alpha/2)$

of quaternions $p_{\mathcal{F}}$ and $p_{\mathcal{R}}$ respectively. Now we can use equation (3) for crossing between coordinate systems $\mathcal{F}$ and $\mathcal{R}$

$$p_{\mathcal{R}} = r p_{\mathcal{F}} r^{-1}, \tag{8}$$

$$p_{\mathcal{F}} = r^{-1} p_{\mathcal{R}} r. \tag{9}$$

3.2 Rigid Body Rotation

Each rigid body is coupled with their body fixed coordinate system $\mathcal{F}$. The origin of $\mathcal{F}$ is put into position of a center of mass of the rigid body and the orientation of $\mathcal{F}$ is represented by the rotational quaternion. The rotation of rigid bodies is governed by the dynamic equation [3]

$$\begin{aligned}
\tau_x^b &= I_{xx}\dot{\omega}_x^b + \omega_y^b \omega_z^b (I_{yy} - I_{zz}),\\
\tau_y^b &= I_{yy}\dot{\omega}_y^b + \omega_z^b \omega_x^b (I_{zz} - I_{xx}),\\
\tau_z^b &= I_{zz}\dot{\omega}_z^b + \omega_x^b \omega_y^b (I_{xx} - I_{yy}),
\end{aligned}$$

where τ_x^b, τ_y^b and ,τ_z^b are components of the total torque τ^b expressed in $\mathcal{F}$, ω_x^b, ω_y^b and ω_z^b are components of the angular velocity ω measured in $\mathcal{F}$ and I_{xx}, I_{yy} and I_{zz} are the moments of inertia in $\mathcal{F}$. For simplicity we will consider $I_{xx} = I_{yy} = I_{zz} = I$ and so we can write dynamic equation

$$\tau^b = I\dot{\omega}^b. \tag{10}$$

We also need to know the time derivative of quaternion [1], [2]

$$\dot{q} = \frac{1}{2} q \Omega, \tag{11}$$

where Ω is quaternion defined as $\Omega = (0, \omega_x^b, \omega_y^b, \omega_z^b)$. There are several algorithms to solve equations (10) and (11) numerically. The very simple algorithm is the explicit leapfrog scheme [4], [5]. It is outlined below.

We can obtain q at the time step $t + \Delta t$

$$
\begin{aligned}
q(t+\Delta t) &= q(t) + \Delta t \dot{q}(t) + \frac{\Delta t^2}{2}\ddot{q}(t) + \mathcal{O}(\Delta t^3) \\
&= q(t) + \Delta t \dot{q}(t + \Delta t/2) + \mathcal{O}(\Delta t^3).
\end{aligned}
$$

The quaternion derivative $q(t + \Delta t/2)$ is needed and equation (11) shows that $q(t + \Delta t/2)$ and $\omega(t + \Delta t2)$ are also required. We can easily calculate

$$q(t + \Delta t/) = q(t) + \dot{q}(t)\Delta t/2,$$

where $\dot{q}(t)$ is given by equation (11) and $\omega(t+\Delta t/2)$ can be obtained using (10)

$$\omega(t + \Delta t/2) = \omega(t - \Delta t/2) + I^{-1}\tau^b(t)\Delta t.$$

To complete the algorithm we have to be able to compute the total torque and therefore forces acting on the each particle are required. There are a lot of potentials (electrostatic, van der Waals, Lennard-Jones,...) which describe acting forces between particles. If we are considering a pair acting potential then the force acting on the particle i can be computed

$$\mathbf{F}_i = \sum_{j=1, j \neq i}^{N} F_P(r_i, r_j),$$

where N is number of particles in simulated system, $F_P(r_i, r_j)$ is the acting force between particles i and j which depends on the chosen potential P and particle's positions r_i and r_j.

4 Conclusion

In this paper we introduced quaternion algebra and showed how quaternion multiplication can describe rotations in three dimensional space. In the second part we illustrated how can be quaternions used in a molecular dynamics algorithm. It should be mentioned that there are different possibilities describing rotations in three dimensional space such as Euler angles or rotational matrices. If we compare quaternions and rotational matrices, quaternions bring benefit in smaller memory requirements and in comparison with the Euler angles, quaternions do not have singularities.

References

1. Evans D.J.: On the representation of orientation space. Mol. Phys., 34 (1977)
2. Evans DJ, Murad S.: Singularity free algorithm for molecular dynamics simulation of rigid polyatomics. Molecular Physics, Vol. 34 (1977)
3. Goldstein H.: Classical Mechanics, 2nd ed. Addison-Wesley (1980)
4. Griebel M., Knapek S., Zumbusch G.: Numerical Simulation in Molecular Dynamics. Springer-Verlag. Berlin Heidelberg New York (2007)
5. Fincham D.: Leapfrog Rotational Algorithms. Molecular Simulation, Vol. 8 (1992)

Application of Multidimensional Data Structures to Indexing XML Data

Filip Křižka and Michal Krátký

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{filip.krizka, michal.kratky}@vsb.cz

Abstract. Holistic approaches are considered as the most robust solution for processing of twig pattern queries requiring no complicated query optimization. Holistic approaches use an abstract data type called a stream which is an ordered set of XML nodes with the same schema node. A straightforward implementation of a stream is a paged array. In this article, we introduce a multidimensional implementation of the stream for path labeling schemes. We also show that this implementation can be extended in such a way that it supports fast searching of nodes with a content. Although many multidimensional data structures have been introduced in recent years, we show that it is necessary to combine two variants of the R-tree (Ordered R-tree and Signature R-tree) for an efficient implementation the stream ADT.

Key words: indexing XML data, stream, stream ADT, path labeling scheme, R-tree, R*-tree, Ordered R-tree, Signature R-tree

1 Introduction

In this article, we introduce a multidimensional implementation of the stream ADT for path labeling schemes. Multidimensional indices have been utilized to indexing XML data in works [8, 11, 5], however there was no a relation to holistic joins. In this article, we introduce the utilization of multidimensional indices in a relation with holistic joins. We show that the stream index can be extended in such a way that it supports fast searching of nodes with a content which is important for an efficient query processing of XPath queries containing a value condition. Since the R-tree is a well-known multidimensional data structure, in this article, we utilize this data structure and its variants. A comprehensive description of multi-dimensional data structures is given in [15].

2 Multidimensional Index for XML

In [3, 1], we introduced an R-tree variant to indexing of multidimensional ordered data called Ordered R-tree. In [13], a signature extension of the R-tree for processing narrow range queries has been introduced.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 405–410.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

In the R-tree [9], tuples are clustered in tree's pages when *MBRs* (*Minimal Bounding Rectangles*) are built. R-tree supports various types of queries, e.g. point and range queries.

It is a hierarchical data structure representing spatial data by the set of nested n-dimensional *minimum bounding rectangles* (MBR). Each MBR is defined by two tuples QL and QH, where $QL^i \leq QH^i$, $1 \leq i \leq n$. If $\mathcal{N}$ is an inner node, it contains pairs (R_i, P_i), where P_i is a pointer to a child of the node $\mathcal{N}$. If R is the inner node MBR, then the boxes R_i corresponding to the children $\mathcal{N}_i$ of $\mathcal{N}$ are contained in R. Boxes at the same tree level may overlap. If $\mathcal{N}$ is a leaf node, it contains pairs (R_i, O_i), so called *index records*, where R_i contains a spatial object O_i. Each node of the R-tree contains between m and M entries unless it is the root and corresponds to a disk page.

A range query is processed by a depth-first search algorithm. If the query rectangle intersects an MBR then the node related to the MBR is retrieved and searched. In general, there is no order; therefore, tuples of a result set are sorted in the same order in which these tuples were inserted in leaf nodes and new MBRs have been inserted in inner nodes.

There are many variants of the R-tree, the following data structures utilize the R*-tree [6]. In [3, 1], Ordered R-tree allowing to index ordered multidimensional tuples has been introduced. A new concept, the *First Tuple* (FT), to each MBR was introduced to support the ordering of tuples in the R-tree. All nodes are then ordered according to their FTs which are utilized during the tree building. In [13], Signature R-tree was introduced for more efficient processing of narrow range queries. Signature R-tree is a variant of the R-tree including multidimensional signatures for a more efficient filtration of irrelevant tree nodes. In this case, the irrelevant node does not contain any tuple of the query rectangle. The multidimensional signature contains a signature of tuples in one node for each dimension.

Evidently, we need a combination of the Ordered and Signature R-trees for a multidimensional implementation of the Stream ADT. In our preliminary results, we show the performance of this data structure.

3 Preliminary Results

In our experiments we used the XMARK collection of scaling factor 45 with 92 mil. nodes and 547 streams using the prefix path streaming scheme and Dewey Order labeling scheme. Basic characteristics of streams are shown in Table 1. The page size is 2kB for all data structures used. If we consider 40 B per one label in average and the 50% page utilization, the average number of labels in one page is 25. Consequently, the average count of pages of one stream is 6,759.1. Although we can increase the page utilization by a bulk-load algorithm up-to 100% [10], the average count of pages in one stream is high again (approximately 3,379.5 in this case). It is clear that reading and searching in the long streams is a time consuming operation.

Table 1. Stream Statistics of XMARK

	All Labels	**Label Length**	
		1–7	**8–14**
XML file size [GB]	5.0	–	–
#Labels	92,430,540	59,374,127	33,056,414
#Streams	547	148	399
Minimum count of labels in one stream	1	1	1
Maximum count of labels in one stream	2,695,696	2,695,696	2,695,696
Average count of labels in one stream	168,977.2	401,176.5	82,848.2

This collection has been indexed by the R*-tree, Ordered R*-tree, and the combination of Ordered R*-tree and Signature R*-tree introduced in the previous section. In Table 2, we see properties of the created R-trees[1]. We see that the page utilization is by 7% higher in the case of the R*-tree than in the case of the ordered variants. This is caused by inserting of labels exported from the streams. It means, all labels of one stream were inserted before all labels of another stream. In the case of the ordered variants, labels are always inserted into the last leaf node, therefore the page utlization is 50%. Evidenly, if labels are inserted as an XML document is read or a rebuild algorithm is used, the utilization will be higher. The lower size of the R*-tree is given by the utilization. As we see, the overhead of the Ordered R*-tree compared to the R-tree is only 1.5%; the size of the First Tuple array is 0.05GB. The overhead of the Signature R*-tree is higher: the size of arrays with multidimensional signatures for two lowest levels of the tree is 2×0.59 GB. It means, the overhead is 31.3% compared to the R*-tree and Ordered R*-tree, however the results show that the Signature R*-tree is the most efficient from the query processing time point of view. In this case, we used the following configuration of multidimensional signatures (for detail see [13]): the signature length for the leaf nodes and super-leaf nodes was 128 and 2048 for each dimension, respectively, the number of hashing functions was 3, and one bit was set for each tuple value.

Labels are variable-length tuples with the maximal length for an XML collection. The maximal length is 14 for this collection. We used an approach introduced in paper [12] where the multidimensional forest indexing variable-length tuples has been introduced. In this case, two multidimensional spaces have been created: the first one includes labels of the length 1–7 and the second one includes labels of the length 8–14. Since our queries retrieved only labels up-to the length 7, we used only the first multidimensional space in our experiments (and in Tables 2–4). In Table 1, we see statistics of streams for labels of the length 1–7 as well. This tree includes 59 mil. labels (see also Table 2) and the average count of labels in one stream is higher than in the whole collection (401,176.5 labels).

[1] All data structures have been implemented in C++

Table 2. R*-trees Statistics

	R*-tree	**Ord. R*-tree**	**Ord-Sig. R*-tree**
Dimension	7		
Paged Size [B]	2,048		
Leaf Node Capacity [items]	63		
#Leaf Nodes	1,624,012	1,855,441	1,855,441
#Inner Nodes	96,350	109,142	109,142
#Tuples	59,374,127		
Avg. #Tuples in Leaf Nodes	36.56	32	32
Page Utilization [%]	57.9	50.9	50.9
Total Tree Size [GB]	3.28	3.31	4.94
R-tree Size [GB]	3.28	3.36	3.75
First Tuple Array [GB]	–	0.05	0.05
Signatures level 0 [GB]	–	–	0.59
Signatures level 1 [GB]	–	–	0.59
Avg. count of pages of one stream	10,973.1	12,536.8	12,536.8

In our experiments, we tested three groups of queries. The first group includes queries of the type 1, for example $(292, *, *, *, *, *, *)$ or $(18, *, *, *, *, *, max)$. The second group includes queries of the type 2, for example $(297, *, *, *, *, *, 860099805)$ or $(18, *, *, *, *, 1718073587, max)$. The third group includes queries of the type 4, for example $(295, 0, 5, 2462, *, *, 860712262)$ or $(18, 0, 4, 134143, *, 1718073587, max)$. The DAC (Disk Access Cost) [14] for these queries is shown in Table 3. We used c_Q (so called *relevance*) introduced in [13] as a quality ratio of range query processing: $c_Q = N_r/N_p$, where N_r is the number of relevant leaf nodes, it means nodes including tuples in the query rectangle, N_p is the number of all accessed leaf nodes. Evidently, if only relevant nodes are accessed during a range query then $c_Q = 1$.

Table 3. c_Q and DAC for Tested Queries

Query Group	**Result Size**	**R*-tree**		**Ordered R*-tree**		**Ord-Sig. R*-tree**	
		c_Q	**DAC**	c_Q	**DAC**	c_Q	**DAC**
1	630,366.8	0.85	71,182.5	1.00	20,863.5	1.00	20,863.5
2	7.3	0.03	58,951.8	0.25	11,157.5	0.83	69.0
3	1	1.00	8.5	0.75	7.5	1.00	7.5

We see that for the Query Group 2, c_Q is rather low for the R*-tree. In the case of the Ordered R*-tree, c_Q is higher (0.25), but it is close to 0 for the most of queries. Whereas c_Q is low for some query groups in the case of the R*-tree and Ordered R*-tree, c_Q is 1 or close to 1 for the Ordered-Signature R*-tree. In Table 4, we see the query processing time for tested queries. We must note that we read all pages from the disk and we did not use any method to efficient reading of pages during the range query processing (see [7]), since, in the case of real XML query processing, pages are not read together.

Table 4. Query Processing Time for Tested Queries [s]

Query Group	Result Size	R*-tree	Ordered R*-tree	Ord-Sig. R*-tree
1	630,366.8	16.44	4.99	4.99
2	7.3	8.43	1.79	0.03
3	1	0.00	0.00	0.00

We must note that the result of a range query is not sorted in the case of the R*-tree, therefore labels of the result must be sorted and it means an additional overhead. Another problem of the R*-tree is that leaf nodes include labels of various streams, whereas ordered variants include labels of one stream in the same leaf nodes. Consequently, it is necessary to consider the combination of Ordered and Signature R*-trees to indexing XML data. Moreover, we see that the average count of pages accessed (measured by DAC) $<<$ the average count of pages of one stream (12,536.8 in this case) for query groups 2 and 3. In the case of the Query group 1, we read all labels in a stream and DAC is approximately equal to the count of pages in the streams.

4 Conclusion

In this article, we introduced a multidimensional implementation of the stream index for the path labeling schemes related to holistic join approaches. We also show that this stream index can be extended in such a way that it supports fast searching in XML nodes with a content. Our preliminary results showed that it is necessary to combine two variants of the R*-tree: Ordered and Signature R*-trees. In our future work, we should compare the performance of our solution with other stream implementations for real XML queries. We must also solve the storage of variable-length labels since the solution using the multidimensional forest is not optimal.

Projects and Publications

Filip Křižka is a member of team solving the grant of GAČR, Czech Republic, No. P202/10/0573, Handling XML Data in Heterogeneous and Dynamic Environments and the grant of SGS, Czech Republic, No. SP/2010138, Datové struktury podporující techniky indexování XML dat and the grant of GAČR, Czech Republic, No. GA201/09/0990, Zpracování XML dat and SGS grant SP2011/172,

Detekce plagiátů v textových dokumentech, řešitel: Jan Martinovič. He is co-author of four publications in conference proceedings [3, 2, 1, 4]. One publication is indexed in Web of Science.

Own References

1. F. Křižka and M. Krátký. On the Efficient Indexing of Ordered Multidimensional Tuples. In *5th International Conference for Internet Technology and Secured Transactions (ICITST 2010), London*. IEEE, 2010.
2. F. Křižka, M. Krátký, and R. Bača. Benchmarking a B-tree Compression Method. In *Proceedings of the Conference on Theory and Practice of Information Technologies (ITAT 2009)*, 2009.
3. F. Křižka, M. Krátky, and R. Bača. On Support of Ordering in Multidimensional Data Structures. In *Proceedings of the East European Conference on Advances in Databases and Information Systems (ADBIS), Novi Sad*. Springer-Verlag, 2010.
4. F. Křižka, M. Krátký, R. Bača, and P. Chovanec. Multidimensional Implementation of Stream ADT. In *Proceedings of The Third International Workshop on XML Data Management (XMLDM 2011)*. Springer–Verlag, 2011, accepted.

Other References

5. M. G. Bauer, F. Ramsak, and R. Bayer. Multidimensional mapping and indexing of xml. In *BTW*, volume 26 of *LNI*, pages 305–323. GI, 2003.
6. N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An Efficient and Robust Access Method for Points and Rectangles. In *Proceedings of the 9th ACM International Conference on Management of Data (SIGMOD 1990)*, 1990.
7. P. Chovanec, M. Krátký, and R. Bača. Optimization of Disk Accesses for Multidimensional Range Queries. In *Proceedings of the 21st International Conference on Database and Expert Systems Applications (DEXA 2010)*, pages 358–367. Berlin: Springer, 2010.
8. T. Grüst. Accelerating XPath Location Steps. In *Proceedings of SIGMOD 2002*, pages 109–120. ACM Press, 2002.
9. A. Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In *Proceedings SIGMOD 1984*, pages 47–57. ACM Press, 1984.
10. I. Kamel and C. Faloutsos. On packing R-trees. In *Proceedings of the Second International Conference on Information and Knowledge Management (CIKM 1993)*, pages 490–499. ACM Press, 1993.
11. M. Krátký, J. Pokorný, and V. Snášel. Implementation of XPath Axes in the Multidimensional Approach to Indexing XML Data. In *Current Trends in Database Technology – EDBT 2004 Workshops*, volume 3268. Springer–Verlag, 2004.
12. M. Krátký, T. Skopal, and V. Snášel. Multidimensional Term Indexing for Efficient Processing of Complex Queries. *Kybernetika, Journal*, 40(3):381–396, 2004.
13. M. Krátký, V. Snášel, P. Zezula, and J. Pokorný. Efficient Processing of Narrow Range Queries in the R-Tree. In *Proceedings of the 10th International Database Engineering and Applications Symposium (IDEAS 2006)*, pages 69–79. IEEE, 2006.
14. S. S. Lightstone, T. J. Teorey, and T. Nadeau. *Physical Database Design: the Database Professional's Guide*. Morgan Kaufmann, 2007.
15. H. Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, 2006.

Multi-Agent Systems and Transparent Intesional Logic

Jakub Macek and Tomáš Frydrych

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jakub.macek1@vsb.cz, frydrych.t@gmail.com

Abstract. In this paper, we focus on the concept of time from the perspective of an agent in a Multi-Agent Systems. We propose a method of modelling dynamic behaviour of these systems using Transparent Intensional Logic. We also describe several issues where the formalization affects the implementation of both the system and the agents, and propose solutions to the problems that may arise.

Keywords: TIL, Transparent Intensional Logic, multiagent system, agent, communication

1 Introduction

Multi-agent system (MAS) is a system of autonomous, intelligent, but resource-bounded agents. The agents are active in their perceiving environment and acting in order to achieve their individual as well as collective goals. To this end the agents have to be able to communicate and collaborate with each other as well as with their environment by exchanging messages. This paper shows the basic possibilities of using Transparent Intensional Logic in Multi-agent system.

The paper is organized as follows. In the next section we propose basic describe of Multi-Agent system. In section 2 we are introducing Transparent Intensional Logic. In the third section we'll discuss the time problem in MAS. Finally, in section 4 we describe the communication of the agents in MAS.

2 Transparent Intensional Logic

Transparent Intensional Logic (TIL)[1] is a logical system founded by Prof. Pavel Tichý[2]. It is a higher-order system primarily designed for the logical analysis of natural language. As an expressive semantic tool it has a great potential to be utilised in artificial intelligence and in general whenever and wherever people need to communicate with the computers in a human-like way.

Due to its rich and transparent procedural semantics, all the semantically salient features are explicitly present in the language of TIL constructions.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 411–416.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

It includes explicit intensionalisation and temporalisation, as well as hyper-intensional level of algorithmically structured procedures (known as TIL constructions) which are of particular importance in the logic of attitudes, epistemic logic and the area of knowledge representation and acquisition.

3 Multi-Agent System design

Simulation of a real system in a virtual one poses several challenges. Usually this is related to the limited computational capacity and thus it is solved by simplifying the problem. In case of MAS, however, there is one aspect of real world that cannot be reproduced in software and necessitates the use of a different approach.

One of the key parts of implementation is the TIL inference machine. For the purpose of this paper, we define inference machine as a black-box component, which takes knowledge base (as a set of TIL constructions) and a query (a TIL construction) as an input and produces positive answer if the query can be inferred from the given knowledge base and negative answer otherwise.

3.1 Time

Time is an essential aspect of the MAS system and there are two possible ways how to handle it. Either it can be a part of implementation or a part of formalization. Since TIL allows very easy usage of time in inference, we prefer the latter. This also transparently prevents possibility of contradictions in the knowledge base due to improper handling of time. For the purposes of a MAS system, we can easily use the type of real numbers to map the time moments, presupposing the hypothesis of continuum.

By including the time aspects into formalization, the agents can effectively take into account both past states of the system and predict future possibilities based on trustworthy facts (e.g. "reliable agent TrafficSign17 sent me message that for the next 20 system ticks the Crossing5 will be closed, so I won't plan my route through there"). Such information can be stored in agent's knowledge base in the form of rules.

3.2 Centralization

In a real world, all systems are decentralized in a sense that no individual part of the system has a complete control over any other part. Thus there is no central dispatcher to control the behaviour of the system. Hence a real-world system is unpredictable and we have to take that into account.

To adduce an example, consider two robotic agents. First robot jumps and lands. Afterwards neither of them need to consult any central authority to determine the jumper's current position; both agents simply accept the fact and store it into their individual knowledge-bases. Furthermore, the actual position is guaranteed to be consistent for both agents, so the stored knowledge depends only on the ability to observe the actual state of the world correctly.

However, this is not possible in virtual systems. To properly model real world systems it is necessary to include a central authority that determines the objective state of the system. We can benefit from implementing this central individual as a passive agent (these are kind of agents with very limited capability, their purpose is to pass along information which they receive from other agents in the system).

Due to the lack of a counterpart for this agent in real world systems and its special position, we refer to it as the *system* ("agent queries the *system*") or more specifically the *system agent* in case of ambiguity.

We are faced with similar problem when dealing with agents' ability to perceive the state of the system. The solution to this problem varies depending on the purpose of the concrete MAS. Generally, the two possibilities are either to feed the information to the agent constantly (being more true to the real world) or to provide sensory information on demand.

First case requires implementation of short-term memory, otherwise the agent would only be able to base its decisions on current sensory input, which usually doesn't correspond to reality. The price for this simpler approach is a rapid expansion of agent's knowledge base, because the agent can be overloaded with facts that it cannot manage.

More complicated is the implementation of a query based perception. Whenever an agent is confronted with a fact that it cannot understand, the agent must put a query to the system (e.g. "Am I seeing a pedestrian?", "Have I ever seen a yellow car?"). These queries can become quite complicated. Moreover, the system cannot give an answer beyond the scope of the querying agent's perception ("Has anyone seen a pedestrian?").

3.3 Active agents

The second constraint on agent's knowledge base is the possibility of false knowledge. This would project into agent's inferred facts and if allowed to spread, it would also break down the entire system. Similar effect can be caused by extending agent's inference beyond deduction (for example with inductive reasoning).

An agent itself cannot directly affect the state of the MAS. The agent can, however, control (to a degree) its own actions. These however must be limited by system's "natural laws" and determination of consequences of all actions fall to the system agent.

To properly formalize this chain of responsibility, we must separate the agents decision to take some action from the occurrence of said action.

Actions in the knowledge base are represented as simple predicate functions that take at least three arguments: world, time and agent taking the action. For example:

$$[[[{}^0JumpTo\ {}^0World]\ {}^0Time1]\ {}^0Agent1\ {}^0Place1]$$

represents that in the MAS given by $World$, the $Agent1/\iota$ has $Jump$ed at $Time1/\tau$ to $Place1/\iota$. The action should be always formalized with the most specific description available. We can also extend system's rules in a way that mentioned construction entails the fact that the agent jumped.

$$[\,[\,[\,{}^0Jump\;{}^0World]\;{}^0Time1]\;{}^0Agent1]$$

3.4 Decisions a simulation

To include the action in system's knowledge base, the agent must communicate its intention to take the action to the system. The system will then proceed with determining the effects of a given decision and inform the agent of these effects.

When there is a decision required of the agent, the system sends it to the agent for notifying. The agent will then use all available decision schemes to generate possible decisions. Then the agent will create sets of decisions that it wants to simulate. This step has a great influence on the efficiency of the agent.

Single decision per simulation is the most simple option, but such an agent would exhibit almost no foresight. Better would be combining more decisions that occur sequentially in time. It's fairly obvious that an unoptimized solution would lead to exponential increase in required computational time. To properly represent reality, we need to allow not only sequential decisions, but also concurrent ones.

Practically, there needs to be a limit on how many decisions can agent make at once and how far into the future. This more or less represents what might be called agent's intelligence.

Each set of decisions will be then added with agent's private knowledge base into a simulation knowledge base. The actual simulation is not a process based on a time-line, as it is usually understood, but merely a computation of desirability of the resulting state of the system. This can be done by inferring how many of agent's goals were achieved (goals are further described in section about formalization).

When finished, the agent will choose set of decisions with most desirable effects and send the decision the system asked for. The agent may simulate its decisions in future, but only the immediate decisions are relevant to the system.

4 Time problem

To show insufficiency of the solution without using temporal aspect of the problem, let's define more simpler game with limited scope. We will use just one agent-player (denoted by Z) and a map consisting of two places (A and B). The player will be able to move between these two places without restriction. We will introduce appropriate rules for the player movements into the system. We also have to include a rule stating that the player can be at one place at a time.

At the beginning, the player will be at place A. Knowledge base of game thus contains these facts and rules:

$$\forall x((LocatedAt(x, A) \wedge MovesTo(x, B)) \supset LocatedAt(x, B))$$
$$\forall x((LocatedAt(x, B) \wedge MovesTo(x, A)) \supset LocatedAt(x, A))$$
$$\forall x(LocatedAt(x, A) \supset \neg LocatedAt(x, B))$$
$$\forall x(LocatedAt(x, B) \supset \neg LocatedAt(x, A))$$
$$LocatedAt(Z, A)$$

And as a result of the last fact, we can also deduce that $\neg LocatedAt(Z, B)$.

Now the player makes a decision to move in the place B: $MovesTo(Z, B)$ and the system infers which is the new position of the player:

$$\forall x((LocatedAt(x, A) \wedge MovesTo(x, B)) \supset LocatedAt(x, B)) \wedge MovesTo(Z, B)$$
$$\models LocatedAt(Z, B)$$

We can now see that this conclusion contradicts the assumption that $\neg LocatedAt(Z, B)$. Obviously, the problem is due to the fact that we did not take into account temporal aspects of agents positions. This can be solved by removing invalid facts from knowledge base, but a more elegant solution is to introduce time to our formal representation of the system.

5 Communication

Basic standards for communication in MAS are given by FIPA (The Foundation for Intelligent Physical Agents). A message, which can take up an arbitrary form, is defined as the basic unit of communication. According to FIPA standards, a message should contain a set of attributes. The most important attributes of communication are performative, content and ontology.

A crucial role among these attributes plays the content of the message. Perfomative denotes a type of the message. Basic performatives are Query, Inform and Request. Content carries the semantic of the message. It can be encoded in any suitable language. We use TIL constructions as a content language of agent messages. Ontology is a vocabulary of the domain-specific terms. These (and only these) terms can be used in the content of the message. More about agent communication and architecture of an intelligent agent in MAS see in[3].

For example we demonstrate simple communication between MAS and agent. We have two agents, first for MAS called Agent0, second to represent figure, called Agent1. In our example Agent1 needs to know his location. So Agent1 sends message to Agent0. Content of this message is query to Agent1's location:

$$\lambda w \lambda t\, [\,{}^0 Message_{wt}\, {}^0 Agent1\, {}^0 Agent0\, \lambda w \lambda t\, [\,{}^0 Query_{wt}\, {}^0 Location]]$$

Agent0 receives message and responds with current location of Agent1:

$$\begin{aligned}&\lambda w \lambda t\, [\,{}^0 Message_{wt}\, {}^0 Agent0\, {}^0 Agent1\\ &\lambda w \lambda t [\,{}^0 Answer_{wt}\, {}^0 Location [[\,{}^0 LocatedAt\; w]\, {}^0 14]\, {}^0 Agent1\, {}^0 Place3]]\end{aligned}$$

6 Conclusion

In this paper, we have demonstrated the basic possibilities of using Transparent Intensional Logic in Multi-agent system. We demonstrated a method for the definition of facts, rules and communication between agents in the multi-agent system that is changing in time and presents solution using Transparent Intensional Logic.

Acknowledgements This research has been supported by the Grant Agency of the Czech Republic, projects 401/09/H007 'Logical Foundations of Semantics' and 401/10/0792 'Temporal aspects of knowledge and information', and also by the internal grant agency of VSB-TU of Ostrava - SP/2010214 Modeling, simulation and verification of software processes II.

References

1. Duží, M., Jespersen, B., Materna, P.: *Procedural Semantics for Hyperintensional Logic (Foundations and Applications of TIL).* Springer. 2010. 17. Logic, Epistemology, and the Unity of Science. ISBN
2. Tichý, P. (2004): *Collected Papers in Logic and Philosophy*, V. Svoboda, B. Jespersen, C. Cheyne (eds.), Prague: Filosofia, Czech Academy of Sciences, and Dunedin: University of Otago Press.
3. Duží, M., Ciprich, N., Košinár, M., Kohut, O., Frydrych, T.: *The Architecture of an Intelligent Agent in MAS.* Waki Print Pia, Kanagawa, Japan. 2008

TIL and Ontology Specification

Marek Menšík[1,2] and Lukáš Vích[1]

[1]Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
[2]Silesian University in Opava, Bezručovo nám. 13, 746 01 Opava
mensikm@gmail.com, vic097@vsb.cz

Abstract. In this paper we introduce Transparent Intensional Logic (TIL) as a tool apt for the specification of ontologies and knowledge representation. Ontology is characterized as the result of conceptualization of the domain under scrutiny. We specify the structure and content of ontology and the notion of ontology is distinguished from the notion of conceptual analysis. We show that TIL can serve as a highly expressive universal framework that makes it possible to rigorously specify the content of ontology and respective knowledge base, including the relations between particular concepts.

1 Introduction

In this paper we introduce an otology and specification language based on the Transparent Intensional Logic (TIL). From the formal point of view, TIL is a hyperintensional, typed partial lambda calculus. A main feature of the λ-calculus is its ability to systematically distinguish between functions and functional values. In this paper we also introduce TIL as an useful tool for ontology specification.

2 Basic principles of TIL

TIL[1] is an overarching semantic theory for all sorts of discourse, whether colloquial, scientific, mathematical or logical. The theory is a procedural one, according to which sense is an abstract, pre-linguistic procedure detailing what operations to apply to what procedural constituents to arrive at the product (if any) of the procedure. Such procedures are rigorously defined as TIL *constructions*. The semantics is entirely anti-contextual and compositional and it is, to the best of our knowledge, the only one that deals with all kinds of context in a uniform way. Thus the sense of a sentence is an algorithmically structured *construction* of the proposition denoted by

[1] For more details see [1].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 417–422.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

the sentence. The denoted proposition is a flat, or unstructured, mapping with domain in a logical space of possible worlds. Our motive for working ‘top-down’ has to do with anti-contextualism.

Intuitively, construction C is a *procedure* (a generalised algorithm). Constructions are *structured* in the following way. Each construction C consists of sub-instructions (constituents), each of which needs to be executed when executing C. Thus a specification of a construction is a specification of an instruction on how to proceed in order to obtain the output entity given some input entities.

There are two kinds of constructions, atomic and compound (molecular). Atomic constructions (*Variables* and *Trivializations*) do not contain any other constituent but themselves; they specify objects (of any type) on which compound constructions operate. The *variables x, y, p, q, ...,* construct objects dependently on a valuation; they v-construct. The *Trivialisation* of an object X (of any type, even a construction), in symbols $^{\mathbf{0}}\boldsymbol{X}$, constructs simply X without the mediation of any other construction. *Compound* constructions, which consist of other constituents as well, are *Composition* and *Closure*. *Composition* $\boldsymbol{[F\ A_1 \ldots A_n]}$ is the operation of functional application. It v-constructs the value of the function f (*valuation-*, or v-, -constructed by F) at a tuple argument A (v-constructed by A_1, …, A_n), if the function f is defined at A, otherwise the Composition is *v-improper*, i.e., it *fails* to v-construct anything.[2] *Closure* $\boldsymbol{[\lambda x_1 \ldots x_n\ X]}$ spells out the instruction to v-construct a function by abstracting over the values of the variables $x_1,\ldots,x_n$ in the ordinary manner of the λ-calculi. Finally, higher-order constructions can be used twice over as constituents of composite constructions. This is achieved by a fifth construction called *Double Execution*, $^{\mathbf{2}}\boldsymbol{X}$, that behaves as follows: If X v-constructs a construction X', and X' v-constructs an entity Y, then ^2X v-constructs Y; otherwise ^2X is *v-improper*, failing as it does to v-construct anything.

TIL constructions, as well as the entities they construct, all receive a type. The formal ontology of TIL is bi-dimensional; one dimension is made up of constructions, the other dimension encompasses non-constructions. On the ground level of the type hierarchy, there are non-constructional entities unstructured from the algorithmic point of view belonging to a *type of order 1*. Given a so-called *epistemic (or objectual) base* of *atomic types* (**ο**-truth values, **ι**-individuals, **τ**-time moments / real numbers, **ω**-possible worlds), the induction rule for forming functional types is applied: where α, $\beta_1,\ldots,\beta_n$ are types of order 1, the set of partial mappings from $\beta_1 \times\ldots\times \beta_n$ to α, denoted ‘$(\alpha\ \beta_1 \ldots \beta_n)$’, is a type of order 1 as well. Constructions that construct entities of order 1 are *constructions of order 1*. They belong to a *type of order 2*, denoted ‘$*_1$’. The type $*_1$ together with atomic types of order 1 serves as a base for the induction rule: any collection of partial mappings, type $(\alpha\ \beta_1 \ldots \beta_n)$, *involving* $*_1$ in their domain or range is a *type of order 2*. Constructions belonging to a type $*_2$ that identify entities of order 1 or 2, and partial mappings involving such constructions, belong to a *type of order 3*. And so on *ad infinitum*.

---

2 As mentioned above, we treat functions as partial mappings, i.e., set-theoretical objects, unlike the *constructions* of functions.

The sense of an empirical expression is a *hyperintension* that is a construction that produces a (possible world) *α-intension*, where α-intensions are members of type (αω), i.e., functions from possible worlds to an arbitrary type α. On the other hand, *α-extensions* are members of a type α, where α is not equal to (βω) for any β, i.e., extensions are functions whose domain is not the set of possible worlds.

Intensions are frequently functions of a type ((ατ)ω), i.e., functions from possible worlds to *chronologies* of the type α (in symbols: $\alpha_{\tau\omega}$), where a chronology is a function of type (ατ).

Notation: An object A of a type α is denoted 'A/α'. That a construction $C/*_n$ *v*-constructs an object of type α is denoted '$C \rightarrow_v \alpha$'. We use variables *w* and *t* as *v*-constructing elements of type ω (possible worlds) and τ (times), respectively. If $C \rightarrow_v \alpha_{\tau\omega}$ *v*-constructs an α-intension, the frequently used Composition of the form [[*Cw*]*t*], the intensional descent of the α-intension, is abbreviated 'C_{wt}'. The analysis of a sentence consists in discovering the logical construction (procedure) encoded by a given sentence. To this end we apply a *method of analysis* that consists of three steps:[3]

1) *Type-theoretical analysis*, i.e., assigning types to the objects that receive mention in the analysed sentence.
2) *Synthesis*, i.e., combining the constructions of the objects *ad* (1) in order to construct the proposition of type $o_{\tau\omega}$ denoted by the whole sentence.
3) *Type-Theoretical checking*.

Let us analyse the sentence "Tom is seeking a parking place".

Ad (1) types: *Tom*/ι; $Seeking/(o\iota(o\iota)_{\tau\omega})_{\tau\omega}$: the relation-in-intension of an individual to a property the instance of which the individual wals to find; $Parking/(o\iota)_{\tau\omega}$: a property of individuals. The whole sentence denotes a proposition, i.e. an object of type $o_{\tau\omega}$.

Ad (2). We aim at *literal analysis* of the sentence. Thus semantically simple expressions are analysed as Trivializations of the objects they denote: ${}^0Tom$, ${}^0Seek$, ${}^0Parking$. Now we need to apply the relation *Seeking* to *Tom* and *Parking*. However, the relation-in-intension is not a type-theoretically proper object to be applied; it must be extensionalised first: $[[{}^0Seeking\ w]\ t] \rightarrow (o\iota(o\iota)_{\tau\omega})$, or ${}^0Seeking_{wt}$ for short. The Composition $[{}^0Seeking_{wt}\ {}^0Tom\ {}^0Parking] \rightarrow_v o$ *v*-constructs the truth-value **T** or **F** according as in a given world *w* and time *t* of evaluation Tom is seeking a parking or not. In order to construct the proposition denoted by the sentence, we must now abstract over the values of *w* and *t*. Thus we have:

$$\lambda w \lambda t\ [{}^0Seeking_{wt}\ {}^0Tom\ {}^0Parking] \rightarrow o_{\tau\omega}$$

Ad (3) Finally we perform a type-theoretical control in order to check whether particular constituents are combined in compliance with the definition of constructions.

$$\begin{array}{l} \lambda w \lambda t\ \ [{}^0Seeking_{wt}\qquad {}^0Tom\qquad {}^0Parking] \\ \qquad \underbrace{(\,(o\iota(o\iota)_{\tau\omega})\qquad \iota\qquad (o\iota)_{\tau\omega}}_{o} \\ o_{\tau\omega} \end{array}$$

3 For details see [2].

3 Ontology content and representation[4]

Formal ontology is a result of the conceptualization of a given domain. It contains definitions of the most important entities, forms a conceptual hierarchy together with the most important attributes and relations between entities. Material individuals are mereological sums of other individuals, but only *contingently* so. Similarly, values of attributes and properties are ascribed to individuals contingently, provided a given property is purely contingent, that is without an essential core. Thus we advocate for a (modest) *individual anti-essentialism*. On the other hand, on the intensional level of propositions, properties, offices and roles, that is entities which we call 'intensions', the most important relation to be observed is that of *requisite.* For instance, the property of being a mammal is a requisite of the property of being a whale. It is an *analytically necessary* relation between intensions that gives rise to the so-called *ISA hierarchy.* Thus on the intensional level we advocate for *intensional essentialism*; an essence of a property is the set of all its requisites. Finally, on the hyper-intensional level of concepts, relations to be observed are *equivalence* (i.e. producing the same entity), *refinement* (a compound concept is substituted for a simpler yet equivalent concept), *entailment* and *presupposition.*

The structure of ontology building starts on the hyper-intensional level with the specification of primitive concepts. Next we specify compound concepts as ontological definitions of entities of a given domain. Having defined entities, we can specify their most important descriptive attributes. The building process continues by specifying particular (empirical) relations between entities and analytical relations of requisites that serve to build up ontological hierarchy. Finally, the most important general rules that govern behaviour of particular agents are specified. Here again we distinguish *analytically necessary* constraints from *nomic* and *common necessities* that are given by laws and conventions, respectively; they are not valid analytically necessary. Summarising, basic parts of a formal ontology should encompass:

(1) Conceptual (terminological) dictionary which contains:
   a) primitive concepts
   b) compound concepts (ontological definitions of entities)
   c) most important descriptive attributes, in particular identification of entities
(2) Relations
   a) empirical relations between entities, in particular the part-whole relation
   b) analytical relations between intensions, i.e., requisites and essence, which give rise to ISA hierarchy
(3) Rules
   a) Analytically necessary rules
   b) Nomologically necessary rules
   c) Common rules of 'necessity by convention'

Ontological definition of an entity is a compound construction of the entity. Such a definition often serves as a refinement of a primitive concept of the entity, which makes it possible to prove some analytic statements about the respective concepts.

[4] There are many frameworks dealing with ontology problematics, more in [3],[4]

3.1 Agents' Ontology

The terminological vocabulary is determined by the set of primitive concepts which are Trivializations of non-constructions. Compound concepts are then ontological definitions of so-constructed objects. Attributes are empirical functions of type $(\alpha\beta)_{\tau\omega}$. If α is a numeric type, then the attribute is descriptive. Consistency constraints are constructions of propositions which determine admissible states of a knowledge base. The degree of necessity of a given constraint is specified as follows. If $C \rightarrow_v$ o v-constructs the respective condition to be met, the basic kinds of constraints ordered from the highest to the lowest are:

a) Analytically necessary rules; these are specified as $\forall w \forall t\ C$.

b) Nomic necessary rules; these are specified as form $\lambda w \forall t\ C$.

c) Common rules of 'necessity by convention'; these are constructions of the form $\lambda w \lambda t\ \forall x\ [C \ldots x \ldots]$.

For instance, the rule that a road consists of at least one and at most eight lanes is empirical necessity by convention. It is this Closure:

$$\lambda w \lambda t\ \forall x\ [[{}^0Road_{wt}\ x] \supset [{}^0\exists \lambda y\ [[{}^0Part\_of_{wt}\ y\ x] \wedge [{}^0Lane_{wt}\ y]] \wedge [[{}^0Card\ \lambda y\ [[{}^0Part\_of_{wt}\ y\ x] \wedge [{}^0Lane_{wt}\ y]]] \leq {}^08]]]$$

On the other hand, the rule that a road is an element of the traffic infrastructure is analytical. It is specified by this construction:

$$[{}^0Req\ {}^0Element\_Infra\ {}^0Road] = \forall w \forall t\ [\forall x\ [[{}^0Road_{wt}\ x] \supset [{}^0Element\_Infra_{wt}\ x]]]$$

We use two specific types which serve for agents' learning, sharing and distribution of knowledge:

- *Unrecognised*/$(o{*}_n)_{\tau\omega}$: the respective message content $\lambda w \lambda t\ [{}^0Unrecognised_{wt}\ {}^0What]$ informs the receiver *Whom* that the concept *What* has not been recognized, it is not contained in the sender *Who*'s ontology. Thus the message is a request for refinement.

- *Refine*/$({*}_n{*}_n)_{\tau\omega}$: the function that assigns to a construction *What* another construction *C*' such that *C*' is equivalent to *What* by constructing the same entity, and *C*' is a more detailed instruction than *What.*
  $\lambda w \lambda t\ [[{}^0Refine_{wt}\ {}^0What] = {}^0C']$.

Here is an example of a scenario. An agent *a* is asking an agent *b* to inform about the action he is just executing; *b* informs *a* that he is moving; However, *a* does not have a primitive concept ${}^0Moving$ in its ontology. For this reason *a* sends a message of the type *Unrecognised* asking for the definition of this concept:

$$\lambda w \lambda t\ [{}^0Unrecognised_{wt}\ {}^{00}Moving]$$

The agent *b* then answers by a message of type *Refine* that contains the respective definition informing *b* that the concept of moving is defined here as going to a specified destination by means of a conveyor: $\lambda w \lambda t\ [[{}^0Refine_{wt}\ {}^{00}Moving] =$

$${}^0[\lambda w \lambda t\ \lambda x\ \exists y\ \exists z\ [[{}^0Going\_to_{wt}\ x\ y\ z] \wedge [{}^0Destination_{wt}\ x\ y] \wedge [{}^0Conveyor_{wt}\ z]]]$$

Supposing that the primitive concepts ^{0}*Going_to,* ${}^0Destination_{wt}$, ^{0}*Conveyor* the agent *a* 'knows' they can store the new piece of knowledge into their ontology and then ask for additional information like, for instance, the type of a conveyor.

Conclusion

In this paper we described basic principles of TIL. Also we specified a typical content of ontology, and the main tenets of building and sharing the ontology by particular agents. The need for a fine-grained logical semantics has been demonstrated. We advocated for TIL with its expressive *procedural semantics*, because this is the way how distinguish between the three kinds of context: extensional, intensional and hyperintensional. Due to its rich ontology, TIL operates on all the three levels of abstraction smoothly, in an extensional way.

Acknowledgements

This research has been supported bby the Grant Agency of the Czech Republic, project No. 401/09/H007 "Logical Foundations of Semantics" and by the project No 401/10/0792 "Temporal aspects of knowledge and information", and also by the internal grant agency of VSB-TU Ostrava, project SP2011/56 „Znalostní přístup při modelování, simulaci a vizualizaci softwarových procesů".

References:

[1] Duží, M., Jespersen, B. and P. Materna: *Procedural Semantics for Hyperintensional Logic*; *Foundations and Applications of Transparent Intensional Logic.* Springer: series for Logic, Epistemology and the Unity of Science, Vol. 17, 2010.

[2] Gordon, M. J. C. and Melham, T. F. (eds.) 1993: Introduction to HOL: A Theorem Proving Environment for higher Order Logic. Cambridge: Cambridge University Press.

[3] Kifer, M., Lausen, G., and James Wu. Logical foundations of object- oriented and frame-based languages. Journal of the ACM, 42(4):741-843, 1995.

[4] Baader, F., Calvanese, D., McGuinness, D., L., Nardi, D., and Patel-Schneider, P., F., editors. The Description Logic Handbook: Theory, Implementation and Application. Cambridge University Press, 2002.

A Note on Realization of A Multi Camera Railway Crossing Safety System

Karel Mozdřeň and Radovan Fusek

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{karel.mozdren, radovan.fusek.st}@vsb.cz

Abstract. Increasingly we witness the accidents, causing not only incalculable damage to railway electrification and security systems, but also losses of lives. The train, derailed at high speed, can destroy great area of those systems, and paralyze the traffic for tens or hundreds hours. An obstacle, that prevents free passage trough the hazardous area, could be the source of such an accident. Cameras are suitable for monitoring such areas and recording them for a long period of time, but in combination with a suitable computer vision algorithm, it can completely autonomously inform us about extraordinary situations. In this paper we examine the problem of monitoring the critical areas using multiple cameras and combination of camera images to obtain large enough precision to determine hazard level of the situation in the secured area. Furthermore, we show an example of possible use of the system in practice and in the end we discuss the results of our work.

1 Introduction

Damage to the traction system have dire consequences. Besides the possible loss of life or bodily injury there could arise considerable financial cost to repair the damage and it is also necessary to provide an alternative traffic at the time of the failure. Many things must planned and executed eg. shutdown, servicing, inspections, investigations and other various activities that follow after an accident. Most companies know what to do after the accident, as they have prepared emergency plans. However, the main effort of every company should be accident prevention. The accident itself is the result of a combination of one or more factors that may not independently cause the accident, but their combination may be crucial for its rise. Prevention of Accidents is the set of actions that minimize the risk of its occurrence, or at least its consequences. We focused on the use of cameras as a tool for minimalization of damage. We are trying to find hazards before an accident occurs.

Most areas that are guarded by a camera system are operated by humans. These days it is required to use the computer vision algorithms to ensure continuous supervision of a supervised area. Examples of such an applications are measurement of trafic parameters, detecting objects and their position etc. [1][3].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 423–429.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

We secure the area near the railway electrification and safety system. In this area obstacles could derail the train and it might harm people or damage installed wiring and safety equipment. In this paper we present a method for monitoring of these areas through a system of cameras and combination of their images using computer vision algorithm in such a way that any obstacle is accurately measured and hazard level is determined. This paper is organized as follows. First, we familiarize readers with the critical area and its classification. Furthermore, we present a method for measuring the size of an obstacle, and finally we show an example of practical application.

2 Critical Area

Critical area is a place where an increased risk of the accident occurs. In our case, the railway area. If an object (obstacle) is located on the railway track, we have to remove it from this area or inform the responsible person as soon as possible because it could cause a train derailment and subsequent damage to the most of the systems. The most critical areas are those where other objects besides trains (cars, pedestrians, etc.) often enter into the track. Examples of such an areas are train stations, railway crossings, overpasses, bridges and other structures that may fall into the track, or areas near other structures from which dangerous objects may fall off. Those are usually secured only by prohibition of entry. For this the signs, trafic lights, and sometimes even camera systems are used.

Camera systems have the advantage that a worker performing a monitoring function does not have to be directly present in the secured area, but still can watch all the action on the screen. In addition, video is stored, which is a good source of information for the investigation of any accident event. We also extend this system by obstacle detection and its measurement from the image. This is important from the point of view, that man, no matter how vigilant and observing he is, can not monitor what is happening in all critical areas at once. Our method allows to inform the security worker in a short time of areas, where something could be wrong.

3 Camera System

Most of the applied camera systems are aimed to cover the largest area possible. However, we try to cover relatively small area with more than one camera. We do this because we want to measure the position and size of the object more precisely. Using only one camera, we get only a little and distorted information about the object. But if we look at it from several angles, we get much more information. In our method, we measure the base of the object, which allows us to find out if it intervenes with the critical region. This would be very difficult without the use of at least two cameras. If we use just one, we see it only from one perspective, and do not know how it looks on the other side. The greater the number of cameras are used, the greater the accuracy of the measurement we get.

Besides the number of cameras it is also necessary to consider their position around the critical area. It makes little sense to have cameras close to them selfs, because they would provide nearly the same image. It is advisable to place the camera against each other. Then the object would be covered from most of the sides. For more than two cameras, the placement should be on a circle that encloses the supervised area and the angular distance between those cameras should be the same (Fig. 1), but it is not necessary. It is also advisable to choose the number of cameras as multiples of two, because it allows us to always have a camera in pairs. It turns out that in the most cases two cameras are sufficient and it is not necessary to place them in exactly the same distance from the critical area. But It is important that all cameras see the whole supervised area.

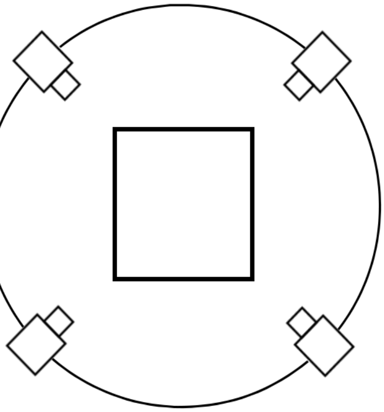


Fig. 1. Example of the camera positioning around the critical area

4 Measurement of the Obstacle Basement

The actual measurement of the size of the obstacle is a task which consists of several sub-tasks. First, we have to obtain images from all cameras. Then, all the images have to be transformed into the same space using top-view transformation. Next we use background detection algorithm on all of the images and using background substracion we get segment, representing the area of the obstacle. The combination of these images is used to find the base of the object. The next section describes all these steps in detail.

4.1 Image acquisition

IP cameras are the best choice for our purpose because of the easy connectability using computer networks. Most of them provide picture using various protocols such as HTTP, MJPEG, and others. Therefore, getting the images from the cameras is relatively simple. The problem that arises is the synchronization of the images. Each camera sends an image slightly time-shifted and it is therefore necessary to synchronize the images. Synchronization is well described in this publication [4], and therefore I will not describe it in detail. Generally it is sufficient to measure the delay of the images provided by all cameras. By measuring the time differences we select only those images that correspond with each other and we use them for further processing.

4.2 Top-View Transformation

To combine the images, we have to transform them into the same space (plane) first. We transform all the images to the view from above. We use geometric transformation known as the homography. It is a transformation converting image from one plane to another. Practically this means marking some corresponding points in all images and then using them for homography to transfer them into the same top-view image space. For simplicity, the points at the corners of the critical area are used. Boundaries of critical areas are chosen in the form quadrangle and so we have to handle four points only. Homography is homogeneous transformation, which generally has the form (1)

$$x' = Hx, \tag{1}$$

where H is the homography transformation matrix, x is the point coordinates vector in the input image, and x' is the coordinates vector in the ouput image after the transformation. We are trying to find transformation matrix H using four correspondence points. So we get eight equations (four points in one and four points in the second picture) and 8 unknowns. This leads to the exact solution and we do not need to use a minimizing scheme. It should be noted that computations of those matrices are performed only once at the start of the application and then it is used to perform the transformation for the whole time. This could have been done only because cameras does not change their possition (they are static). For a deeper understanding of the homography we refer to this publication [2].

4.3 Background substraction

This step marks the picture pixels of the obstacle. This is where methods of background substraction are used. Those are based on modeling of the background from the video sequence and finding the difference of this model with the current frame obtained from camera. As a result we get affiliation of the picture areas to the background or foreground. This means that image points marked as background have values approaching to zero and points marked as foreground have values approaching to one.

There are many methods for modeling of the background like the observation of average intensity of each image point, a mixture of Gaussians (MOG), or a methods based on the analysis of principal components (PCA). Methods with the smallest error are MOG and PCA, but considering the conclusions from the authors of the publication [5], where they proposed the idea that the results does not depend that much on the method for background detection as much as on the preprocessing and postprocessing of the image (for example image filtering and mathematical morphology). In our case we have used the method known as the running average with selectivity [6] for its speed and simplicity. This method also does not require retention of several recent images in the memory, and background modeling is determined by following formula (2)

$$B_{t+1} = \alpha F_t(x, y) + (1 - \alpha)B_t(x, y), \tag{2}$$

in the case that the point $F_t(x, y)$ is marked as a backround, and in the case it is marked as a foreground by this formula (3)

$$B_{t+1} = B_t(x, y). \tag{3}$$

4.4 Cameras Images Combination and Hazard Level Estimation

Measurement of the base of an object is a crucial part of our method. The base could be understood as an area in the image covered by the object when viewed from above. Best results would be achieved with a camera mounted above the critical area so it would see it from above. In many cases it is not technically feasible and the camera views the region from the side. This considerably complicates the determination of the size and position of the object. A control of evacuation of the critical area, based only on images from single a camera leads to false alarms caused by nearby objects. This alarm arises when the object stands between the camera and the critical area. When the object is sufficiently high, it covers greater area. However, using multiple cameras images reduces this error to minimum because it reveals the other side of the object. This could be observed in following figures (Fig. 2).

Measurement of the base itself is based on the principle of background visibility. In order to mark the area as a base of an object, it has to be invisible to each camera. Knowing that the pixels in the area of the picture marked as the foreground (detected object) have values aproaching to one, and the background pixels have values aproaching to zero, the object base area in the output image G is corresponding to minimal functional values of the background images set B (4). This is based on fuzzy-logical operation called intersection of sets.

$$G(x, y) = min(B_1(x, y), ..., B_n(x, y)) \tag{4}$$

To measure the danger level of the critical area situation it is necessary to define another image function H determining the level of danger for each part of the critical area. This function returns values from zero to one, where zeros indicate the safe parts of the critical area and ones the dangerous parts. The danger level of the situation is then assessed as the maximal value of the intersection of image functions H and G based on the same formula as (4), but only for this two image fuctions. Then the result is a single number that gives the information about the level of danger in the critical area. So if the unambiguous answer to the question If the area is secure or not is required, it is necessary to set a threshold for the calculated danger level value. Then the tresholding would give us the "yes" or "no" answer.

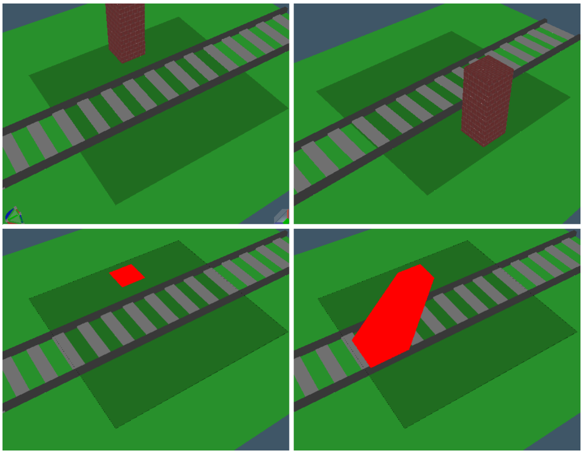


Fig. 2. First row: Images from both cameras, Bottom left: Base detection using images from multiple cameras, Bottom right: Detected area of an obstacle using only one camera

5 Conclusion

In this paper we have proposed a method which allows to measure the level of danger in the area secured by multiple cameras minimizing the occurrence of false alarms. For this purpose a method to measure the base of an object was designed. We used homography to get top-view image, backgroud substraction to find an object in the image and visibility across all cameras images to find the object base. For its low computational cost it is suitable for use in the monitoring of hazardous areas where immediate response is required. In our case it was securing of the railway electrification and safety systems, where the camera is used as a preventative measure and minimizes the accidents occurrence or its consequences. This method is relatively cheap and easily applicable for other purposes like securing of the crossroads.

6 Acknowledgement

This paper summarizes the results from papers presented on recent workshops and conferences [7–9]. One paper was also chosen as a "Best Paper" in ESMO 2011 conference [8]. This work was supported by Ministry of Industry and Trade of The Czech Republic - TIP no. FR-TII/027.

References

1. Beymer, D.; McLauchlan, P.; Coifman, B.; Malik, J., *A real-time computer vision system for measuring traffic parameters*, 1997. Proceedings., 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol., no., pp.495-501, 17-19 Jun 1997 doi: 10.1109/CVPR.1997.609371
2. Hartley, R. I. and Zisserman, A., *Multiple View Geometry in Computer Vision*, Second edition, 2004, Cambridge University Press, ISBN: 0521540518
3. Jinman Kang; Cohen, I.; Medioni, G., *Continuous tracking within and across camera streams*, 2003. Proceedings. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, no., pp. I-267- I-272 vol.1, 18-20 June 2003 doi: 10.1109/CVPR.2003.1211363
4. Moore, C.; Duckworth, T.; Aspin, R.; Roberts, D., *Synchronization of Images from Multiple Cameras to Reconstruct a Moving Human*, 2010 IEEE/ACM 14th International Symposium on Distributed Simulation and Real Time Applications (DS-RT), vol., no., pp.53-60, 17-20 Oct. 2010 doi: 10.1109/DS-RT.2010.15
5. Parks, D.H.; Fels, S.S., *Evaluation of Background Subtraction Algorithms with Post-Processing*, 2008. AVSS '08. IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, vol., no., pp.192-199, 1-3 Sept. 2008 doi: 10.1109/AVSS.2008.19
6. Piccardi, M., *Background subtraction techniques: a review*, 2004 IEEE International Conference on Systems, Man and Cybernetics, vol.4, no., pp. 3099- 3104 vol.4, 10-13 Oct. 2004 doi: 10.1109/ICSMC.2004.1400815
7. Mozdren, K., Stýskala, V., *Cameras as a Tool for Minimalization of Damage on Railway Electrification and Safety System.* In EPE 2011, 12th International Scientific Conference on Electric Power Engineering, 2011, ISBN 978-80-248-2393-5
8. Mozdren, K., Stýskala, V., *Multiple Camera System for Collision Prevention in Industry.* In ESMO 2011, IX International conference of students and young researchers, Kremenchug:KNU, 2011, 80-81, ISSN 2079-5106
9. Mozdren, K., Stýskala, V., *Kontrola vyklizení oblasti hlídané kamerami.* In XLIV. Sešit katedry elektrotechniky, Ostrava:VŠB - Technická univerzita Ostrava, 2011, ISBN 978-80-248-2378-2

Using Social Network Analysis Metrics to Study Elearning Collaboration

Gamila Obadi and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vaclav.snasel@vsb.cz

Abstract. Many organizations are forming virtual teams of geographically distributed knowledge members to collaborate on a variety of tasks. But how effective are these virtual teams compared to traditional face-to-face groups? This paper describes a study about two networks. The first is an explicit network of a consulting company employee, and the second is an implicit elearning network. The main objective is to find a quality knowledge and trustworthy collaborators based on methods of social network analysis

1 Introduction

Social networking sites has grown tremendously in popularity over the past few years. Since the early 1990s , the majority of web users were consumers of content created by a relatively small amount of publishers. From the early 2000s, user-generated content has become increasingly popular on the web.This change in behavior of the Web users has improved the level of knowledge exchange and collaboration among them. However, in some circumstances it is difficult to achieve efficient and effective knowledge sharing due to the following two barriers: (1) the difficulty in finding quality knowledge, and (2) the difficulty in finding trustworthy collaborators to interact with [1]. In that case, models of credibility which are used extensively on search engine research and information retrieval can be used in order to evaluate the trustworthiness of the actor's' knowledge.

Several graph theoretic models of credibility rely strongly on the consideration of the in-degree of the node so as to extract importance and trustworthiness. However, there are social activities derive much of their credibility by their productions. In that case, the in-degree cannot provide input to evaluate the importance of that entity and therefore an alternative evaluation is needed, which has to consider the outputs of the entity [2, 3].

Social networks on the web can be explicit or implicit. Explicit social networks are the kind of networks where a person is defined by the people s/he connects to (face book, my space), the ties between the actors are limited to the people they happen to know. The power of the explicit network appears in the trust that the contacts represents. While in implicit social networks a person is defined by his/her interests (amazon). Implicit social circles are groups of people who

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 430–435.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

are connected and share particular interests, but don't know they are actually connected to each other. Many of the new web 2.0 applications implicitly contain a lot of valuable information about how people are connected. For example, in a blogging system, if A and B both comment in many identical blogs, they are likely to share similar interests [6].

Social network analysis (SNA) has become an increasingly popular way to approach research problems and describe social processes. The availability of computing resources to individual researchers and the creation and standardization of software analysis packages, and more recently by graphical packages that visualize networks motivated researchers to apply it to examine the characteristics and structure of online communities. In [7] SNA techniques and visualizations were used to examine interactions in an online asynchronous forum. In [8] SNA was used to study of the growth of community and use of Internet resources among a class of 15 distance learners.In [9] researchers presented some reflections about the business model of the association of Brazilian universities and its members actuation based on methods of social network analysis.

In this paper, the authors studied two different social networks, one was constructed explicitly to provide consulting for the employee of a company, and the other was derived implicitly from the log files of an elearning system. The authors' main objective is to study the behavior of the members of these two networks using some social networks analysis measures to determine expert users.

2 Explicit Social network

2.1 Dataset Description

This dataset contains two networks from a consulting company (46 employee). In the first network, the ties are differentiated on a scale from 0 to 5 in terms of frequency of information or advice requests (Please indicate how often you have turned to this person for information or advice on work-related topics in the past three months). 0: I Do Not Know This Person; 1: Never; 2: Seldom; 3: Sometimes; 4: Often; and 5:Very Often.

In the second network, ties are differentiated in terms of the value placed on the information or advice received (For each person in the list below, please show how strongly you agree or disagree with the following statement: In general, this person has expertise in areas that are important in the kind of work I do.). The weights in this network is also based on a scale from 0 to 5. 0: I Do Not Know This Person; 1: Strongly Disagree; 2: Disagree; 3: Neutral; 4: Agree; and 5: Strongly Agree .

2.2 Social Network Analysis

Frequency of Information Request Network Analysis: To show the overall structure of the frequency of consulting requests network NetDraw was used Fig 1.

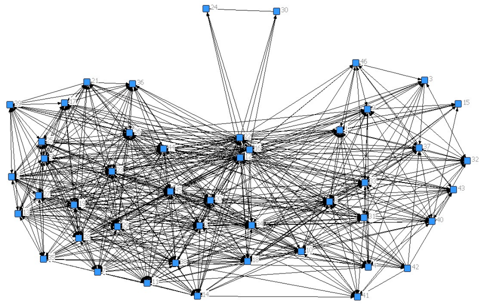


Fig. 1. Frequency of consulting requests weighted graph

To work with a binary matrix , the researchers considered to take the value of 3 (Sometimes) as a cut line. Above this value the connection between two employee receives the value 1, otherwise 0. The graph density is .08 and the overall structure of this network is shown in Fig 2. The employee in-degree centrality was measured by using UCINET 6, and the results showed that Actors 16, 24, 30, 37 and 43 appears as isolates (with centerality = 0.0) while actors 6,19, 26, and 45 were at the center of the network (with centrality = 9.0).

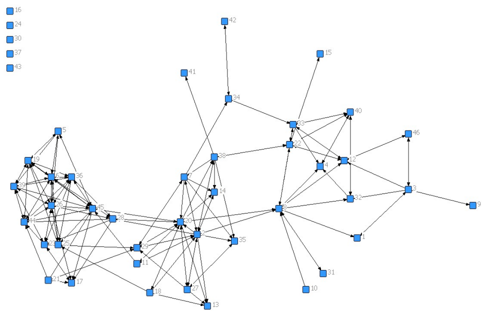


Fig. 2. Frequency of consulting requests binary graph

Value of Information Network Analysis: The number of consulting requests that an actor gets is not the only factor that determine his level of trust, to calculate how trustworthy is he/she another test were conducted by the researchers. In this case the in-degree centrality of the value of information was measured. Each employee were asked to evaluate the value of experience of his colleagues on a scale of 0 to 5,and UCINET 6 was used to analyze the data. The results showed that even though actor 20 has less requests of information (indegree =7) than actor 26 (indegree =9) ,his colleagues considered him more experienced (indegree = 129)than actor 26 (indegree = 89).

2.3 Other Measures of Centrality

Degree centrality measures might be criticized because they only take into account the immediate ties that an actor has, or the ties of the actor's neighbors, rather than indirect ties to all others [4]. One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighborhood. Therefore We have implemented three other measures of centrality: Closeness, Betweenness and Eigenvector centrality . All these measures are valid ways of describing a nodes centrality. However, each value measures a nodes centrality in different ways.

Closeness centrality approaches emphasize on the distance of an actor to all others in the network by focusing on the distance from each actor to all others. It is an inverse measure of centrality: the larger the numbers, the more distant an actor is, and the less central [4, 5]

Betweenness Centrality is defined as the number of shortest paths that pass through a node. It is the number of "times" that any node needs go through a given node to reach any other node by the shortest path. A high Betweenness Centrality score can indicate which nodes control the flow of information through the network [4, 5].

Eigenvector centrality is based on the idea that a node is central to the extent that it is connected to others who are central. An actor who is high on eigenvector centrality is connected to many actors who are themselves connected to many actors

	Degree	Closeness	Betweenness	Eigenvector
8	22.222	14.331	30.832	5.985
18	8.889	13.433	1.961	9.308
19	17.778	12.821	0.156	40.689
20	26.667	14.754	27.207	21.844
21	8.889	12.894	0.424	15.629

Table 1, Frequency of Information Request Network Centrality

The findings showed that actor 19 who has smaller degree centrality than actor 20, has higher egienvector centrality than actor 20, which implies that actor 19 is connected to persons who are centrals by them selfs. Actors 8 and 20 have the highest betweenness centrality, so they control the majority of information flow through the network

3 Implicit Social Network

3.1 Dataset Description

This dataset contains a network extracted from the log files of the Moodle elearning management system used at Silesian University. This network consists of 162 students and the ties are differentiated on a scale from 0 to 1 in terms of the similarity of the students' blogging behavior. The results from UCINET show that the density value for the valued matrix of similarity for the students' network is .125 with .278 standard deviation.The .125 value for the density of the valued matrix represents the average level of the blogging behavior similarity between any couple of students . To calculate the density of this network, the adjacency matrix was dichotomized (0.5 similarity was selected to be the cut line). When that was done the value for the density was found to be 0.132.

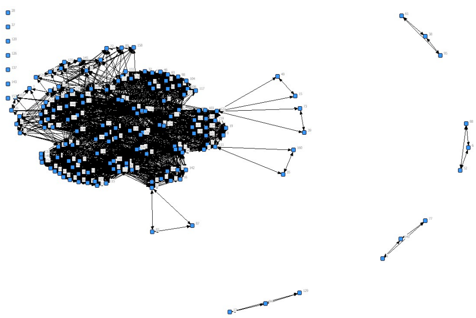


Fig. 3. Students' blogging similarity weighted graph

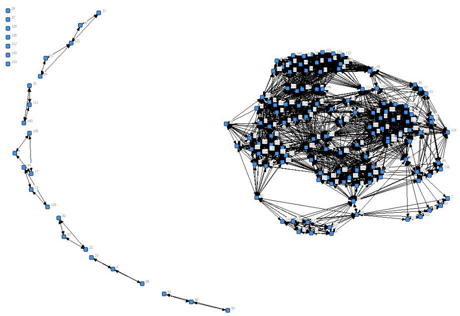


Fig. 4. Students' blogging similarity binary graph

3.2 Social Networks Analysis

The measures of centrality was calculated for the students' network, and the results were reported in table 2. Even though student 132 has the highest degree and betweenness centrality measures, s/he reported lower eigenvector centrality as compared to that reported by student 56 .

		Degree	Closeness	Betweenness	Eigenvector
38	37	0.000		0.000	0.000
57	56	34.161	3.161	3.558	25.274
132	131	17.391	3.128	0.047	15.405
133	132	34.161	3.162	5.027	19.210
134	133	14.286	3.140	0.093	7.731

Table 2,Students' blogging similarity Network Centrality

4 Conclusion

The Findings of this study illustrated that in these two different networks actors showed different styles of behavior. Some actors were isolates while some others were at the center of the network, receiving information requests from their colleagues and answering them efficiently. Besides, the study proved that actors acquires their power and importance in the network not only by the number of ties that they have, but the level of centrality of their contacts, their distance from the others and the number of paths passes through them. The study showed that social network analysis metrics provide useful information about virtual interactions in both implicit and explicit networks.

References

1. Yang, S. J. H., Chen, I. Y. L., Kinshuk Chen, N.-S. (2007). Enhancing the Quality of e-Learning in Virtual Learning Communities by Finding Quality Learning Content and Trustworthy Collaborators. Educational Technology Society, 10(2), 84-95.
2. Korfiatis,N., Poulos,M.,Bokos,G.(2oo6). Evaluating Authoritative Sources in Collaborative Editing Environments: An insight from Wikipedia.
3. BORODIN, A.,ROBERTS, G.,ROSENTHAL,.,TSAPARAS, P.(2oo5). Link Analysis Ranking: Algorithms, Theory, and Experiments.ACM Transactions on Internet Technology, Vol. 5, No. 1, February 2005, Pages 231297.
4. Hanneman, Robert A. and Mark Riddle. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/ hanneman/)
5. http://www.analytictech.com/networks/centrali.htm
6. http://www.sapir.eu/papers/deliverables/sapir
7. Pedro A. Willging (2005). Using Social Network Analysis Techniques to Examine Online Interactions. US-China Education Review,ISSN1548-6613,USA.
8. Caroline Haythornthwaite (1998). A social network study of the growth of community among distance learners. Information Research, Vol. 4 No. 1, July 1998
9. Mauro Faccioni Filho, Rafael vila Faraco, Youssef Ahmad Youssef . Social network analysis of an association of Brazilian universities.

Expectation-Maximization Method for Boolean Factor Analysis

Pavel Yu. Polyakov[1], Dušan Húsek[2], and Alexander A. Frolov[3]

[1]Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavel.mipt@mail.ru
[2]Institute of Computer Science,Academy of Sciences of the Czech Republic, Praha, Czech Republic
dusan@cs.cas.cz
[3]Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Moscow, Russia
aafrolov@mail.ru

Abstract. Boolean factor analysis is one of the most efficient methods to reveal and to overcome informational redundancy of high-dimensional binary signals. In the present study, we introduce new Expectation–Maximization method which maximizes the likelihood of Boolean Factor Analysis solution. Using the so-called bars problem benchmark [2], we compare efficiencies of the proposed method with Dendritic Inhibition neural network [7].

1 Introduction

Factor analysis is one of the most powerful statistical method to reveal and reduce information redundancy in high dimensional signals. Boolean Factor Analysis (BFA) as a special case of factor analysis implies that components of original signals, factor loadings and factor scores are binary values. Each binary signal can be interpreted as a representation of presence or absence of attributes (binary signal components) in the pattern. The number of considered attributes is the dimensionality of the signal space, the appearance of an attribute is encoded as One, and absence as Zero. The patterns are assumed to be composed of many "objects" in different combinations. We define an object as a collection of highly correlated attributes and suppose that objects are relatively independent of one another. Hence the attributes of different objects are only slightly correlated. In terms of BFA, objects are factors, attributes constituting the object are factor loadings, and the presence or absence of an object in the pattern is identified by the value of the factor score (One or Zero). Correlation between the attributes constituting each factor can be revealed by statistics over large data set constituted by patterns that contain each factor many times in different combinations with other factors. The aim of BFA is to detect this hidden structure of the signal space and to form a representation in which these independent objects are presented explicitly.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 436–441.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

In spite of the fact that binary data representation is typical for many fields, including social science, marketing, zoology, genetics and medicine, BFA methods are rather moderately developed. To overcome this lack we propose an Expectation–Maximization (EM) approach to BFA [5]. This method maximizes likelihood of BFA solution appropriate to general BFA generative model.

The well-known benchmark for learning of objects from complex patterns is the Bars Problem (BP)[2]. Using this benchmark, we compare the performance of the proposed method with Dendritic Inhibition (DI) neural network [7] which supposed to be most efficient in solving the bars problem [7–9]. We used information gain provided by BFA as a measure of performance of BFA methods. It was shown previously [3] that this measure is sensitive to both noise in signals and errors in BFA results, and thus is a reliable basis for comparison of different BFA methods.

1.1 Boolean Factor Analysis Generative Model

In terms of BFA, each pattern of a signal space is defined by a binary row vector $\mathbf{x}$ with dimensionality N equal to the total number of attributes. Every component of $\mathbf{x}$ takes value One or Zero, depending on the presence or absence of the related attribute. Each factor $\mathbf{f}_i$ is a binary row vector of dimensionality N whose One valued entries correspond to highly correlated attributes of the i-th object. Although the probability of the object's attribute to appear in a pattern simultaneously with its other attributes is high, it is not obligatory equal to 1. We denote this probability as p_{ij}, where j is the index of an attribute and i is the index of a factor. For attributes constituting the factor, that is for attributes with $f_{ij} = 1$, probability p_{ij} is high, and for other attributes (with $f_{ij} = 0$) it is zero.

As in linear factor analysis, we suppose that additionally to common factors $\mathbf{f}_i$ that influence to more than one attribute, each signal also contains some specific or unique factors that influence to only particular attributes. The contribution of specific factors is defined by a binary row vector $\mathbf{n}$, which we call "specific noise". Each specific factor n_j is characterized by a probability q_j that j-th component of vector $\mathbf{n}$ takes One.

As a result, any vector $\mathbf{x}$ can be presented in the form

$$\mathbf{x} = [\bigvee_{i=1}^{L} s_i \wedge \mathbf{f}'_i] \vee \mathbf{n}, \tag{1}$$

where $\mathbf{s}$ is a binary row vector of factor scores of dimensionality L, L is the total number of factors, $\mathbf{f}'_i$ is a distorted version of factor $\mathbf{f}_i$ and $\mathbf{n}$ is a specific noise defining the influence of specific factors. Factor distortion implies that some Ones of the i-th factor become Zeros with probability $1-p_{ij}$, that is components constituting the factor can take Zeros but neither component not constituting the factor (for components not constituting the factor $p_{ij} = 0$) can take One in distorted version of factor. Nevertheless it can be One in the input signal due to the presence of other factors or specific noise. We assume that factors appear in patterns (that is related scores s_i take Ones) independently with probabilities π_i $(i = 1, \ldots, L)$.

The aim of Boolean Factor Analysis is to find the parameters of generative model $\mathbf{\Theta} = (p_{ij}, q_j, \pi_i, i = 1, \ldots, L, j = 1, \ldots, N)$ and the factor scores $s_{mi}, m = 1, \ldots, M$ for each of patterns of the observed data set $\mathcal{X}$ containing M signals. However, it is supposed that the found factors could also be detected in any arbitrary pattern $\mathbf{x} \notin \mathcal{X}$ if generated by the same model.

We suppose that each component of the common factor is distorted independently of the presence of other factors in the input signal and independently of specific noise. Thus, given generative parameters $\mathbf{\Theta}$ and factor scores $\mathbf{s}$, the probability of the j-th component of $\mathbf{x}$ to take the value x_j is

$$P(x_j|\mathbf{\Theta}, \mathbf{s}) = x_j - (2x_j - 1)(1 - q_j) \prod_{i=1}^{L} (1 - p_{ij})^{s_i}, \tag{2}$$

We suppose also that, given generative parameters and factor scores, different components of $\mathbf{x}$ are also statistically independent. Thus

$$P(\mathbf{x}|\mathbf{\Theta}, \mathbf{s}) = \prod_{j=1}^{N} P(x_j|\mathbf{\Theta}, \mathbf{s}). \tag{3}$$

Since we assume that factors appear in patterns independently then

$$P(\mathbf{s}|\mathbf{\Theta}) = \prod_{i=1,L} \pi_i^{s_i} (1 - \pi_i)^{1-s_i}. \tag{4}$$

2 Expectation-Maximization Method — EMBFA

The Expectation-Maximization (EM) method [1] allows for finding parameters $\mathbf{\Theta}$ of a given probabilistic generative model to maximize the likelihood of the observed dataset $\mathcal{X}$. Since it is applied here to BFA generative model we call it EMBFA. The EM method maximizes the likelihood of the observed data by maximizing the free energy

$$\mathcal{F}(\mathbf{\Theta}, \mathbf{g}) = \sum_{m=1}^{M} \sum_{\mathbf{s}} g_m(\mathbf{s}) [\log P(\mathbf{x}_m|\mathbf{s}, \mathbf{\Theta}) + \log P(\mathbf{s}|\mathbf{\Theta})] + \sum_{m=1}^{M} H(g_m(\mathbf{s})), \tag{5}$$

where $g_m(\mathbf{s})$ is the expected distribution of factor scores for the m-th pattern, $H()$ denotes Shannon entropy. The iterations of EM alternatively increase $\mathcal{F}$ with respect to the distributions g_m, while holding $\mathbf{\Theta}$ fixed (the E-step), and with respect to parameters of the model $\mathbf{\Theta}$, while holding g_m fixed (the M-step).

At the E-step, when $\mathbf{\Theta}$ is fixed, the distributions g_m which maximize $\mathcal{F}(\mathbf{\Theta}, \mathbf{g})$ are calculated according to the following equation

$$g_m(\mathbf{s}|\mathbf{\Theta}) = \frac{P(\mathbf{s}|\mathbf{\Theta}) P(\mathbf{x}_m|\mathbf{s}, \mathbf{\Theta})}{\sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{\Theta}) P(\mathbf{x}_m|\mathbf{s}, \mathbf{\Theta})},$$

where $P(\mathbf{x}_m|\mathbf{s},\mathbf{\Theta})$ is given by (3) and $P(\mathbf{s}|\mathbf{\Theta})$ is given by (4). The obtained distributions g_m provide the expected likelihood of the observed data over factor scores given parameters of the generative model [6].

At the M-step, when distributions g_m are fixed, π_i can be estimated as $(1/M)\sum_{m=1}^{M} s_{mi}$, where $s_{mi} = \sum_{\mathbf{s}} g_m(\mathbf{s}|\mathbf{\Theta})s_i$. Respectively, p_{ij} and q_j can be found by steepest ascent maximization of $\mathcal{F}(\mathbf{\Theta},\mathbf{g})$:

$$\Delta p_{ij} = \gamma\frac{\partial\mathcal{F}}{\partial p_{ij}}, \qquad \Delta q_j = \gamma\frac{\partial\mathcal{F}}{\partial q_j} \tag{6}$$

where γ is a learning rate.

Since we assume that probabilities p_{ij} are sufficiently high for components constituting the i-th factor ($f_{ij} = 1$) and equal to zero for other components ($f_{ij} = 0$), at each iteration step we put $p_{ij} = 0$ if the j-th component satisfies inequality

$$p_{ij} < 1 - \prod_{l\neq i}(1 - \pi_l p_{lj}), \tag{7}$$

where the right side of the inequality is the probability that the j-th attribute appears in the pattern due to other factors except $\mathbf{f}_i$. It is interesting to note that without this threshold truncation of p_{ij}, EM does not converge because of uncertainties arising from the competition between factors defined by p_{ij} and noise defined by q_j. The obtained values of p_{ij}, q_j and π_i are used as the input for the next E-step. EM iterative procedure terminates once all values $\sqrt{\sum_j(p_{ij}^{old} - p_{ij}^{new})^2/\sum_j p_{ij}^{old}}$ $(i = 1\ldots L)$ remained smaller than $2.5\cdot 10^{-3}$ for 20 sequential iterations.

When procedure converges, the final values s_{mi} are estimates of the factor scores. To satisfy the generative model, we binarized those values. The binarization threshold was chosen to maximize BFA information gain [3].

We restricted EMBFA algorithm to the case of sparse scores, when only a small number of factors (no more then 3) are supposed to be mixed in the observed patterns. In this case, summation over $\mathbf{s}$ in the above formulas is reduced to

$$\sum_{\mathbf{s}}(...) = (...)_{\mathbf{s}=0} + \sum_i(...)_{\mathbf{s}=\mathbf{s}_i} + \sum_{i<j}(...)_{\mathbf{s}=\mathbf{s}_{ij}} + \sum_{i<j<k}(...)_{\mathbf{s}=\mathbf{s}_{ijk}}, \tag{8}$$

where $\mathbf{s}_i$ is the vector of factor scores with all zeros except s_i, $\mathbf{s}_{ij}$ is the vector of factor scores with all zeros except s_i and s_j, and $\mathbf{s}_{ijk}$ is the vector of factor scores with all zeros except s_i, s_j and s_k. An increase of the number of terms in (8) leads to the considerable rise in computational complexity.

3 Performance of BFA Methods in Solving Bars Problem

In this section we compare efficiencies of EMBFA and DI. DI was developed in [8] for finding parts-based decompositions of images. The method is based on a feed-forward neural network with lateral inhibition. The main idea of the method is to use lateral inhibition of individual synapses instead of total inhibition of a neuron. As a result of network learning, neurons of the output layer acquire specific sensitivity to factors constituting patterns of the data set: the appearance of a factor in the pattern presented to the input layer leads to the strong activation of the related neuron at the output layer. Output neurons can be activated when factors are partially distorted and in the presence of noise. Thus, activity of each neuron at the output layer provides a gradual estimation of the confidence that this pattern contains the related factor. To transform gradual activity of output layer to binary vector of factor scores we used the same binarization procedure as for EMBFA. Namely we used the binarization threshold maximizing the BFA information gain.

The performance of the methods are compared using information gain G [4]. The calculation of G for given dataset $\mathcal{X}$ and set of factor scores $\mathcal{S}$ can be performed using special procedure [3] which also derives the parameters of generative model $\mathbf{\Theta}$.

For EMBFA, as well as for DI, the number of desired hidden factors has to be set in advance. In the majority of computer experiments performed by Spratling [7], this number was taken twice higher than the actual number of factors. In our experiments with EMBFA and DI, the predefined number of hidden factors was also taken twice higher than the actual number of factors.

To start the EM procedure, we set $\pi_i = 1/32$, $q_j = 0$ and initialized p_{ij} with random values uniformly distributed in the range from 0.3 to 0.8.

Figure 1 demonstrates sensitivity of BFA methods to noise. The noise was assumed to be distributed uniformly over signal components and factors so that $q_j = q$ for any j, and $p_{ij} = p$ for any i and j. We tested BFA methods using data sets of the size $M = 800$.

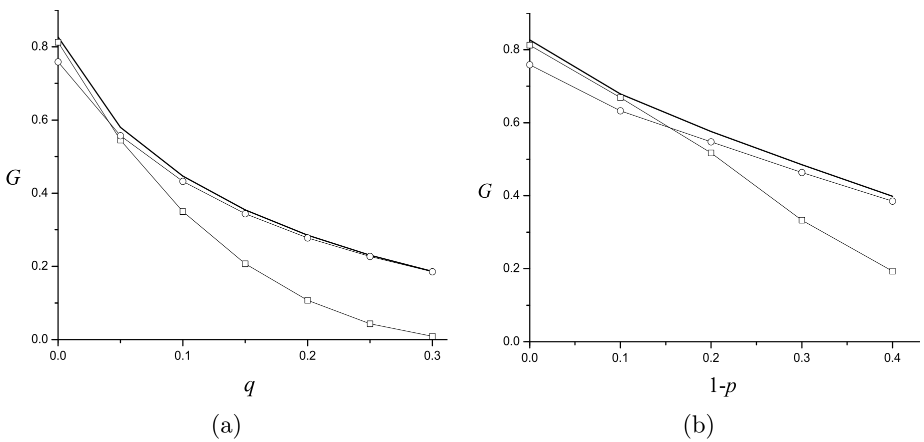


Fig. 1. Information gain G in dependence on q for $p = 1$ (a) and in dependence on p for $q = 0$ (b). $\circ$ – EMBFA, $\square$ – DI. Thick line – "ideal" solution.

In Fig. 1 “ideal” solution corresponds to the case when all scores are exactly the same as those used in the generated data set. As expected information gain decreases when noise increases even for “ideal” solution. And for very noisy data G tends to zero that means BFA has no sense.

As shown in Fig. 1(a) the performance of EMBFA is better for large q (regarding “ideal” solution) and slightly worsen for small q. In contrast DI demonstrates strong sensitivity to noise: its performance is almost ideal for $q = 0$ but it is getting worse for noisy data. EMBFA and DI exhibits similar sensitivity to factor distortion as to specific noise (Fig. 1(b)).

4 Conclusion

The proposed Expectation-Maximization (EMBFA) method is proved to be suitable for Boolean Factor Analysis. In comparison with Dendritic Inhibition (DI) neural network [7], it demonstrates slightly worse performance when noise is absent. But in contrast to DI, it exhibits a low sensitivity to noise (both factor distortion and specific noise).

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(1), 1–38 (1977)
2. Foldiak, P.: Forming sparse representations by local anti-hebbian learning. Biological Cybernetics 64, 165170 (1990)
3. Frolov, A., Husek, D., Polyakov, P.Y.: New measure of boolean factor analysis quality. Adaptive and Natural Computing Algorithms pp. 100–109 (2011)
4. Frolov, A.A., Husek, D., Polyakov, P.Y.: Bulevskij faktornyj analiz na osnove attraktornoj nejronnoj seti i nekotoryje ego prilozenija [Boolean factor analysis by means of attractor neural network and some its applications]. Neirokomputery: Razrabotka, Primenenie (issn 1999-8554) 1, 25–46 (2011)
5. Frolov, A.A., Husek, D., Polyakov, P.Y.: Expectation-Maximization Approach to Boolean Factor Analysis. In: Proceedings of International Joint Conference on Neural Network. San Jose, California (july 2011)
6. Neal, R.M., Hinton, G.E.: A view of the EM algorithm that justifies incremental, sparse, and other variants. Learning in graphical models 89, 355–368 (1998)
7. Spratling, M.W.: Learning image components for object recognition. Journal of Machine Learning Reasearch 7, 793–815 (2006)
8. Spratling, M.W., Johnson, M.H.: Preintegration lateral inhibition enhances unsupervised learning. Neural Computation 14(9), 2157–2179 (2002)
9. Spratling, M.W., Johnson, M.H.: Exploring the functional significance of dendritic inhibition in cortical pyramidal cells. Neurocomputing 52(54), 389–395 (2003)

Bayesian Spam Filtering and Normalized Compression Distance

Michal Prílepok

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.prilepok@vsb.cz

Abstract. Unsolicited or undesired email messages (i.n. spam) has become a major problem lately. One of the major obstacles is to decrease a rate of false reports, detected in incoming email. This work concerns additional comparison of incoming email with previously reported spam messages. To achieve this goal the normalized compression distance was used, which further contributed to the decrement of rate of false reports.

Keywords: spam, Bayesian spam filter, NCD, compression

1 Introduction

Nowadays, new methods for spam recognition have to be developed, as spammer are trying hard to fill our mailboxes with undesired email messages (e.g. advertisement, phishing, ... As the amount of unwanted email messages increases, the existing filtering methods have to adapt to fulfill new requirements for better performance and efficiency. Basic techniques for fighting spam can be divided into the following categories:

- analysis of the sender,
- analysis of the content of the message,
- analysis of sponsor.

However, neither of these categories is a self-sufficient. Current state of the art algorithms use combination of all three methods, which proven to be the best way of fighting spam [2].

The filters based on content analysis of email messages are the most effective. These filters utilizes multiple content analysis algorithms to successfully detect spam. Bayesian spam filtering (BSF) is one class of such a filters using statistical learning as a tool for content analysis.

Filters based on statistical learning are often trained on some well defined set of emails, where messages are already marked as spam or ham. In learning process word counts are extracted and each word is assigned with probability of occurrence in spam and ham. This information is than stored in the database for further use. The most commonly used method in this process is distribution

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 442–447.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

of email to words (or other parts of the text). Then the algorithm statistically calculates for every word the probability of email to be spam of ham.

In detection mode, Bayesian filters utilizes information from database to decide whether the incoming email is spam or ham. This decision is based on probability distribution of word occurrence in the document. The probability that email is a spam is than produced and some threshold value have to be chosen to deduce about result. As mentioned beforehand, the most commonly used learning filters are based on Bayes theorem. These types of algorithms could be incorporated in server layer, where they can analyze the global aspect of data, as well as on end-user computer where the user decision could be added to filtering result.

As mentioned beforehand, the most commonly used learning filters are based on Bayes theorem. These types of algorithms could be incorporated in server layer, where they can analyze the global aspect of data, as well as on end-user computer where the user decision could be added to filtering result [1, 2].

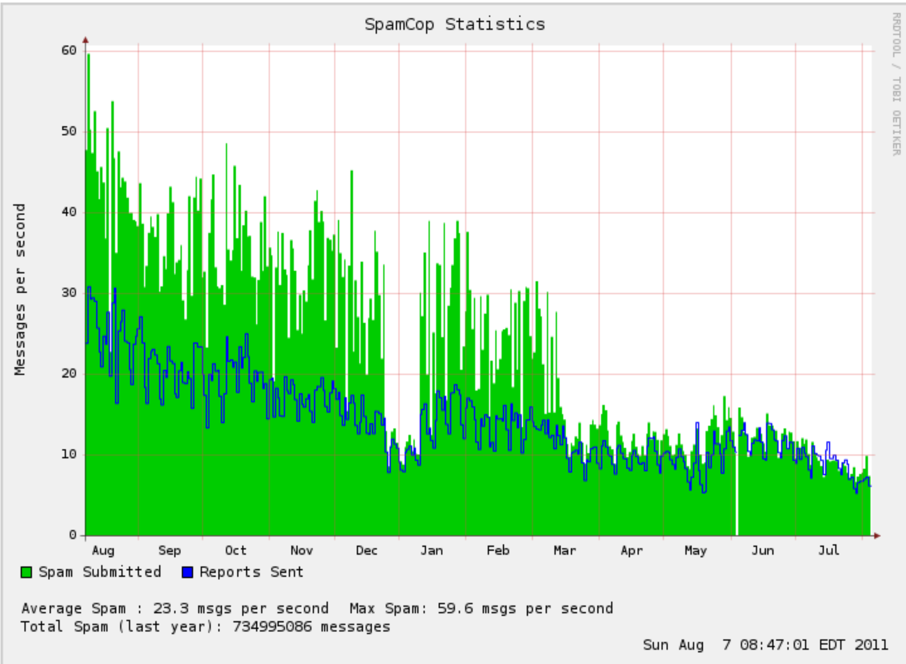


Fig. 1. Total spam report volume last year [12]

2 Bayesian spam filtering (BSF)

BSF is a statistical learning technique for filtering emails, which uses naive Bayesian classifiers. Bayesian classifiers works directly on words in email messages. The relations among the words are calculated based on Bayes probability theorem. After that this probability is used to form the decision of what type of message incoming email is.

If the message contains the specific word, the probability of being spam can be calculated using the formula (1): [3–6, 10]:

$$Pr(S|W) = \frac{Pr(W|S) * Pr(S)}{Pr(W|S) * Pr(S) + Pr(W|H) * Pr(H)} \tag{1}$$

where:

- $Pr(S|W)$ is the probability of message S contains the word W,
- $Pr(S)$ is the probability that message is spam,
- $Pr(W|S)$ is the probability that the word W appears in spam message
- $Pr(H)$ is the overall probability that message is not spam
- $Pr(W|H)$ is the probability that the word W appears in ham message.

Probability $Pr(W|S)$ and $Pr(W|H)$ are determined in the learning phase of filter. But recent statistics indicates that actual probability of the email being spam is approximately 80% $((Pr(S) = 0.8, Pr(H) = 0.2))$. In this case Bayesian filters assume that 80% of checked email is spam and only remaining 20% is ham.

However, the most of BSF does not assume that there is a-priori reason for an incoming message to be spam. In this case the probabilities are set to be equal $(Pr(S) = 0.5, Pr(H) = 0.5)$.

This type of filters are considered impartial, so they have no prejudice against incoming messages. However this assumption allows us to simplify the Bayes formula (2):

$$Pr(S|W) = \frac{Pr(W|S)}{Pr(W|S) + Pr(W|H)} \tag{2}$$

A probability values $Pr(W|S)$ and $Pr(W|H)$ are set up in learning phase. Therefore, data set for the filter initialization should contain about the same number of ham and spam messages.

The decisions based on only one word would be very prone to error. So it makes sense that final conclusion would be based on more words. Therefore, it is worthy of the Bayesian filter determine the probability of resulting message examined from more words.

BSF consider the words in document as independent items, but this assumptions does not hold for natural language. In natural language the words are depending on each other so that the probability of one affects the other. With respect to this, one could derive the formula (3) from Bayes theorem:

$$p = \frac{p_1 p_2 \cdots p_N}{p_1 p_2 \cdots p_N + (1 - p_1)(1 - p_2) \cdots (1 - p_N)} \tag{3}$$

where:

- p is the probability that message is a spam,

- p_1 is the probability $p(S|W_1)$ that spam message contains a first word,
- p_2 is the probability $p(S|W_2)$ that spam message contains a second word,
- $\ldots$,
- p_N is the probability $p(S|W_N)$ that spam message contains a Nth word.

The result p is usually compared to a given threshold to decide whether the message is spam or not, If p is lower than the threshold, the message is considered to be a ham, otherwise it is considered to be a spam.

3 Normalized Compression Distance (NCD)

The normalized compression distance (NCD) is a mathematical tool to determine if objects are similar. The object similarity is measured by compression algorithm and is based on algorithmic complexity developed by Kolmogorov, called Normalized Information Distance (NID). NCD can be used to determine similar objects of any type, such as music, texts, or sequences of genes. Lately NCD has been also applied to other domains such as plagiarism detection or visual data mining [7–9].

The resulting degree of similarity is calculated by the following formula (4):

$$NCD = \frac{C(xy) - \min(C(x), C(y))}{\max(C(x), C(y))} \tag{4}$$

NCD produce a result in interval $0 \leq NCD(x, y) \leq 1$, results close to 0 mean that x and y are almost identical.

4 The combination BSF and NCD

To achieve even better rate of spam detection, one can combine BSF with other remedial techniques. One possibility is combining BSF and NCD, where NCD is used to further refine results by comparing output of BSF with training set.

This method computes the additional similarity with NCD file, which has approximately the same size as the tested message. Files with different size after compression lower chance to be similar to tested message.

The resulting value from NCD is in between 0 and 1, where 0 means maximum similarity to test message. To be able to combine probabilities from BSF and NCD, one has to use formula (5) to which has been derived from Bayes theorem.

$$Pr(NCD) = 1 - NCD \tag{5}$$

Now $Pr(NCD)$ has the same meaning as the BSF probability.

The final result is now obtained with formula (6):

$$P = \frac{P_B * P_{NCD}}{P_B * P_{NCD} + (1 - P_B)(1 - P_{NCD})} \tag{6}$$

where:

- P is final probability of message,
- P_B is probability of BSF,
- P_{NCD} is modified probability of NCD.

5 Conclusion

The algorithms were tested on a dataset with 270 045 emails [11], which consists of 62.23% spam messages and 36.77% ham messages.

Three variants of algorithms were tested:

1. classical BSF without any modifications
2. combination of BSF and NCD, where NCD checks all emails with BSF value in $\langle 0.5, 1\rangle$
3. combination of BSF and NCD, where NCD checks all emails with BSF value in $\langle 0.5, 0.75\rangle$.

Classical BSF successfully identified 94.50% spam emails with 4.60% of spam emails passed through and 9.81% ham messages falsely marked as spam. Total number of emails marked as spam was 6.52% and processing speed was 288 email per second.

The second variant of algorithm successfully identified 99.49% of spam emails. Unidentified spam emails were 0.51% and the number of ham emails marked as spam raised to the value 12.66%. The overall error rate of algorithm is 4.98% incorrectly detected emails. Average speed of testing emails was 33.83 emails per second.

Last modification of BSF with more limited additional control by NCD, has successfully marked 99.49% of spam emails. Rate of unidentified spam was 0.51% and ham rate of messages identified incorrectly 12.67%. Average speed of testing emails was 192 emails per second.

Variants utilizing both BSF and NCD proven to be more successful in spam detection, however with more ham marked as spam and with much slower processing speed. This algorithm would need further research to decrease the amount of false-positive finding to be massively adapted in consumer email clients. This is left for further research. The results of experiments are summarized in table 1.

Table 1. Comparison of the results

		BSF	BSF width NCD (>0.5)	BSF width NCD (0.5-0.75)
Spam	**Success rate**	95.40%	99.49%	99.49%
	Error rate	4.60%	0.51%	0.51%
Ham	**Success rate**	90.19%	87.64%	87.33%
	Error rate	9.81%	12.66%	12.67%
Overall error rate		6.52%	4.98%	4.98
Throughput		288.36	32.93	192.59

References

1. Ahmed Khorsi:*An Overview of Content-Based Spam Filtering Techniques*.Informatica (Slovenia) (INFORMATICASI) 31(3):269-277 (2007).
2. Paul Wolfe, Charlie Scott, Mike W. Erwin:*Antispam : metody, nastroje a utility pro ochranu pred spamem*.Computer Press 2004.
3. Yang Song, Aleksander Kolcz, C. Lee Giles:*Better Naive Bayes classification for high-precision spam detection.*. Softw., Pract. Exper. (SPE) 39(11):1003-1024 (2009).
4. Beatrice Cynthia Dhinakaran, Dhinaharan Nagamalai, Jae-Kwang Lee:*Bayesian Approach Based Comment Spam Defending Tool.* ISA 2009:578-587.
5. Youcef Begriche, Ahmed Serhrouchni:*Bayesian statistical analysis for spams.* LCN 2010:989-992.
6. Tiago A. Almeida, Akebo Yamakami, Jurandy Almeida:*Evaluation of Approaches for Dimensionality Reduction Applied with Naive Bayes Anti-Spam Filters*.ICMLA 2009:517-522.
7. Paul M. B. Vitnyi:*Universal Similarity*.CoRR abs/cs/0504089 (2005).
8. Nicholas Tran:*The Normalized Compression Distance and Image Distinguishability*.19th IS&TSPIE Symposium on Electronic Imaging Science and Technology, January 2007, San Jose, USA.
9. Gary Robinson: *Spam Detection*.
   URL: http://radio-weblogs.com/0101454/stories/2002/09/16/spamDetection.html
10. *SpamLinks*.URL: http://spamlinks.net/
11. NIST:*Spam Track - email datasets*.http://trec.nist.gov/data/spam.html.
12. SpamCop.net: *Statistics on Spam trends*.http://spamcop.net/spamstats.shtml

Semi-Smooth Newton Method for 2D Contact Problems with Tresca and Coulomb Friction

Kristina Rádková and Radek Kučera

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{kristina.radkova.st, radek.kucera}@vsb.cz

Abstract The contribution deals with contact problems for two elastic bodies with friction. After the description of the problem we present its discretization based on linear or bilinear finite elements. The semi–smooth Newton method is used to find the solution. We present active sets algorithms. The non-symmetric and symmetric case is distinguished with using BiCGSTAB and CGM, respectively. Finally we will arrive at the globally convergent dual implementation of the algorithm.

Keywords: contact problems, semi–smooth Newton method, dual formulation

1 Introduction

Let us consider two homogeneous isotropic elastic bodies represented by bounded domains $\Omega^k \subset \mathbb{R}^2$ with sufficiently smooth boundaries $\partial\Omega^k$, $k = 1, 2$. Each boundary consists of three disjoint parts Γ_u^k, Γ_p^k, and Γ_c^k open in $\partial\Omega^k$ so that $\partial\Omega^k = \overline{\Gamma}_u^k \cup \overline{\Gamma}_p^k \cup \overline{\Gamma}_c^k$ and $\overline{\Gamma}_u^k \neq \emptyset$; The zero displacements are prescribed on Γ_u^k while surface tractions $\boldsymbol{p}^k \in (L^2(\Gamma_p^k))^2$ act on Γ_p^k. The bodies may get into contact on the *contact interface* given by Γ_c^1 and Γ_c^2, where we consider three contact conditions: the non-penetration of bodies, the transmission of contact stresses, and the Coulomb friction law. Elastic properties of Ω^k are described by the Lamè constants $\lambda^k, \mu^k > 0$. Finally, assume the volume forces $\boldsymbol{f}^k \in (L^2(\Omega^k))^2$.

Our aim is to find an equilibrium state of Ω^1 and Ω^2. By the solution of this problem we mean displacement vector fields $\boldsymbol{u}^k = (u_1^k, u_2^k)^\top$, $k = 1, 2$, satisfying the equilibrium equations and the Dirichlet and Neumann conditions.

To solve the contact problem with the Coulomb friction we use the method of successive approximations (fixed–point approach) and in each step we solve the problem with the Tresca friction (according to a predefined mapping).

The formulation of the contact problem with Coulomb friction results in an *implicit* variational inequality of the elliptic type. If we replace the slip bounds (of the Coulomb friction law) by an à-priori given positive function g defined on Γ_c^1, we arrive at the contact problem with Tesca friction. This well known problem is represented by the variational inequality of the second type for which there is a unique solution (for an appropriate g).

VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Discrete problems

After the discretization of our problem we arrive at

$$\boldsymbol{K}\boldsymbol{u} - \boldsymbol{f} + \boldsymbol{N}^\top \boldsymbol{\lambda}_\nu + \boldsymbol{T}^\top \boldsymbol{\lambda}_t = \boldsymbol{0}, \tag{2.1}$$

$$\boldsymbol{N}\boldsymbol{u} - \boldsymbol{d} \leq \boldsymbol{0},\ \boldsymbol{\lambda}_\nu \geq \boldsymbol{0},\ \boldsymbol{\lambda}_\nu^\top(\boldsymbol{N}\boldsymbol{u} - \boldsymbol{d}) = \boldsymbol{0}, \tag{2.2}$$

$$\left.\begin{array}{l} |\lambda_{t,i}| \leq \mathcal{F}_i\lambda_{\nu,i} \\ |\lambda_{t,i}| < \mathcal{F}_i\lambda_{\nu,i} \Rightarrow u_{t,i} = 0 \\ |\lambda_{t,i}| = \mathcal{F}_i\lambda_{\nu,i} \Rightarrow \exists c_t \geq 0:\ u_{t,i} = c_t\lambda_{t,i} \end{array}\right\} \quad i = 1,\dots,m. \tag{2.3}$$

We use two Lagrange multipliers $\boldsymbol{\lambda}_\nu, \boldsymbol{\lambda}_t \in \mathbb{R}^m$ that are the opposite of the discrete relative normal and tangential contact stresses, respectively. The stiffness matrix and the load vector are represented by $\boldsymbol{K}$ and $\boldsymbol{f}$, respectively. Matrices $\boldsymbol{N}$ and $\boldsymbol{T}$ are associated with the contact nodes in normal and tangential direction, respectively. $\boldsymbol{\mathcal{F}}$ stands for the coeficient of the Coulomb friction. Our unknown vector of the nodal displacement is $\boldsymbol{u}$. As this is just short report we have omitted the corresponding dimensions.

In order to obtain the discrete contact problem with Tresca friction, we replace $\mathcal{F}_i\lambda_{\nu,i}$ in (2.3) by the entries g_i of $\boldsymbol{g} \in \mathbb{R}^m$, $\boldsymbol{g} \geq \boldsymbol{0}$.

Next we use the equivalent formulation of the previous problems as the systems of non-smooth equations. We introduce the projection mappings

$$\boldsymbol{P}_{\mathbb{R}^m_+} : \mathbb{R}^m \mapsto \mathbb{R}^m_+ \quad \boldsymbol{P}_{\boldsymbol{\Lambda}(\boldsymbol{r})} : \mathbb{R}^m \mapsto \boldsymbol{\Lambda}(\boldsymbol{r})$$

with $\mathbb{R}^m_+ := \{\boldsymbol{\mu} \in \mathbb{R}^m : \boldsymbol{\mu} \geq \boldsymbol{0}\}$, $\boldsymbol{\Lambda}(\boldsymbol{r}) := \{\boldsymbol{\mu} \in \mathbb{R}^m : |\boldsymbol{\mu}| \leq \boldsymbol{r}\}$, respectively and $\boldsymbol{r} \in \mathbb{R}^m$, $\boldsymbol{r} \geq \boldsymbol{0}$. The definitions of the components of $\boldsymbol{P}_{\mathbb{R}^m_+}$, $\boldsymbol{P}_{\boldsymbol{\Lambda}(\boldsymbol{r})}$ are based on the max-function in $\mathbb{R}^1$:

$$(\boldsymbol{P}_{\mathbb{R}^m_+})_i(\boldsymbol{\mu}) = \max\{0, \mu_i\}, \tag{2.4}$$

$$(\boldsymbol{P}_{\boldsymbol{\Lambda}(\boldsymbol{r})})_i(\boldsymbol{\mu}) = \max\{0, \mu_i + r_i\} - \max\{0, \mu_i - r_i\} - r_i. \tag{2.5}$$

Denote $\boldsymbol{y} := (\boldsymbol{u}^\top, \boldsymbol{\lambda}_\nu^\top, \boldsymbol{\lambda}_t^\top)^\top \in \mathbb{R}^{2n+2m}$ and consider a parameter $\rho > 0$. The discrete contact problem with Coulomb friction is equivalent to the equation

$$\boldsymbol{G}(\boldsymbol{y}) = \boldsymbol{0} \tag{2.6}$$

where $\boldsymbol{G} : \mathbb{R}^{2n+2m} \mapsto \mathbb{R}^{2n+2m}$ is defined by

$$\boldsymbol{G}(\boldsymbol{y}) := \begin{pmatrix} \boldsymbol{K}\boldsymbol{u} - \boldsymbol{f} + \boldsymbol{N}^\top\boldsymbol{\lambda}_\nu + \boldsymbol{T}^\top\boldsymbol{\lambda}_t \\ \boldsymbol{\lambda}_\nu - \boldsymbol{P}_{\mathbb{R}^m_+}(\boldsymbol{\lambda}_\nu + \rho(\boldsymbol{N}\boldsymbol{u} - \boldsymbol{d})) \\ \boldsymbol{\lambda}_t - \boldsymbol{P}_{\boldsymbol{\Lambda}(\boldsymbol{\mathcal{F}}\boldsymbol{\lambda}_\nu)}(\boldsymbol{\lambda}_t + \rho\boldsymbol{T}\boldsymbol{u}) \end{pmatrix}.$$

2.1 Algorithms

As we use the semi-smooth Newton method (SSNM) our algorithms are based on the following iterative scheme:

$$\boldsymbol{F}^o(\boldsymbol{y}^{(k-1)})\boldsymbol{y}^{(k)} = \boldsymbol{F}^o(\boldsymbol{y}^{(k-1)})\boldsymbol{y}^{(k-1)} - \boldsymbol{F}(\boldsymbol{y}^{(k-1)}), \quad k = 1, 2, \dots, \tag{2.7}$$

where $\boldsymbol{F} : \mathbb{R}^{2n+2m} \mapsto \mathbb{R}^{2n+2m}$ is slantly differetiable and $\boldsymbol{F}^o(\boldsymbol{y})$ is a slanting function to $\boldsymbol{F}$ at $\boldsymbol{y} \in \mathbb{R}^{2n+2m}$. It is well known that sequence $\{\boldsymbol{y}^{(k)}\}$ generated by (2.7) converges superlinearly to the solution of $\boldsymbol{F}(\boldsymbol{y}) = \boldsymbol{0}$ when the initial iterate $\boldsymbol{y}^{(0)} \in \mathbb{R}^{2n+2m}$ is a sufficiently accurate approximation of the solution.

We will show an implementation of SSNM that is equivalent to an active set algorithm. Firstly, let us introduce notations. Let $\mathcal{M} = \{1, 2, \dots, m\}$ be the set of all indices and let $\boldsymbol{y} = (\boldsymbol{u}^\top, \boldsymbol{\lambda}_\nu^\top, \boldsymbol{\lambda}_t^\top)^\top \in \mathbb{R}^{2n+2m}$ be given. The active set $\mathcal{A}_\nu := \mathcal{A}_\nu(\boldsymbol{y})$ corresponding to the non-penetration condition is defined by

$$\mathcal{A}_\nu(\boldsymbol{y}) = \{i \in \mathcal{M} : \lambda_{\nu,i} + \rho(\mathbf{Nu} - \mathbf{d})_i > 0\}$$

and the respective inactive set is its complement $\mathcal{I}_\nu := \mathcal{I}_\nu(\boldsymbol{y}) = \mathcal{M} \setminus \mathcal{A}_\nu(\boldsymbol{y})$. For $\mathcal{S} \subseteq \mathcal{M}$ we introduce the diagonal matrix

$$\boldsymbol{D}_{\mathcal{S}} = diag(s_1, \dots, s_m) \in \mathbb{R}^{m\times m}, \quad s_i = \begin{cases} 1 & \text{for } i \in \mathcal{S}, \\ 0 & \text{for } i \notin \mathcal{S}. \end{cases}$$

Non–symmetric case. First we propose the algorithm for direct solving discrete contact problems with Coulomb friction. To this end we introduce two inactive sets $\mathcal{I}_t^+ := \mathcal{I}_t^+(\boldsymbol{y})$, $\mathcal{I}_t^- := \mathcal{I}_t^-(\boldsymbol{y})$ corresponding to the condition of Coulomb friction:

$$\begin{aligned} \mathcal{I}_t^+(\boldsymbol{y}) &= \{i \in \mathcal{M} : \lambda_{t,i} + \rho(\mathbf{Tu})_i > \mathcal{F}_i\lambda_{\nu,i}\}, \\ \mathcal{I}_t^-(\boldsymbol{y}) &= \{i \in \mathcal{M} : \lambda_{t,i} + \rho(\mathbf{Tu})_i < -\mathcal{F}_i\lambda_{\nu,i}\}. \end{aligned}$$

The respective active set is $\mathcal{A}_t := \mathcal{A}_t(\boldsymbol{y}) = \mathcal{M} \setminus (\mathcal{I}_t^+(\boldsymbol{y}) \cup \mathcal{I}_t^-(\boldsymbol{y}))$. After suitable adjustments we arrive at the following algorithm.

Algorithm *ActiveSetCoulomb1*

(0) Set $k := 1$, $\rho > 0$, $\varepsilon_u > 0$, $\boldsymbol{u}^{(0)} \in \mathbb{R}^{2n}$, $\boldsymbol{\lambda}_\nu^{(0)}, \boldsymbol{\lambda}_t^{(0)} \in \mathbb{R}^m$.

(1) Define the active and inactive sets at $\boldsymbol{y} = ((\boldsymbol{u}^{(k-1)})^\top, (\boldsymbol{\lambda}_\nu^{(k-1)})^\top, (\boldsymbol{\lambda}_t^{(k-1)})^\top)^\top$:

$$\mathcal{A}_\nu = \mathcal{A}_\nu(\boldsymbol{y}),\ \mathcal{I}_\nu = \mathcal{I}_\nu(\boldsymbol{y}),\ \mathcal{A}_t = \mathcal{A}_t(\boldsymbol{y}),\ \mathcal{I}_t^+ = \mathcal{I}_t^+(\boldsymbol{y}),\ \mathcal{I}_t^- = \mathcal{I}_t^-(\boldsymbol{y}).$$

(2) Solve:

$$\begin{pmatrix} \boldsymbol{K} & \boldsymbol{N}^\top & \boldsymbol{T}^\top \\ \boldsymbol{D}_{\mathcal{A}_\nu}\boldsymbol{N} & \boldsymbol{D}_{\mathcal{I}_\nu} & \boldsymbol{0} \\ \boldsymbol{D}_{\mathcal{A}_t}\boldsymbol{T} & \boldsymbol{\mathcal{F}}(\boldsymbol{D}_{\mathcal{I}_t^-} - \boldsymbol{D}_{\mathcal{I}_t^+}) & \boldsymbol{D}_{\mathcal{I}_t^+\cup\mathcal{I}_t^-} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^{(k)} \\ \boldsymbol{\lambda}_\nu^{(k)} \\ \boldsymbol{\lambda}_t^{(k)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{D}_{\mathcal{A}_\nu}\boldsymbol{d} \\ \boldsymbol{0} \end{pmatrix}.$$

(3) Set $err^{(k)} := \|\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k-1)}\|/\|\boldsymbol{u}^{(k)}\|$. If $err^{(k)} \le \varepsilon_u$, return $\boldsymbol{u} := \boldsymbol{u}^{(k)}$, $\boldsymbol{\lambda}_\nu := \boldsymbol{\lambda}_\nu^{(k)}$,
and $\boldsymbol{\lambda}_t := \boldsymbol{\lambda}_t^{(k)}$.

(4) Set $k := k + 1$ and go to step (1).

As we deal in step (2) with non–symmetric matrices with generalized saddle–point structure we use BiCGSTAB.

Symmetric case. The only difference when we want to propose the algorithm for solving the discrete problem with Tresca friction is in the definition of inactive sets when $\mathcal{F}_i \lambda_{\nu,i}$ is substituted by g_i. The algorithm then differs only in the step (2). The matrix now can be made symmetric and we can use CGM. The algorithm stays the same and the step (2) is presented below.

Algorithm *ActiveSetTresca*
(2) Solve:

$$\begin{pmatrix} \boldsymbol{K} & \boldsymbol{N}_{\mathcal{A}_\nu}^\top & \boldsymbol{T}_{\mathcal{A}_t}^\top \\ \boldsymbol{N}_{\mathcal{A}_\nu} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{T}_{\mathcal{A}_t} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^{(k)} \\ \boldsymbol{\lambda}_{\nu,\mathcal{A}_\nu}^{(k)} \\ \boldsymbol{\lambda}_{t,\mathcal{A}_t}^{(k)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f} - \boldsymbol{T}_{\mathcal{I}_t^+}^\top \boldsymbol{g}_{\mathcal{I}_t^+} + \boldsymbol{T}_{\mathcal{I}_t^-}^\top \boldsymbol{g}_{\mathcal{I}_t^-} \\ \boldsymbol{d}_{\mathcal{A}_\nu} \\ \boldsymbol{0} \end{pmatrix},$$

and set $\boldsymbol{\lambda}_{\nu,\mathcal{I}_\nu}^{(k)} = \boldsymbol{0}$, $\boldsymbol{\lambda}_{t,\mathcal{I}_t^+}^{(k)} = \boldsymbol{g}_{\mathcal{I}_t^+}$, $\boldsymbol{\lambda}_{t,\mathcal{I}_t^-}^{(k)} = -\boldsymbol{g}_{\mathcal{I}_t^-}$.

In order to solve the discrete contact problem with Coulomb friction, one can use the algorithm *ActiveSetTresca* in each successive approximation. In order to achieve high computational efficiency, we prefer the inexact implementation in which one iteration of the algorithm *ActiveSetTresca* is performed. In other words, we modify the algorithmic scheme *ActiveSetTresca* so that we take $\boldsymbol{g} := \boldsymbol{\mathcal{F}} \max\{\boldsymbol{\lambda}_\nu^{(k-1)}, 0\}$ in the beginning of each iteration.

3 Implementation

In this section, we interpret each step (2) of algorithm *ActiveSetTresca* as the minimization of a strictly quadratic objective function in terms of $\boldsymbol{\lambda}$, for that one can use the conjugate gradient method. We will arrive at the globally convergent dual implementation of the algorithm; see [3].

We start with notations.

$$\mathcal{A} = \{i|\ i \in \mathcal{A}_\nu\} \cup \{i+m|\ i \in \mathcal{A}_t\} \ \text{ and } \ \mathcal{I} = \{1,2,\ldots,2m\} \setminus \mathcal{A}.$$

Denote

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{T} \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{0} \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_\nu \\ \boldsymbol{\lambda}_t \end{pmatrix},$$

and

$$q(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{\lambda}^\top \boldsymbol{A}\boldsymbol{\lambda} - \boldsymbol{\lambda}^\top \boldsymbol{b},$$

where $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{K}^{-1}\boldsymbol{B}^\top$, $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{K}^{-1}\boldsymbol{f} - \boldsymbol{c}$. Finally note that the gradient to q at $\boldsymbol{\lambda}$ reads as

$$\nabla q(\boldsymbol{\lambda}) = \boldsymbol{A}\boldsymbol{\lambda} - \boldsymbol{b}.$$

The linear system in step (2) of algorithm *ActiveSetTresca* reads as

$$\begin{pmatrix} \boldsymbol{K} & \boldsymbol{B}_{\mathcal{A}}^\top \\ \boldsymbol{B}_{\mathcal{A}} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^{(k)} \\ \boldsymbol{\lambda}_{\mathcal{A}}^{(k)} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{f}} \\ \boldsymbol{c}_{\mathcal{A}} \end{pmatrix}, \tag{3.8}$$

where $\widehat{\boldsymbol{f}} = \boldsymbol{f} - \boldsymbol{T}_{\mathcal{I}_t^+}^\top \boldsymbol{g}_{\mathcal{I}_t^+} + \boldsymbol{T}_{\mathcal{I}_t^-}^\top \boldsymbol{g}_{\mathcal{I}_t^-}$ and the remaining components of $\boldsymbol{\lambda}^{(k)}$ are given by

$$\boldsymbol{\lambda}^{(k)}_{\nu,\mathcal{I}_\nu} = \boldsymbol{0},\ \boldsymbol{\lambda}^{(k)}_{t,\mathcal{I}_t^+} = \boldsymbol{g}_{\mathcal{I}_t^+},\ \boldsymbol{\lambda}^{(k)}_{t,\mathcal{I}_t^-} = -\boldsymbol{g}_{\mathcal{I}_t^-}. \tag{3.9}$$

The following lemma interprets the solution.

Lemma 1. *(i) Let $\mathcal{A} = \emptyset$. Then $\boldsymbol{\lambda}^{(k)}$ is fully determined by (3.9) and*

$$\boldsymbol{u}^{(k)} = \boldsymbol{K}^{-1}\widehat{\boldsymbol{f}} = \boldsymbol{K}^{-1}(\boldsymbol{f} - \boldsymbol{B}^\top \boldsymbol{\lambda}^{(k)}).$$

(ii) Let $\mathcal{A} \neq \emptyset$. Then $\boldsymbol{\lambda}^{(k)}_{\mathcal{A}}$ is the minimizer to the problem

$$\min \frac{1}{2}\boldsymbol{\lambda}_{\mathcal{A}}^\top \boldsymbol{A}_{\mathcal{A}\mathcal{A}} \boldsymbol{\lambda}_{\mathcal{A}} - \boldsymbol{\lambda}_{\mathcal{A}}^\top \widehat{\boldsymbol{b}}_{\mathcal{A}}, \tag{3.10}$$

where $\widehat{\boldsymbol{b}}_{\mathcal{A}} = \boldsymbol{B}_{\mathcal{A}}\boldsymbol{K}^{-1}\widehat{\boldsymbol{f}} - \boldsymbol{c}_{\mathcal{A}}$. The remaining components of $\boldsymbol{\lambda}^{(k)}$ are given by (3.9) and

$$\boldsymbol{u}^{(k)} = \boldsymbol{K}^{-1}(\widehat{\boldsymbol{f}} - \boldsymbol{B}_{\mathcal{A}}^\top \boldsymbol{\lambda}^{(k)}_{\mathcal{A}}) = \boldsymbol{K}^{-1}(\boldsymbol{f} - \boldsymbol{B}^\top \boldsymbol{\lambda}^{(k)}).$$

In the next lemma, we show how to define the active and inactive sets without the knowledge of $\boldsymbol{u}^{(k)}$.

Lemma 2. *Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k)}$. It holds:*

$$\begin{aligned}
\mathcal{A}_\nu &= \{i \in \mathcal{M} : \lambda_{\nu,i} - \rho\nabla_i q(\boldsymbol{\lambda}) > 0\}.\\
\mathcal{I}_t^+ &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho\nabla_{i+m} q(\boldsymbol{\lambda}) > g_i\},\\
\mathcal{I}_t^- &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho\nabla_{i+m} q(\boldsymbol{\lambda}) < -g_i\}.
\end{aligned}$$

Finally, we show how to determine $\boldsymbol{\lambda}^{(k)}$ via the constrained minimization problem.

Lemma 3. *Let $\mathcal{A} \neq \emptyset$. Then $\boldsymbol{\lambda}^{(k)}$ determined in step (2) of algorithm ActiveSetTresca is the minimizer of the problem:*

$$\min q(\boldsymbol{\lambda}) \tag{3.11}$$

$$\text{subject to } \boldsymbol{\lambda}_{\nu,\mathcal{I}_\nu} = \boldsymbol{0},\ \boldsymbol{\lambda}_{t,\mathcal{I}_t^+} = \boldsymbol{g}_{\mathcal{I}_t^+},\ \boldsymbol{\lambda}_{t,\mathcal{I}_t^-} = -\boldsymbol{g}_{\mathcal{I}_t^-}. \tag{3.12}$$

Algorithm *ActiveSetTresca* (Dual Version: Implementation 1)

(0) Set $k := 1$, $\rho > 0$, $\varepsilon_\lambda > 0$, $\boldsymbol{\lambda}^{(0)} \in \mathbb{R}^{2m}$.

(1) Define the active and inactive sets at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k-1)}$:

$$\begin{aligned}
\mathcal{A}_\nu &= \{i \in \mathcal{M} : \lambda_{\nu,i} - \rho \nabla_i q(\boldsymbol{\lambda}) > 0\},\\
\mathcal{I}_t^+ &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho \nabla_{i+m} q(\boldsymbol{\lambda}) > g_i\},\\
\mathcal{I}_t^- &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho \nabla_{i+m} q(\boldsymbol{\lambda}) < -g_i\},\\
\mathcal{I}_\nu &= \mathcal{M} \setminus \mathcal{A}_\nu,\ \mathcal{A}_t = \mathcal{M} \setminus (\mathcal{I}_t^+ \cup \mathcal{I}_t^-)
\end{aligned}$$

and $\mathcal{A}$, $\mathcal{I}$.

(2) If $\mathcal{I} \neq \emptyset$, set $\boldsymbol{\lambda}_{\mathcal{I}}^{(k)}$ so that $\boldsymbol{\lambda}_{\nu,\mathcal{I}_\nu}^{(k)} = \mathbf{0}$, $\boldsymbol{\lambda}_{t,\mathcal{I}_t^+}^{(k)} = \boldsymbol{g}_{\mathcal{I}_t^+}$, $\boldsymbol{\lambda}_{t,\mathcal{I}_t^-}^{(k)} = -\boldsymbol{g}_{\mathcal{I}_t^-}$.
If $\mathcal{A} \neq \emptyset$, find

$$\boldsymbol{\lambda}_{\mathcal{A}}^{(k)} := \arg\min \frac{1}{2} \boldsymbol{\lambda}_{\mathcal{A}}^\top \boldsymbol{A}_{\mathcal{A}\mathcal{A}} \boldsymbol{\lambda}_{\mathcal{A}} - \boldsymbol{\lambda}_{\mathcal{A}}^\top \widehat{\boldsymbol{b}}_{\mathcal{A}}.$$

(3) Set $err^{(k)} := \|\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^{(k-1)}\| / \|\boldsymbol{\lambda}^{(k)}\|$. If $err^{(k)} \leq \varepsilon_\lambda$, return

$$\boldsymbol{u} := \boldsymbol{K}^{-1}(\boldsymbol{f} - \boldsymbol{B}^\top \boldsymbol{\lambda}^{(k)}),\ \boldsymbol{\lambda} := \boldsymbol{\lambda}^{(k)}$$

and stop.

(4) Set $k := k + 1$ and go to step (1).

At this moment we implement CGM in step (2) of this last algorithm. Some numerical experiments have been done.

References

1. Z. Dostál. *Optimal Quadratic Programming Algorithms, with Applications to Variational Inequalities*. Springer, 2008, New York, 1st ed.
2. G. H. Golub, C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, 1996, Baltimore, 3rd ed.
3. R. Kučera, K. Rádková. *An Experimental Study of the Semi-Smooth Newton Method for Solving 2D Contact Problems with Tresca and Coulomb Friction*, in preparation
4. J. Nocedal, S. J. Wright. *Numerical Optimization*. Springer, 2006, New York, 2nd ed.
5. K. Rádková. *Newton–like methods for nonlinear programmig*. Master thesis, UP Olomouc, 2010.

Using Support Vector Machines in IDS Systems

Peter Scherer

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
peter.scherer@vsb.cz

Abstract. Intrusion Detection System (IDS) is a system, that monitors network traffic and tries to detect suspicious activity. In this paper we discuss the possibilities of application of Support Vector Machines (SVM) for use in the IDS. There we used SVM as classification SVM of type 1, known too as C-SVM. By appropriate choosing of kernel and SVM parameters we achieved improvements in detection of intrusion to system. Finally, we experimentally verified the efficiency of applied algorithms in IDS.

Key words: Intrusion Detection System, SVM, kernel

1 Introduction

Three criteria are important for computer systems security: confidentiality, integrity and availability. Computer security is defined as a protection against threads for these criteria. The major manners of computer security are techniques like user authentication, data encryption, avoiding programming errors and firewalls. They are known as first line of defense. The last line of defense is used *Intrusion Detection System* (IDS). An Intrusion Detection System is software application (device respectively) that monitors network and system activities for malicious attempts, threads or policy violations and produces reports and statistics. Several machine-learning paradigms including soft computing approach [2], neural networks and fuzzy inference system [11], genetic algorithms [14], Bayesian network, matrix factorization approach [16], multivariate adaptive regression splines etc. have been investigated for the design of IDS. In this paper we investigate and evaluate the performance of classification via Support Vector Machines. The motivation for using SVM is to improve the accuracy of the Intrusion Detection System.

2 Classification SVM of type 1 (C-SVM) and their parameters

Support Vector Machine (SVM) is a preferably technique for linear binary data classification. In [10] authors state that a classification task usually involves

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 454–459.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

separating data into training and testing sets. Each instance in the training set contains one target value (i.e. the class labels) and several attributes (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

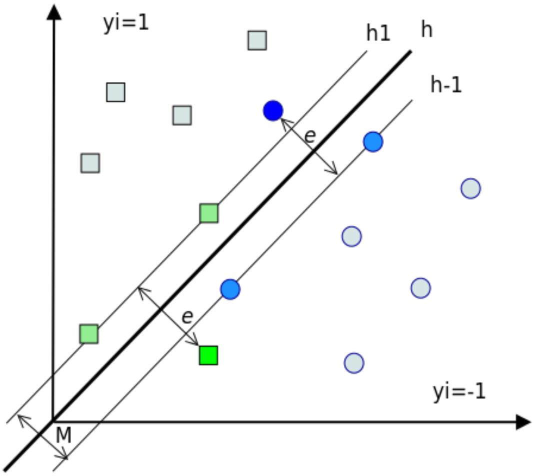


Fig. 1. General linear binary classification case.

Given a binary training set $(\boldsymbol{x}_i, y_i)$, $\boldsymbol{x}_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$, $i = 1, \ldots, m$, the basic variant of the SVM algorithm attempts to generate a separating hyper-plane in the original space of n coordinates (x_i parameters in vector $\boldsymbol{x}$) between two distinct classes, Fig. 1. During the training phase the algorithm seeks for a hyper-plane which best separates the samples of binary classes (classes 1 and -1). Let $h_1 : \boldsymbol{wx} + b = 1$ and $h_{-1} : \boldsymbol{wx} + b = -1$ ($\boldsymbol{w}, \boldsymbol{x} \in \mathbb{R}^n, b \in \mathbb{R}$) be possible hyper-planes such that majority of class 1 instances lie above h_1 and majority of class -1 fall below h_{-1}, whereas the elements coinciding with h_1, h_{-1} are hold for Support Vectors. Finding another hyper-plane $h : \boldsymbol{wx} + b = 0$ as the best separating (lying in the middle of h_1, h_{-1}), assumes calculating $\boldsymbol{w}$ and b, i.e. solving the nonlinear convex programming problem. The notion of the best separation can be formulated as finding the maximum margin M that separates the data from both classes. Since $M = 2\left\|\boldsymbol{w}\right\|^{-1}$, maximizing the margin cuts down to minimizing $\|\boldsymbol{w}\|$ Eq.(1).

$$\min_{\boldsymbol{w},b} \frac{1}{2} \left\|\boldsymbol{w}\right\|^2 + C \sum_i \varepsilon_i \tag{1}$$

with respect to: $1 - \varepsilon_i - y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \leq 0$, $-\varepsilon_i \leq 0$, $i = 1, 2 \ldots, m$

Regardless of having some elements misclassified (Fig. 1) it is possible to balance between the incorrectly classified instances and the width of the separating margin. In this context, the positive slack variables ε_i and the penalty parameter C are introduced. Slacks represents the distances of misclassified points to the initial hyper-plane, while parameter C models the penalty for misclassified training points, that trades-off the margin size for the number of erroneous classifications (bigger the C smaller the number of misclassifications and smaller the margin). The goal is to find a hyper-plane that minimizes misclassification errors while maximizing the margin between classes. This optimization problem is usually solved in its dual form (dual space of Lagrange multipliers):

$$\boldsymbol{w}^* = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \tag{2}$$

where $C \geq \alpha_i \geq 0$, $i = 1, \ldots, m$, and where $\boldsymbol{w}^*$ is a linear combination of training examples for an optimal hyper-plane. However, it can be shown that $\boldsymbol{w}^*$ represents a linear combination of Support Vectors $\boldsymbol{x}_i$ for which the corresponding α_i Langrangian multipliers are non-zero values. Support Vectors for which $C > \alpha_i > 0$ condition holds, belong either to h_1 or h_{-1}. Let x_a and x_b be two such Support Vectors ($C > \alpha_a, \alpha_b > 0$) for which $y_a = 1$ and $y_b = -1$. Now b could be calculated from $b^* = 0.5\boldsymbol{w}^*(\boldsymbol{x}_a + \boldsymbol{x}_b)$, so that classification (decision) function finally becomes:

$$f(\boldsymbol{x}) = \operatorname{sgn} \sum_{i=1}^{m} \alpha_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^* \tag{3}$$

To solve non-linear classification , one can propose the mapping of instances to a so-called feature space of very high dimension: $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^d$, $n \ll d$ i.e. $\boldsymbol{x} \rightarrow \varphi(\boldsymbol{x})$. The basic idea of this mapping into a high dimensional space is to transform the non-linear case into linear and then use the general algorithm already explained above Eqs. (1), (2), and (3). In such space, dot-product from Eq. (3) transforms into $\varphi(\boldsymbol{x}_i) \cdot \varphi(\boldsymbol{x})$. A certain class of functions called *kernels* [6] for which $k(\boldsymbol{x}, \boldsymbol{y}) = \varphi(\boldsymbol{x}) \cdot \varphi(\boldsymbol{y})$ holds, are called kernels. They represent dot-products in some high dimensional dot-product spaces (feature spaces), and yet could be easily recomputed into the original space. As example was chosen a Radial Basis Function Eq. (4), also known as Gaussian kernel [1], and was one of implemented kernels in the experimenting procedure.

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\gamma \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2) \tag{4}$$

Now Eq. (3) becomes:

$$f(\boldsymbol{x}) = \operatorname{sgn} \sum_{i=1}^{m} \alpha_i y_i k(\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^* \tag{5}$$

After removing all training data that are not Support Vectors and retraining the classifier, the same result would be obtained [6] by applying the function above. Thus, one depicted, Support Vectors could replace the entire training set, which is the central idea of SVM implementation.

3 Experiments

The data used for training and testing was prepared by the Agency DARPA intrusion detection evaluation program in 1998 at MIT Lincoln Labs [13]. Experiments were performed on a collection containing five pairs of data sets: the learning set (5092 vectors of 42 attributes) and testing set (6890 vectors of 42 attributes). Each pair represents a learning and testing data for one type of five classes of network attacks. Individual vectors describing the network traffic are described by 41 attributes (range $0 - 1$, is therefore not necessary to normalization). The 42^{nd} attribute was used in learning process. The attribute determines type of network attack in the question. In the case of testing, the existence of the attribute was neglected. We measure only classification accuracy of the vector, that describes the network attack.

3.1 Classification Using SVM type 1 (C-SVM)

It is necessary to determine the appropriate combination of parameters C and γ for better efficiency. In our experiment, the parameter C is in the range of 2^{-5} and 2^{15} in increments of powers of 2 and a parameter γ is in the range of 2^{-15} and 2^{3} in increments of powers of 2. We used 110 combinations of parameters C γ in total. In the case of same results of prediction with different parameters C and γ, the combination of parameters with the lowest time-intensive calculation model was chosen. In Tables 1,2, 3, and 4 is possible to see the best result combination.

The four most utilized kernel functions (linear, polynomial, RBF and sigmoid) was used for process of learning. As technology, we used library LibSVM [5].

Table 1. Classification using linear kernel.

Attack type	Training time (s)	C	γ	Accuracy (%)
Normal	0.71	2^{-1}	2^{-1}	99.55
Probe	0.25	2^{3}	2^{-1}	99.81
DOS	0.35	2^{7}	2^{-3}	99.81
U2R	0.17	2^{3}	2^{-3}	99.80
R2L	0.35	2^{5}	2^{-5}	99.64

4 Conclusion

In this paper we have described the method for the illustrated prediction accuracy by using SVM in the IDS. The best average success in SVM algorithm is more than 99%. Best of all kernels is RBF kernel, it has a success rate 99.846%. It will be useful to compare these kernel methods on other document collections. In our future work we will investigate other kernel functions to search for better attacks prediction in the IDS and SVM paralelization.

Table 2. Classification using polynomial kernel.

Attack type	Training time (s)	C	γ	Accuracy (%)
Normal	0.78	2^{13}	2^{-7}	99.83
Probe	0.24	2^{-3}	2^{-1}	99.81
DOS	0.47	2^{9}	2^{-5}	97.18
U2R	0.16	2^{15}	2^{-5}	99.80
R2L	0.24	2^{15}	2^{-5}	99.71

Table 3. Classification using RBF kernel.

Attack type	Training time (s)	C	γ	Accuracy (%)
Normal	0.88	2^{1}	2^{-3}	99.87
Probe	0.26	2^{5}	2^{-5}	99.90
DOS	0.29	2^{15}	2^{-7}	99.88
U2R	0.18	2^{9}	2^{-3}	99.83
R2L	0.37	2^{13}	2^{-7}	99.75

Table 4. Classification using sigmoid kernel.

Attack type	Training time (s)	C	γ	Accuracy (%)
Normal	0.95	2^{5}	2^{-5}	99.58
Probe	0.38	2^{7}	2^{-5}	99.88
DOS	0.43	2^{15}	2^{-9}	99.83
U2R	0.20	2^{5}	2^{-3}	99.83
R2L	0.42	2^{11}	2^{-7}	99.65

Table 5. Classification using SVM.

Attack type	SVM kernel			
	linear	polynomial	RBF	sigmoid
Normal	99.550	99.830	99.870	99.580
Probe	99.810	99.810	99.900	99.880
DOS	99.810	97.180	99.880	99.830
U2R	99.800	99.800	99.830	99.830
R2L	99.640	99.710	99.750	99.650
Average	99.722	99.266	99.846	99.754

5 Publications

1. Peter Scherer, Martin Vicher, Pavla Dráždilová, Jan Martinovič, Jiří Dvorský, and Václav Snášel: Using SVM and Clustering algorithms in IDS systems. Inproceedings of the Dateso 2011 Annual International Workshop on Databases, Texts, Specifications and Objects, p. 108-119, 2011.

Acknowledgment

This article is based on paper published on Dateso 2011.

References

1. S. Abe. Support Vector Machines for pattern classification. London, Springer, 2005.
2. A. Abraham and R. Jain. Soft Computing Models for Network Intrusion Detection Systems. Classification and Clustering for Knowledge Discovery Studies in Computational Intelligence, p. 191–207, 2005.
3. B. Al-Shboul and S.-H. Myaeng. Initializing k-means using genetic algorithms, 2009.
4. P. Berkhin. A Survey of Clustering Data Mining Techniques. Grouping Multidimensional Data, p. 25–71, 2002.
5. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001 `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
6. N. Cristiani and J. Shawe-Taylor. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge, Cambridge University Press, 2000.
7. R. Duda and P. Hart. Pattern Classification and Scene Analysis. John Wiley & Sons, New York, 1973.
8. D. H. Fisher. Knowledge Acquisition Via Incremental Conceptual Clustering. Kluwer Academic Publisher, 1987.
9. G. Gan, C. Ma, and J. Wu. Data Clustering Theory. Algorithms and Applications. ASASIAM, 2007.
10. C. Hsu, C. Chang and C. Lin. A Practical Guide to Support Vector Classification, journal Bioinformatics, 2003.
11. S. Chavan, K. Shah, N. Dave, S. Mukherjee, A. Abraham and S. Sanyal. Adaptive Neuro-Fuzzy Intrusion Detection Systems, International Conference on Information Technology: Coding and Computing (ITCC'04), 2004.
12. A. K. Jain and P.J. Flynn. Image segmentation using clustering. In Advances in Image Understanding: A Festschrift for Azriel Rosenfeld, IEEE Press, 65–83, 1996.
13. MIT Lincoln Laboratory `http://www.ll.mit.edu/IST/ideval/`
14. S. Owais, V. Snasel, P. Kromer and A. Abraham. Survey: Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques, p. 300–307, Computer Information Systems and Industrial Management Applications, 2008.
15. N. Sahoo. Incremental hierarchical clustering of text documents. adviser: Jamie Callan, 2006.
16. V. Snasel, J. Platos, P. Kromer, A. Abraham. Matrix Factorization Approach for Feature Deduction and Design of Intrusion Detection Systems, p. 172–179, The Fourth International Conference on Information Assurance and Security, 2008.

Robust library for Robot Soccer Game

Václav Svatoň

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vaclav.svaton@vsb.cz

Abstract. Robot Soccer is a very attractive platform in terms of research. It contains a number of challenges in the areas of robot control, artificial intelligence and image analysis. This article presents a look at the overall architecture of the game and describes some results of our experiments, mainly in analysis and optimization of strategies.

1 Introduction

Robot soccer is interesting mainly for its multi-agent research [1], including real-time image processing [2] and control, path planning, obstacle avoidance and machine learning. The robot soccer game presents an uncertain and dynamic environment for cooperating agents [3].

In robot soccer, the game situation in the playground is typically read in terms of the robot postures and ball position. Using real-time information of this dynamically changing game situation, the system of robot soccer team would need to continually decide the action of each team robot, and to direct each robot to perform a selected action.

The very objective of the game is as simple as in real football. Win the game over an opponent by higher number of scored goals. To achieve this goal it is necessary the best possible cooperation of team players and adaptability to the opponent's actual strategy. However, this often is not quite as easy, because your opponent is obviously trying the same thing.

2 Architecture

The main objective was to create a unified and robust library for robot soccer game, which will be used for the game of real robots and also for the game running in the simulator. The library consists of a number of interconnected modules that contain the functionality required for prediction, image analysis and derivation of the opponent's strategy or robot control. Such architecture brings many advantages. In particular, the possibility to experiment with different methods used to select the best winning strategy, or even create partially simulated game containing real and simulated robots.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 460–464.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

One of the key elements of the library is *storage*. *Storage* is used as a global repository of information about the current game. It contains the coordinates and speed of the robots of both teams and a ball. Information about the robot also includes a so-called *MoveTo* coordinates, which are calculated by strategies and that tell the simulator or physical robots, where should the robot move in next steps.

Our library also includes a multi-threaded support which means that the simulator engine and strategies for both teams now runs in separate threads. In practice, it looks like this. Simulator engine at regular intervals pulls current *MoveTo* coordinates of all robots from the *storage*, then calculates the new coordinates of real objects on the gamefield, redrawn the gamefield area of simulator and fill the newly calculated coordinates of all objects back into *storage*. At the same time there runs algorithms in separate threads for strategies of both teams. These algorithms can pull the current real coordinates of all objects on the gamefield from *storage*, according to them selects the appropriate rule from strategy and after tactics, fills the new *MoveTo* coordinates for each robot back into Storage. This cycle is repeated until the end of the game.

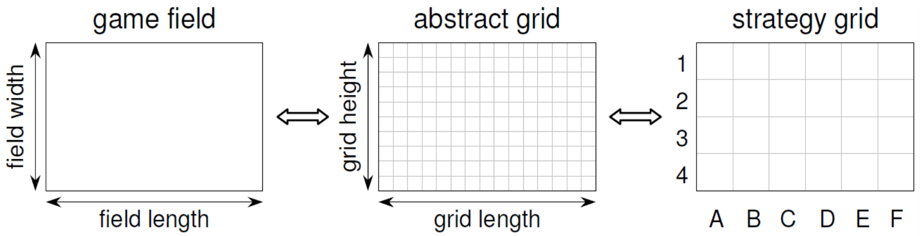


Fig. 1. Inner game representation

3 Strategies and Tactics

The strategy is a list of rules that describe the behavior of robots due to the current situation on the gamefield. Since a real coordinates of robots in the physical scene would be too detailed, here we introduce the grid coordinates (Fig. 1), which are also much more illustrative. The gamefield is logically divided into 6 x 4 squares and these grid coordinates are used by strategies. Each rule thus contains the grid coordinates of the ball, my and opponent's robots and also coordinates of where to move my robots. A more detailed description of strategies and rules can be found in [4].

Tactics are located under the strategies in the term of hierarchy so tactics are performed after strategies. The rule of strategy determines into which square (grid coordinate) robot has to move but only tactics determines the way how to get there. Tactics affect the behavior of the robot at a lower level than strategy, i.e. whether it can shoot, pass the ball, or to go toward the goal with the ball.

3.1 Generation of Offensive strategies

The main idea here is to generate an offensive strategy. One possible way is by analyzing the log from the recorded game. For this purpose we can use our simulator. In every step of the game, simulator records the coordinates of all robots on the game-field, coordinates of the ball and the rules of the strategies that were at the moment chosen for both teams. From this log we are able to easily reconstruct the course of the game and find out the last selected rule from opponent's strategy that has led to his scoring. This rule is extracted and put it into the set of rules of the other team to improve their strategy. The results of these experiments show that adding this rule, and then playing the same game, it is possible to substantially alter the development of the whole match, or even change the overall result for benefit of the other team.

In future work we would like to focus on generating defensive strategy that works similarly to the above mentioned case. From the last rule before the goal, we create defensive rule, which prevents the original rule to be executed on the opponent's side. These experiments are currently based on information from the log but for future practical use we plan to generate offensive and defensive strategies in real-time during game.

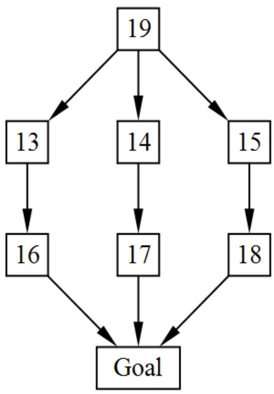


Fig. 2. Hub rule 19

3.2 Hub strategies

Before we discuss the problem of the *hub strategy,* firstly we introduce ranking model of strategies. K. Bryan and T. Leise [5] rank the importance of web pages according to an eigenvector of a weighted link matrix. This model is helpful to rank strategies. But strategy ranking is different to page ranking. In [5], a core idea in assigning an importance score to any given web page is that the page's score is derived from the links made to that page from other web pages. The links to a given page are called the backlinks for that page. The web thus becomes a democracy where pages vote for the importance of other pages by linking to them.

However, in strategy ranking, the strategy's score is derived from the links made from that strategy to other strategies. The links from a given strategy are called the *forwardlinks* for that strategy. The strategy thus becomes a power where strategies influence other strategies by linking to them. The strategy that has more links to other strategies is more important because that strategy has more influence on the process of game, the game situation is mostly depend on the follow-up execution of that strategy. In addition, this kind of strategy is a *hub* because many other strategies have to be achieved through it (Fig. 2). If we can destroy opponent's *hub*, then we eliminate most of the follow-up rules of opponent. Also to protect our strategy, we can find our *hub* and replace it with other rules that do not have this one common element by creating so-called *parallel strategy*.

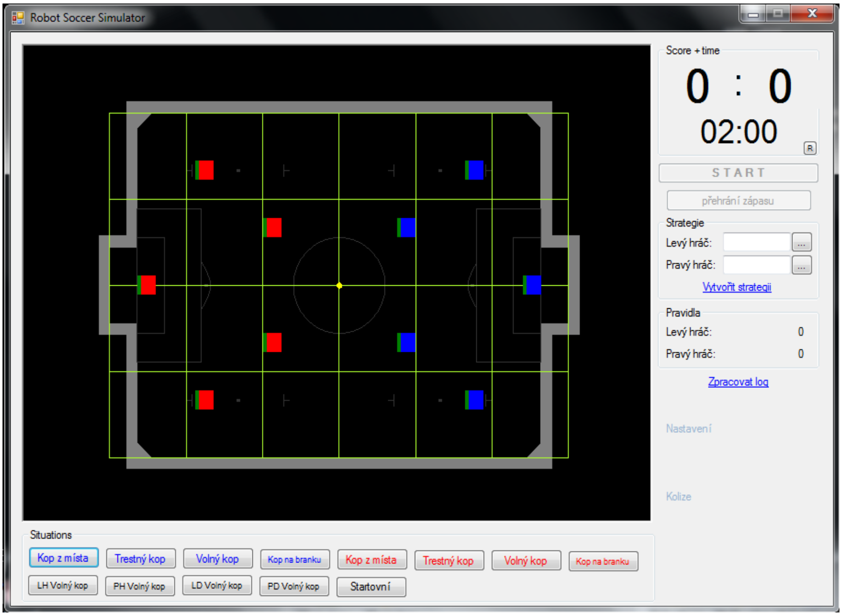


Fig. 3. Main screen of Robot Soccer Simulator

4 Robot Soccer Simulator and Strategy Creator

One part of the library is a *2D Simulator* that has been mentioned several times before (Fig. 3). Its use in the robot soccer game is almost necessary, especially for testing new strategies and tactics, or simply if you do not have a sufficient number of physical robots. The simulator also allows us to replay a recorded game from log, which is perfect for subsequent analysis of the previously played match.

An integral part of the simulator is an application called *Strategy Creator*, which is used for easy and intuitive creation of our strategies for the above-mentioned robot soccer library. The strategy is created by insertion of rules that are created based on the current situation on the gamefield. *Strategy Creator* also includes area called *the region of interest* (represented by a square around the ball), which prioritizes robots that are located within this area, when selecting the best rule from the strategy.

5 Conclusion

Description of strategy as well as the very architecture of robot soccer game is still an open question. Many researchers propose different approaches how to describe strategies or design architecture. We present one of many possible solutions, architecture based on several interconnected modules that contain the necessary functionality. This architecture allows us to experiment with different modules independently of the rest of the library. This article describes some methods for improvement of this architecture on which we are currently working on. Due to constantly improving the functionality of these modules, library is in constant development.

6 Publications

The following list contains only sent articles:

Experimental Soccer Robot Identification Using Light-Emitting Diodes
Martinovič J., Ochodková E., Svatoň V. and Kocyan T.

Improvement of Hub Strategy in Robot Soccer Game
Wu J., Snášel V., Martinovič J., Ochodková E., Svatoň V. and Abraham A.

Generation of Offensive Strategies in Robot Soccer Game
Wu J., Snášel V., Martinovič J., Ochodková E. and Svatoň V.

References

1. Bezek A.: Discovering strategic multi-agent behavior in a robotic soccer domain. In: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. ACM (2005)
2. Gomez V. M., Canas J. M., Martin F. S. and Matellan V.: Vision-based schemas for an autonomous robotic soccer player. In: IV Workshop de Agentes Fsicos WAF (2003)
3. Bezek A., Gams M., and Bratko I.: Multi-agent strategic modeling in a robotic soccer domain. In: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems. ACM (2006)
4. Martinovič J., Snášel V., Ochodková E., Zołtá L., Wu J., and Abraham A.: Robot soccer - strategy description and game analysis. In: Modelling and Simulation, 24th European Conference ECMS 2010, Kuala Lumpur, Malaysia (2010)
5. Bryan K. and Leise T.: The $25,000,000,000 eigenvector. The linear algebra behind google. In: SIAM Review, vol. 48, no. 3 (2006)

Quality Measuring of Segmentation Evaluation Methods

Štěpán Šrubař

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
stepan.srubar@vsb.cz

Abstract. Image processing can consists of segmentation. It could be performed by a human or an algorithm. In both cases, results can be evaluated by different methods. Such evaluation compares two segmentations, typically. This could serve for evaluation of quality of current segmentation or whole algorithm. Still, the results are based on one of many evaluation methods. Poor quality of evaluation method will lead to misleading results of quality of the segmentation algorithm. Therefore, evaluation methods should be also evaluated. Following article will present some evaluation methods, evaluation methodology and a data set. Results will show the quality of each evaluation method.

Keywords: Segmentation evaluation, Segmentation Difference.

1 Evaluation methods

Evaluation of segmentations is often done by humans. The main advantage is natural view on results. On the other hand, the evaluation is very slow and costly. Many algorithms for automatic evaluation of segmentations were proposed. They differ by limitations on segmentations, speed and quality. The first two can be and was easily measured but the third was omitted. Still, it is crucial to know the quality of evaluation algorithm. Results of low quality evaluation are useless. This article tries to describe and present evaluation of evaluation methods. Results are then used for classification of presented algorithms. Following section provides overview of segmentation evaluation algorithms, evaluation methodology, testing data set and the results of quality evaluation.

Detailed description of all methods would be too extensive, so only references will be provided. Few methods had to be extended because they were proposed for comparison of two segments only. Such extensions will be described.

Methods evaluating segmentations can be split into two different categories. Methods from the first category evaluate segmentation using the original image. In the second category, methods do not use the original image, thus they evaluate two segmentations to each other.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 465–470.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

1.1 Image-Segmentation Methods

First category consists of methods evaluating a segmentation using the original image. They include measuring of mean color, deviation of color in segments and contrast on segment borders, typically.

Goddness Uniformity (GU) [14] uses only the deviation of color. The same approach is used in Figure of Certainty (FOC) [14]. Variation of color in segments, as well as number of segments, was proposed Liu and Yang (LY) [7]. Borsotti et al. (BC) [1] used the same approach but penalized equally sized segments.

On the other hand, Levine and Nazif (LN) [6] proposed measuring of contrast of neighboring segments weighted by the length of the common border. The only method, that uses variation of color in segments, as well as contrast between neighboring segments, was proposed in Zeboudj contrast (ZC) [2]. Using the smallest neighborhood could lead to very high sensitivity to noise.

1.2 Segmentation-Segmentation Methods

Pair of Multi-class Errors (M_I, M_{II}) were proposed by W. A. Yasnoff et al. [14]. They require the same number of segments and corresponding segments have to have the same index. The same requirement occurs in Symmetric Divergence (SYD) [14]. D. Martin et al. proposed three quite similar methods (GCE, LCE, BCE) [5, 11]. They are independent on number of segments and does not need a priori knowledge of correspondence. Using the similarity of these methods, another method was artificially created ($GBCE$) [12].

One of many methods proposed by Q. Huang and B. Dom is called Hamming distance (HD) [5]. They proposed also 4 other methods without any naming. The only of them, which can be easily and practically used, is called $H_{2\mu}$ [4]. It is second proposed method that uses mean distances.

The simplest method denoted as Partition Distance (PD) [3] sums only sizes of intersections of corresponding segments. On the other hand, it requires equally numbered corresponding segments. Such condition is fulfilled by preprocessing of both segmentations. Same formula is proposed by Meilă and Heckerman (MH) [10]. The correspondence is found in original segmentations, therefore the results would totally differ from PD. Larsen (L) [9] uses relative size of intersections and takes the largest one only. The last presented method that uses intersections was proposed by Van Dongen (VD) [5]. It is one of few methods which are metrics.

Following methods are based on distance measuring. Such evaluations are slower than previous methods which used only sizes of intersections. On the other hand, distance of pixels to a single set can be computed in $O(n)$ time which is practically still fast enough. Yasnoff et al. (YD) [13] proposed sum of square of distances. Unfortunately, the results cannot be normalized. Figure of Merit (FOM) [14] multiplies the square by a parameter, one is added and the reciprocal value is taken. The results are higher than zero but lower or equal to one, evidently. Monteiro and Campilho (MC) [11] used combination of distance and logarithm of distance. Again, the results cannot be normalized.

Another group of methods compute number of couples of pixels. Both pixels of each couple can occur either in the same segment, or in the different segments. In case of two segmentations, each couple can have one of four possible properties according to being/not being in the same segment in first/second segmentation. Number of such couples are denoted naturally as N_{11}, N_{10}, N_{01} and N_{00}. Following methods use these four variables only.

Jaccard (JC) [5] divided N_{11} by sum of N_{11}, N_{10} and N_{01}. Fowlkes and Mallows (FM) [9] used geometric mean in the numerator. Wallace (W) [5] used similar formula as Jaccard, but he used either N_{01} or N_{10} to obtain two symmetrical results. Another metric was defined by Mirkin (M) [9]. We could obtain method proposed by Rand [5] from M by simple linear modification. Therefore, the quality evaluation of both methods will be the same. Rand method was extended for fuzzy segmentations (PRI) [5]. Still, the results for discrete segmentations are the same.

Last group of methods is based on statistical approach. Object Count Agreement (OCA) [14] evaluates the number of segments only. It uses limit, integral and gamma function. Classical entropy of information is used in Normalized Mutual Information (NMI) [5]. The basis was taken and another metric has arisen. It is called Variation of information (VI) [9].

1.3 Extension of Methods

Some methods evaluate only single segment but we need evaluation of whole segmentations. Methods M_I, M_{II} and FOM are modified and denoted as SM_I, SM_{II} and $SFOM$. Modification can be expressed by the following formula:

$$SX(S_1, S_2) = \frac{\sum_{s \in S_1} |s| \cdot X(s, corr(s))}{\sum_{s \in S_1} |s|}, \tag{1}$$

where X is original method and $corr(s)$ is segment corresponding to segment s.

1.4 Segmentation Difference

Following proposed method belongs to segmentation-segmentation category. It will be presented in more detail. Segmentation Difference (SD) was already proposed and presented in [12]. Basis of the method was retained but slight modifications were made.

$$SD(S_1, S_2) = (GD(S_1, S_2), BD_x(S_1, S_2)), \tag{2}$$

$$GD(S_1, S_2) = \frac{\sum_i |c_i| \cdot g(i)}{\sum_i |c_i|}, \tag{3}$$

$$g(i) = -\log_2 \left(\left[\sum_{j \in c_{i1}} \left(\frac{|s_{1j}|}{\sum_{k \in c_{i1}} |s_{1k}|} \right)^2 \right] \cdot \left[\sum_{j \in c_{i2}} \left(\frac{|s_{2j}|}{\sum_{k \in c_{i2}} |s_{2k}|} \right)^2 \right] \right), \tag{4}$$

$$BD_1(S_1,S_2) = \frac{\sum_i \left(\sum_{j\in c_{i1}} d(j,c_{i2}) + \sum_{j\in c_{i2}} d(j,c_{i1})\right)}{9\cdot w\cdot h\cdot \sqrt{w\cdot h}}, \tag{5}$$

$$BD_2(S_1,S_2) = \frac{2\cdot\sum_i |c_i| \left(\frac{\sum_{j\in bord(c_{i1})} d(j,bord(c_{i2}))}{|bord(c_{i1})|} + \frac{\sum_{j\in bord(c_{i2})} d(j,bord(c_{i1}))}{|bord(c_{i2})|}\right)}{3\cdot\sqrt{w\cdot h}\cdot\sum_i |c_i|} \tag{6}$$

$$|c_i| = \sum_{j\in c_{i1}} |s_{1j}| + \sum_{j\in c_{i2}} |s_{2j}|, \tag{7}$$

$$d(j,x) = \min_{k\in x} ||j-k||_2, \tag{8}$$

where s_{12} is the second segment from the segmentation S_1, c_{i1} is segment from i-th correspondence and segmentation S_1, $d(j,x)$ is distance of a pixel j to a set of pixels x, w and h are width and height, respectively, and $bord(x)$ is set of border pixels of a segment x.

Correspondence virtually connects segments from both segmentations. It could be categorized according to number of segments. One-one correspondence includes only one segment from each segmentation. One-many can contain more than one segment but either from the first or the second segmentation. Finally, many-many can include arbitrary number of segments from both segmentations.

The results of a method can differ greatly according to used type of correspondences. Segmentation Difference can process many-many correspondences. Still, we could use one-many only. Two different BD variants were presented. In conjunction with two possible types of correspondence, we could use four different variants of Segmentation Difference (see table 1). All four variants were evaluated.

	BD_1 (5)	BD_2 (6)
one-many	SD_1	SD_3
many-many	SD_2	SD_4

Table 1. Table of variants of Segmentation Difference using different types of correspondence and different definition of Border Distance.

2 Methodology and Results

Image segmentation database was created in Berkeley [8]. It consists of 300 images and each image was segmented by four humans at least. In case of image-segmentation evaluation methods, each method should evaluate a segmentation belonging to the correspondent image by a lower number and the segmentation from different image by a higher number. If we obtain results of a single method using all the images and segmentations, we could define a threshold that will determine which numbers are low and which are high. Practically, there will be some results, which will belong to the higher numbers even if it represents segmentation corresponding to the image. This is denoted as false rejection (FR). Symmetrically, low result which represents segmentation from different image is denoted as false acceptation (FA). Rates of both types of errors could be computed from the whole set of results. Naturally, we search for the optimal threshold, which will decrease the total error rate to the minimum. Such optimal threshold error rate will represent quality of an image-segmentation algorithm. Similarly, we could obtain results from evaluation of couples of segmentations using segmentation-segmentation algorithms. The quality of an algorithm is the error rate of optimal threshold. Number of results of image-segmentation algorithm was 980700. Number of results of symmetrical segmentation-segmentation algorithms was 5341546, while for asymmetrical algorithms it was 10683092 results.

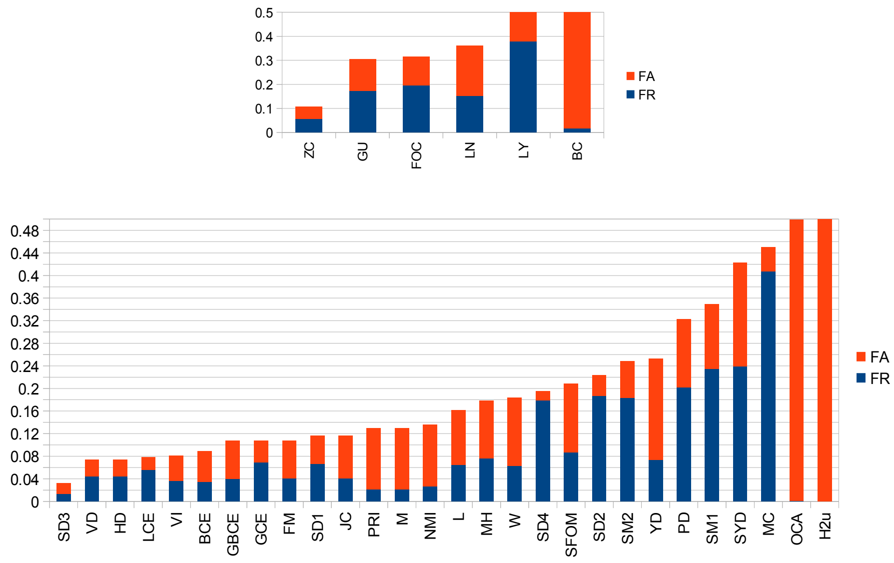


Fig. 1. Results of quality evaluation of image-segmentation and segmentation-segmentation algorithms.

Image-segmentation and segmentation-segmentation algorithms are evaluated separately due to different basis of evaluation approaches and practical use. Error rates of optimal thresholds can be seen in graphs in figure 1. Error rate is presented by fractional numbers. The lower the error is, the higher the quality of an evaluation algorithm is.

3 Conclusion

Presented methods were implemented and evaluated using large image segmentation data set. Quality evaluation showed great differences in presented methods. Best image-segmentation method was Zeboudj contrast with total error around 10%. In case of segmentation-segmentation algorithms, the best was proposed SD_3 algorithm with average error lower than 3.3%.

References

1. M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recogn. Lett.*, 19:741–747, June 1998.
2. Jean-Pierre Cocquerez and Sylvie Philipp-Foliguet. *Analyse d'images : filtrage et segmentation.* Masson, Paris, 1995.
3. Jaime dos Santos Cardoso and Luís Corte-Real. Toward a generic evaluation of image segmentation. *Image Processing, IEEE Transactions on*, 14(11):1773–1782, 2005.
4. Qian Huang and Byron Dom. Quantitative methods of evaluating image segmentation. In *Image Processing, 1995. Proceedings., International Conference on*, volume 3, pages 53–56 vol.3, 1995.
5. Xiaoyi Jiang, Cyril Marti, Christophe Irniger, and Horst Bunke. Distance measures for image segmentation evaluation. *EURASIP J. Appl. Signal Process.*, 2006:209–209, January 2006.
6. Martin D. Levine and Ahmed M. Nazif. Dynamic measurement of computer generated image segmentations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-7(2):155 –164, March 1985.
7. J. Liu and Y. H. Yang. Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:689–700, 1994.
8. David R. Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Technical Report UCB/CSD-01-1133, EECS Department, University of California, Berkeley, January 2001.
9. Marina Meilă. Comparing Clusterings by the Variation of Information. *Learning Theory and Kernel Machines*, pages 173–187, 2003.
10. Marina Meilă and David Heckerman. An experimental comparison of model-based clustering methods. *Mach. Learn.*, 42:9–29, January 2001.
11. Fernando Monteiro and Aurélio Campilho. Performance evaluation of image segmentation. In Aurélio Campilho and Mohamed Kamel, editors, *Image Analysis and Recognition*, volume 4141 of *Lecture Notes in Computer Science*, pages 248–259. Springer Berlin / Heidelberg, 2006.
12. Štěpán Šrubař. Survey and comparison of image segmentation evaluation algorithms. In *Annual Doctoral Workshop on Mathematical and Engineering Methods in Computer Science (MEMICS)*, 2010.
13. William A. Yasnoff, Jack K. Mui, and James W. Bacus. Error measures for scene segmentation. *Pattern Recognition*, 9(4):217 – 231, 1977.
14. Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, August 1996.

Hierarchical Acceleration of Mean-Shift Segmentation Methods

Milan Šurkala

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
milan.surkala.st@vsb.cz

Abstract. The Mean-Shift segmentation method is very popular for its good segmentation and filtration results. Many variations of Mean-Shift methods appeared in the last years in order to improve segmentation quality and, especially, the speed. In this paper, a hierarchical approach to the Mean-Shift segmentation is discussed. We present our methods called Hierarchical Blurring Mean-Shift and Hierarchical Evolving Mean-Shift.

Keywords: mean-shift, segmentation, filtration, hierarchy, acceleration

1 Introduction

In 1975, the Mean-Shift (MS) algorithm was presented [2]. It is a clustering algorithm often used for filtration and segmentation of various datasets. In this paper, we will focus on filtering and segmenting digital images represented by a set of pixels. Original MS has quite good results in respect of a filtration quality but has a disadvantage in an insufficient speed. Therefore, in the last years many new variations of Mean-Shift methods appeared.

Mean-Shift is an iterative algorithm that search for the position with the highest density of data points. Each data point is shifted according to the computed mean-shift vector toward the density maximum. This maximum to which many of similar data points converge, is called an attractor. The similarity of points is given by a parameter σ that limits the size of a neighbourhood (the area and brightness range in which the mean-shift method is computed).

2 Mean-Shift Methods

Consider $X = \{x_n\}_{n=1}^{N} \subset R^d$ as a dataset of N points in the d-dimensional space. The *Kernel density estimator* with a kernel $K(x)$ can be defined as

$$p(x) = \frac{1}{N} \sum_{n=1}^{N} K \left(\left\| \frac{x - x_n}{\sigma} \right\|^2 \right). \quad (1)$$

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 471–476.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

In this equation, N is the number of pixels, x is a pixel being processed and x_n are the pixels in a neighbourhood of the processed pixel. The term denoted by σ is a bandwidth, which limits the size of the neighbourhood. We can distinguish between two types of bandwidths. The spatial limit (in x and y axis) is denoted by σ_s and the range limit (luminance or colour) is denoted by σ_s.

Kernel function $K(x)$ is used as a weighting function. Many various types of kernels exist, we will focus on *Epanechnikov* kernel, which is denoted by equation

$$K(x) = \left\{ 1 \, , if \left\| 1 - x^2 \right\| \leq 10, if \left\| x \right\| > 1 \; . \right. \tag{2}$$

If we use previously mentioned Epanechnikov kernel in the original version of MS, mean shift vector would be represented by the equation

$$m_{\sigma,k}(x) = \frac{\sum_{i-1}^{N} x_i k \left(\left\| \frac{x - x_i}{\sigma} \right\|^2 \right)}{\sum_{i-1}^{N} k \left(\left\| \frac{x - x_i}{\sigma} \right\|^2 \right)} - x, \tag{3}$$

where the term x is the former position of the processed pixel and new computed position of this pixel is represented by the first term. It is computed as a weighted mean of pixels which belong to the searching window (sphere with dimensions given by σ_s and σ_r respectively).

Blurring Mean-Shift (BMS) [1] is an another method of Mean-Shift with a small modification of equation leading to a significantly different method of computation. MS does not change the original dataset (the computation is done with the original data), BMS changes the dataset in each iteration with computed values. Data modified in the previous stage are used as the source data for the next iteration. It is proven that BMS has faster convergence in comparison to MS. The main advantage of BMS is that it reduces the number of iterations. The equation of BMS can be written as

$$m_{\sigma,k}(x) = \frac{\sum_{i-1}^{N} x_i k \left(\left\| \frac{x_m - x_i}{\sigma} \right\|^2 \right)}{\sum_{i-1}^{N} k \left(\left\| \frac{x_m - x_i}{\sigma} \right\|^2 \right)} - x. \tag{4}$$

Evolving Mean-Shift (EMS) was presented in 2009 [6]. The mean shift vector is considered as an energy. The energy between two pixels is given by the equation

$$E_{x_i x_j} = f(x_i) \left(k(0) - k \left(\frac{x_i - x_j}{h_{x_i}} \right) \right). \tag{5}$$

The term f_{x_i} is a shrinking factor which can be used for modifying the size of the kernel. We set this value to 1, because we will use new hierarchical approach to modify the size. Energy of the dataset is a sum of all energies between all pixels.

At last, Hierarchical Mean-Shift (HMS) was also presented in 2009 [3]. It utilizes multiple Mean-Shift steps (we call it stages). Each stage uses different kernel size. Benefits of the hierarchical approach will be cleared in Section 3 devoted to our HBMS and HEMS method.

3 Hierarchical Blurring and Hierarchical Evolving Mean-Shift

In this section we will present our methods Hierarchical Blurring Mean-Shift [5] and Hierarchical Evolving Mean-Shift [4]. Number of computation is not only dependent on the size of a source image but also on the size of the searching window. In MS and BMS, number of steps is given by $n\pi\sigma^2 r$, where n is the number of pixels, σ is a radius of the searching window ($\pi\sigma^2$ is its area or number of pixels in a neighbourhood) and r is number of iterations. It is obvious that we can only lower number of iterations (which BMS does in comparison with MS) and the size of the neighbourhood. Smaller neighbourhood (lower value of σ_s) significantly improves the speed, but on the other hand, the number of segments rises and the segmentation is worse (more smaller segment emerge).

Idea of the hierarchical approach is that we can process this segmentation (using small σ_s) very fast and we obtain lowered number of small segments that is significantly smaller than original number of pixels. If we consider these segments as source pixels for the next mean-shift stage (weighted according to the number of pixels that the segments contains), we lower the number of source points n. So we can enlarge the size of the searching window in the next stage, while a complexity of computation will remain almost the same due to the reduced dataset. In the next stages, smaller number of larger segments is formed. Many stages can be run with increasing searching window and all of these stages run in very short time. BMS is more suitable for hierarchical approach as it forms more uniform segments lower in number than MS. Therefore, the reduction of the dataset and computational time is more significant. We proved it in [5].

4 Experiments

This section presents the experiments with HBMS and HEMS algorithms. We will concentrate on the segmentation quality, although this property can hardly be numerically evaluated as it is partly subjective. As a source image, the famous Lena gray-scale photo in 256×256 pixels was used. We also used picture from The Berkeley Segmentation Dataset. This picture has 481×321 pixels.

Figure 1(a) show filtered results of MS, BMS and EMS methods as well as of the 2-stage and the 3-stage configurations. We used $\sigma_r = 24$ for all algorithms and stages. Parameter σ_s was set to 40 for the final stage. In the 2-stage configuration, $\sigma_s 1 = 5$. We used $\sigma_s 1 = 3$ and $\sigma_s 1 = 9$ in the 3-stage configuration.

MS, BMS and EMS show nice filtered result with dithered gradients. Therefore, filtered images are quite nice for a human eye, but they do not form clearly defined segments. It obviously depends on the use of segmented images, whether it is appropriate or not. All 2-stage versions have clearly defined boundaries. These images have also good filtration, although slightly bigger difference between the original (presented in [5]) and filtered image is visible. 3-stage methods sometimes form more artifacts, but the advantage is in much higher speed. We can achieve $20 - 50\times$ speedup with the 2-stage algorithm and sometimes more than 100× speedup with the 3-stage variant. The speedup is highly dependent on a choice of kernel sizes (σ_s in early stages). Study of the optimal choice of the window size is a subject to further investigation.

(a) MS, $t = 1588s$ (b) HMS_2, $t = 50s$ (c) HMS_3, $t = 10.5s$

(d) BMS, $t = 161s$ (e) $HBMS_2$, $t = 6.2s$ (f) $HBMS_3$, $t = 3.6s$

(g) EMS, $t = 480s$ (h) $HEMS_2$, $t = 22.3s$ (i) $HEMS_3$, $t = 12.4s$

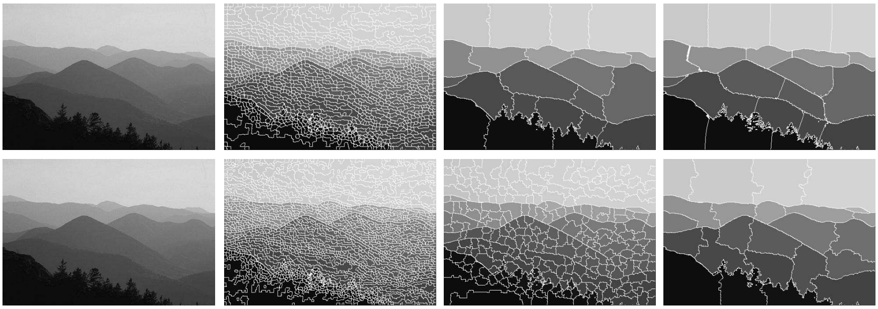
Fig. 1. Segmentation output from each method we compared.

Figure 2 shows evolution of segmentation in the 2-stage and the 3-stage HBMS. Kernel sizes were experimentally set to the values which led to the best computational time. The dataset originally contains 154,401 pixels (segments).

Hierarchical approach decreased computational time approximately 55 times in comparison with BMS. The 3-stage configuration decreased computational time more than 90 times. The numbers of segments in each stage are presented in the Table 1. The 2-stage HBMS has very similar segmentation quality as original BMS in fraction of original time. The 3-stage HBMS achieved greater speedup, but segmentation is visually slightly worse in respect of boundary shapes.

	1st stage	2nd stage	3rd stage	all stages	1st stage	2nd stage	3rd stage
BMS	851.86 s	0 s	0 s	851.86 s	62	-	-
HBMS2	9.12 s	6.29 s	0 s	15.41 s	2749	52	-
HBMS3	5.93 s	2.26 s	1.3 s	9.49 s	5004	459	31

Table 1. Computational time of BMS and HBMS with final searching window $\sigma_s = 50$.



The original image *After 1st stage* *After 2nd stage* *After 3rd stage∗*

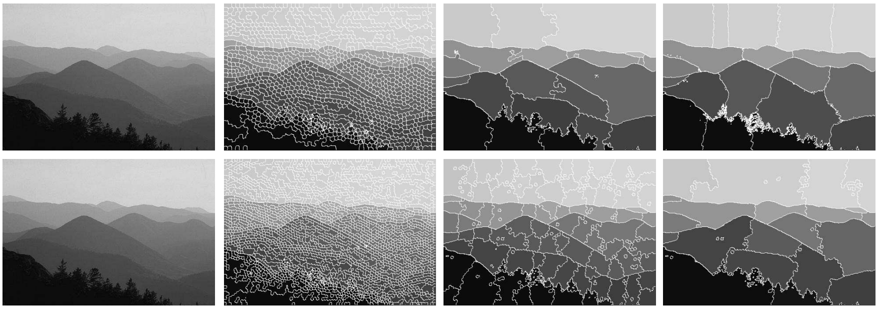
Fig. 2. Segmentation output after each HBMS stage ($\sigma_{s1} = 4$, $\sigma_{s2} = 50$ in the first row and $\sigma_{s1} = 3$, $\sigma_{s1} = 10$, $\sigma_{s2} = 50$ in the second row.
* BMS at the top

In a case of HEMS, significant improvement in the speed is also obvious. Even the 2-stage configuration achieved 64× speedup. We were able to achieve 95× speedup in the 3-stage variant of algorithm.

	1st stage	2nd stage	3rd stage	all stages	1st stage	2nd stage	3rd stage
BMS	2727.28 s	0 s	0 s	2727.28 s	25	-	-
HBMS2	19.63 s	22.5 s	0 s	42.13 s	1953	36	-
HBMS3	10.02 s	8.95 s	9.71 s	28.68 s	4477	183	36

Table 2. Computational time of BMS and HBMS with final searching window $\sigma_s = 50$.

Hierarchical variants of EMS have slightly larger number of segments. Nevertheless, segments have adequately shaped boundaries according to significant gradients and only weak post-processing can be needed. Subjectively, the filtration and segmentation quality has been lowered a little, but the improvement in the speed is a huge compensation. EMS is slightly less appropriate for the hierarchical approach than BMS in respect of image quality.



The original image *After* 1*st stage* *After* 2*nd stage* *After* 3*rd stage*∗

Fig. 3. Segmentation output after each HEMS stage ($\sigma_{s1} = 4.5$, $\sigma_{s2} = 50$ in the first row and $\sigma_{s1} = 3$, $\sigma_{s1} = 16$, $\sigma_{s2} = 50$ in the second row.
*EMS at the top

5 Conclusion

Hierarchical approach applied to various Mean-Shift methods proved to be very efficient way to significant increase of the computational speed without decreasing the segmentation quality (HBMS) or only small degradation of the filtration and segmentation quality in the case of HEMS. If we experimentally set the best σ values for each stage, we were able to get 60× and more than 90× speedup with 150.000 pixel image in the 2-stage and the 3-stage configuration respectively.

References

1. Carreira-Perpiñán, M.: Fast nonparametric clustering with Gaussian blurring mean-shift. In: Proceedings of the 23rd international conference on Machine learning. pp. 153–160. ICML '06, ACM, New York, NY, USA (2006)
2. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory 21(1), 32–40 (January 1975)
3. Vatturi, P., Wong, W.K.: Category detection using hierarchical mean shift. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 847–856. KDD '09, ACM, New York, NY, USA (2009)
4. Šurkala, M., Mozdřeň, K., Fusek, R., Sojka, E.: Hierarchical Blurring Mean-Shift. ACIVS 2011 (2011)
5. Šurkala, M., Mozdřeň, K., Fusek, R., Sojka, E.: Hierarchical Evolving Mean-Shift. MEMICS 2011 (2011)
6. Zhao, Q., Yang, Z., Tao, H., Liu, W.: Evolving mean shift with adaptive bandwidth: A fast and noise robust approach. In: ACCV (1). pp. 258–268 (2009)

Learning Styles in Adaptive Teaching

Ondřej Takács[1], Jana Šarmanová[1], and Kateřina Kostolányová[2]

[1]Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{ondrej.takacs, jana.sarmanova}@vsb.cz
[2]University of Ostrava, Pedagogical faculty, Department of Information and Communication Technologies, Dvorakova 7, 701 03, Ostrava, Czech Republic
katerina.kostolanyova@osu.cz

Abstract. In adaptive e-learning we try to make learning more efficient by adapting the process of learning to students' individual needs. To make this adaptation possible, we need to know key students characteristics – his motivation, group learning preferences, sensual type and various learning styles. One of the easiest ways to measure these characteristics is to use questionnaires. New questionnaire was created because there was no questionnaire to measure all these characteristics at once. This questionnaire was filled by 500 students from different fields of study. These results were analyzed using clustering, decision tree and principal component analysis. Several interesting dependencies between students' properties were discovered using this analysis.

Keywords: learning styles, e-learning, data mining analysis

1 Introduction

In [1] we outlined behavior of intelligent, virtual teacher who would adapt their teaching to both static and dynamically changing characteristics of students. Key parts of static student characteristics are learning styles, which we will analyze further on.

For definition of learning styles we will use [2]: Learning styles are delicate manifestations of human individuality in different learning situations. They are procedures of learning which person prefers in given period. Procedures original by their orientation, motivation, structure, succession, depth, elaboration and flexibility. Man uses them in most pedagogical situations; they are less dependent on content of learning. Learning styles have character of meta strategy, which is grouping activities of lower order: individual learning strategies, learning tactics in them and parts of learning tactics – learning operations. Monitoring them, evaluating them and orienting them in specific direction. Regulating them according to conditions of learning, course of learning, results of learning and social context of learning.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 477–482.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Properties of the Student

By analyzing and studying the already published classifications by various authors we have come to the following characteristics that may be taken into account in the area of e-learning. We have divided these characteristics into several groups:

The group of characteristics called the sensual perception describes the form of information most convenient for the student – visual type, auditive type, kinaesthetic type and verbal type. For this group of characteristics we used the VARK questionnaire [3].

Another group of characteristics called the social aspects deals with the company most convenient to the student within the learning process – whether the student prefers learning with his classmates, with a teacher or by himself. For his group of characteristics the LSI questionnaire was used [4].

The group of characteristics called the affective characteristics deals with student's feelings and attitude which influence the learning process. The most important characteristic of this category is motivation. For this group a part of the LSI questionnaire was used [4].

The most extensive group of characteristics is called the tactics of learning. These describe the „technology“ of the way the student learns. The method of the student's learning is described by the order of learning, which can be done either subsequently in mutually related blocks (pole rule) or in almost random manner, without connections, with large skipping (pole latitude). To describe this characteristic a part of the ILS questionnaire was used [5].

According to the way of learning we divide the tactics to theoretical deduction and experimentation. To describe these characteristics a part of the ILS questionnaire was used [5].

According to the order of learning there can be either detailistic or holistic tactics. For this characteristic a part of the TSI questionnaire was used [6].

The conception of student's learning can be divided into three levels: deep, strategic and surface. This characteristic is tested by ASSIST questionnaire [7].

The learning auto-regulation defines, to which extent the student is able to manage his learning independently. This implies his need of external supervision of the processes of learning where, on one side, there are those who appreciate punctual instructions and on the other side those who prefer to manage the learning processes themselves. For this characteristic a part of the LSI questionnaire was used [4].

2.1 Measure of Properties

To measure aforesaid characteristics, it is best to use questionnaires. In the pilot phase we used the above mentioned questionnaires, but their combination wasn't applicable due to their length. This led to inaccurate filling of the questionnaires, shown by the results of their analysis [8]. These reasons led to compiling a new shorted questionnaire LPQ [9]. The usual duration of filling the LPQ is only ten minutes while measuring all properties needed.

Most questions in the LPQ questionnaire are made of statements together with scales of agreement to this statement. These scales are usually from 1 (I don't agree) to 5 (I agree). Some questions have four variants, one for each sensual type.

This questionnaire was converted into electronic form. Before filling the questionnaire, its purpose was introduced to the students. A group of university students from different fields of study were asked to fill this questionnaire, as well as a group of high school students. Total of 508 students filled this questionnaire, 190 from those being grammar school students, 196 pedagogical university students, 68 high school students and 62 informatics university students. The duration of filling the questionnaires varied between two minutes and half an hour. To exclude suspiciously fast filling students we sat aside 45 questionnaires filled in less than five minutes.

From filled questionnaires we saved data about students' answers and questionnaire results that were computed according to a scoring key of this questionnaire. Values of all properties ranged from 0 to 100, zero meaning that the student does not possess this characteristic and one-hundred meaning that he does.

For the data analysis association rules method, which looks for dependencies between attributes and conveys them in form of rules, was used. Each rule consists of a condition and a conclusion and is assigned with reliability and support. *Reliability* provides information about how many percent of records found with this condition being true meet also the conclusion. *Support* provides information about for how many percent of records the condition is true.

Decision tree analysis was also performed. This analysis finds numbers of interesting rules for one goal attribute. Rules are then represented in a form of a tree. Leafs gives value of the goal attribute, nodes are conditions of rules and branches are values of conditions. Finally, a cluster analysis was performed, which tries to find groups of similar objects that are different from other group's objects in data.

3 Results of Analysis

3.1 Association Rules

Association rules found many dependencies between properties. We were looking for rules with minimal support of 10% and minimal reliability of 70% and found total of 147 rules. Majority of these rules have very low level of depth learning in conclusion. Conclusions also contained visual type, strategic learning and motivation. Some of these rules are shown in Table 1. Most interesting are the rules with high support. For example almost half of non-visual type of students mostly don't possess depth learning style (see rule number 1). Unfortunately 70% of all students don't possess depth learning style, so many of these rules are deformed by this fact. Because of this we can't make any valid conclusions based on this analysis. But we can use its results to determine goal attributes for the decision tree analysis: depth and strategic learning, visual perception and motivation.

Table 1: Picked rules with 70% minimal *reliability* and 10% minimal *support*, ordered by *support*

Num	Condition	Conclusion	Support	Reliability
1	visual = very low	depth = very low	43%	71%
2	strategic = very low	depth = very low	38%	76%
3	social (alone - group) = middle	depth = very low	38%	72%
4	verbal = low	depth = very low	31%	73%
5	theoretical – experimental = high	depth = very low	31%	76%
…	…	…	…	…
32	motivation = high, theoretical – experimental = high	depth = very low	17%	73%
33	social (alone - group) = middle, verbal = low	depth = very low	17%	75%
34	auditive = middle	visual = very low	16%	80%
…	…	…	…	…
69	depth = very low, systematic (sequential - random) = very low	strategic = very low	13%	72%
70	visual = very low, motivation = high, strategic = very low	depth = very low	13%	73%
…	…	…	…	…
87	depth = very low, visual = very low, field = University, pedagogical	motivation = high	12%	70%
88	strategic = very low, visual = middle	depth = very low	12%	79%
…	…	…	…	…
107	depth = very low, surface = very low, autoregulation = low	visual = very low	11%	71%
…	…	…	…	…
147	autoregulation = low, social (alone - group) = low	depth = very low	10%	87%

3.2 Decision Tree Analysis

This method was used for four goal attributes based on the association rule analysis results: depth and strategic learning, visual perception and motivation. Further on we describe results of these analyses.

By analyzing depth learning (see Fig. 1) we found out that holistic, group learning students mostly have depth learning style. Other results derived from this tree can't be unfortunately used because 70% of students in this data sample don't have depth learning style.

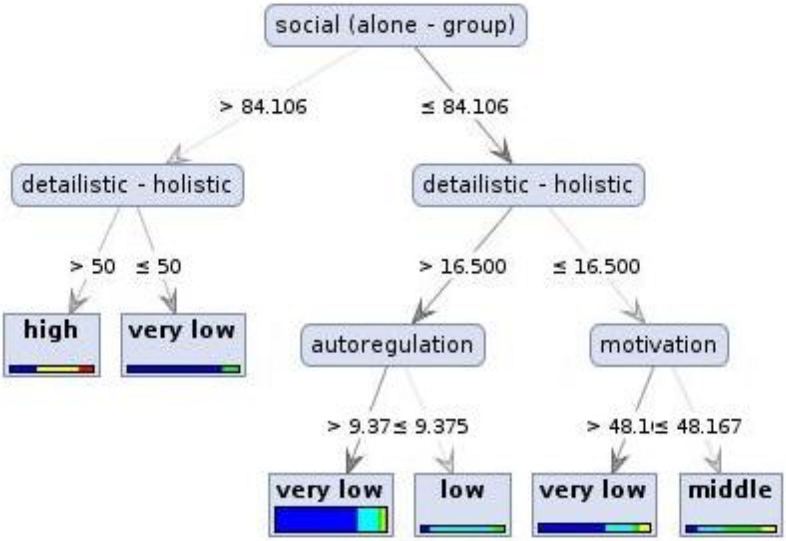


Figure 1: Decision tree for depth learning

The analysis of strategic learning style revealed that most experimenting students, not learning alone and randomly don't possess the strategic learning style. Other rules provided by this tree cannot unfortunately be used because of their low reliability.

By analyzing motivation we discovered that students have high motivation if they:

- don't possess surface learning style,
- possess surface learning but don't have high auto-regulation,
- possess neither surface learning style, nor high auto-regulation nor strategic learning style.

Through the analysis of visual perception we found out that students don't possess visual perception if they:

- are auditive types,
- are not auditive and are kinesthetic types,
- are neither auditive nor kinesthetic, but are verbal types.

3.3 Cluster Analysis

Agglomerative clustering didn't find any significant clusters. The out-coming dendogram only showed isolated points that gradually clustered to one global cluster. If there were several big clusters, it would mean that there are often only some combinations of students' properties and majority of these combinations are very rare. But there are no such clusters so we can't simply set small number of learning types and assign students to them.

4 Conclusion

Newly created LPQ questionnaire has many advantages over previously used group of questionnaires. Its main benefit is lower filling time that led to more honest filling. Analysis of filled questionnaires found only minor dependencies between chosen learning styles. That’s why it’s suitable to use formerly defined learning styles. One potential problem is that very few students possess the depth learning style. This can be explained by ill questions in the questionnaire or by the composition of the set of the tested students.

Another important conclusion derived from the cluster analysis is that students can’t be divided in several groups that have similar properties. To the contrary, we must deal with every student individually because his properties are very different from the properties of other students. This is an important result for the design of the adaptive learning system. We can’t adapt learning to only several groups and simply test every student and assign him to one of the groups.

LPQ questionnaire proved to be suitable for using directly in adaptive learning process. It will be presented to untested students at the beginning of the lecture. We plan to do more analysis of results of this questionnaire on students that will participate in adaptive learning.

These results were published at [10].

References

1. Kostolányová, K., Šarmanová, J. Intelligent individualization of study through e-learning. Ostrava: Ostravská univerzita, pp. 136--141. (2008)
2. Mareš, J. Styly učeni žáků a student, Prague: Portal (1998)
3. Fleming, N. D, Mills, C.: Not Another Inventory, Rather a Catalyst for Reflection, To Improve the Academy, vol. 11, pp. 137. (1992)
4. Mareš, J., Skalská, H.: Česká verze dotazníku LSI, zjišťující styly učení u žáků základních škol. Pedagogický výzkum a transformace české školy, pp. 54. (1993)
5. Felder, M. R., Soloman, A. B.: Index of Learning Styles, http://www.ncsu.edu/felder-public/ILSpage.html (2009)
6. Sternberg, J. R Thinking styles, Cambridge: University Press, pp. 192 (1999)
7. Entwistle, N, Tait, H.: Approaches and Study Skills Inventory for Students, Centre for Research on Learning and Instruction, Edinburgh (1996).
8. Takács, O., Šarmanová, J., Kostolányová, K., Results of analysis of learning styles. Information and Communication Technology in Education. Ostrava: Ostravská univerzita, pp. 205--210. (2009)
9. Novotný, J. S.: Individualization of teaching through e-learning: Development of Students? Learning Profile Questionnaire, Theoretical and Practical Aspects of Distance Learning, Cieszyn (2010)
10. Takács, O., Kostolányová, K., Šarmanová J. Adaptation of Presentation and Structure of e-Learning Material. IASK, 2010. pp. 215--222. (2010)

Parallel SOM Learning

Lukáš Vojáček

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.vojacek@vsb.cz

Abstract. The paper is oriented to the problem of clustering for large datasets with high-dimensions. Based on the learning phase using artificial neural network, especially Self organizing map, which we find as a suitable method for the reduction of the problem complexity. Due to the fact, that the learning phase of artificial neural networks can be time-consuming operation (especially for large high-dimensional datasets), we decided to accelerate this phase using parallelization to improve the computational efficiency.

1 Introduction

Recently, the problem of high-dimensional data clustering is arising together with the development of the information and communication technologies which supports growing opportunities to process the large data collections.

With increasing data dimensionality, there are commonly discussed two main challenges. The first is the dimensionality, which rapidly increases the computational complexity with respect to the number of dimension. Therefore, this problem makes some common algorithms computationally impracticable in many real applications. The second challenge is the specific similarity representation in the high-dimensional space.

The authors propose the effective clustering algorithm exploiting the features of neural networks, especially Self organizing maps (SOM), for the reduction of data dimensionality. The problem of computational complexity is solved using the parallelization of the basic SOM algorithm.

2 Self Organizing Maps

Self organizing map (SOM), also called Kohonen map, is a type of artificial neural network invented by professor Teuvo Kohonen in 1982 [8]. The input space of the training samples is represented in a low dimensional (often two-dimensional) space, called map. The model is capable of projecting the high-dimensional space to the lower-dimensional space [9] and is efficient in the structure visualization due to its feature of the topological preservation using a neighborhood function. The obtained low-dimensional map is often used for pattern detection, clustering, or for characterization and analysis of the input space.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 483–488.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

SOM technique has been applied in many spheres like speech recognition [11, 3], image classification [7, 2], document clustering [6, 4] etc. The detailed description of the SOM application is provided in [3].

The model of SOM consists of two layers of nodes. The input layer for receiving and transmitting the input information and the output layer called the map represented the output characteristics. The output layer is commonly organized as the two-dimensional map of nodes, but there are known extensions as, for example, hexagonal grid of output layer. The both layers are feed-forward connected. It is known, that the maps with the smaller grid of the output layer have the behavior similar to K-means clustering [1]. Thr larger output maps have the ability to describe the topological characteristics of the input data collection (often with using U-Matrix for the interpretation of the distance between the nodes).

The SOM input layer is given by input vectors $\boldsymbol{x} \in \mathbb{R}^n$. Each node of this layer is then connected with one of the output nodes K by means of a weight of reference vector $\boldsymbol{w}_k \in \mathbb{R}^n, k = 1, \ldots, K$. During the learning phase, the weight vector $\boldsymbol{w}_k(t)$ is computed for the network at time t, where $t = 0, 1, \ldots$ is discrete time index for each input vector $\boldsymbol{x}(t)$. The passage through the network at time t is an epoch. The learning (training) phase is performed through competitive learning, where for each training example $\boldsymbol{x}(t)$ is computed similarity to all weight vectors $\boldsymbol{w}_k(t)$. The output neuron with the most similar vector is then called as the *best matching unit* (BMU). The weights of the winning neuron and the neurons in the closest neighborhood are then updated and adjusted for the appropriate input vector. The weight vector initialization is commonly assigned randomly, or by using other data mining methods. Concrete implementation of the SOM depends on the method used for the weight vectors' actualization during the training phase.

2.1 Parallel SOM Algorithms

Till lately, most of the conventional algorithms were designed as sequential. The sequential algorithms were well suited to the past generation of computers, which basically performed the operations in the sequential fashion. With the development of the parallel computation, where the operations were performed simultaneously, there is growing the necessity to redesign the serial algorithms to their parallel implementations.

The parallelization of SOM learning algorithms can be implemented by the network partitioning. The *network partitioning* is the implementation, where the neural network is partitioned among the processors. Then, each input data sample is processed by its assigned processor or the parallel task. The network partitioning was implemented by several authors [5, 12].

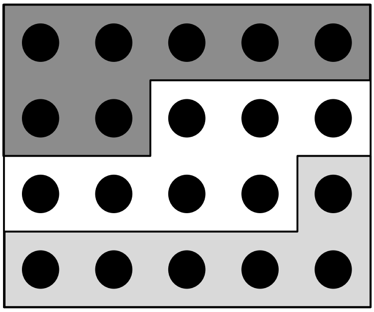


Fig. 1. SOM Map Division

The main principle of our parallel implementation is based on division of the neural network into the parts, where each part is assigned to one processor. This division is shown in the Fig. 1, where we have the map of 5×4 nodes. This map is divided into 3 parts which are associated with 3 processors. Not always there is the possibility to divide the map into the identical parts. In these cases, there is the neural network (map) divided using the principle, that the parts of the network differ in at the most one neuron.

The training phase is based on the serial version of the SOM algorithm. Each process finds its own BMU in its part of the map; this node is then compared with other BMU obtained by other processes. The information about the BMU of the whole network is then transmitted to all the processes, which in accordance with this information update weights of the appropriate nodes in their parts of the network.

Extended Parallel SOM Algorithms During the competitive learning phase, where for each input vector $\boldsymbol{x}(t)$ is computed similarity to all weight vectors $\boldsymbol{w}_k(t)$, and where the best matching unit (BMU) is founded, we have notified the problem with high computation complexity while working with higher dimensions. Instead of standard method for calculation of similarity with Euclidean distance, see Eq. (1), we have used its modified version using multiplication, see Eq. (2).

$$d_k(t) = \sqrt{\sum_{i=1}^{n}(x_i - w_i)^2}, \tag{1}$$

where n is the dimension of input vector $\boldsymbol{x}$, and $\boldsymbol{w}$ is the neuron's weight.

$$d_k(t) = \sqrt{\sum_{i=1}^{n} x_i^2 - 2x_i w_i + w_i^2} \tag{2}$$

Due to this modification we can calculate x_i^2 and w_i^2 at the beginning, and then during the computation of euclidean distance we can compute only $2 * x_i * w_i$. Than, in the actualization phase, we must recalculate only the part w_i^2, because only this value is changing during the weight actualization.

During the learning process, the standard approach is, that there are changed all the neuron weights. In our case, where we are working with spars data collection, there can be changed only the weight values, which are different from zero[10].

This method brings certain level of error, which can be eliminated using two approaches:

- We can do standard actualization in periodic epochs.
- We can do standard actualization only if any of weight component exceeds the limit value (threshold).

3 Experiments

The experiments was oriented to the acceleration of the SOM algorithm. As mentioned above, we have tested the parallel implementation of the SOM algorithm training phase for various dimensions of the SOM map and the input vector.

All the experiments were performed on Windows HPC server 2008 with 6 computing nodes, where each node has 8 processors with 12 GB of memory.

3.1 Standard SOM Acceleration

The first experiment was provided on the training set of 300 2-dimensional samples, while the dimension of the neural network (SOM map) was changed. The outputs are presented in the Table 1, where the records with asterisk (*) were provided only by one computing node. In these cases, there is not provided the network communication between processes and due to this fact is the computation faster.

Table 1. Computing Time Dependence on Map Dimension and Number of Processors

Processors	Computing Time [hh:mm:ss]		
	Map Dimension 100×100	Map Dimension 300×300	Map Dimension 500×500
1*	0:01:20	0:12:57	0:35:57
8*	0:00:16	0:01:56	0:21:43
16	0:00:14	0:01:06	0:03:18
24	0:00:13	0:00:50	0:02:05
32	0:00:15	0:00:48	0:01:37
40	0:00:15	0:00:38	0:01:21

The second experiment was provided on the map with selected dimension of 100×100 nodes, while the input vector dimension of the training data set was changed. There was used the training set of 150 records. The outputs are presented in the Table 2, where the records with asterisk (*) were provided only by one computing node. In this cases, there is not provided the network communication between processes; due to this fact is the computation faster.

Table 2. Computing Time Dependence on Input Vector Dimension and Number of Processors

Processors	Computing Time [hh:mm:ss]	
	Input Vector Dimension 4	Input Vector Dimension 8
1*	0:06:29	0:12:08
8*	0:01:01	0:01:48
16	0:00:36	0:01:00
24	0:00:27	0:00:44
32	0:00:25	0:00:38
40	0:00:23	0:00:34

As we can see from Tables 1 and 2, the acceleration of the SOM algorithm is appreciable. With growing number of processors is increasing the computation effectiveness, and the computational time is sufficiently reducing.

3.2 Extended SOM Acceleration

We have used sparse data collection of 8707 input vector dimension, the data collection consisted on 1033 records. We have set the SOM map of 25 × 25 neurons, for learning phase were processed 50 epochs. For the approach with hybrid algorithm was set the threshold for weight actualization to 0.3.

Table 3. Final Error and Total Time for Different Experiments

Usage of Euclidean distance	Threshold	Error	Time (sec)
All epochs (standard algorithm)		3.0875	185.649
First and last epoch		3.7578	7.351
First and last epoch	yes	3.2512	18.002
Every tenth epoch		3.7523	7.835
Every tenth epoch	yes	3.2124	23.128
Every even epoch		3.6535	14.962
Every even epoch	yes	3.1200	53.747

4 Publications

1. Lukas Vojacek, Jan Martinovic, Katerina Slaninova, Pavla Drazdilova, Jiri Dvorsky: Combined Method for Effective Clustering based on Parallel SOM and Spectral Clustering. In *proceedings of the Dateso 2011 Annual International Workshop on Databases, Texts, Specifications and Objects*, p. 120-131, 2011.

2. Lukas Vojacek, Jan Martinovic, Katerina Slaninova, Jiri Dvorsky, Ivo Vondrak: Parallel Hybrid SOM Learning on High Dimensional Sparse Data. In *The 10th International Conference on Computer Information Systems and Industrial Management Applications (CISIM 2011)*.[1]

Acknowledgment

This article is based on paper published on Dateso 2011.

References

1. F. Bacao, V. Lobo, and M. Painho. Self-organizing maps as substitutes for k-means clustering. In *Computational Science - ICCS 2005, Pt. 3, Lecture Notes in Computer Science*, pages 209–217, 2005.
2. H. Bekel, G. Heidemann, and H. Ritter. Interactive image data labeling using self-organizing maps in an augmented reality scenario. *Neural Networks*, 18(5-6):566–574, June-July 2005.
3. B. Gas, M. Chetouani, J.-L. Zarader, and C. Charbuillet. Predictive kohonen map for speech features extraction. In W. Duch, J. Kacprzyk, E. Oja, and S. Zadrozny, editors, *Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005*, volume 3697 of *Lecture Notes in Computer Science*, pages 793–798. Springer Berlin / Heidelberg, 2005.
4. A. Georgakis and H. Li. An ensemble of som networks for document organization and retrieval. In *Proceedings of AKRR'05, International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning*, pages 141–147, 2005.
5. W. Gropp, E. Lusk, and A. Skjellum. *Using MPI: portable parallel programming with the message-passing inferace.* MIT Press, 1999.
6. K. Kishida. Techniques of document clustering: A review. *Library and Information Science*, 35(1):106–120, January 2005.
7. O. Kohonen, T. Jaaskelainen, M. Hauta-Kasari, J. Parkkinen, and K. Miyazawa. Organizing spectral image database using self-organizing maps. *Journal of Imaging Science and Technology*, 49(4):431–441, July-August 2005.
8. T. Kohonen. *Self-Organization and Associative Memory*, volume 8 of *Springer Series in Information Sciences.* Springer, Berlin, Heidelberg, 1984. 3rd ed. 1989.
9. T. Kohonen. *Self Organizing Maps.* Springer-Verlag, 3rd edition, 2001.
10. H. C.-R. J. Machon-Gonzalez, I. Lopez-Garcia. A hybrid batch som-ng algorithm, 2010.
11. S. Meenakshisundaram, W. L. Woo, and S. S. Dlay. Generalization issues in multiclass classification - new framework using mixture of experts. *Wseas Transactions on Information-Science and Applications*, 4:1676–1681, Dec 2004.
12. C. H. Wu, R. E. Hodges, and C. J. Wang. Parallelizing the self-organizing feature map on multiprocessor systems. *Parallel Computing*, 17(6–7):821–832, September 1991.

[1] Submited

Communicating Sequential Processes

Robert Vajdík

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vaj049@vsb.cz

Abstract. This article deals with issues of communication sequential processes. The first part is devoted to the process algebra in general. Further in text language of CSP is described in more details. This article is focused on familiarization with the structure of language of CSP, such as sequential communication, recursion, selection, parallel communication, hiding, renaming, and behavioral processes. There is described the usage of behavioral diagrams for visual record of CSP language in the last section.

1 Introduction

With development of advanced computing architectures, the emphasis is put on new optimized algorithms. These algorithms use parallelism and concurrency. Processes which implement algorithms decompose into more parallel processes which cooperate, communicate and synchronize. For modeling and verification of such systems of simultaneous processes are strictly necessary formal means. For a description of parallel and distributed systems is used process algebra. Process algebra sees the system as a network of processes. Process can be any system whose behavior is expressed by discrete actions. Actions are either interactions between processes (communication) or may be internal actions of processes that are independent on other processes. Working with processes can be also affected by operations. On the basis of defined semantic processes and complicated operations, different processes session equality (congruence due to the comparison, weak bi-simulation, strong bi-simulation) can be defined. If we use the specifications and implementations to describe the same process algebra, we should be able to show that implantation meets specification proving relationship of equality. Unfortunately, most of the problems formulated by means of process algebras is un-decidable in principle. The process algebra includes: CCS (Calculus of Communicating Systems) [2,3], ACP (Algebra of Communicating Processes) [4], CSP (Communicating Sequential Processes) [1]. Further in this text CSP are described.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 489–494.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

1.1 CSP

Communicating Sequential Processes (CSP) is based on the mathematical theory of concurrent processes. This language was created in 1978. The current form of the language went through many changes. The mathematical theory of concurrent processes deals with the simultaneously of calculations and affect each other. This language is constantly evolving and improving. Huge contributes was in 1968 by Carl Adam Petri with his research with Place / Transition Net, also known under the name of Petri nets.
CSP is an advanced language for modeling concurrent systems. CSP and PN modeling languages are based on the same mathematical theory of concurrency process. Here I would like to refer to the interesting work Stochastic Analysis of CSP Specifications Using CSP and-a-Petri Net Translation Tool: CSPN [5]
CSP is an advanced language for modeling concurrent systems based on Petri nets. CSL is a formal language for describing and modeling of parallel systems which communicate with each other via messages. Communication is meant only communication between systems, but it can also be man-channel communication system. CSP can be used in practice where there is a parallel communication. Eg. to verify the parallelism in the electronic banking system or to describe security protocols. By the process algebra CSP is constantly evolving and improving, leading to increasing the usability of applications.

2 History

CSP was first described in 1978 by a British professor of CAR Hoare [1]. This person is known for his algorithm such as QuickSort or Hoare logic, which is a set of logical rules considering correctness of programming languages.3 Structure

3.1 Base elements

CSP processes communicate with each other or with an observer events or actions. They are visible, and describe the process behavior. The process is located at a particular time in any state. An example might be throwing coins into the vending machine for coffee and choc vending machine after release. Events are thus:

- **Coins**: Coin for vending machine
- **Choc**: choc form vending machine

The set of all (visible) events is called the alphabet and is marked with the Greek letter Σ (sigma large). It is associated with the process and tightly set. The process cannot perform events outside of alphabet. There can be also invisible internal process events. Those internal actions are marked with τ (tau small).

Processes are marked with capital letters, events are marked with small. The simplest process is the process STOP. It is behaves in a way that does not perform any action. Although it initially does not appear, it may be useful as an equivalent of such a process, which has no action and does not accept any communication - for example, it can happen when the system is stuck in the deadlock. How importance has critical states will be described below.

3.2 Communications

As already mentioned, the CSP processes communicate with each other. To select the communication between processes in CSP is a prefix operator. If we consider the event u a process P, syntax notation looks like this u → P means that the program performs at an event and then behaves exactly as described in process P. Our model example of a vending machine for choc it could be written as follows:

$$\text{coin} \rightarrow \text{STOP}$$

This example shows that, after inserting coins into the vending machine, nothing happens. The same events can be repeated, as the following example, the vending machine after inserting two coins will stop working.

$$\text{coin} \rightarrow \text{coin} \rightarrow \text{STOP}$$

Processes can also be characterized by a tree structure, consisting of circles connected by arrows. In the terminology of automata, the wheel represents the state of the process and the arrow represents the transaction between the states. The root wheel tree (usually the highest) symbolizes the start condition, which varies in the direction of the arrows down. Each arrow is describing an event that occurs during a transaction. The arrows leading from the same wheels must have a different label.

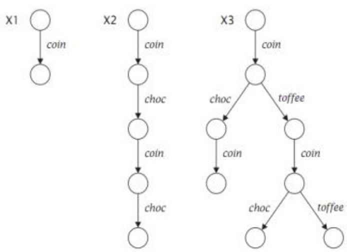


Fig. 1. Select choc

In these three examples, each branch of the tree stem end with STOP, it represents the wheel with no outgoing arrows.

3.3 Recursion and its problems

To describe repetitive behaviors we use recursion. Thus, we can create a loop of the commands. As an example I give a vending machine for candies.

$$\text{VENDING MACHINE} = \text{coin} \rightarrow \text{choc} \rightarrow \text{VENDING MACHINE}$$

The above example shows that the vending machine, after inserting a coin gives a choc and then waits for the insertion of additional coins. Recursion can be also written like this:

$$AUTOMAT = \mu\ X{:}\{coin, choc\} \bullet ((coin \rightarrow(choc \rightarrow X))$$

This is the identical record as the first record of that recursion. But this is a process description in abbreviated form.

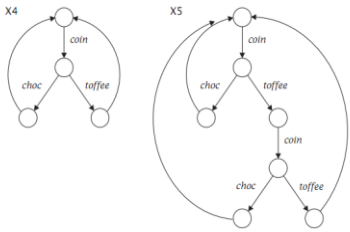


Fig. 2. Recursive selection of candies

To write down the recursion we use unnamed arrow that leads from the lowest process to the root process.
These two images symbolize the same sequence of processes. This is one of the weaknesses of images, because it is hard to prove congruence of transactions drawn.
Another problem about these images is, that they cannot demonstrate processes with very large or infinite number of states.

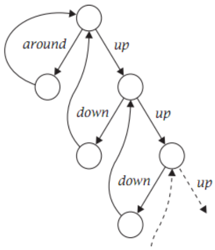


Fig. 3. Problem
In this example will never be enough room to draw the whole picture.

3.4 Parallel Processes

So far, attention has focused on modeling only sequential processes. When communicating parallel processes, we consider the situation that when communicating between Adele and Barbara and third party listens.
In parallel communication processes interact. Parallel communication operator, denoted P | | Q This notation means that the parallel communication can occur only if a ∈ Σ, which can process and execute the process P and process Q
This example is better understood with knowledge of sequential processes, where there is a withdrawal from the intersection of sets (alphabets) A ∩ B. Then the system behaves as P(a) ‖ Q(a). Parallel processes can also be synchronized in a process algebra based on defined events.

4 Overview operator and their importance

$\alpha \rightarrow P$ - prefix operator, performs the input or output event α and then goes into the process P
$(\alpha \rightarrow P) \square (\beta \rightarrow Q)$ - a deterministic choice of branching according to the course of events (it must be $\alpha \neq \beta$),
$(\alpha \rightarrow P) \circ (\beta \rightarrow Q)$ - nondeterministic choice of branching according to the course of events (it can be $\alpha = \beta$)
$P \,|||\, Q$ - operator of the leading, independent parallel composition,
$P \parallel \{\alpha\} \parallel Q$ - operator of a parallel interface, parallel composition with a shared event α,
$(\alpha \rightarrow P) \setminus \{\alpha\}$ - operator of hiding (hiding) α events.

5 Behavioral diagrams

Behavioral diagrams describing the system behavior based on the UML specification. These diagrams in UML describe the behavior of states and events that trigger the transition from one state to another. To do this we need to create a new process in the CSP for each of the states in the diagram. So we cannot use standard notation, but adapted. It is a combination of UML diagrams(such activity diagram) and diagrams used for the definition of algebra CSP.

Example:
VENDING MACHINE = coin $\rightarrow$ (choc $\rightarrow$ VENDING MACHINE | coffee VENDING MACHINE)
CUSTOMER = coin $\rightarrow$ coffee $\rightarrow$Skip
SYSTEM = VENDING MACHINE $_A\|_A$ VENDING MACHINE
A = {coin, choc, coffee}

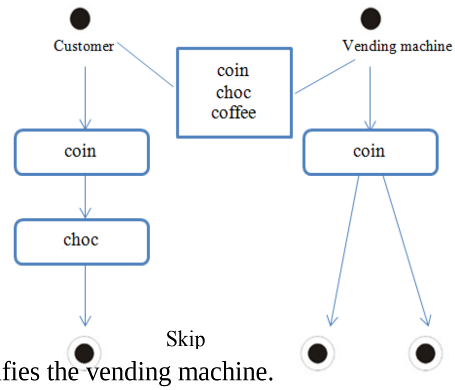


Fig. 3. Diagram specifies the vending machine.

5.1 CSP and BPMN

Interesting study carried out by Wong in Article: A Process Semantics for BPMN.[6] BPMN is bridge between design and implementation of business processes. This article describes how BPMN (like an activity diagram) can subset of CSP. This semantics can formally developers to analyze and compare BPMN diagrams.

6 Conclusions

Demands on computer systems are extremely increasing. System failure could lead to huge losses. Therefore new systems require reliability. The main roles of increasing system reliability are the usages of formal methods. They can extremely enhance security systems and errors can be detected in the earlier stages of design.
Unfortunately, formal methods are usually mathematically based methods. Most of developers don't know much about it. In practice mostly project managers skip formal specification and verification, because this study leads to an increase in time and human resources in the project development cycle.

References

1. Hoare C. A. R: Communicating Sequential Processes, C. A. R. Hoare, 2004
2. Milner, R.: Communication and Concurrency. Prentice-Hall, 1989.
3. Brim, L. Komunikace a paralelismus, 2001. Materials for class communication and parallelism, Faculty of informatics, Masaryk university in Brno
4. Bergstra, J. A., Klop, J. W.: Process Algebra for Synchronous Communication. Information and Control, Amsterdam, 1984.
5. Sheldon, Stochastic Analysis of CSP Specifications Using a CSP-to-Petri Net Translation Tool: CSPN [cit.2011-22-05]. Accessible from (http://www.csm.ornl.gov/~sheldon/bucket/IPDS96.pdf)
6. Y.H.Wong P, Gibbons, J: A Process Semantics for BPMN [cit.2011-22-05]. (http://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxwZXRlcnlod29uZ3xneDo1NDQ1MDc2NTRiZjJhYzcx)

IP Nearshoring: Questionnaire Analysis of Software Quality Improvement in Nearshoring Context

Oldřich Vašut

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
oldrich.vasut@vsb.cz

Abstract. This paper summarises results of the trial academic IP Nearshoring Project and it extends previous papers which introduce and present first outcomes. As long as project lasted for 3 years, challenge of comparing final results has arisen. Discussion about issues influencing nearshoring software services is the main concern of our investigation. Questions we wanted to be answered were extracted from participants' feedback data and questionnaires. Results are affected by real factors teams had to face. Project also involved issues such as a different technical background, cultural differences, various attitudes, communication and linguistic skills.

1 Introduction

Various studies have been written about outsourcing, offshoring and nearshoring as they play important role in developing products or providing services in today's world. Their use reduces production costs and sometimes also brings quality improvement. Let us shortly clarify these concepts. Outsourcing is one of the modern successful approaches for reducing costs. It refers to contracting-out of business processes to third-party domestic provider. Generally it involves transfer of one or more business functions previously performed in-house to external provider. The contracting-out of services to providers in other countries is called offshore outsourcing or offshoring, for short. Furthermore, nearshoring is one type of offshoring and refers to the outsourcing of business or IT processes to providers in nearby countries [1].

Further text discuss quality of deliverables produced together with communication and project management methods involvement when nearshore services. The main idea is to show the interdependences within several aspects observed in this limited experimental project.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 495–500.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Project Overview

IP Nearshoring was a project under the auspices of the Life Long Learning Program of EU as the Intensive Programme. Teams from six different European universities were participating in here. Countries involved were Netherlands, Poland, Romania, Czech Republic, Finland and Sweden. Each team was given one of two case studies as an assignment. Software Development Process was adjusted by dividing into two parts where two teams were involved in one process. Teams prepared specification and analysis of the information system based on entry cases. Then teams exchanged created specifications and every team had to implement product or prototype based on specification obtained from partner team. Final project stage consists of two weeks of presenting teams' outputs, evaluations and lectures. Consequential part of every presentational session included team's evaluation of each deliverable.

Each university was represented by one team of circa 10 students, supervised by 2-3 professors. During each year few IT projects were proposed to teams. Every team was given one case assignment. Firstly, team made specification on basis of case obtained from client. In the first phase team played role of company that uses outsourcing services, titled as contractor. When first phase of creation specification was finished, contractor passed documentation to partner team. Second phase begun and it embraced implementation of the final product following the specification received. Partner team was acting as an outsourcing service provider, titled as subcontractor [2].

3 Evaluation Techniques

The additional part of the project included presentations and discussions on work done. Groups debated mostly about specification or product issues and how product complied with specification and the suitability of employed project management methodologies. Last part of the day contained evaluation where teams provided feedback by given questionnaire for every specification and product parts. Teams answered 7 questions, but 4 of them were filtered out of the usage in our analysis. Filtering number of questions was necessary due to data incompleteness. Questions 1,6 and 7 has been chosen for processing. Further conclusions are mainly based on questionnaires provided.

Questions given in team feedback forms:

1. What do you think of the quality of the deliverables?
2. Is the method/technique used appropriate?
3. Is the method/technique usable in nearshoring context?
4. What are the advantages of the methods/techniques?
5. In which way can consistency be improved?
6. In which way can communication be improved?
7. What do you think about the project management?

All answers to questions had textual character. Thus it was necessary to categorize answers for easier statistical processing. We have set up 5 category levels demonstrating quality levels of certain issue.

Quality category levels are:

- 1 – poor;
- 2 – satisfactory;
- 3 – good;
- 4 – very good;
- 5 – excellent.

4 Outcomes

Each team had free hands in the terms of managing their work. Professors' supervision had minimal impact and it should not affect overall results. Teams could have used any project management method, specification tools and programming languages or other tools for development.

Leading professors were observing few key aspects throughout whole programme period [2]

- Ability each team define user requirements, system requirements and software requirements
- Method used for specification and analysis
- Ration of how much of functional and non-functional requirements were actually implemented
- If subcontractor team followed specification
- Level of communication between contractor and subcontractor teams

4.1 Quality of deliverables

First two charts (Fig. 1and Fig. 2) visualize dependency between specification quality and product quality over 3 years. Trend of specification quality and trend of product quality are recognizable. First three handovers FI → RO, RO → PL, NL → CZ follow the growing trend.

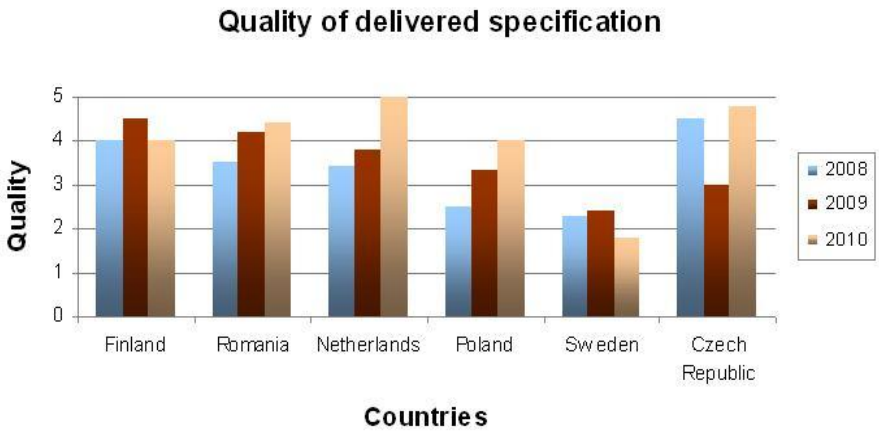


Fig. 1. Quality of delivered specification

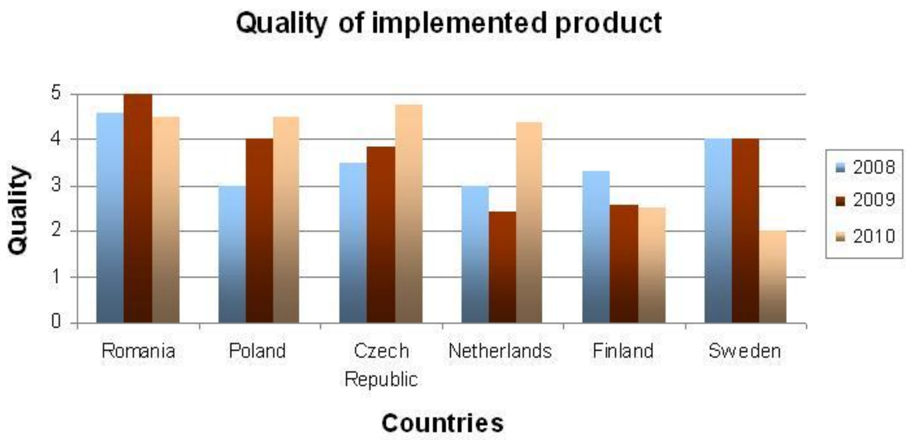


Fig. 2. Quality of implemented product

4.2 Level of communication

Communication quality level was evaluated as lower average most of the time. Quality of communication affects relations between specification and product quality. Chart on Fig. 3 proves hypothesis that deliverables with higher quality evaluation involved more frequent and most efficient communication. This outcome points to importance of communication between contractor and subcontractor.

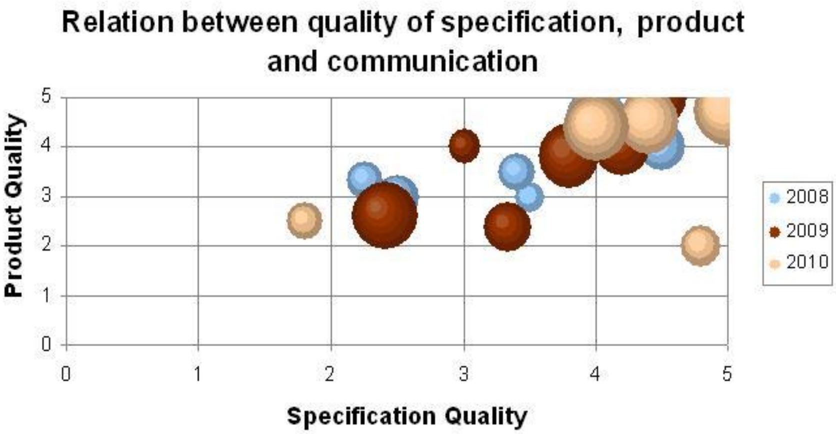


Fig. 3. Relation between quality of specification, product and communication

4.3 Project management methods

Several project management methods and methods for organization team work were selected by throughout whole programme period. Nine teams did not use any specific method labelled as common sense in further text. The rest of the teams took advantage of one of well known methods such as RUP (Unified Process), Prince2, DSDM, Scrum, XP or their combination. Teams reached significantly better results when one of the standardized methods was employed.

Evaluation of methods does not reflect interdependencies between questions given in questionnaire form. This lack in evaluation criteria is due to fact that questionnaires were not designed from statistical point of view [3]. However task of the feedback questionnaires was fulfilled as far as they gathered general feedback on work of each team and highlighted issues teams were facing.

5 Conclusions

IP Nearshoring as experimental academic project demonstrated working procedures in nearshoring context around Europe. This experiment pointed out issues and risks when part of IT businesses are outsourced. Nearshoring brings well-funded advantages compare to offshoring such as lower impact of cultural differences, working in the same or very close time zone, language tightness, higher level or standard educational system. Thus these benefits are balanced with higher software costs.

We can highlight several aspects observed in our project. The quality of produced documentation during specification phase has significant impact on the quality product built on its basis. Teams using some defined project management method or method for organization team work achieved better results than teams using ad hoc methods. Questionnaire outcomes also showed that teams were result driven when providing feedback. Communication as problematic related to project management and team work appeared to be very meaningful. Lack of communication between contractor and subcontractor was higher for most of the teams. Defining milestones during the process seems to be necessary in order to present work progress. That is why agile and iterative approaches come into scene in nearshoring. Technical differences could have influenced analysis as far as university teams took slightly different subjects. In the other hand cultural differences had not practical impact due to geographical closeness when nearshored.

We consider IP programmes as an interesting source of statistical experiments were several aspects can be observed. The design of questionnaires must be emphasized, in order not to affect the outputs of individual questions.

References

1. Meyer, T.: 'Nearshoring to Central and Eastern Europe DB Research', available: http://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD0000000000201757.pdf (2006) [accessed 2011-04-06]
2. Moldoveanu, A. D. B., Asavei, V., Moldoveanu, F., Morar, A. A., Egner, A. I., Boiangiu, C. A.: 'Turning nearshoring into a success managing technical background differences', available: http://www.freepatentsonline.com/article/Annals-DAAAM-Proceedings/246013801.html (2010) [accessed 2011-04-06]
3. Berdie, D. R., Anderson, J. F., Niebuhr, M. A.: Questionnaires: design and use, Metuchen, N.J.: Scarecrow Press (1986)

Fast Encoding and Decoding Algorithms for Variable-length Codes

Jiří Walder and Michal Krátký

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jiri.walder,michal.kratky}@vsb.cz

Abstract. Data compression has been widely applied in many data processing areas. Compression methods utilize variable-length codes with the shorter codes assigned to symbols or groups of symbols that appear in the data frequently. Fibonacci, Elias-delta, and Elias-Fibonacci codes are representatives of these codes, and are often utilized for the compression of numbers. Time consumption of encoding as well as decoding algorithms is important for some applications in the data processing area. In this case, efficiency of these algorithms is extremely important. Fast encoding and decoding algorithms Fibonacci, Elias-delta, and Elias-Fibonacci codes are presented in this paper. Our approach is up-to 7.9× more efficient than the conventional bit-oriented algorithms.
Keywords: compression, encoding, decoding, Elias-delta, Fibonacci, Elias-Fibonacci

1 Introduction

Data compression has been widely applied in many data processing areas. Various compression algorithms have been developed for processing text documents, images, video, etc. In particular, data compression is of the foremost importance and has been well researched as it is presented in excellent surveys [18, 24].

The time consumption of decompression is sometimes more critical than the time of compression; therefore, efficient decompression algorithms were studied in many works for the decompression of data structures [20, 11, 3] or text files [17, 9]. Pages are decompressed during every reading from the secondary storage into the main memory or items of a page are decompressed during every access to the page. In the case of a physical implementation of database systems, retrieval of compressed data structure's pages can be more efficient than retrieval of uncompressed pages due to the fact that the cost of decompression is lower than the cost of page accesses to the secondary storage [3, 1]. If insert or update operations are considered, data compression becomes more significant.

Fast encoding/decoding algorithms have been intensively studied in a connection with video and audio encoders/decoders. These studies often integrate the variable-length codes in encoders and decoders. The result is the following recommendations: T.81 for JPEG [21] and H.261 for MPEG encoder [13]. These encoders use the Exponential Golomb [22] or Huffman codes. Many works

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 501–506.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

and patents (e.g. [10, 14, 23]) deals with fast encoders/decoders for these codes. These algorithms use very short fixed translation tables for encoding/decoding of 8 bit-length numbers. These encoders are therefore not useful for general universal codes where the number domain can be large (e.g. 32 bit-length numbers). These large number domains are often used in the data processing area or in the text files compression; therefore, other algorithms using more general variable length codes are investigated.

The first effort to develop fast decoding algorithms for Fibonacci codes of order ≥ 2 has been proposed in [15, 16]. The fast encoding algorithms for Fibonacci, Elias-delta, and Elias-Fibonacci codes have not yet been studied.

In this article, we present an overview of the work related to fast encoding and decoding algorithms for Elias-delta, Elias-Fibonacci, and Fibonacci of order 2 and 3 codes. In Section 2, we provide a brief description of techniques applied in fast encoding and decoding algorithms. Some applications of variable-length codes are described in Section 3. In Section 4, experimental results are put forward; the proposed fast algorithms are compared to the conventional approaches.

2 Fast Encoding and Decoding Algorithms

In our work we developed fast encoding and decoding algorithms for Elias-delta, Fibonacci of order 2 and 3 and Elias-Fibonacci codes. We put forward the detailed overview of conventional encoding algorithms in our article [7] and for decoding algorithms in [5].

The main issue of the conventional encoding/decoding algorithm is handling encoded numbers in the bit-by-bit manner. To design a fast encoding/decoding algorithms, encoded numbers are separated into segments larger than one bit. These principles have been utilized in both our works for decoding [5] and fast encoding [7] algorithms. We refer reader to these articles for more details. In the following text we only summarize all important facts of these articles.

The fast encoding algorithms are based on precomputed tables. This tables allow encoding individual numbers directly into the code. The encoding requires the number segmentation which is carried out by right shifting. In case of Elias-delta and Elias-Fibonacci code we use binary shifting whereas in all Fibonacci codes we use newly introduced Fibonacci right shift operation. In our article we presented theoretical background on Fibonacci right shift operation and also the algorithm for effective Fibonacci right shift computation.

The fast decoding algorithms are based on a finite automaton with a precomputed mapping table for each state of the automaton. The precomputed table included in each automaton state allows converting segments of the input stream's bytes directly into decoded numbers for each state. The mapping table also defines the new automaton state for each segment. The algorithm for mapping table building is based on a conventional bit-oriented algorithm. During the building of the mapping table we must consider all states and possibilities when we read the current segment. Decoding also requires shifting, but in case of decoding we need left binary operation for Elias-Fibonacci and Elias-delta codes. For Fibonacci code we presented the theoretical background of new Fibonacci left shift operation and we also presented the effective Fibonacci left shift computation.

3 Applications of Fast Decoding and Encoding Algorithms

Various codes have been applied for the data compression [19]. In contrast with fixed-length codes, statistical methods utilize variable-length codes, with the shorter codes assigned to symbols or groups of symbols that have a higher probability of occurrence.

In our works [1–3], we utilized variable-length codes in two applications. In first we applied these codes for the compression of XML node streams with the following advantages: we can decrease the size of the data file and, therefore, decrease the number of disk accesses and we can store variable-length tuples. Tuples in a regular stream block are stored in an array with the fixed items' size.

In second case we utilized variable-length codes for the compression the R-tree [12, 8]. We applied the difference technique [18] to coordinates in minimum bounding boxes because they are close to each other. After that we assigned the smaller differences by shorter codes. The nodes are decompressed always when they are read from the disk. As result, we again decreased the size of the data file and, therefore, decreased the number of disk accesses and achieved the faster R-tree query processing. This compression is described in [3, 2].

4 Experimental Results

The proposed Fast encoding/decoding algorithms have been tested and compared with the conventional algorithms. The algorithms' performance has been tested for various collections. The tests were performed on a PC with dual core Intel 2.4GHz, 3 GB RAM using Windows 7 32-bit.

The test collections used in experiments had the same size: $10,000,000$ numbers. The proposed algorithms are universal and may be applied for arbitrary numbers > 0. However, we worked with numbers $\leq 4,294,967,295$, it means the maximal value is the value of the 32 bit-length binary number.

We tested uniformly, normally, exponentially random distributed 8, 16, 24 and 32-bit numbers. The results for different distributions are presented in Figure 1 and Tables 1 and 2.

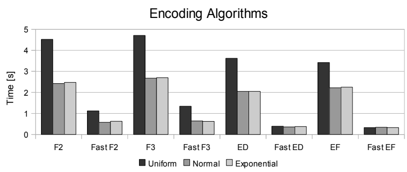
All fast algorithms achieved better encoding/decoding times than conventional algorithms. The speedup ratio for encoding ranges from $3.8\times$ for Fibonacci or order 3 code to $7.9\times$ for Elias-Fibonacci code and for decoding ranges from $5.6\times$ for Fibonacci or order 3 code to $6.4\times$ for Elias-Fibonacci code.

Table 1. Times of conventional and fast encoding algorithms for different distributions.

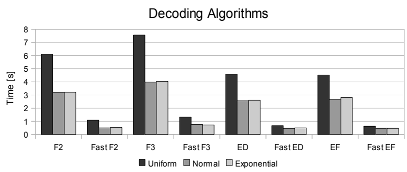
Conventional Algorithm [s]				
	Fibonacci 2	**Fibonacci 3**	**Elias-delta**	**Elias-Fibonacci**
Uniform	4.524	4.711	3.619	3.417
Normal	2.420	2.673	2.047	2.220
Exponential	2.471	2.699	2.045	2.248
Avg	3.14	3.36	2.57	2.63
Fast Algorithm [s]				
	Fibonacci 2	**Fibonacci 3**	**Elias-delta**	**Elias-Fibonacci**
Uniform	1.122	1.343	0.390	0.328
Normal	0.578	0.641	0.358	0.343
Exponential	0.624	0.621	0.374	0.328
Avg	0.77	0.87	0.37	0.33

Table 2. Times of conventional and fast decoding algorithms for different distributions.

Conventional Algorithm [s]				
	Fibonacci 2	**Fibonacci 3**	**Elias-delta**	**Elias-Fibonacci**
Uniform	6.099	7.565	4.586	4.524
Normal	3.183	3.979	2.559	2.652
Exponential	3.214	4.040	2.606	2.808
Avg	4.17	5.19	3.25	3.33
Fast Algorithm [s]				
	Fibonacci 2	**Fibonacci 3**	**Elias-delta**	**Elias-Fibonacci**
Uniform	1.092	1.327	0.671	0.624
Normal	0.499	0.764	0.468	0.468
Exponential	0.530	0.718	0.500	0.467
Avg	0.71	0.94	0.55	0.52



(a)



(b)

Fig. 1. Times of conventional and fast encoding algorithm (a) and times of conventional and fast decoding algorithms (b) for various number distributions

5 Conclusion

In this article, we presented fast encoding/decoding algorithms for Elias-delta, Fibonacci of orders 2 and 3, and Elias-Fibonacci codes. Fast encoding/decoding algorithms are based on the precomputed mapping tables and utilize segmentation instead of bit-by-bit conventional algorithms. The Fast Fibonacci encoding/decoding algorithm utilizes the newly proposed Fibonacci shift operation, whereas Elias-delta and Elias-Fibonacci utilize a common shift operation. In our experiments, all fast decoding algorithms are up-to 6.4× faster than the conventional algorithms and up-to 7.9× for encoding algorithms. We also describe few applications of these codes. In our future work, we want to develop fast algorithm for other variable-length codes. We also plan to utilize universal codes in more applications.

Projects and Publications

Jiří Walder is a member of team solving the grant of GAČR, Czech Republic, No. P202/10/0573, Handling XML Data in Heterogeneous and Dynamic Environments, the grant of SGS, Czech Republic, No. SP/2010138, Datové struktury podporující techniky indexování XML dat and SGS grant SP2011/172, Detekce plagiátů v textových dokumentech, řešitel: Jan Martinovič. He is co-author of five publication in conference proceedings [1, 3, 4, 6, 2], one accepted journal article [5] and one submitted journal article [7]. Two publications are indexed in Web of Science.

Own References

1. R. Bača, J. Walder, M. Pawlas, and M. Krátký. Benchmarking the Compression of XML Node Streams. In *In Database Systems for Advanced Applications - 15th International Conference, DASFAA 2010, International Workshops: GDM, BenchmarX, MCIS, SNSMW, DIEW, UDM, Tsukuba, Japan.* Springer-Verlag, LNCS 6193/2010.
2. P. Chovanec, M. Krátký, and J. Walder. Lossless R-tree Compression using Variable-Length Codes. In *In Proceedings of The 5th International Conference for Internet Technology and Secured Transactions (ICITST).* London, UK, IEEE, 2010.
3. J. Walder, M. Krátký, and R. Bača. Benchmarking Coding Algorithms for the R-tree Compression. In *Proceedings of the Dateso 2009 Annual International Workshop on Databases, Texts, Specifications and Objects*, pages 32–43. CEUR Workshop Proceedings, Volume: 471, 2009.
4. J. Walder, M. Krátký, and R. Bača. Using Self Organizing Map for Power Outage Data Analysis. In *Proceedings of ELNET, Ostrava*, pages 66–80, 2009.
5. J. Walder, M. Krátký, R. Bača, J. Platoš, and V. Snášel. Fast Decoding Algorithms for Variable-Lengths Codes. *Information Science*, Submitted: February 16 2010, Revised: December 13 2010, Accepted: June 15 2011.
6. J. Walder, M. Krátký, and J. Platoš. Fast Fibonacci Encoding Algorithm. In *Proceedings of the Dateso 2010 Annual International Workshop on Databases, Texts, Specifications and Objects*, pages 72–83. MATFYZPRESS publishing house of Faculty of Matematics and Physics Charles University,CEUR Workshop Proceedings, Volume: 310, 2010.
7. J. Walder, M. Krátký, and J. Platoš. Fast Encoding Algorithms for Some Variable-length Codes. *Submitted in IEEE Transactions on Information Theory*, Submitted: April 14 2011.

Other References

8. N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM SIGMOD*, pages 322–331. ACM Press, 1990.
9. T. C. Bell and I. H. Witten. *Text Compression.* Prentice Hall, 1990.
10. H. Fujiwara, T. Sakaguchi, A. Shimatzu, M. ting Sun, K. hu Tzou, and K. min Yang. Variable-length codeword encoder, 1995.
11. J. Goldstein, R. Ramakrishnan, and U. Shaft. Compressing Relations and Indexes. In *Proceedings of the 14th International Conference on Data Engineering, ICDE 1998*, page 370, Los Alamitos, CA, USA, 1998. IEEE Computer Society.
12. A. Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, Annual Meeting, Boston, USA*, pages 47–57. ACM Press, June 1984.
13. R. H.261. Video codec for audiovisual services at p x 64 kbit/s itu-t, 1993.
14. J. Jeon, Y. Park, and H. Park. A Fast Variable-Length Decoder Using Plane Separation. *CirSysVideo*, 10(5):806–812, August 2000.
15. S. T. Klein. Fast Decoding of Fibonacci Encoded Texts. In *Proceedings of the International Data Compression Conference, DCC'07*, page 388, Washington, DC, USA, 2007. IEEE Computer Society.
16. S. T. Klein and M. K. Ben-Nissan. Using Fibonacci Compression Codes as Alternatives to Dense Codes. In *Proceedings of the International Data Compression Conference, DCC'08*, pages 472–481, Washington, DC, USA, 2008. IEEE Computer Society.
17. H. Plantinga. An Asymmetric, Semi-adaptive Text Compression Algorithm. In *Proceedings of the International Data Compression Conference, DCC 1994.* IEEE Computer Society, 1994.
18. D. Salomon. *Data Compression The Complete Reference.* Third Edition, Springer–Verlag, New York, 2004.
19. D. Salomon. *Variable-length Codes for Data Compression.* Springer-Verlag, 2007.
20. H. Samet. Data Structures for Quadtree Approximation and Compression. *Communications of the ACM archive*, 28(9):973–993, September 1985.
21. R. T.81. Digital compression and coding of continuous-tome still images., 1992.
22. J. Teuhola. A compression method for clustered bit-vectors. *Inf. Process. Lett.*, 7(6):308–311, 1978.
23. M. ting Sun, K. hu Tzou, A. Shimatzu, M. ting Sun, K. hu Tzou, and K. min Yang. High-speed flexible variable-length-code decoder, 1992.
24. I. H. Witten, A. Moffat, and T. C. Bell. *Managing Gigabytes, Compressing and Indexing Documents and Images, 2nd edition.* Morgan Kaufmann, 1999.

Optimization of Strategies Set in Robot Soccer Game

Jie Wu, Václav Snášel, Eliška Ochodková, Jan Martinovič, and Václav Svatoň

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
defermat2008@hotmail.com, {vaclav.snasel, eliska.ochodkova, jan.martinovic, vaclav.svaton.st}@vsb.cz

Abstract. The robot soccer game is full of challenging in the field of robot and artificial intelligence. Strategy is a kernel subsystem of robot soccer game. In our work, we propose a ranking model of strategies, based on which we can detect hub strategy and solve its problem. Additionally, we also present a method to generate offensive strategies. The experiment that eliminates the impact of external factors at mostly shows the effectiveness of strategic description and validity of our method.

1 Introduction

Robot soccer is a challenging platform for robotics and artificial intelligence. Through the game of robot soccer, soccer robotics studies how multiple soccer-playing robots on each team could be built to cooperate in an adversarial environment to achieve specific objectives. In order to cooperate well, soccer robots must execute a series of actions according to a certain strategy, which is hidden behind the series of actions, so that the team robots could achieve their common targets. Hence, the strategies play an important role in robot soccer game.

Description of Strategies is the basis of the robot soccer game. Many different forms of strategic description have been developed to support corresponding decision-making system, such as multi-agent strategic modeling [1, 2], strategy-based decision tree [3], and case-based reasoning [4], etc. All of these approaches contain two common elements, i.e. grid positions and possession of ball. These common elements are also reflected in our description of strategies [5].

In [5], we present that strategy can be expressed easily as (*M, O, B, D*), where M is the teammates' positions of *mine*, O, *opponents'* positions, B, *ball* position, and D, my teammates' *destination* grids. Then the case in Fig. 1 can be described as "**If** (M_1, M_2, M_3, M_4) is close to $(A1, A2, B1, B2)$, **and if** (O_1, O_2, O_3, O_4) is close to $(B1, B2, C1, C2)$, **and if** B is close to $(B2)$, **then** (M_1, M_2, M_3, M_4) go to $(A1, B2, C1, C2)$". Now if we represent the grid position by using digital coordinates, the case in Fig. 1 can be denoted as follows.

- ○ *Mine* 11 12 21 22 ;
- ○ *Oppo* 21 22 31 32 ;
- ○ *Ball* 22 ;
- ○ *Dstn* 11 22 31 32 .

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 507–512.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

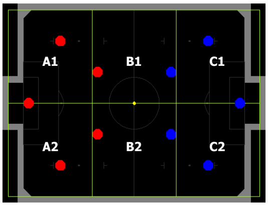


Fig. 1. An example case.

The strategies set contains many strategies. The ability of robot team is represented by the whole strategies set. Therefore, how to optimize strategies set is an important issue in robot soccer game. In strategies set, there exist hub strategy. Many other strategies have to be achieved through the hub strategy. If we can destroy opponent's hub strategy, then we eliminate most of the follow-up rules of opponent, and vice versa. Additionally, the offensive strategy could be generated by game simulation.

This paper is organized as follows. Section 2 introduces the ranking model by which we could detect hub strategy. Based on the model, in section 3 we propose a method to solve the problem of hub strategy. In section 4 we present a method to generate offensive strategies and prove its validity by simulation game of robot soccer. Finally, section 5 draws the conclusions.

2 Ranking Model of Strategy

K. Bryan and T. Leise [6] rank the importance of web pages according to an eigenvector of a weighted link matrix. This model is helpful to rank strategies. But strategy ranking is different to page ranking.

In [6], a core idea in assigning an importance score to any given web page is that the page's score is derived from the links made to that page from other web pages. The links *to* a given page are called the *backlinks* for that page. The web thus becomes a democracy where pages vote for the importance of other pages by linking to them. While in strategy ranking, the strategy that has more links to other strategies is more important because that strategy has more influence on the process of game, the game situation is mostly depend on the follow-up execution of that strategy. In addition, this kind of strategy is a hub because many other strategies have to be achieved through it. If we can destroy opponent's hub strategy, then we eliminate most of the follow-up rules of opponent, and vice versa.

Fig. 2 shows an example of a strategies set with only four strategies. An arrow from strategy A to strategy B indicates a link from strategy A to strategy B. For strategy 1, we have $x_1 = x_2/1 + x_3/3 + x_4/2$, since strategy 2, strategy 3 and strategy 4 are *forwardlinks* for strategy 1 and strategy 2 contains only one link, strategy 3 assembles three links, while strategy 4 receives two links. Similarly, $x_2 = x_3/3 + x_4/2$, $x_3 = x_1/2$, $x_4 = x_1/2 + x_3/3$. These linear equations can be

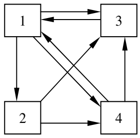


Fig. 2. An example of a strategies set with only four strategies.

written as

$$\begin{bmatrix} 0 & 1 & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \tag{1}$$

In this case we obtain $x_1 = 0.3750$, $x_2 = 0.1875$, $x_3 = 0.1875$, $x_4 = 0.2500$. Thus strategy 1 gets the highest importance score.

It should be noted that the node with highest importance score may be foible of the system. Opponent may hinder the implementation of our strategy by attacking this node, and vice versa.

3 Hub Strategy

Robot soccer simulator is a good test bed for strategies set. In the simulator, two teams such as red and blue team will share the same strategies selection algorithm, prediction of movement, tactics, and so on, which would eliminate the impact of the simulator, and distinctly display differences between the two strategies sets. According to the game simulation result, we can find the better strategies set, because the better one could achieve better competition result.

In our method, firstly we detect the hub strategy based on the ranking model, then we replace the hub strategy by other new strategies in order to decrease the influence of hub strategy to follow-up strategies. In our experiment, there are three strategies set, named RedI, RedII and Blue, where RedI contains 19 strategies, RedII and Blue 18 ones. The last 7 strategies of RedI are displayed in Table 1, and last 6 strategies of RedII and Blue in Table 2 and Table 3 respectively. These strategies are the difference of the three strategies set.

Fig. 3(a) corresponds to the structure of the last 7 strategies in RedI. Fig. 3(b) shows the structure of the last 6 strategies in RedII. Apparently, in RedI strategy 19 is a hub strategy, while in RedII they are *parallel strategies*. Due to space limitation, we don't list the common 12 strategies of all three strategies set.

Table 4 lists ten games' results of RedI vs Blue, where RedI team got 3 wins, 3 defeats and 4 ties. Table 5 lists ten games' results of RedII vs Blue, where RedII team got 4 wins, 2 defeats and 4 ties. Obviously, the strategies set of RedII team is better than RedI. The reason is that strategies set Blue hinders the implementation of hub strategy 19 in RedI, while in RedII there isn't hub

Table 1. Last 7 Strategies of RedI.

Rule	*Mine*	*Oppo*	*Ball*	*Dstn*
13	31344243	42435154	42	41435253
14	31344243	42435154	42	42435253
15	31344243	42435154	42	32435253
16	41435253	52536263	52	52536263
17	42435253	52536263	52	52536263
18	32435253	52536263	52	42536263
19	21243233	42435154	33	31344243

Table 2. Last 6 Strategies of RedII.

Rule	*Mine*	*Oppo*	*Ball*	*Dstn*
13	31344243	42435154	42	41435253
14	32334243	42435154	42	42435253
15	22324243	42435154	42	32435253
16	41435253	52536263	52	52536263
17	42435253	52536263	52	52536263
18	32435253	52536263	52	42536263

Table 3. Last 6 Strategies of Blue.

Rule	*Mine*	*Oppo*	*Ball*	*Dstn*
13	32332223	41443233	32	42432233
14	42432233	52534243	42	52533243
15	52533243	52534243	52	62634253
16	31323323	41443233	32	41424322
17	41424322	52534243	42	51525332
18	51525332	62635253	52	62635242

Table 4. RedI vs Blue.

Team	*Score*									
RedI	1	0	1	2	0	0	1	2	3	1
Blue	1	0	1	0	1	0	0	0	4	3

Table 5. RedII vs Blue.

Team	*Score*									
RedII	1	0	1	2	1	1	1	0	2	0
Blue	0	1	0	3	1	1	0	0	1	0

strategy and all strategic implementation are resultful. Therefore, the problem of hub strategy could be resolved by replacing hub strategy by parallel strategies.

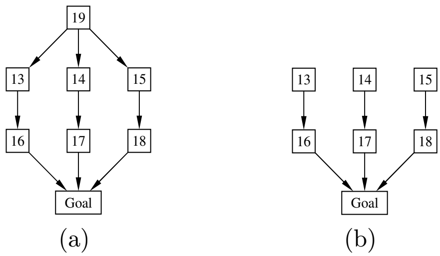


Fig. 3. (a) Graph of the last 7 strategies in RedI; (b) Graph of the last 6 strategies in RedII.

4 Generation of Offensive Strategies

According to the game simulation result, it's easy to know which strategies set is better, because the better strategies set could achieve better competition result. In the game log file, we can find the process of goal, which is useful to improve the generation of strategies set. For example, we can find the last action before goal score in the log file of game, and then transform it to be an offensive strategy.

In our experiment there are two teams, i.e. red and blue team, whose strategies set is named Redset and BluesetI, respectively. Due to space limitation, we don't list the content of strategies set. In the strategies set of red team, there are 10 strategies, while 38 strategies in blue team. Table 6 lists ten games' results of Redset vs BluesetI, where blue team got three wins, five defeats and two ties, with 5 gains and 7 loss balls.

In the first game, the last action before red team made score can be represented as follow,

◦ *Mine* 42 53 62 63 ;
◦ *Oppo* 13 42 42 53 ;
◦ *Ball* 63 ;
◦ *Dstn* 42 53 63 63 .

which can be added as a new strategy to the strategies set, and then blue team get a new strategies set BluesetII. Table 7 lists ten games' results of Redset vs BluesetII, where blue team got eight wins and two ties, with 11 gains and only 2 loss balls. Apparently, the strategies set of blue team is improved greatly by the last strategy that greatly enhances blue team's offensive.

Table 6. Redset vs BluesetI.

Team	*Score*									
Red	1	0	1	0	0	2	0	2	1	0
Blue	0	0	0	1	1	1	2	0	0	0

Table 7. Redset vs BluesetII.

Team	*Score*									
Red	0	1	0	0	0	0	0	1	0	0
Blue	1	2	1	1	0	2	1	1	1	1

5 Conclusions

In this paper, based on the ranking model, we propose a method to solve the problem of hub strategy in robot soccer game. Firstly we detect the hub strategy by ranking model, then replace the hub strategy by parallel strategies. Now we get a new strategies set RedII. If the game result between RedII and Blue is better than that of RedI and Blue on the statistical level, we can say the new strategies set RedII is better than RedI, and also we can say the method to improve hub strategy is effective. Additionally, we present a method to generate offensive strategies. We extract one action from the game log file which results in a goal score, then add it as a new offensive strategy to the team's strategies set whose score falls behind. Now we get a new strategies set. If the new strategies set could vanquish the former preponderant set on the statistical level, we can say the added new offensive strategy is effective. We test the method in simulation game, because in the simulator everything is same for two teams except strategies, which can eliminate the impact of external factors. The differences of the game results simply reflect the performance of strategies set.

References

1. Bezek, A.: Discovering strategic multi-agent behavior in a robotic soccer domain. In: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. ACM (2005) 1177–1178
2. Bezek, A., Gams, M., and Bratko, I.: Multi-agent strategic modeling in a robotic soccer domain. In: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems. ACM (2006) 457–464
3. Huang, H.P. and Liang, C.C.: Strategy-based decision making of a soccer robot system using a real-time self-organizing fuzzy decision tree. Fuzzy Sets and Systems, Vol. 127. Elsevier (2002)
4. Ros, R., Arcos, J.L., Lopez de Mantaras, R., and Veloso, M.: A case-based approach for coordinated action selection in robot soccer. Artificial Intelligence, Vol. 173. Elsevier (2009) 1014–1039
5. Martinovič, J., Snášel, V., Ochodková, E., Zołtá, L, Wu, J., and Abraham, A.: Robot Soccer - Strategy Description And Game Analysis. In: Modelling and Simulation, 24th European Conference ECMS 2010. Kuala Lumpur, Malaysia (2010)
6. Bryan, K. and Leise, T.: The $25,000,000,000 Eigenvector: The Linear Algebra behind Google. SIAM Review, Vol. 48 (2006) 569–581

Efficient TFETI Based Solver for Elasto-plastic Problems of Mechanics

Martin Čermák

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
martin.cermak@vsb.cz

Abstract. This paper illustrates how to implement effectively solvers for elasto-plastic problems. We consider the time step problems formulated by nonlinear variational equations in terms of displacements. To treat nonlinearity and nonsmoothnes we use semismooth Newton method. In each Newton iteration we have to solve linear system of algebraic equations and for its numerical solution we use TFETI algorithm. In our benchmark we demonstrate our approach on von Mises plasticity with isotropic hardening and use return mapping concept.

Key words: Elasto-plasticity, Domain decomposition, TFETI

1 Introduction

The paper is organized as follows. We briefly review the TFETI methodology that transforms the large primal problem of elastostatics in terms of displacements into the smaller and better conditioned dual one in terms of the Lagrange multipliers (pressures) whose conditioning is further improved by using the projectors defined by the natural coarse grid. Further we briefly review the elasto-plasticity methodology for von Mises plasticity with isotropic hardening. We illustrate the efficiency of our algorithm on the solution of 3D elasto-plastic model benchmark and give encouraging results of numerical experiments.

2 TFETI domain decomposition

To apply the TFETI domain decomposition, we tear the body from the part of the boundary with the Dirichlet boundary condition, decompose the body into subdomains, assign each subdomain a unique number, and introduce new "gluing" conditions on the artificial intersubdomain boundaries and on the boundaries with imposed Dirichlet data.

More specifically, the original body Ω is decomposed into a system of s homogeneous isotropic elastic bodies, each of which occupies, in a reference configuration, a subdomain Ω^p in $\mathbb{R}^d, d = 2, 3,\ p = 1, \ldots, s$. After decomposition each boundary Γ^p of Ω^p consists of three disjoint parts Γ^p_U, Γ^p_F, and Γ^p_G,

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 513–518.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

$\Gamma^p = \overline{\Gamma}_U^p \cup \overline{\Gamma}_F^p \cup \overline{\Gamma}_G^p$, with the corresponding displacements $\mathbf{U}^p$ and forces $\mathbf{F}^p$ inherited from the originally imposed boundary conditions on Γ. For the artificial intersubdomain boundaries, we use the following notation: Γ_G^{pq} denotes the part of Γ^p that is glued to Ω^q and Γ_G^p denotes the part of Γ^p that is glued to the other subdomains. Obviously $\Gamma_G^{pq} = \Gamma_G^{qp}$. An auxiliary decomposition of the problem with renumbered subdomains and artificial intersubdomain boundaries is in Fig. 1.

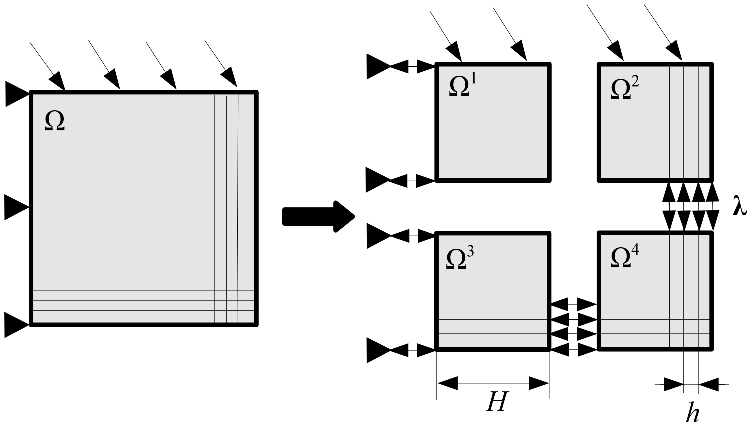


Fig. 1. TFETI domain decomposition with subdomain renumbering and traces of discretization

The gluing conditions require continuity of the displacements and of their normal derivatives across the intersubdomain boundaries. The mechanical properties of Ω^p are defined by the Young modulus E^p, the Poisson ratio ν^p, and the density ρ^p.

Let $c_{ijk\ell}^p$ and $\mathbf{g}^p$ denote again the entries of the elasticity tensor and a vector of body forces, respectively. For any sufficiently smooth displacement $\mathbf{u} : \overline{\Omega}^1 \times \ldots \times \overline{\Omega}^s \to \mathbb{R}^d$, the total potential energy is defined by

$$\mathcal{J}(\mathbf{u}) = \sum_{p=1}^{s} \left\{ \frac{1}{2} a^p(\mathbf{u}^p, \mathbf{u}^p) - \int_{\Omega^p} (\mathbf{g}^p)^\top \mathbf{u}^p d\Omega - \int_{\Gamma_F^p} (\mathbf{F}^p)^\top \mathbf{u}^p d\Gamma \right\}, \tag{1}$$

where

$$a^p(\mathbf{u}^p, \mathbf{v}^p) = \int_{\Omega^p} c_{ijk\ell}^p e_{ij}^p(\mathbf{u}^p) e_{k\ell}^p(\mathbf{v}^p) d\Omega \quad \text{and} \quad e_{k\ell}^p(\mathbf{u}^p) = \frac{1}{2}\left(\frac{\partial u_k^p}{\partial x_\ell^p} + \frac{\partial u_\ell^p}{\partial x_k^p}\right). \tag{2}$$

Let us introduce the product Sobolev space

$$\mathcal{V} = H^1(\Omega^1)^d \times \ldots \times H^1(\Omega^s)^d, \tag{3}$$

and let

$$\mathcal{K} = \{\mathbf{v} = (\mathbf{v}^1, \dots, \mathbf{v}^s) \in \mathcal{V} : \ \mathbf{v}^p = \mathbf{U}^p \ \text{ on } \ \Gamma_U^p, \ \ \mathbf{v}^p = \mathbf{v}^q \ \text{ on } \ \Gamma_G^{pq}\}$$

be its non-empty, closed, and convex subset. The displacement $\mathbf{u} \in \mathcal{K}$ of the system of subdomains in equilibrium satisfies

$$\mathcal{J}(\mathbf{u}) \leq \mathcal{J}(\mathbf{v}) \ \text{ for any } \ \mathbf{v} \in \mathcal{K}. \tag{4}$$

The finite element discretization of $\overline{\Omega} = \overline{\Omega}^1 \cup \ldots \cup \overline{\Omega}^s$ with a suitable numbering of nodes results in the quadratic programming (QP) problem

$$\min_{\mathbf{u}} \frac{1}{2}\mathbf{u}^\top \mathbf{K}\mathbf{u} - \mathbf{f}^\top \mathbf{u} \quad \text{subject to } \ \mathbf{B}\mathbf{u} = \mathbf{c}, \tag{5}$$

where $\mathbf{K} = \mathrm{diag}(\mathbf{K}_1, \dots, \mathbf{K}_s)$ denotes a symmetric positive semidefinite block-diagonal stiffness matrix of order n, $\mathbf{B}$ denotes an $m \times n$ full rank constraint matrix, $\mathbf{f} \in \mathbb{R}^n$ is a load vector, and $\mathbf{c} \in \mathbb{R}^m$ is a constraint vector.

The diagonal blocks $\mathbf{K}_p$ that correspond to the subdomains Ω^p are positive semidefinite sparse matrices with known kernels, the rigid body modes. This is a great advantage because all blocks can be effectively regularized and then decomposed using any standard sparse Cholesky type factorization method for nonsingular matrices [2, 3].

The matrix $\mathbf{B}$ with the rows $\mathbf{b}_i$ and the vector $\mathbf{c}$ with the entries c_i enforce the prescribed displacements on the part of the boundary with imposed Dirichlet condition and the continuity of the displacements across the auxiliary interfaces.

A parallel and numerically scalable algorithm for the numerical solution of (5) is introduced in [3] with scalability demonstrated up to 315 millions of unknowns and 4800 cores.

3 Elasto-plasticity

Elasto-plastic problems are the so-called quasi-static problems where the history of loading is taken into account. We consider the von Mises elasto-plasticity with the strain isotropic hardening and incremental finite element method with the return mapping concept [1].

The elasto-plastic deformation of an body Ω after loading is descibed by the Cauchy stress tensor $\boldsymbol{\sigma}$, the small strain tensor $\boldsymbol{\varepsilon}$, the displacement $\mathbf{u}$, and the nonnegative hardening parameter $\boldsymbol{\kappa}$. Symmetric tensor is represented by the vector and its deviatoric part is denoted by the symbol *dev*.

Let us denote the space of continuous and piecewise linear functions constructed over a regular triangulation of $\overline{\Omega}$ with the discretization norm h by $V_h \subset V$, where $V = \left\{\mathbf{v} \in H^1(\Omega)^d : \ \ \mathbf{v} = 0 \text{ on } \Gamma_U\right\}$. Let

$$0 = t_0 < t_1 < \ldots t_k < \ldots < t_N = t^* \tag{6}$$

be a partition of the time interval $[0, t^*]$. Then the solution algorithm after time and space discretizations has the form:

Algorithm 3.

1. Initial step: $\mathbf{u}_h^0 = 0,\ \boldsymbol{\sigma}_h^0 = 0,\ \boldsymbol{\kappa}_h^0 = 0,$
2. **for** $k = 0,\ \ldots, N-1$ **do** (load step)
3. From previous step we know: $\mathbf{u}_h^k,\ \boldsymbol{\sigma}_h^k,\ \boldsymbol{\kappa}_h^k$ and compute $\triangle\mathbf{u}_h,\ \triangle\boldsymbol{\sigma}_h,\ \triangle\boldsymbol{\kappa}_h$

$$\triangle\boldsymbol{\varepsilon}_h = \varepsilon(\triangle\mathbf{u}_h), \quad \triangle\mathbf{u}_h \in V_h, \tag{7}$$

$$\triangle\boldsymbol{\sigma}_h = T_\sigma(\boldsymbol{\sigma}_h^k,\ \boldsymbol{\kappa}_h^k,\ \triangle\boldsymbol{\varepsilon}_h), \tag{8}$$

$$\triangle\boldsymbol{\kappa}_h = T_\kappa(\boldsymbol{\sigma}_h^k,\ \boldsymbol{\kappa}_h^k,\ \triangle\boldsymbol{\varepsilon}_h). \tag{9}$$

4. Solution $\triangle\boldsymbol{\sigma}_h(\boldsymbol{\sigma}_h^k,\ \boldsymbol{\kappa}_h^k, \varepsilon(\triangle\mathbf{u}_h))$ is substituted into equation of equilibrium:

$$\int_\Omega \triangle\boldsymbol{\sigma}_h^T(\boldsymbol{\sigma}_h^k,\ \boldsymbol{\kappa}_h^k, \varepsilon(\triangle\mathbf{u}_h))\varepsilon(\mathbf{v}_h)dx = \langle\triangle\mathbf{f}_h^k,\ \mathbf{v}_h\rangle, \quad \forall\mathbf{v}_h \in V_h. \tag{10}$$

   This leads to a nonlinear system of equations with unknown $\triangle\mathbf{u}_h$ which is solved using the Newton method. The linearized problem arising in each Newton step is solved by the TFETI algorithmic scheme [3, 7].
5. Then we compute values for the next step: $\mathbf{u}_h^{k+1} = \mathbf{u}_h^k + \triangle\mathbf{u}_h,\ \boldsymbol{\sigma}_h^{k+1} = \boldsymbol{\sigma}_h^k + \triangle\boldsymbol{\sigma}_h,\ \boldsymbol{\kappa}_h^{k+1} = \boldsymbol{\kappa}_h^k + \triangle\boldsymbol{\kappa}_h.$
6. **enddo**

Basic relations for the return mapping concept can be found in [7].

4 Numerical experiments

Described algorithms were implemented in `MatSol` library [4] developed in Matlab environment and tested on the solution of 3D problems.

Let us consider a 3D plate with a hole in the center (due to symmetry only a quatre of the whole structure is used). The geometry of the body with traces of decomposition and discretization is depicted in Fig. 2. In Fig. 3 we see a zoom of Fig. 2 near the hole. Symmetry conditions are prescribed on the left and lower sides of Ω. The surface load $g(t) = 450\sin(2\pi t)$ [MPa], $t \in [0, \frac{1}{4}]$ [sec], is applied to the upper side of Ω. The elasto-plastic material parameters are $E = 206900$ [MPa], $\nu = 0.29$, $Y = 450$, $H_m = 100$ and the time interval $[0, \frac{1}{4}]$ [sec] is divided into 50 steps. We consider a mesh with 4450 nodes and 19008 tetrahedrons.

In the nth Newton iteration we compute an approximation $\triangle\mathbf{u}^n$ by solving the constrained linear problem of the form

$$\min_{\mathbf{B}\triangle\mathbf{u}^n=\mathbf{o}} \frac{1}{2}\left(\triangle\mathbf{u}^n\right)^\top \mathbf{K}^n\triangle\mathbf{u}^n - \left(\triangle\mathbf{u}^n\right)^\top \triangle\mathbf{f}^n$$

using the scalable TFETI algorithmic scheme proposed in [3]. We stop the Newton method in every time step if $\|\triangle\mathbf{u}^{n+1} - \triangle\mathbf{u}^n\|/\left(\|\triangle\mathbf{u}^{n+1}\| + \|\triangle\mathbf{u}^n\|\right)$ is less than 10^{-9}.

Notice that the maximum number of the Newton iterations is small for all time steps, therefore the method is suitable for the problem. In the following figures, we depict plastic and elastic elements and von Mises stress in the xy plane with the z coordinate 0 [mm]. In Figs. 4 and 5, we can see which elements are plastic (gray color) and which are elastic (white color) in chosen time steps. Particularly, in time steps 1-12 we observe only elastic behavior, and in time steps 13-50 plastic behavior of some elements. The maximum value of hardening at each time step is depicted in Fig. 6. The von Mises stress distribution on deformed mesh is showed in Fig. 7.

5 Conclusions and goals

We have presented an efficient algorithm for the numerical solution of elasto-plastic problems. These problems lead to the quasi-static problems, where each nonlinear and nonsmooth time step problem is solved by the semismooth Newton method. In each Newton iteration we have to solve an auxiliary (possibly of large size) linear system of algebraic equations. We proposed a new approach how to solve such system efficiently using in a sense optimal algorithm based on our Total-FETI variant of FETI domain decomposition method. We illustrated the efficiency of our algorithm on the solution of 3D elasto-plastic model benchmark and gave results of numerical experiments. The results indicate that the algorithm may be efficient.

Nowadays we adapt this approach to the solution of contact problems.

References

1. R. Blaheta, *Numerical methods in elasto-plasticity.* Documenta Geonica 1998, PERES Publishers, Prague, 1999.
2. T. Brzobohatý, Z. Dostál, P. Kovář, T. Kozubek, A. Markopoulos, *Cholesky decomposition with fixing nodes to stable evaluation of a generalized inverse of the stiffness matrix of a floating structure*, IJNME, DOI: 10.1002/nme.3187
3. T. Kozubek, V. Vondrak, M. Mensik, D. Horak, Z. Dostal, V. Hapla, P. Kabelikova, M. Cermak, *Total FETI domain decomposition method and its massively parallel implementation*, Computers and Structures, 2011, submitted.
4. T. Kozubek, A. Markopoulos, T. Brzobohatý, R. Kučera, V. Vondrák, Z. Dostál, *MatSol - MATLAB efficient solvers for problems in engineering.* http://matsol.vsb.cz/.
5. M. Čermák, *Newton method for elasto-plasticity problems*, Proceedings of ČVUT 2010
6. M. Čermák, T. Kozubek, A. Markopoulos, *An efficient solution of elasto-plastic problems in mechanics*, SNA 2011, ISBN 978-80-86407-19-7
7. M. Čermák, T. Kozubek, A. Markopoulos, *An efficient parallel solver for elasto-plastic problems of mechanics*, PARENG 2011, ISBN 978-1-905088-43-0

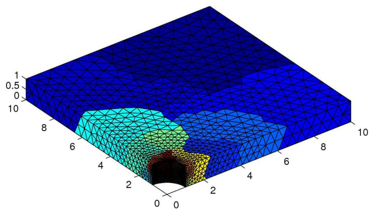


Fig. 2. Geometry in [mm] with traces of decomposition and discretization

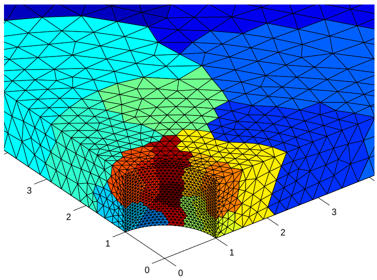


Fig. 3. Zoom of Fig. 2 near the hole

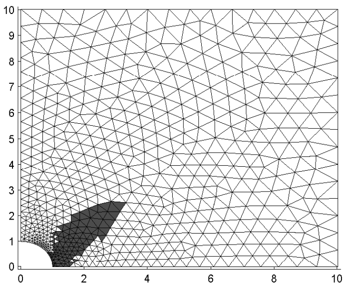


Fig. 4. Plastic and elastic elements after 35 time steps

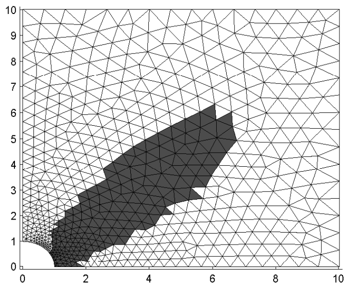


Fig. 5. Plastic and elastic elements after 50 time steps

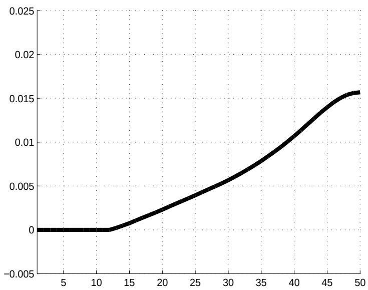


Fig. 6. Maximum values of hardening in time iterations

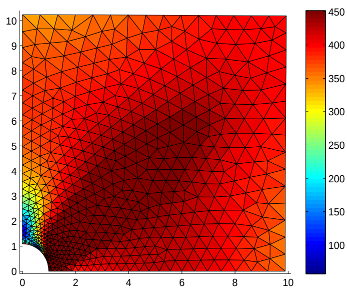


Fig. 7. Von Mises stress distribution on the deformed mesh (scaled 10x)

Parallelization Strategies of the Total FETI Coarse Space Projector

Vaclav Hapla and David Horak

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vaclav.hapla@vsb.cz

Abstract. This paper deals with discussion of several variants of parallelization of the projections to the natural coarse space in the TFETI algorithm. A 2D elasticity benchmark was implemented using PETSc parallelization library and scalability was demonstrated up to 201 mil. of unknowns and 3072 cores.

Keywords: domain decomposition method, FETI, parallel implementation, coarse problem, scalability.

1 Introduction

FETI (Finite Element Tearing and Interconnecting) domain decomposition methods are based on "divide and conquer" strategy. They are suitable for parallel solution of elliptic partial differential equations arising from many technical problems. The original FETI-1 method was proposed by Farhat and Roux [1]. It is based on the decomposition of the spatial domain into non-overlapping subdomains that are "glued" by Lagrange multipliers.

Application of the duality principle not only reduces the problem size and the condition number but also simplifies constraints so that efficient algorithms that utilize cheap projections and other tools may be used. Later, Farhat, Mandel, and Roux [2, 3] introduced a "natural coarse problem" whose solution was implemented by auxiliary projectors so that the resulting algorithm became in a sense optimal.

In our approach, we use the Total-FETI (TFETI) [4] variant which simplifies the construction of the generalized inverse of the stiffness matrix by using Lagrange multipliers not only for gluing the subdomains along the auxiliary interfaces, but also to enforce the Dirichlet boundary conditions. Basis of the kernel of the stiffness matrix is then apriori known and can be constructed easily. In addition, this fact allows easy yet powerful regularization of the stiffness matrix and computation of its stable generalized inverse.

The generalized inverse has a block diagonal layout corresponding to the decomposition into subdomains so that its application can be carried out in parallel without any communication and high parallel scalability is enjoyed. For fixed discretization, the increasing number of subdomains decreases the subdomain problem size. This results in shorter times for the subdomain stiffness

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 519–524.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

matrix factorizations and subsequently forward and backward substitutions during stable generalized inverse application. On the other hand, the dual dimension (the number of Lagrange multipliers) and the coarse problem size increase which results in longer times for all dual vector operations and orthogonal projectors applications. In this paper, different parallelization strategies of the coarse problem solution are proposed and their efficiency is demonstrated on a 2D linear elasticity benchmark up to 201 mil. of unknowns and 3072 cores.

2 Coarse space projector parallelization

Let us start with the discrete primal TFETI formulation of the linear elasticity problem

$$\min_{\mathbf{u}} \frac{1}{2}\mathbf{u}^\top \mathbf{K}\mathbf{u} - \mathbf{f}^\top \mathbf{u} \quad \text{subject to} \quad \mathbf{B}\mathbf{u} = \mathbf{o}, \tag{1}$$

and observe the primal data distributions, i.e. the distributions of the stiffness matrix $\mathbf{K}$, its generalized inverse $\mathbf{K}^\dagger$, its kernel basis $\mathbf{R}$, the constraint matrix $\mathbf{B}$, and the right hand side $\mathbf{f}$ (see [I] for more details). Let N_d denote the dual dimension and N_c the number of cores being at disposal for our computations.

The distribution of the primal data is quite straightforward as every block corresponds to a subdomain. The critical point of the TFETI method is the application of the projector $\mathbf{Q_G} = \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}$, with $\mathbf{G} = \mathbf{R}^T\mathbf{B}^T$. The action time and the level of communication depend first of all on $\mathbf{G}$ matrix distribution and implementation of the coarse problem solution, i.e.,

- whether and how $\mathbf{G}$ should be distributed,
- how action of $(\mathbf{G}\mathbf{G}^T)^{-1}$ and hence $\mathbf{Q_G}$ application should be implemented,
- whether $\mathbf{G}$ should be orthonormalized so that $(\mathbf{G}\mathbf{G}^T)^{-1} = \mathbf{I}$.

Each of cores works with local part associated with its subdomains. Current core is specified by *rank* ($rank = 0, ..., N_c - 1$). Some operations require communication through the vector transfers.

We suggest seven ways of handling the matrix $\mathbf{G}$ and subsequent realizations of the solution of the coarse problem $\mathbf{G}\mathbf{G}^T\mathbf{x} = \mathbf{b}$. These strategies of the projector application can be viewed from two points:

I. how $\mathbf{G}$ is *distributed* and its *action carried out* (see Figs. 1(a) and 1(c)):
   A) horizonatal blocks,
   B) vertical blocks.
II. how *the coarse problem is solved* which implies *the level of preprocessing of* $\mathbf{G}$ *and* $\mathbf{G}\mathbf{G}^T$ (see Figs. 1(d) and 1(b)):
   1) iteratively using CG by the master process,
   2) directly using the Cholesky factorization by the master process,
   3) applying explicit inverse of $\mathbf{G}\mathbf{G}^T$ in parallel,
   4) the coarse problem is eliminated provided that the rows of $\mathbf{G}$ are orthonormalized.

The natural coarse space matrix $\mathbf{G}$ is computed in a way where each of cores owns sparse sequential matrices $\mathbf{R}_{[rank]}$ and $\mathbf{B}_{[rank]}$, so that this core computes local block $\mathbf{G}_{[rank]} = \mathbf{R}_{[rank]}^T \mathbf{B}_{[rank]}^T$ of $\mathbf{G}$ matrix without any communication.

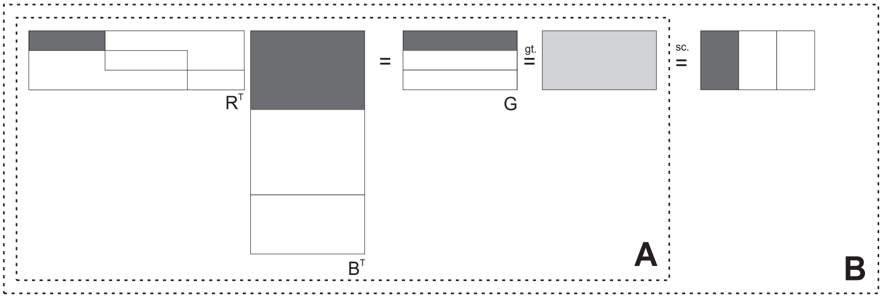
Distribution into horizontal blocks in A cases leads to enormous increase of communication because sequential dual vectors have to be scattered to the zeroth core, added together and broadcasted. Also the reduction of computational times is not so significant. Measured times during numerical experiments using A distribution and detailed analysis confirmed its unsuitability to reach good parallel scalability.

So the only way how to run huge jobs on massively parallel computers is to distribute $\mathbf{G}$ matrix into vertical blocks and parallelize all dual vectors and operations with them – B cases - this is a big advantage. Taking into account the decomposition of huge problems into the large number of subdomains resulting in large dual dimension and defect of $\mathbf{K}$, this can significantly reduce the computational time. Preprocessing of $\mathbf{G}$ comprises a redistribution of horizontal blocks $\mathbf{G}_{[rank]}$ into vertical (i.e. horizontal $\mathbf{G}_{[rank]}^T$).

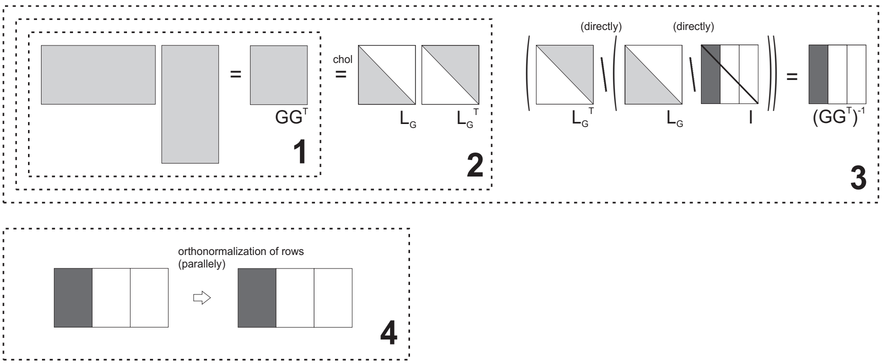
In cases 1 and 2 it is necessary to transfer the whole $\mathbf{G}$ matrix to the zeroth core or in cases 3 to all cores. Master core or each of cores then computes sequential product $\mathbf{G}\mathbf{G}^T$.

In cases 1 we employ an iterative solver for the coarse problem solution. In cases 2 and 3 any Cholesky type factorization of $\mathbf{G}\mathbf{G}^T$ and the triangular solve has to be performed. In case 2 we use the finite solver for the coarse problem solution on the zeroth core. In case 3 we use it only during preprocessing phase for the computation of the i-th column of $(\mathbf{G}\mathbf{G}^T)^{-1}$, i.e. inverse of $\mathbf{G}\mathbf{G}^T$, on each of cores assigning the i-th column of identity matrix to the right hand side $\mathbf{b}$, so that the inverse can be efficiently reached and applied in parallel if we assign N_d/N_c corresponding columns of identity matrix to the right hand side vector. Coarse problem is then solved by means of ordinary matrix-vector multiplication by the distributed $(\mathbf{G}\mathbf{G}^T)^{-1}$ matrix.

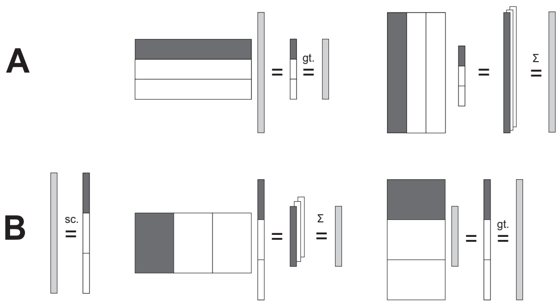
In case $B4$ rows of $\mathbf{G}$ are orthonormalized resulting in a matrix $\bar{\mathbf{G}}$. This way has a big advantage - we eliminate the coarse problem ($\mathbf{G}\mathbf{G}^T = (\mathbf{G}\mathbf{G}^T)^{-1} = \mathbf{I}$) completely. We have to redistribute first the sequential blocks $\mathbf{G}_{[rank]}$ into the parallel dense matrix $\bar{\mathbf{G}}^T$ and then perform orthonormalization of its columns, i.e. rows of $\bar{\mathbf{G}} = \mathbf{T}\mathbf{G}$ described by a nonsingular matrix $\mathbf{T}$. For this purpose the classical Gram-Schmidt algorithm was chosen, that appears more suitable for the parallelization of this problem than the modified or iterated classical Gram-Schmidt algorithm. Let us recall that the algorithm requires half of floating-point operations, on parallel computers it can be much faster than the modified one, and its parallel efficiency equals that of iterated classical one. The columns of the matrix $\mathbf{G}^T$ are copied into the array of parallel vectors $\mathbf{g}[\,]$ (local size N_d/N_c, global size N_d) and the process of orthonormalization is performed in parallel. The obtained vectors form columns of the matrix $\bar{\mathbf{G}}^T$. Considering this type of distribution, this process requires only transfers of dot products and can be very efficient [7]. Another possibility is to use forward substitution of factorized $\mathbf{G}\mathbf{G}^T$ applied to N_d/N_c columns of the original $\mathbf{G}$ matrix as the right hand side.



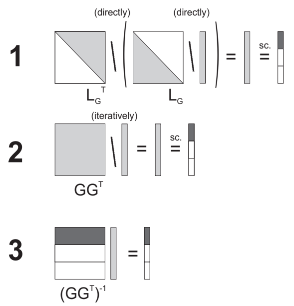
(a) Parallelization of **G**



(b) Preprocessing of **G** and $\mathbf{GG}^T$



(c) Action of **G** depending on (a)



(d) Action of $(\mathbf{GG}^T)^{-1}$ depending on (b)

Fig. 1. Various ways of realization of the $\mathbf{Q}_\mathbf{G}$ action

3 Numerical experiments

Described algorithms were implemented using the standard numerics parallelization tool PETSc (Portable Extensible Toolkit for Scientific Computation), version 3.1.014, developed by Argonne National Laboratory [6] and tested on the solution of a 2D linear elasticity problem. For these computations we used *HECToR* phase 2b system at EPCC in Edinburgh [5] offering 1856 XE6 compute nodes, each with two AMD 2.1 GHz 12-core processors giving a total of 44,544 cores and with 32 GB of main memory. The processors are connected with a high-bandwidth interconnect using Cray Gemini communication chips.

As a benchmark we used a 2D elastostatic problem of the steel traverse represented by a domain $\Omega = (0, 600) \times (0, 200)$ with the sizes given in [mm]. The material constants are defined by the Young modulus $E = 2.1 \cdot 10^5$ [MPa], the Poisson ratio $\nu = 0.3$, and the density $\rho = 7.85 \cdot 10^{-9}$ [ton/mm^3]. The body is fixed in all directions along the left side $\overline{\Gamma}_U = \{0\} \times [0, 200]$ and loaded by gravitational forces with the gravity acceleration $g = 9800\text{mm/s}^2$.

To illustrate the efficiency of the different strategies of the coarse problem solution we decomposed the body Ω into identical square subdomains with the side length H. We gradually chose decompositions into 8×24, 16×48, 24×72, and 32×96 boxes. All subdomains were further discretized by the same uniform triangular meshes characterized by the discretization parameter h and the ratio $H/h = 180$.

In Table 1, we report the computational times for preprocessing and $\mathbf{Q_G}$ action of the parallel strategies for the coarse problem solution described in Section 2. Obviously the best strategy corresponds to the case $B4$, where $\mathbf{G}$ is distributed in vertical blocks and orthonormalized.

Table 1. Performance of the coarse problem solution for varying strategy, decomposition and discretization

Number of subdom.	192	768	1728	3072
Number of cores	192	768	1728	3072
Primal dim.	12,580,224	50,320,896	114,476,544	201,283,584
Dual dim.	129,984	537,216	1,228,464	2,183,424
Kernel dim.	576	2304	5184	9216
$\mathbf{G}_{[rank]}$	1.001e-02	1.152e-02	1.489e-02	1.527e-02
broadcast of $\mathbf{G}$ to all cores	9.102e-02	3.710e-01	8.353e-01	1.389e+00
B1,2,3: $\mathbf{GG}^T$ assembling	6.710e-02	2.469e-01	7.155e-01	1.203e+00
B2,3: $\mathbf{GG}^T$ Chol. fact.	8.090e-03	1.042e-01	8.108e-01	2.004e+00
$B3$: inverse	1.767e-01	1.149e+00	6.401e+00	9.264e+00
$B4$: orthonormalization	9.669e-02	5.983e-01	3.262e+00	4.629e+00
$B1$: $\mathbf{Q_G}$ action	1.070e-02	6.934e-02	3.204e-01	6.424e-01
$B2$: $\mathbf{Q_G}$ action	8.046e-03	5.404e-02	2.321e-01	4.576e-01
$B3$: $\mathbf{Q_G}$ action	5.822e-03	3.742e-02	1.760e-01	3.621e-01
$B4$: $\mathbf{Q_G}$ action	6.096e-03	2.694e-02	6.424e-02	9.874e-02

4 Conclusions and goals

This paper analyzes seven parallelization strategies for the coarse problem solution in the TFETI algorithm and the projector application. The data distributions are discussed with the conclusion that all B variants behave much better in parallel up to thousands cores reducing significantly times of all dual actions (axpy, dot).

Further work should involve fine tuning and optimization of the presented actions, theoretical performance analysis, more holistic analyses of FETI-based algorithms, comparison of times presented in this paper (pure MPI approach) with hybrid parallelization approaches, experiments with numerics parallelization frameworks other than PETSc, e.g. Trilinos.

Main results

[I] T. Kozubek, V. Vondrak, M. Mensik, D. Horak, Z. Dostal, V. Hapla, P. Kabelikova, M. Cermak, "Total FETI domain decomposition method and its massively parallel implementation", Advances in Engineering Software, submitted
[II] D. Horak, V. Hapla, "TFETI-1 coarse problem parallelization strategies", to appear in proceedings of 9TH INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING AND APPLIED MATHEMATICS 2011, accepted

References

1. C. Farhat, F-X. Roux, "An unconventional domain decomposition method for an efficient parallel solution of large-scale finite element systems", SIAM J. Sci. Stat. Comput. 13, 1992; 379–396.
2. C. Farhat, J. Mandel, F-X. Roux, "Optimal convergence properties of the FETI domain decomposition method", Comput. Methods Appl. Mech. Eng. 115, 1994
3. F-X. Roux, C. Farhat, "Parallel implementation of direct solution strategies for the coarse grid solvers in 2-level FETI method", Contemporary Math. 218, 1998
4. Z. Dostál, D. Horák, R. Kučera, "Total FETI - an easier implementable variant of the FETI method for numerical solution of elliptic PDE", Communications in Numerical Methods in Engineering, 22(2006), 12, pp. 1155-1162
5. HECToR Home Page, `http://www.hector.ac.uk/`
6. S. Balay, J. Brown, K. Buschelman, W.D. Gropp, D. Kaushik, M.G. Knepley, L. Curfman McInnes, B.F. Smith, H. Zhang", PETSc Web page, http://www.mcs.anl.gov/petsc, 2011.
7. F. J. Lingen, Efficient Gram-Schmidt orthonormalization on parallel computers, Research report, Department of Aerospace Engineering, Delft University of Technology, 1999.

Numerical Experiments with Gradient Based Methods for Calibration of a Heat Conduction Models

Rostislav Hrtus

Institute of Geonics of the Academy of Sciences of the Czech Republic,
Ostrava, Studentská 1768, 708 00 Ostrava - Poruba
Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
hrtus@ugn.cas.cz

Abstract. The paper provides a numerical experiments with techniques suitable for calibration of a heat conduction or similar models based on PDE and discretized by the FE method. Especially, an application of gradient based optimization methods is discussed. A special model problem, originated from rock mechanics, is used for comparison of the solution methods.

1 Introduction

Problems of calibration of models by identification of parameters appears in many fields of science and engineering. The main purpose of this paper is formulation of such calibration/identification problems and description of numerical optimization methods for their solution. The gradient-based optimization methods were implemented and tested on calibration of a nonstationary heat conduction model arising in geomechanics.

2 Identification problems - nonstationary case

Let us consider the parabolic heat flow problem in the form

$$c\,\frac{\partial u}{\partial t} - \text{ div } (k \text{ grad } u) = f \quad \text{in } \Omega \times (0,T) \quad \text{in } \Omega\,, \tag{1}$$

$$u = \hat{u} \text{ on } \Gamma_D \quad \text{and } \; k\,\text{grad} u \cdot n = \hat{q} \quad \text{on } \Gamma_N, \tag{2}$$

where $\partial\Omega = \Gamma_D \cup \Gamma_N$. We also assume discretization by the FEM given by division $\mathcal{T}_h$ of Ω into triangular/tetrahedral elements and considering the space V_h of continuous piecewise linear functions.

As a time discretization scheme we choose the implicit Euler method using uniform time discretization given by $t_i = i\Delta t,\ i = 1, \dots, T/n_t$.

Further, the coefficients k and c are assumed to be constant on elements $E_i \in \mathcal{T}_h$ or on larger subdomains $\tilde{E}_j \in \mathcal{T}_H$. Here, each $\tilde{E}_j\,(j = 1, \dots, s)$ is

VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

a union of some elements from $\mathcal{T}_h$. Thus $s \leq nel = \text{card}(\mathcal{T}_h)$, but frequently $s \ll nel$. We can construct the set of parameter vectors K_H,

$$K_H = \{k = k(x),\ c = c(x) :\ \exists p \in R^{2s} :\ k\,|_{\tilde{E}_j} = p_j\,, \quad c\,|_{\tilde{E}_j} = p_{j+s} \quad \forall\, j = 1, \ldots, s\}. \tag{3}$$

The *parameter identification problem* considers the coefficients k, c represented by the vector $p \in R^{2s}$ as unknown and seek for optimal $p \in R^{2s}$ to fit the observation data

$$d_j^k = u_{obs}(x_j, t_k),\ \ x_j \in N_{obs} \subset N_h,\ \ \text{card}(N_{obs}) = m, \quad t_k \in T_{obs} \subset \{t_i : i = 1, \ldots, n_t\},\ \ \text{card}(T_{obs}) = m_t,$$

using the corresponding objective function

$$F_{obs} = \frac{1}{2} \sum_{x_j \in N_{obs}, t_k \in T_{obs}} |\ d_j^k - \bar{u}_j^k(p)\ |^2\ . \tag{4}$$

The identification problem can be then formulated as

$$\begin{gathered} \text{find }\ (p, \bar{u}) :\ p \in R^{2s},\ \bar{u} \in (R^n)^{n_t} \\ \widetilde{F}(p) = F(p, \bar{u}(p)) = F(p, \bar{u})\ = \min F(q, \bar{v}), \quad \bar{q} \in R^s,\ \bar{v} \in (R^n)^{n_t}. \end{gathered} \tag{5}$$

Above function $F(q, \bar{v})$ is defined as follows

$$F(q, \bar{v}) = F_{obs}(q, \bar{v}) + \beta F_{reg}(q),$$

where the construction of the regularization part F_{reg} is discussed e.g. in [6], [7], [5] and in our case, we do not use regularization part.

3 Unconstrained minimization techniques

The objective function $\widetilde{F}(p)$ from (5) has the nonlinear least squares structure given by (4). Generally, without regularization, we have

$$\widetilde{F}(p) = \frac{1}{2} R(p)^T R(p), \quad p \in R^{2s}, d \in R^{m \cdot m_t}, \bar{u} \in R^{n \cdot n_t}. \tag{6}$$

The residual $R(p)$ is defined as

$$R : R^{n \cdot n_t} \to R^{m \cdot m_t},\ R(p) = S\bar{u}(p) - d, \tag{7}$$

where $S :\ R^{n \cdot n_t} \to R^{m \cdot m_t}$ is the observation operator.

The optimization problem can be solved by direct methods like Nelder–Mead simplex method or genetic algorithms [2]. But here, we consider methods exploiting smoothness of the mapping $p \to R(p)$, i.e. the existence of Jacobian

$$J(p) = D_p R(p) = (J_{ij}(p))\,,\ J_{ij}(p) = \frac{\partial R_i(p)}{\partial p_j}\,.$$

Having the Jacobian $J(p)$, the expressions

$$\mathrm{Grad}\widetilde{F}(p) = J(p)^T R(p), \tag{8}$$

$$\mathrm{Hess}\widetilde{F}(p) = J(p)^T J(p) + (D_p^2 R(p))R(p) \sim J(p)^T J(p) = H_{GN}(p) \tag{9}$$

gives possibility of using iterative optimization methods as steepest descent (SDM) or Gauss–Newton (GN) and Gauss–Newton Levenberg–Marquardt (GN–LM) type iterations. A general solution algorithm can be written in the form

$$p^0 \text{ given } ;\ p^{i+1} = p^i - \alpha_i z^i,\ i \geq 0, \tag{10}$$

where

$$z^i = g^i \text{ (SDM) } \text{ or } z^i = H_i^{-1} g^i, \quad g^i = J(p^i)^T R(p^i), \tag{11}$$

$$H_i = J(p^i)^T J(p^i) \text{ (GN) } \text{ or } H_i = \nu_i I + J(p^i)^T J(p^i) \text{ (GN-LM)}\ . \tag{12}$$

Note that $\alpha_i = 1$ or α_i are given by a golden section linesearch (GSL) or backtracking with Armijo rule (AR). We may choose ν_i as e.g.
$\nu_i = \min\{1;\ c \parallel J(p^i)^T R(p^i) \parallel\},\ c > 0$, see e.g. [3].

The Jacobian $J(p)$ can be computed by forward differences.

4 Model problem and numerical results

The in-situ Äspö Pillar Stability Experiment (APSE) has been performed to determine THM processes and take the data (Fig. 1). The measured data are now used for validation of mathematical models for rock mass prediction within the DECOVALEX 2011 international project. To determine accurately the temperature changes, a heat flow model is formulated and monitored temperatures are used for model calibration. More details can be found in [2].

A FEM model of APSE, realized by GEM software [1], considers domain of $105 \times 125 \times 118$ m and tetrahedral 3D mesh with $99 \times 105 \times 59$ nodes. The grid is refined around the pillar, see Fig. 1.

There are 14 temperature monitoring positions and temperatures are measured in 12 time moments. Altogether 168 values of temperature measurement (vector d) are used for calibration, $m \cdot m_t = 168$.

The material parameters are different conductivity k and heat capacity c. The simplest decomposition $\mathcal{T}_H^{(2)}$ decompose the model domain just into two parts (dry and wet according to Fig.1 right). Convergence of function values of objective function are shown by red graph and to the right from this graph there

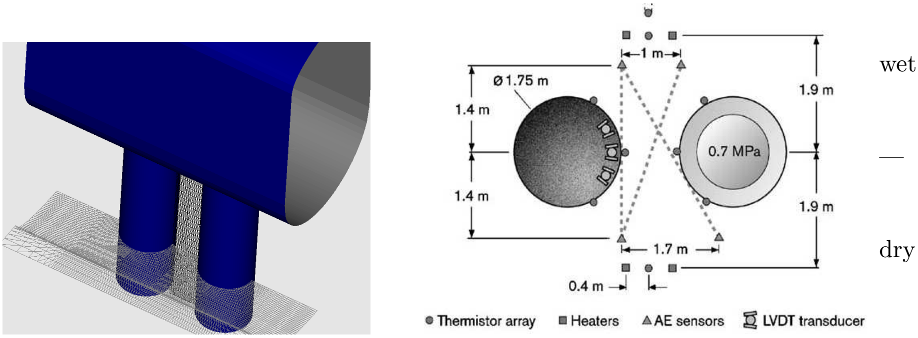


Fig. 1. The APSE model - detail of the FE grid around the pillar (GEM software [1]) and ground view on the pillar, holes, location of heaters and points of temperature measurement.

are four graphs showing convergence of parameters. In horizontal axis in every one graph there is number of state problem evaluations needed for computing one iteration of algorithm included linesearch.

Figure 2 shows convergence characteristics for solving $\mathcal{T}_H^{(2)}$ problem by SDM and damping parameters are determined by a GSL.The same situation we can see in Figure 3, convergence characteristics for solving our problem by SDM with AR backtracking, which starts with $\alpha_i = 1$ and halves α_i if $F(p^{i+1}) < F(p^i)$ is violated. We can see quite oscillatory behaviour of convergence of parameters in both cases. Gauss-Newton iterations are shown in Figure 4 with GSL and in Figure 5 with AR backtracking. There is obvious smoother behaviour of convergence of this method. Note that the minimum found by this method is lower than in previous SDM case. Third method GN with Levenberg-Marquardt modification is shown in Figure 6 with GSL, which is slowly converging, but founded minimum is quite satisfactory, and in Figure 7 with AR backtracking, which works well, but not better than pure GN.

5 Conclusions

We tested three methods for calibration of a heat conduction model. The most efficient method was the Gauss-Newton one. Levenberg-Marquardt method causes light smoothing of the convergence speed. Steepest descent method was not as suitable as previous mentioned methods. Our future aims includes solution of problems involving more parameters, application of parallel computing, semianalytic computation of gradients and Jacobians and some theoretical convergence analysis.

Acknowledgements

The work was conducted within the international DECOVALEX-2011 Project (DEvelopment of COdes and their VALidation against EXperiments). Author

Fig. 2. Convergence of SDM with GSL on the left and convergence of four parameters on the right (left top - k_{dry}, right top c_{dry}, left bottom k_{wet}, right bottom c_{wet}).

Fig. 3. Convergence of SDM with AR on the left and convergence of four parameters on the right (left top - k_{dry}, right top c_{dry}, left bottom k_{wet}, right bottom c_{wet}).

would like to thank Radim Blaheta for valuable discussion and comments and Roman Kohut for his kind help and provided software [1].

References

1. R. Blaheta, O. Jakl, R. Kohut, J. Starý, GEM - a Platform for Advanced Mathematical Geosimulations. In: Proceedings of the conference Parallel Processing and Applied Mathematics, PPAM 2009, to appear
2. R. Blaheta, R. Kohut, O. Jakl, Solution of Identification Problems in Computational Mechanics - Parallel Processing Aspects, PARA 2010, LNCS Springer accepted for publication, 9pp.
3. J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Classics in Applied Mathematics 16, SIAM, Philadelphia 1996
4. C.T. Kelley, Iterative Methods for Optimization, SIAM, Philadelphia 1999
5. T.K. Nilssen, X.C. Tai, Parameter estimation with the augmented Lagrangian method for a parabolic equation, J. Optim. Theory Appl 124(2005), pp. 435-455
6. C.R. Vogel, Computational Methods for Inverse Problems SIAM, Frontiers in Applied Mathematics no.23, Philadelphia 2002
7. C.R. Vogel, J.K. Wade, Analysis of costate discretizations in parameter estimation for linear evolution equations, SIAM J. Control Optim. 33(1995), pp. 227-254

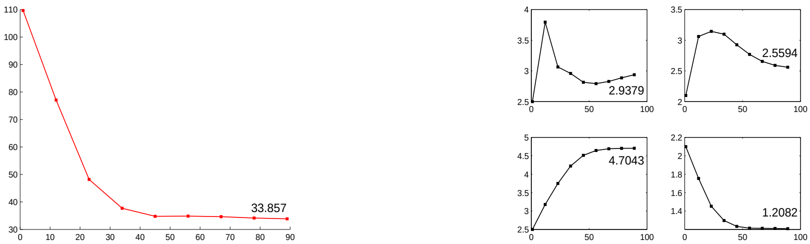


Fig. 4. Convergence of GN with GSL on the left and convergence of four parameters on the right (left top - k_{dry}, right top c_{dry}, left bottom k_{wet}, right bottom c_{wet}).

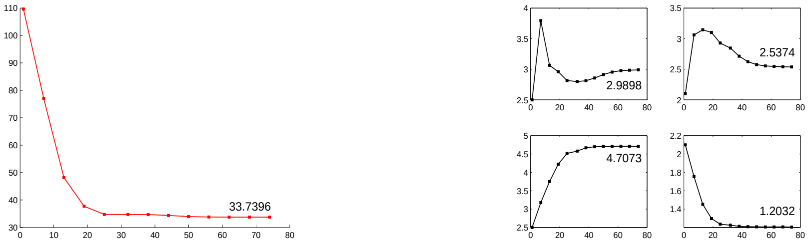


Fig. 5. Convergence of GN with AR on the left and convergence of four parameters on the right (left top - k_{dry}, right top c_{dry}, left bottom k_{wet}, right bottom c_{wet}).

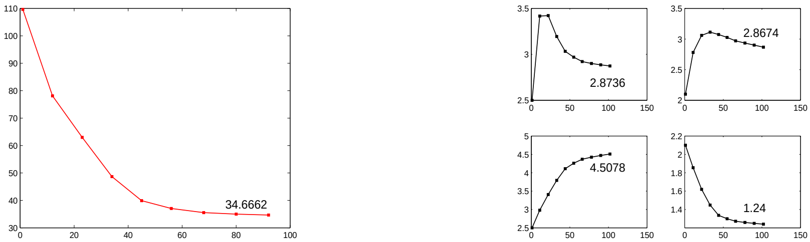


Fig. 6. Convergence of GN-LM with GSL on the left and convergence of four parameters on the right (left top - k_{dry}, right top c_{dry}, left bottom k_{wet}, right bottom c_{wet}).

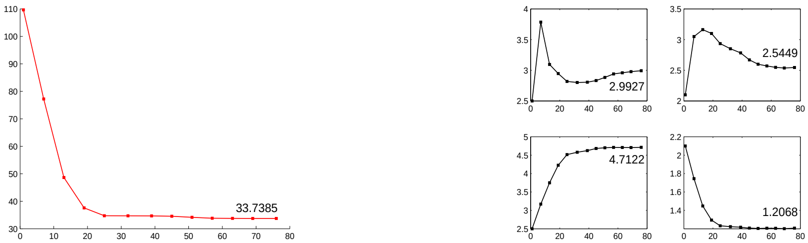


Fig. 7. Convergence of GN-LM with AR on the left and convergence of four parameters on the right (left top - k_{dry}, right top c_{dry}, left bottom k_{wet}, right bottom c_{wet}).

Survival Analysis: Comparison of Two Different Surgery Techniques for Minimization of Postoperative Complications

Kateřina Janurová

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
katerina.janurova@vsb.cz

Abstract. This paper deals with evaluation of medical survival censored data of 844 patients, who underwent surgery of colon and rectum cancer in the University Hospital of Ostrava. There are used two different surgery techniques for this operation: either classical (open) or laparoscopic. Basic question which arises generally at surgeries with several possible techniques is which type of operation to choose to guarantee longer overall survival time. For the comparison was used the non-parametric approach which results from Kaplan-Meier estimates of the survival function. The survival curve was constructed for each group of patients, i.e. for classical (open) and laparoscopic type of operation. Final survival curves were compared and evaluated using advanced methods of statistical inference, as is hypothesis testing (e.g. log-rank test).

Keywords: censored medical survival data, comparison of surgery techniques, survival curve, log-rank test

1 Introduction

Minimally invasive surgery is generally associated with lower operative stress and more favorable post-operative course. On the other hand there are many negative factors in using laparoscopic techniques in colorectal surgery, which can participate in morbidity in large measure (e.g. the risk of capnoperitoneum, longer operative time and extreme positioning of patients). The comparison of morbidity and mortality after both types of surgeries is frequently published result of numerous medical investigations. For example, the consensus of European association of endoscopic surgery for colon carcinoma mentions, that there is no difference between morbidity of laparoscopic and open operations of colon [1].

2 Kaplan-Meier estimator of survival function

Long-term survival analysis has been performed by using the Kaplan-Meier method. Let us suppose randomly right censored survival data. We can obtain from our experiment *n* independent observations in the form as follows:

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 531–536.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

$$(t_1, c_1), \ldots, (t_n, c_n),$$

where t_i is either a time of death or a time in which the observation of *i-th* patient is stopped (withdrawn) and $c_i = 1$ (resp. $c_i = 0$) is censoring indicator, accordingly to the death (resp. stopping time) occurred first.
Assume that we may obtain from the experiment among n observations $m \le n$ observed number of death and $n - m$ withdrawn values and that in the sample $t_1, \ldots, t_n$ no conformity has occurred. Then we create rank-ordered sample $t_{(1)} < \ldots < t_{(n)}$. Let $c_{(i)}$ be the indicator corresponding to $t_{(i)}$, $i = 1, \ldots, n$, then the Kaplan-Meier estimator of survival function $\hat{R}(t)$ is:

$$\hat{R}(t) = \prod_{t_{(i)} \le t} \frac{n_i - d_i}{n_i}, \tag{1}$$

with the convention that $\hat{R}(t) = 1$ if $t < t_{(1)}$ and $\hat{R}(t) = 0$ if $t > t_{(n)}$.

Notation: n_i …number of patients at risk of dying until $t_{(i)}$ (the time $t_{(i)}$ is not included).
d_i …number of observed deaths

Kaplan-Meier estimator has an asymptotic normal distribution. Once we obtain the survival function estimate we would like to be informed about the dispersion around $\hat{R}(t)$. This dispersion is determined by the variance or standard deviation of the estimator. Asymptotic variance is known as Greenwood's formula and we derive it by the delta method which is based on a first-order Taylor series expansion [2]

$$\hat{\mathrm{Var}}\left(\hat{R}(t)\right) = \left(\hat{R}(t)\right)^2 \sum_{t_i \le t} \frac{d_i}{n_i (n_i - d_i)}.$$

The lower and upper bounds for pointwise confidence interval estimates are based on the log-minus-log survival function $\ln\{-\ln[\hat{R}(t)]\}$, its variance (or standard deviation) and a quantile of the standard normal distribution (as a consequence of asymptotically normally distributed Kaplan-Meier estimator). Following expression gives us the bounds of a 100(1 - α) percent confidence interval for the log-minus-log function

$$\ln\left[-\ln\left(\hat{R}(t)\right)\right] \pm z_{1-\alpha/2}\, \hat{\mathrm{SE}}\left\{\ln\left[-\ln\left(\hat{R}(t)\right)\right]\right\}, \tag{2}$$

where $z_{1-\alpha/2}$ is the upper α/2 percentile of the standard normal distribution and $\hat{\mathrm{SE}}\{\ \}$ denotes the estimated standard error of the log-minus-log survival function, which is the positive square root of the estimator of variance of the log-minus-log survival function, defined as follows

$$\hat{\mathrm{Var}}\left\{\ln\left[-\ln\left(\hat{R}(t)\right)\right]\right\} = \frac{1}{\left[\ln\left(\hat{R}(t)\right)\right]^2} \sum_{t_i \le t} \frac{d_i}{n_i\left(n_i - d_i\right)}.$$

By denoting the lower and upper bounds of the confidence interval of (**2**) as $\hat{c}_l$ and $\hat{c}_u$ we may obtain the lower and upper bounds of the confidence interval for the survival function respectively as

$$\exp\left[-\exp\left(\hat{c}_u\right)\right] \text{ and } \exp\left[-\exp\left(\hat{c}_l\right)\right].$$

3 Log-rank test

It is frequently of interest to determine whether two subgroups of samples could arise from identical survival functions. First step we can do to solve this task is graphical display of the Kaplan-Meier estimator of the survival function for each of the groups. Generally we can say that if one survival function lies completely above another, than the proportion of subject estimated to be alive at any point of time for this group is greater than for the other group represented by the lower survival function. The main question is whether the observed difference, what we can see in graph is statistically significant.

One possibility to answer this question is log-rank test. This test is especially good when the ratio of hazard functions in the populations we compare is approximately constant. It is constructed by calculating the number at risk and the number of observed deaths in one of the groups at each observed survival time $t_{(i)}$, if we assume that the survival function is the same in each of the two groups (we mark the groups as Group 0 and Group 1). This produces the estimator of expected number of deaths at time $t_{(i)}$ (for example using Group 1):

$$\hat{e}_{1i} = \frac{n_{1i} d_i}{n_i}$$

and estimator of variance of d_{1i}, with hypergeometric distribution:

$$\hat{v}_{1i} = \frac{n_{1i} n_{0i} d_i \left(n_i - d_i\right)}{n_i^2\left(n_i - 1\right)}.$$

The log-rank test is defined as follows:

$$Q = \frac{\left[\sum_{i=1}^{m} \left(d_{1i} - \hat{e}_{1i} \right) \right]^2}{\sum_{i=1}^{m} \hat{v}_{1i}} .$$

Notation: n_{0i}, n_{1i} … number at risk at observed survival time $t_{(i)}$ in Group 0 and Group 1 respectively
d_{0i}, d_{1i} … number of observed deaths in each group respectively
n_i … total number at risk
d_i … total number of observed deaths

When we use the null hypothesis in the form that two survival function are the same, we may acquire the p-value for Q by the application of the chi-square distribution with one degree of freedom ($p = \Pr(\chi^2(1) \geq Q)$), if we assume that censoring is independent of the group, and that the total number of observed events and the sum of the expected number of events is large.

4 Results

Data from colectomy operations have been used to construct Kaplan-Meier estimator of survival functions $\hat{R}(t)$ according to formula (**2**), see Figure 1, including 95% confidence limits.

Kaplan-Meier Estimates:
Laparoscopic technique: MST = 61.185 SE = 2.516
Open technique: MST = 52.938 SE = 2.405

Table 1. Comparison of groups

Group	Total	Dead	Withdrawn	Withdrawn [%]
Lpsk	455	160	295	64.84
Open	387	205	182	47.03
Total	842	365	477	56.65

Log-rank test result:
Chi-square = 8.070, P-value = 0.005
Since the P-value is less than 0.01, there is a statistically significant difference between the groups at the 99% confidence level.

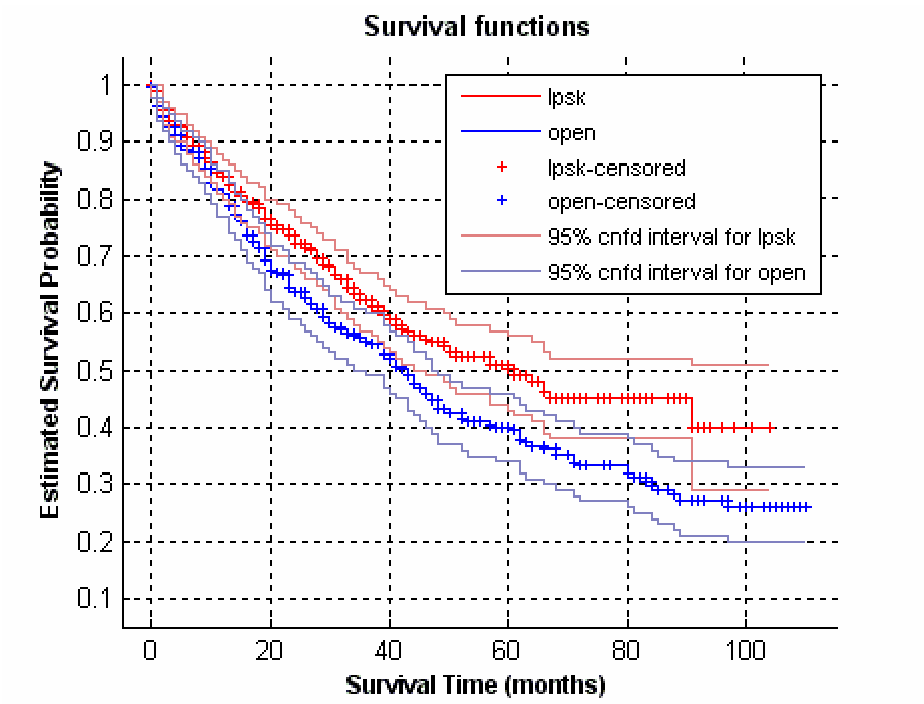


Fig. 1. Survival functions for both operations with 95% confidence intervals

5 Conclusion

We confirmed by using non-parametric approach that laparoscopic surgery for cancer of both rectum and colon is minimally equivalent alternative to open surgery.
We can conclude that there is a statistically significant difference between the survival functions associated to both operation types at the 99% confidence level (survival function associated to the laparoscopic data is significantly greater). Results were presented on the SSARS Conference (Summer Safety and Reliability Seminars 2011).

Acknowledgement

This work was funded by grant SGS SP2011/100 VŠB-TU Ostrava.

References

1. Briš R., Praks P., Janurová K., Martínek L.. Survival analysis on data of different surgery techniques to evaluate risk of postoperative complications. *SSARS 2011 - Summer Safety and Reliability Seminars, Volume 1*. Gdansk 2011. ISBN 978-83-925436-2-6. Pg. 33-38

2. Hosmer D. W., Lemeshow S., May S. *Applied Survival Analysis*. Wiley, 2008 New Jersey. ISBN 978-0-471-75499-2.

Automated Region of Interest Retrieval in Metallographic Images in Matlab with Application to Quality Scoring

Petr Kotas[1], Ladislav Válek[2], and Pavel Praks[1]

[1]Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{petr.kotas, pavel.praks}@vsb.cz
[2]Research – Production Technology, ArcelorMittal Ostrava plc, Ostrava, Czech Republic
ladislav.valek@arcelormittal.cz

Abstract. The aim of the research is development and testing of new methods to assess the quality of digital metallographic images introduced in the quality of steel with high added value. In this paper, we will address the development of methods to assess the quality of metallographic samples, including slabs with the main emphasis on the quality of the image center. For this reason, we introduce an alternative method for automated region of interest retrieval. In the first step, the metallographic image is segmented using both spectral method and thresholding. Then, the extracted macrostructure of the metallographic image is automatically analyzed by statistical methods. Finally, automatically extracted region of interests are compared with results of human experts.

Keyword: Hard industry, Materials science and technology, Quality control, Expert systems, Control system human factors, Information retrieval, Image segmentation, Image classification, Monitoring,

1 Introduction

The aim of the research is the development and testing of methods for automated analysis of metallographic samples. For quality scoring of metallographic images, it is interesting to retrieve a single subregion of the image, leaving other regions unchanged. This can be referred as region of interest (ROI) retrieval [3,6]. The problem of automated retrieval of region of interest in real metallographic images from the steel plant of ArcelorMittal Ostrava plc (Ostrava, Czech Republic) is addressed in this paper. The objective is to monitor the process quality of a part steel plant production. For this reason two different image segmentation algorithms are used and compared in this contribution. The first one computes the histogram of gray levels of the image, while in the second one the spectral segmentation is applied. Then, the extracted metallographic image is automatically analyzed by statistical methods. The aim is to retrieve region

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 537–542.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

of interest, which contains small dots. These small dots significantly influence so called the central quality of the metal sample. The paper is structured as follows. Section II describes the image segmentation. In Section III, automated method for retrieval of region of interest is presented. Experimental comparison of these two approaches for image segmentation follows in Section IV. Finally, Section V closes the paper.

2 Image Segmentation

In this paper we used well known threshold based image segmentation and compare it to previously used spectral segmentation [2]. We present images for which spectral segmentation fails comparing to threshold method as well as we show, that spectral method is significantly slower than threshold method.

It's easy to see that choice of threshold is crucial for successful extraction. The typical histogram of input metallographic data is seen on Figure1.

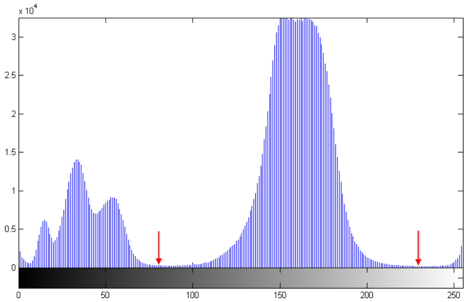
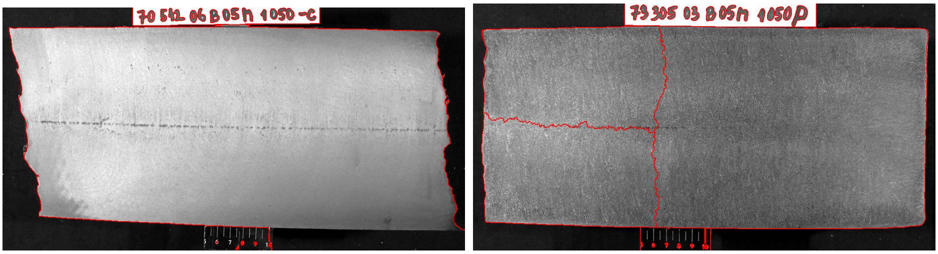


Fig. 1: Histogram of input metallographic data

Obvious choice for threshold would be one of the values marked by red arrows in Figure 1. We refer to [1] and to the knowledge of structure of used images and choose first of two values. There has to be done some morphological post processing because of heterogeneity of background and noise from the camera. This is done by morphological operations open and fill. Detailed description of these operations can be found in [1].

Figure 2 illustrates examples of spectral segmentation failure: the presence of the noise and certain material structure.

The threshold segmentation method shows to be superior in reference to the quality and speed efficiency of the algorithm. The comparison results presented in Table 1 indicate also a possibility of the real-time image segmentation for the threshold based method.



(a) Thresholding (b) Spectral method

Fig. 2: Insufficient results of the spectral segmentation.

Segmentation method	Computation time (secs)
Threshold	1.9
Spectral	69.9

Table 1: Comparison of computation time for thresholding and spectral segmentation

3 Automated Region Of Interest Retrieval

In this section we present simple, but effective method for automated ROI detection. Method is based on moving averaging window that moves down along the rows of an image. In this way, we find one dimensional vector that describes all rows of an image, see Algorithm 2.

In order to achieve successful localization of the ROI there has to be done an intensity pre-processing stage that enables better dark-white contrast highlight. We call it contrast normalization and it is done by the matrix formula

$$I = \frac{average\, intensity\, of\, input\, image}{input\, image}$$

Algorithm 1 Region of interest detection

```
I      = input image
I      = contrastNormalization(I)
window = window height
out    = output vector with averages of  moving window
for all rows do step window do
           tmp       = I(row:row+window,:)
           out(row)  = average(tmp)
end
maxs = findLocalMaximas(out)
```

The window's height has been experimentally estimated to be 100 pixels. This window dimension enables acceptably smooth graph of average values. At the final step we used an explicit knowledge of the problem structure. We deleted all false positive matches located too far from the center of the metal sample.

4 Results Of Automated Region Of Interest Retrieval

Figures 3 present graphs of moving window averages. It is obvious that our ROI is in center of the image with some more false-positive hits. Information from these figures is displayed in Figures 4 and 5, where blue lines represent detected region of interests. Metal-sample in Figure 5 has some minor residue from the segmentation, but this residue does not have significant impact to the successful ROI localization. And the final step of our algorithm is presented in Figure 6, where the final ROI has been successfully detected.

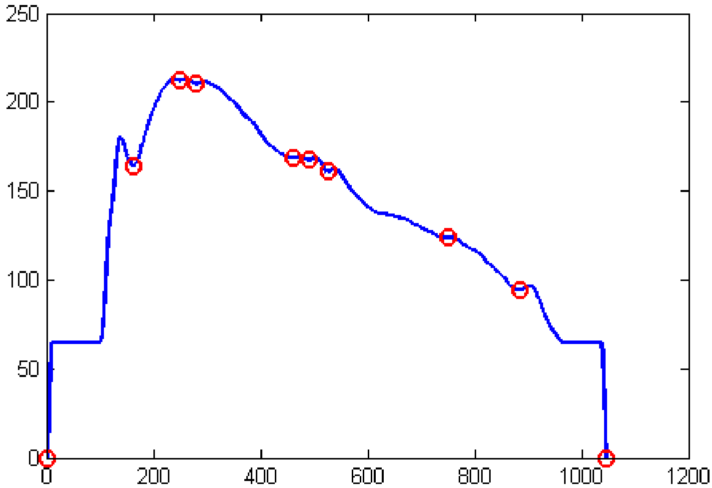


Fig. 3: Metal-Sample A: Mean values of each row in analyzed image. Detected local maxima (expressed by red circles).

5 Conclusions

The objective of this research is to the quality control process in the steel plant. The experimental metallography images have been provided by the ArcelorMittal Ostrava plc (Ostrava, Czech Republic).

We presented comparison of two algorithms for automatic segmentation of real metallographic images in this paper. We introduced the histogram-based

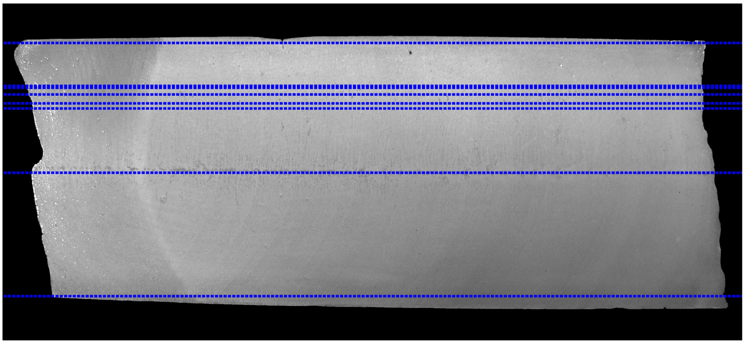


Fig. 4: Metal-Sample A: Identification of region of interest.

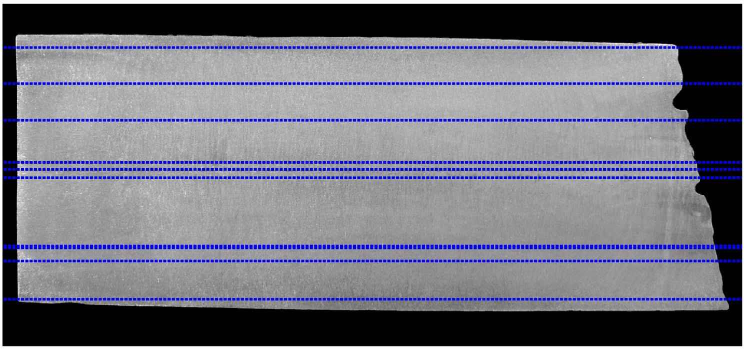


Fig. 5: Metal-Sample B: Identification of region of interest with several false-positives.

Fig. 6: Metal-Sample B: Final ROI detected

segmentation technique for metallographic images and we described the method for automated estimation of the region of interest. The obtained results prove high performance of the histogram segmentation technique.

We successfully automatically detected the region with dots that influence the quality of the metal samples. The classification results obtained by the proposed methods are close to the classification of human experts. There are some false positives matches in the intermediate stages of the algorithm that are deleted in the final step of the proposed methods. Figure 4 and Figure 3 with corresponding local maximum, illustrate the presence of the false positive matches. Final localized ROI could be seen in Figure 6. This final automatically extracted region of interest is consistent with the human expert dots detection.

The obtained results are promising and indicate that the proposed method for ROI can be used for future evaluation of crosscut macro structure of slab samples. In order to achieve more precise evaluation results, individual areas of a sample crosscut images should be analyzed in more details. In our future research we plan to use techniques for automated extraction of dots [7] from the segmented region of interest. The detailed analysis of extracted dots will also be important for future search of metallurgical relations in images [5], which are hidden in the image database.

Acknowledgment

A This work was supported by the project FR—TI1/432 of the Ministry of Industry and Trade of the Czech Republic. Special thanks to Ruzica Vlajkovic and Milos Zeljkovic.

References

1. 2002. *Digital Image Processing, 2nd Edition.* Prentice Hall, R.C. Gonzales, R.E. Woods.
2. L. Grady. *Space-Variant Computer Vision: A Graph-Theoretic Approach.* PhD thesis, Boston University, Boston, 2004.
3. Goldmann T Görtler J. Kayser G. Vollmer E. Kayser K., Schultz H. Automated region of interest retrieval and classification using spectral analysis. *Diagnostic Pathology*, 2009.
4. Zeljkovic Vesna Valek Ladislav Kotas Petr, Praks Pavel. Automated region of interest retrieval of metallographic images for quality scoring estimation. 2010.
5. R. Moravec L. Válek P. Praks, M. Grzegorzek and E. Izquierdo. Wavelet and eigenspace feature extraction for classification of metallography images. *Modelling and Knowledge Bases XIX by H. Jaakkola, Y. Kiyoki and T. Tokuda*, 2008.
6. Mörzinger R. Thallinger G. Sorschag, R. Automatic region of interest detection in tagged images. *Proceedings of the IEEE International Conference on Multimedia and Expo, 1st Workshop on Internet Multimedia Search and Mining*, 2009.
7. Vincelette R. Tameze C. Válek L. Zeljkovic V., Praks P. Pattern classification of real metallographic images. *IEEE Industry Applications Society Annual Meeting, IAS 2009, Houston, TX, USA*, 2009.

Identification of Severity Scores of Atopy Investigation by Logistic Regression

Pavlína Kuráňová

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavlina.kuranova@vsb.cz

Abstract. This paper presents the use of logistic regression for identification of serious diseases influencing the results of Phadiatop test. Our data come from patients who underwent Phadiatop test in the University Hospital of Ostrava, The Clinic of Occupational and Preventive Medicine. In this model, we focus on personal anamneses of each patient and we create a model on the basis of symptoms and severity of diseases (asthma, allergic rhinitis, eczema and others). In addition, we also include the effects of smoking on the results of Phadiatop test. The aim of the analysis is to predict the results of the expensive Phadiatop test, which is the categorical dependent variable of the model. The developed model predicts the results of the Phadiatop test with 75% or even higher probability of success, which may represent significant financial savings for the diagnosis of atopy.

Keywords: Logistic regression, atopy, Phadiatop test, medical data.

1 Introduction

Professional skin diseases belong among the most common occupational diseases. Important group is created by allergic contact dermatosses. In the developed countries, among which also the Czech Republic belongs, the number of atopy patients, who have genetic predispositions to allergy illnesses, is increasing. A higher occurrence of atopy and allergy is among children and youngsters in comparison with older generation. These individuals are nowadays coming to a work environment with the occurrence of potential contact allergens. These individuals are more sensitive to typical symptoms of asthma, eczema, etc. The Phadiatop test is used as a measure of atopy.

Phadiatop test results are divided into the six following groups: groups 0 and I indicate none or weak form of atopy and the remaining groups (II, III, IV, V, VI) indicate increasing severe forms of atopic symptoms. Unfortunately, the Phadiatop test is expensive, so we try to predict the results of the test on the basis of a detailed family

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 543–548.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

and personal anamneses. The knowledge of the results of the Phadiatop test is very important especially for the diagnosis of allergic dermatosses and also for the professional medical care for travelers [1], [2].

Information obtained from personal and family anamneses were used for the presence of asthma, allergic rhinitis, eczema or other forms of allergy (contact allergy, food, etc.) and for the smoke status for each patient. Family and personal anamneses of each patient were evaluated. Then, we created and verified a mathematical model for the accurate classification of the patients into one of the two groups: Group 0 for Phadiatop test results 0 and I, Group 1 for Phadiatop test results II, III, IV, V and VI.

This paper presents an update version of [4], [5]. In this paper, we added the smoke status and provided the analysis on the basis of complete personal anamnesis only.

2 Discrimination Problem

A common problem is to classify objects into one of the two given groups. Each object is described by attributes. The aim of the task is to assign a new object into one of the groups. Assume that the object belongs to one of the two groups (labeled as 0 and 1). The discriminatory problem will be solved on the basis of the logistic regression model.

Generally, we have n objects with p measured attributes. But in case of some of the objects, we do not know whether the object is a member of the group. The measured attributes are represented as p-dimensional random vectors $X_1, \ldots, X_n$.

The classification of the i-th object is expressed by random variable Y_i, which has the value 0 or 1 depending on their membership in the group.

The logistic regression was not originally created for the purpose of discrimination, but it can be successfully applied for this kind of analyses [3], [6].

3 Problem Solutions

The tested biomedical data are taken from the University Hospital of Ostrava, Department of Occupational and Preventive Medicine, Ostrava, the Czech Republic. The logistic regression is used in order to predict the results of the Phadiatop test. The medical database for predicted Phadiatop test contained the information on 1027 patients.

Patients in Group 0 have Phadiatop test results 0 or I (no visible symptoms), so no treatment was necessary. The remaining patients with Phadiatop test results II – VI are members of Group 1. For these patients, a medical treatment is necessary.

We have one dependent variable Y, Phadiatop (*Ph*), which depends on two independent variable personal anamnesis (*OA*) and family anamnesis (*RA*). Variable Y can carry the value 0 or 1, according to the membership of a patient in Group 0 or Group 1. Values of these independent factors were obtained from medical experts. The expert severity scores for personal and family anamneses are presented in Table 1. Here, the category of “Others” represents the score of various kinds of allergies (food allergies, etc).

Table 1. Expert severity scores for diseases.

Factor	Score
Asthma	10
Allergic rhinitis	8
Eczema	6
Others	4

4 Predictions of the Results of Phadiatop Test

In the first step, we tried to create a regression model by assuming all the supplied data. Then, the developed model was tested using the same data. We used an earlier model for the variables of OA (personal anamnesis) and RA (family anamnesis) [4]. We used the data corresponding to all 1027 patients. We obtained the following logistic model:

$$\ln\left(\frac{Ph}{1-Ph}\right) = -1.54347 + 0.212376 \cdot OA + 0.0146104 \cdot RA$$

Prediction results of the model are summarized in Table 4, column Case A.

We evaluated coefficients of OA and RA using the Pearson Chi-squared test, see Table 2. Variable OA (personal anamnesis) is statistically significant at 95% confidence level, but RA variable (family anamnesis) is larger than 0.05. Thus, RA variable is not statistically significant and may be excluded from the model.

Table 2. Test of statistical significance.

Factor	Chi-Square	Df	P-Value
OA	189,544	1	0,0000
RA	1,2909	1	0,2559

4.1 New Access to the Model

On the basis of previous results we tried to exclude variable RA (family anamnesis) and we created a new model on the basis of the evaluated individual symptoms of atopy. Our new model still has one dependent variable Ph- result of Phadiatop test, but this time it has 4 independent variables: asthma, allergic rhinitis, eczema and others.

Example of the input database is shown in Table 3. Column “Number” denotes the number of patients. In other columns; symbol 1 denotes that the patient has specific symptoms and symbol 0 denotes that the patient is without these symptoms. The column PhTest includes the results of the real Phadiatop test. The column Ph includes the results of models (in two groups) of the Phadiatop test.

Table 3. Example of the input medical database.

Number	asthma	allegic rhinitis	Eczma	others	smoke	PhTest	Ph
1020	0	1	1	1	1	0	0
1021	0	1	1	1	0	2	1
1022	0	0	0	1	1	2	1
1023	0	0	0	1	0	5	1

On the basis of medical database we tried to create even better and more precise model including the effect of smoking on a patient with atopy. In our statistical file, the variable smoke was divided as value 0 for non-smokers, value 1 for smokers and value 2 for ex-smokers or occasional smokers.

We obtained the following model:

$$\ln\left(\frac{Ph}{1-Ph}\right) = -1.6327 + 0.1537 \cdot asthma + 0.3202 \cdot allergic\ rhinitis + 0.0815 \cdot eczema + \\ + 0.1516 \cdot others + 0.3221 \cdot smoke$$

Prediction results of the model are summarized in Table 4, column Case B.
This more precise model gives good results for classifying of the patients into one of the two Phadiatop groups.

In the second step, we created a learning group as a random sample from 90% of the database (926 patients). To verify the correctness of the model assumptions, the logistic model was created using the learning group. We obtained the following updated model:

$$\ln\left(\frac{Ph}{1-Ph}\right) = -1.727 + 0.1554 \cdot asthma + 0.3312 \cdot allergic\ rhinitis + 0.1093 \cdot eczema + \\ + 0.1257 \cdot others + 0.5095 \cdot smoke$$

For testing this updated model, we analyzed the remaining data set, i.e. 10% of the database (102 patients), which were not assumed in the training phase.
Prediction results of the model are summarized in Table 4, column Case C.
In summary, the results of Phadiatop test were incorrectly predicted for 23 patients, which gave us the following prediction error rate of the model: $\frac{23}{101} = 0{,}227$.

Table 4. Predictions results of logistic regression models.

	Case A	Case B	Case C
Correctly classified patients	807	807	78
Incorrectly classified patients	220	220	23
Patients predicted for Group 1	233	209	20
Real number of patients in Group 1	331	331	33

Table 5. Analysis of variance.

Source	Deviance	Df	P_value
Model	235,662	5	0,0000
Residual	922,502	917	0,4428
Total	1158,16	922	

Table 6. Test of statistical significance.

Factor	Chi-Square	Df	P-Value
Asthma	24,909	1	0,0000
Allergic rhinitis	5,424	1	0,0199
Eczema	5,030	1	0,0249
Others	175,668	1	0,0000
Smoke	9,328	1	0,0023

We also provided the analysis of a variance for the last model, see Table 5. Since p-value is less than 0.01, there is a statistically significant relationship between variables at 99% confidence level.

We evaluated all coefficients using Person Chi-square test, see Table 6. P-value for every coefficient is less than 0.05. All these variables are statistically significant and cannot be excluded from the model.

5 Conclusions

The Phadiatop test is an expensive medical procedure. For this reason, it would be very interesting to predict the patient´s diagnosis by assuming personal anamnesis, which can be easily obtained. In order to predict the results of the Phadiatop test, biomedical database of the Clinic of Occupational and Preventive Medicine, University Hospital of Ostrava, is analyzed by the multivariate logistic regression analysis. On the basis of the previous results, a new model was used including only personal anamneses. In personal anamneses the values of asthma, allergic rhinitis, eczema, others and smoke are included. The resulting model was developed using 90% of the data and the updated model was successfully tested using remaining 10% of the data. It was statistically proved that each variable from the model is statistically significant, and cannot be excluded from the model. The simplified model predicts the results of the Phadiatop for every fourth or fifth patient incorrectly. It is a very promising premise for future biomedical analyses. In the future research, we would like to divide the results of predicted diseases severity of the Phadiatop test into more groups.

Acknowledgements: This work was supported by the FEECS VŠB – Technical University of Ostrava (Project No. SP2011/100) and also by the Ministry of Education, Youth and Sport of the Czech Republic (Project No. 1M06047).

References

1. Hajduková, Z. & Vantuchová, Y. & Klimková, P. & Makhoul, M. & Hromádka R.: Atopy in patiens with allergic contact dermatitis, Jurnal of Czech Physicians: Occupational therapy, No.2, 2009, pp. 69-73.
2. Hajduková, Z. & Pólová, J. & Kosek V.: The importance of Atopy Investigation in the Department of Travel Medicine, Allergies: Magazine for continuous education in allergy and clinical immunology, No.2, 2005, pp. 106-109.
3. Hosmer, D.W. & Lemeshow S., Applied Logistic Regression. NewYork: Wiley- Interscience, 20004.
4. Kuráňová P., Hajduková Z. & Praks P.: Logistic regression as a tool for atopy investigation. Mendel- 16th International Conference on Soft Computing 2010, pp.187-190.
5. Kuráňová P.: Logistic regression as a relevant statistical tool for medical data investigation and evaluation. SSARS 2011 – 5th Summer Safety & Reliability Seminars, Vol 1, 2011, pp.137-141
6. Rabasová M., Briš R. & Kuráňová P.: Modified logistic regression as a tool for discrimination. Reliability, Risk, and Safety: Theory and Applications. R. Briš, C. Guedes Soares, S. Martorell (eds.); Vol 3, 2010, pp.1967-1972

Massively Parallel Implementation of Total FETI

Martin Menšík, Tomáš Kozubek, and Zdeněk Dostál

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
martin.mensik.st1@vsb.cz

Abstract. We describe an efficient massively parallel implementation of our variant of the FETI type domain decomposition method called Total FETI with a lumped preconditioner. Both numerical and parallel scalability of the proposed TFETI method are demonstrated on a 2D elastostatic benchmark up to 314,505,600 unknowns and 4800 cores. The results are also important for implementation of scalable algorithms for the solution of nonlinear contact problems of elasticity by TFETI based domain decomposition method.

Keywords: domain decomposition method, FETI, parallel implementation, matrix regularization, coarse problem, scalability.

1 Introduction

Let us briefly introduce the concept of the standard domain decomposition method - FETI-1 as it is often called. The strategy here is to decompose the computational domain Ω into the set of n non-overlapping subdomains $\Omega_1, \dots, \Omega_n$ so that

$$\bar{\Omega} = \bigcup_{i=1}^{n} \bar{\Omega}_i$$

In the beginning we can assemble the subdomain stiffness matrices $\mathbf{A_i}$ and the subdomain load vectors ($\mathbf{b_i}$) totally independently and get n linear systems that can be written together as:

$$\begin{bmatrix} \mathbf{A_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A_n} \end{bmatrix} \begin{bmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u_n} \end{bmatrix} = \begin{bmatrix} \mathbf{b_1} \\ \vdots \\ \mathbf{b_n} \end{bmatrix}.$$

It is worth mentioning that not all of the local matrices are invertible. Let us gather the vectors of the null space of matrix $\mathbf{A}$ in a block matrix $\mathbf{R}$.

So far there is no coupling between the subdomains. For this reason we introduce a constraint matrix $\mathbf{B}$ (also called "Jump matrix") to enforce the continuity of the solution $\mathbf{u}$ on the interfaces of the subdomain. It is easy to observe that the entries of B can consist of $0, -1, 1$. The "Total" variant of the

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 549–554.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

FETI-1 employs this matrix $\mathbf{B}$ also for enforcing the Dirchlet condition on the boundary. That eliminates extra effort when dealing with the rigid motions of the subdomains.

The primal formulation is then

$$\min_{\mathbf{Bu}=\mathbf{o}} \frac{1}{2}\mathbf{u}^T\mathbf{A}\mathbf{u} - \mathbf{u}^T\mathbf{b}. \tag{1}$$

Such constrained minimization can by solved by applying the duality theory of convex programming where all the constraints are enforced by the Lagrange multipliers $\boldsymbol{\lambda}$. The Lagrangian associated with problem (1) is

$$L(\mathbf{u},\boldsymbol{\lambda}) = \frac{1}{2}\mathbf{u}^\top\mathbf{A}\mathbf{u} - \mathbf{f}^\top\mathbf{u} + \boldsymbol{\lambda}^\top(\mathbf{B}\mathbf{u} - \mathbf{c}). \tag{2}$$

It is well known that solution of (1) is equivalent to the saddle point problem

$$L(\overline{\mathbf{u}},\overline{\boldsymbol{\lambda}}) = \sup_{\boldsymbol{\lambda}} \inf_{\mathbf{u}} L(\mathbf{u},\boldsymbol{\lambda}). \tag{3}$$

The solution of (3) leads to the equivalent saddle-point problem to find $(\bar{\mathbf{u}},\bar{\boldsymbol{\lambda}}) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying:

$$\mathcal{A}\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\lambda}\end{pmatrix} = \begin{pmatrix}\mathbf{f}\\ \mathbf{c}\end{pmatrix} \tag{4}$$

with

$$\mathcal{A} := \begin{pmatrix}\mathbf{A} & \mathbf{B}^\top\\ \mathbf{B} & \mathbf{O}\end{pmatrix}.$$

We suppose that (4) is uniquely solvable which is guaranteed by the following necessary and sufficient conditions :

$$Ker\,\mathbf{B}^\top = \{\mathbf{o}\}, \tag{5}$$
$$Ker\,\mathbf{A} \cap Ker\,\mathbf{B} = \{\mathbf{o}\}. \tag{6}$$

Notice that (5) is the condition on the full row-rank of $\mathbf{B}$. Let us mention that a basis of $Ker\,\mathbf{A}$ can be constructed directly using subdomain rigid body modes and assume that its vectors are columns of the matrix $\mathbf{R} \in \mathbb{R}^{n\times l}$, $l = n - rank(\mathbf{A})$. More precisely, in 2D each subdomain Ω^p is assigned three columns with the sections

$$\begin{bmatrix} -y_i & 1 & 0\\ x_i & 0 & 1\end{bmatrix}$$

and the zero matrix $\mathbf{O} \in \mathbb{R}^{2\times 3}$ associated with each vertex $V_i \in \overline{\Omega}^p$ and $V_j \notin \overline{\Omega}^p$, respectively.

The first equation in (4) is satisfied iff

$$\mathbf{f} - \mathbf{B}^\top\bar{\boldsymbol{\lambda}} \in Im\,\mathbf{A} \tag{7}$$

and

$$\bar{\mathbf{u}} = \mathbf{A}^{\dagger}(\mathbf{f} - \mathbf{B}^{\top}\bar{\boldsymbol{\lambda}}) + \mathbf{R}\bar{\boldsymbol{\alpha}} \tag{8}$$

for an appropriate $\bar{\boldsymbol{\alpha}} \in \mathbb{R}^{l}$ and arbitrary generalized inverse $\mathbf{A}^{\dagger}$ satisfying $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}$. Moreover, (7) can be equivalently written as

$$\mathbf{R}^{\top}(\mathbf{f} - \mathbf{B}^{\top}\bar{\boldsymbol{\lambda}}) = \mathbf{o}. \tag{9}$$

Further substituting (8) into the second equation in (4) we arrive at

$$-\mathbf{B}\mathbf{A}^{\dagger}\mathbf{B}^{\top}\bar{\boldsymbol{\lambda}} + \mathbf{B}\mathbf{R}\bar{\boldsymbol{\alpha}} = \mathbf{c} - \mathbf{B}\mathbf{A}^{\dagger}\mathbf{f}. \tag{10}$$

Summarizing (9) and (10) we find that the pair $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\alpha}}) \in \mathbb{R}^{m} \times \mathbb{R}^{l}$ satisfies:

$$\mathcal{S}\begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{e} \end{pmatrix}, \tag{11}$$

where

$$\mathcal{S} := \begin{pmatrix} \mathbf{B}\mathbf{A}^{\dagger}\mathbf{B}^{\top} & -\mathbf{B}\mathbf{R} \\ -\mathbf{R}^{\top}\mathbf{B}^{\top} & \mathbf{O} \end{pmatrix}$$

is the (negative) *Schur complement* of $\mathbf{A}$ in $\mathcal{A}$, $\mathbf{d} := \mathbf{B}\mathbf{A}^{\dagger}\mathbf{f} - \mathbf{c}$, and $\mathbf{e} := -\mathbf{R}^{\top}\mathbf{f}$. As both $\mathcal{S}$ and $\mathcal{A}$ are simultaneously invertible, we can compute first $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\alpha}})$ by solving (11) and then we obtain $\bar{\mathbf{u}}$ from (8). Let us note that (11) has formally the same saddle-point structure as that of (4), however, its size is considerably smaller and the diagonal block $\mathbf{B}\mathbf{A}^{\dagger}\mathbf{B}^{\top}$ is much better conditioned than $\mathbf{A}$.

Before discussing the solution method for (11) we introduce new notation

$$\mathbf{F} := \mathbf{B}\mathbf{A}^{\dagger}\mathbf{B}^{\top}, \quad \mathbf{G} := -\mathbf{R}^{\top}\mathbf{B}^{\top}$$

which changes (11) into

$$\begin{pmatrix} \mathbf{F} & \mathbf{G}^{\top} \\ \mathbf{G} & \mathbf{O} \end{pmatrix}\begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{e} \end{pmatrix}. \tag{12}$$

Now we shall split (12) using the orthogonal projector $\mathbf{P}_{\mathbf{G}}$ onto the so-called natural coarse space $Ker\,\mathbf{G}$. As (6) implies that $\mathbf{G}$ is of full row-rank, we can identify $\mathbf{P}_{\mathbf{G}}$ with the following matrix:

$$\mathbf{P}_{\mathbf{G}} := \mathbf{I} - \mathbf{Q}_{\mathbf{G}} \;\text{ and }\; \mathbf{Q}_{\mathbf{G}} := \mathbf{G}^{\top}(\mathbf{G}\mathbf{G}^{\top})^{-1}\mathbf{G},$$

with $\mathbf{Q}_{\mathbf{G}}$ being the orthogonal projector onto the image space of $\mathbf{G}^{\top}$. Applying $\mathbf{P}_{\mathbf{G}}$ on the first equation in (12) we obtain that $\bar{\boldsymbol{\lambda}}$ satisfies:

$$\mathbf{P}_{\mathbf{G}}\mathbf{F}\boldsymbol{\lambda} = \mathbf{P}_{\mathbf{G}}\mathbf{d}, \quad \mathbf{G}\boldsymbol{\lambda} = \mathbf{e}. \tag{13}$$

In order to arrange (13) as one equation on the vector space $Ker\,\mathbf{G}$ we decompose the solution $\bar{\boldsymbol{\lambda}}$ into $\bar{\boldsymbol{\lambda}}_{Im} \in Im\,\mathbf{G}^{\top}$ and $\bar{\boldsymbol{\lambda}}_{Ker} \in Ker\,\mathbf{G}$ as

$$\bar{\boldsymbol{\lambda}} = \bar{\boldsymbol{\lambda}}_{Im} + \bar{\boldsymbol{\lambda}}_{Ker}. \tag{14}$$

Since $\bar{\boldsymbol{\lambda}}_{Im}$ is easily available via

$$\bar{\boldsymbol{\lambda}}_{Im} = \mathbf{G}^\top(\mathbf{G}\mathbf{G}^\top)^{-1}\mathbf{e},$$

it remains to show how to get $\bar{\boldsymbol{\lambda}}_{Ker}$. Substituting (14) into (13) we can see that $\bar{\boldsymbol{\lambda}}_{Ker}$ satisfies:

$$\mathbf{P_G}\mathbf{F}\boldsymbol{\lambda}_{Ker} = \mathbf{P_G}(\mathbf{d} - \mathbf{F}\bar{\boldsymbol{\lambda}}_{Im}), \quad \boldsymbol{\lambda}_{Ker} \in Ker\,\mathbf{G}. \tag{15}$$

Let us note that this equation is uniquely solvable, as $\mathbf{P_G}\mathbf{F} : Ker\,\mathbf{G} \mapsto Ker\,\mathbf{G}$ is symmetric positive definite and invertible if $\mathcal{A}$ is invertible. Finally note that, if $\bar{\boldsymbol{\lambda}}$ is known, the solution component $\bar{\boldsymbol{\alpha}}$ is given by

$$\bar{\boldsymbol{\alpha}} = (\mathbf{G}\mathbf{G}^\top)^{-1}\mathbf{G}(\mathbf{d} - \mathbf{F}\bar{\boldsymbol{\lambda}}). \tag{16}$$

Action of the matrix $\mathbf{F}$ on a vector is perfectly parallelizable because of the block diagonal structure of $\mathbf{A}^\dagger$ and the block structure of $\mathbf{B}$ respecting the decomposition into subdomains. On the other hand, action of the projector $\mathbf{P_G}$ is not directly parallelizable. Let us algorithmically summarize the previous results.

Algorithmic scheme

Step 1.a: Assemble $\mathbf{G} := -\mathbf{R}^\top\mathbf{B}^\top$, $\mathbf{d} := \mathbf{B}\mathbf{A}^\dagger\mathbf{f} - \mathbf{c}$, and $\mathbf{e} := -\mathbf{R}^\top\mathbf{f}$.
Step 1.b: Compute $\bar{\boldsymbol{\lambda}}_{Im} := \mathbf{G}^\top(\mathbf{G}\mathbf{G}^\top)^{-1}\mathbf{e}$.
Step 1.c: Assemble $\tilde{\mathbf{d}} := \mathbf{d} - \mathbf{F}\bar{\boldsymbol{\lambda}}_{Im}$.
Step 1.d: Compute $\bar{\boldsymbol{\lambda}}_{Ker}$ by solving $\mathbf{P_G}\mathbf{F}\boldsymbol{\lambda}_{Ker} = \mathbf{P_G}\tilde{\mathbf{d}}$ on $Ker\,\mathbf{G}$.
Step 1.e: Assemble $\bar{\boldsymbol{\lambda}} := \bar{\boldsymbol{\lambda}}_{Im} + \bar{\boldsymbol{\lambda}}_{Ker}$.
Step 2: Compute $\bar{\boldsymbol{\alpha}} := (\mathbf{G}\mathbf{G}^\top)^{-1}\mathbf{G}(\mathbf{d} - \mathbf{F}\bar{\boldsymbol{\lambda}})$.
Step 3: Compute $\bar{\mathbf{u}} := \mathbf{A}^\dagger(\mathbf{f} - \mathbf{B}^\top\bar{\boldsymbol{\lambda}}) + \mathbf{R}\bar{\boldsymbol{\alpha}}$.

Finally, we introduce the projected conjugate gradient method with preconditioning (PCGP) that we use for computing $\bar{\boldsymbol{\lambda}}_{Ker}$ in Step 1.d of Algorithmic scheme. Thus we want to compute $\bar{\boldsymbol{\lambda}}_{Ker}$ by solving the system $\mathbf{P_G}\mathbf{F}\boldsymbol{\lambda}_{Ker} = \mathbf{P_G}\tilde{\mathbf{d}}$ on $Ker\,\mathbf{G}$ with the cheap lumped preconditioner $\overline{\mathbf{F}^{-1}}$ to $\mathbf{F}$. Let us note that for the orthonormal matrix $\mathbf{B}$ we have

$$\overline{\mathbf{F}^{-1}} = \mathbf{B}\mathbf{A}\mathbf{B}^\top.$$

2 Numerical experiments

Described algorithms were implemented in standard numerics parallelization tool PETSc (Portable Extensible Toolkit for Scientific Computation), version 3.1.014, developed by Argonne National Laboratory [5] and tested on the solution of 2D linear elasticity problems. We varied decomposition and discretization parameters in order to demonstrate the scalability of our method. For these computations we used *HECToR* phase 2b system at EPCC in Edinburgh [4] offering 1856 XE6 compute nodes. Each compute node contains two AMD 2.1 GHz 12-core processors giving a total of 44,544 cores. Theoretical peak performance is around 373 Tflops. There is presently 32 GB of main memory available per node, which is shared between its 24 cores, the total memory is 58 TB. The processors are connected with a high-bandwidth interconnect using Cray Gemini communication chips. The Gemini chips are arranged on a 3 dimensional torus.

As a benchmark we used a 2D elastostatic problem of the steel traverse represented by a domain $\Omega = (0, 600) \times (0, 200)$ with the sizes given in [mm] (see Fig. 1). The material constants are defined by the Young modulus $E = 2.1 \cdot 10^5$ [MPa], the Poisson ratio $\nu = 0.3$, and the density $\rho = 7.85 \cdot 10^{-9}$ [ton/mm^3]. The body is fixed in all directions along the left side $\overline{\Gamma}_U = \{0\} \times [0, 200]$ and loaded by gravitational forces with the gravity acceleration $g = 9800\text{mm/s}^2$.

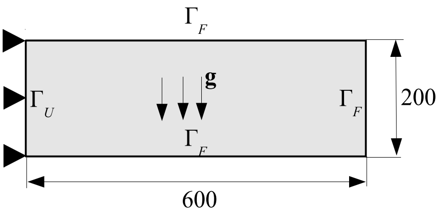


Fig. 1. Benchmark geometry

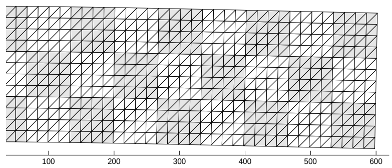


Fig. 2. Displacements with traces of decomposition (scaled 6000x)

To illustrate both the efficiency of the different strategies of the coarse problem solution and the scalability of the Total FETI we decomposed the body Ω into identical square subdomains with the side length H (see Fig. 2). We gradually chose decompositions into 8×24, 16×48, 24×72, 32×96, and 40×120 boxes. All subdomains were further discretized by the same uniform triangular meshes characterized by the discretization parameter h and the ratio $H/h = 180$. An example of the deformed body together with the traces of decomposition for the choice of parameters $h = 16.7\text{mm}$ and $H = 66.7\text{mm}$ is depicted in Fig. 2.

The Total FETI scalability results are summarized in Table 1. We can observe that the number of matrix–vector multiplications stays constant and solution and preprocessing times increase only moderately in agreement with the theory.

3 Conclusions and goals

We described an efficient massively parallel implementation of our Total FETI method. We highlighted all important implementation ingredients. Both the numerical and parallel scalability were illustrated on the solution of 2D elastostatic model benchmark up to 315 millions of unknowns and 4800 cores. The scalability results indicate that the algorithm may be useful for large size engineering problems.

Table 1. Performance of the Total FETI implementation for varying decomposition and discretization

Number of subdom.	192	768	1728	3072	4800
Number of cores	192	768	1728	3072	4800
Primal variables	12,580,224	50,320,896	114,476,544	201,283,584	314,505,600
Dual variables	129,984	537,216	1,228,464	2,183,424	3,422,400
Kernel dimension	576	2304	5184	9216	14,400
PCGP iterations	42	42	42	42	42
Preprocessing time	66.80468	68.03465	71.769	73.57246	78.20962
Solution time	26.7946	27.2203	28.9077	31.7057	40.5153
Total time	93.5992	95.2549	100.6767	105.2782	118.7249

Acknowledgments

This paper is an extract from the original paper [1].

The work was carried out under the grants GA CR 101/08/0574 and the Ministry of Education of the Czech Republic No. MSM6198910027, the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission - Capacities Area - Research Infrastructures. This work made use of the facilities of HECToR, the UK's national high-performance computing service, which is provided by UoE HPCx Ltd at the University of Edinburgh, Cray Inc and NAG Ltd, and funded by the Office of Science and Technology through EPSRC's High End Computing Programme.

References

1. T. Kozubek, V. Vondrak, M. Mensik, D. Horak, Z. Dostal, V. Hapla, P. Kabelikova, M. Cermak, "Total FETI domain decomposition method and its massively parallel implementation", to appear in Advances in Engineering Software – PARENG 2011 Special Issue, Elsevier, submitted
2. T. Brzobohatý, Z. Dostál, P. Kovář, T. Kozubek, A. Markopoulos, "Cholesky decomposition with fixing nodes to stable evaluation of a generalized inverse of the stiffness matrix of a floating structure", IJNME, DOI: 10.1002/nme.3187
3. Z. Dostál, T. Kozubek, A. Markopoulos, M. Mensik, "Cholesky decdomposition and a generalized inverse of the stiffness matrix of a floating structure with known null space", *Applied Mathematics and Computation* 2011; **217**:6067–6077.
4. HECToR Home Page, `http://www.hector.ac.uk/`
5. S. Balay, J. Brown, K. Buschelman, W.D. Gropp, D. Kaushik, M.G. Knepley, L. Curfman McInnes, B.F. Smith, H. Zhang", PETSc Web page, http://www.mcs.anl.gov/petsc, 2011.
6. M. Menšik: "The Inexact Solution of the Primal Part in the Domain Decomposition Method", in Proceedings of the Second International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering, P. Ivnyi and B.H.V. Topping, (Editors), Civil-Comp Press, Stirlingshire, United Kingdom, paper 2, ISBN 978-1-905088-44-7, 2011.

Efficient Solvers for Linear Elasticity Problems Based on the Fictitious Domain Approach

Lukas Mocek

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.mocek@vsb.cz

Abstract The paper deals with the numerical solution of elliptic boundary value problems for 2D linear elasticity using the fictitious domain method in combination with the discrete Fourier transform and the FETI domain decomposition. We briefly mention the theoretical background of these methods, introduce resulting solvers, and demonstrate their efficiency on model benchmarks.

Key words: linear elasticity, fictitious domain method, saddle-point system, discrete Fourier transform, Total-FETI decomposition, Schur complement reduction, orthogonal projector

1 Introduction

In this paper we introduce two fast solvers based on the discrete Fourier transform and the FETI domain decomposition to find a solution of elliptic boundary value problem arising in linear elasticity using the fictitious domain method.

The idea of the fictitious domain method is as follows. A boundary value problem formulated in a domain ω is replaced by a new one defined in a domain Ω having a simple shape (for example a box) and domain ω is embedded to Ω, see Fig. 1. The new problem in Ω is defined in such a way that its solution when restricted to ω matches with the solution of the original problem. The domain Ω is called fictitious domain. A possible way to formulate the new problem in Ω is

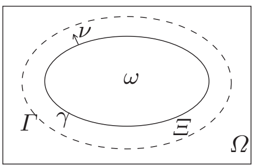


Figure 1. FDM

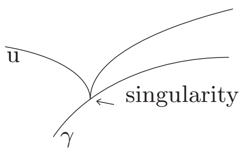


Figure 2. Singularity on γ

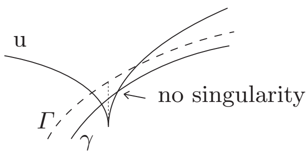


Figure 3. No singularity on γ

based on the use of Lagrange multipliers. The imposed conditions on boundary γ

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 555–560.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

of the original domain ω can be viewed as a constraint. This constraint is enforced by Lagrange multipliers defined on γ. Thus the new formulation in Ω involves two unknowns introduced as the primal variable $u \in V$ and the corresponding Lagrange multiplier $\lambda \in \Lambda$ enforcing imposed boundary conditions on γ which leads to the singularity of u concentrated on γ, see Fig. 2. To improve the convergence rate a modified (smooth) approach was proposed [4]. The idea is to move singularity away from γ, introducing new auxiliary boundary Γ, where the new controls are defined to imposed boundary conditions on γ, see Fig. 1 and 3.

The goal is to find a pair $(u, \lambda) \in \mathbb{R}^{2n} \times \mathbb{R}^{2m}$ that solves a linear system of algebraic equations resulting from the finite element discretization of fictitious domain formulation of the given problem. This system is called the generalized saddle-point system:

$$\left(\begin{array}{c|c} A & B_1^T \\ \hline B_2 & 0 \end{array}\right) \left(\begin{array}{c} u \\ \hline \lambda \end{array}\right) = \left(\begin{array}{c} f \\ \hline g \end{array}\right), \tag{1}$$

where the $(2n \times 2n)$ diagonal block A is singular, the $(2m \times 2n)$ off-diagonal blocks B_1, B_2 are highly sparse, they have full row-ranks and the vectors f and g are of order $2n$ and $2m$, respectively. Due to the structure of matrices we focus on methods based on the Schur complement reduction for solving (1).

The first solver introduced in this paper is based on the Schur complement reduction and the modified fictitious domain formulation and the discrete Fourier transform which is used to the spectral decomposition of the stiffness matrix A [4,7,6]. From that reason entries of B_1 and B_2 are determined by the geometries of Γ and γ, respectively, they are highly sparse and $B_1 \neq B_2$. The second solver which is introduced later is based on Total-FETI domain decomposition [5]. In that case we consider classical fictitious domain formulation of the given problem, it means that A is singular, symmetric positive semidefinite and $B_1 = B_2$.

2 Formulation of linear elasticity problem

Let us consider an elastic body which is represented by a domain $\omega \subset \mathbb{R}^2$ with Lipschitz continuous boundary γ. This boundary is devided into two disjoint parts γ_u and γ_p. The zero displacement is imposed on γ_u while the surface tractions of density $p \in (L^2(\gamma_p))^2$ on γ_p. We finally prescribe the forces of density $f|_\omega$ where $f \in (L^2_{loc}(\mathbb{R}^2))^2$ in ω. We can formulate the equilibrium equations as:

$$\left.\begin{aligned} -div\,\sigma(u) &= f \quad \text{in} \quad \omega, \\ u &= 0 \quad \text{on} \quad \gamma_u, \\ \sigma(u)\nu &= p \quad \text{on} \quad \gamma_p, \end{aligned}\right\} \tag{2}$$

where $\sigma(u)$ is the stress tensor in ω and $\nu = (\nu_1, \nu_2)$ is the unit outward normal vector to γ. The stress tensor $\sigma(u) := \lambda\, tr(\varepsilon(u))I + 2\mu\, \varepsilon(u)$, where $\varepsilon(u) := 1/2(\nabla u + \nabla^T u)$ and "tr" denotes the trace of matrices, $I \in \mathbb{R}^{2\times 2}$ is the identity matrix and $\lambda, \mu > 0$ are the Lamè constants.

3 Solver for linear elasticity problem based on the Schur complement reduction and DFT

The discretization of the modified fictitious domain formulation based on the finite element method leads to the algebraic saddle point system (1), where $A \in R^{2n\times 2n}$ is the stiffness matrix, the matrices $B_1 = B_\Gamma \in R^{2m\times 2n}$ and $B_2 = B_\gamma = (B_{\gamma_u}, C_{\gamma_p})^T \in R^{2m\times 2n}$ are determined by geometries of Γ and γ, respectively, and by the imposed boundary conditions, they have full row-ranks and also they are highly sparse. The vectors f and g are given by $f \in R^{2n}$ and $g = (0^T, p^T)^T \in R^{2m}$, respectively.

To solve (1) we can use an algorithm based on the combination of the Schur complement reduction and the discrete Fourier transform. From the reason that the stiffness matrix A is singular, the first component u of (1) cannot be completely eliminated. It follows that the Schur complement reduction leads to another algebraic system with two unknowns. The first uknown λ from the previous saddle point system and new unknown α, which corresponds to the null space of A. We can formulate this new algebraic system with unkowns (λ, α):

$$\begin{pmatrix} F & G_1^T \\ G_2 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} d \\ e \end{pmatrix}, \tag{3}$$

where $F := B_\gamma A^\dagger B_\Gamma^T$, $G_1 := -N^T B_\gamma^T$, $G_2 := -N^T B_\Gamma^T$, $d := B_\gamma A^\dagger f - g$, $e := -N^T f$ and the first unknown u of the algebraic system (1) is given as $u = A^\dagger(f - B_\Gamma^T \lambda) + N\alpha$. Now we define two orthogonal projectors P_1 and P_2 onto the null spaces of G_1 and G_2, respectively. The first projector splits the saddle-point algebraic structure of the reduced system, the second projector decomposes the unknown $\lambda \in R^{2m}$ into two components λ_{Im} and λ_{Ker} as $\lambda := \lambda_{Im} + \lambda_{Ker}$, where λ_{Im} belongs to the range space of G_2^T and λ_{Ker} belongs to the null space of G_2. Then λ is the first component of the solution to the algebraic system (3) if $\lambda_{Im} = G_2^T(G_2 G_2^T)^{-1} e$ and λ_{Ker} satisfies the following equation $P_1 F \lambda_{Ker} = P_1(d - F\lambda_{Im})$. The component λ_{Ker} is solved by a projected Krylov subspace method for non-symmetric operators such as BICGSTAB(see [4]). Finally the second component of the algebraic system (3) is given by $\alpha = (G_1 G_1^T)^{-1} G_1(d - F\lambda)$.

For $V = H^1_{per}(\Omega)$ containing functions with periodic boundary condition on $\partial\Omega$, the matrix A is singular but the advantage is that A has a block circulant structure which allows to use the highly efficient solver based on the discrete Fourier transform (DFT). We can use DFT for the spectral decomposition of the stiffness matrix A [6] and after that easily evaluate $A^\dagger y$ by Fast Fourier transform without storing A and it is big advantage against other competative solvers. We denote $A^\dagger$ as a generalized inverse of A and $y \in R^{2n}$. More details you can find in [4].

3.1 Numerical example

Let us define the domain ω as interior of the circle $\omega = \{(x, y) \in R^2 | (x - 0.5)^2 + (y - 0.5)^2 < 0.25^2\}$, which is embeded into the fictitious domain $\Omega = (0, 1) \times (0, 1)$

(see Fig. 2(a)). The righthand sides of (2) are $f = -div\,\sigma(\hat{u})$, $c = \hat{u}|_{\gamma_u}$ for $u = c$ on γ_u, and $p = \sigma(\hat{u})\nu$, where $\hat{u}(x, y) = (0.1xy, 0.1xy)$, $(x, y) \in R^2$. The auxiliary boundary Γ is constructed by shifting γ $6h$ units in the direction of outward normal vector ν, where h is the discretization norm. Computational results are depicted in Fig. 4(b) and reported in Table 1.

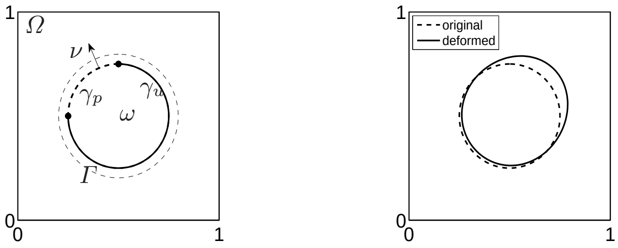


Figure 4. (a) Geometry of ω (b) Original and deformed geometry

Table 1. Computational results

Step h	prim./dual.	Iter	Time(s)	$E_{rel,(L_2(\omega))^2}$	$E_{rel,(H^1(\omega))^2}$
1/64	4225/16	14	0.156	5.6346e-004	8.2138e-001
1/128	16641/28	32	0.624	1.0526e-004	3.4955e-001
1/256	66049/48	70	4.181	3.9067e-005	2.1574e-001
1/512	263169/92	171	54.66	8.4399e-006	1.0043e-001
1/1024	1050625/160	472	623.6	5.9328e-006	8.4156e-002

4 Solver for linear elasticity problem based on Total-FETI

This section deals with the numerical solution of the linear elasticity problem by the FETI based domain decomposition method [5] in combination with the classical fictitious domain approach. The main idea of this method is decomposing of the computational domain into non-overlapping subdomains. These subdomains are glued together again by Lagrange multipliers.

We focus our attention on our new variant of FETI called Total FETI. The main idea is to simplify work with pseudoinversions of the subdomain stiffness matrices by using Lagrange multipliers not only for gluing interconnected subdomains but also for enforcing the Dirichlet boundary conditions [5], see Fig. 5(a) and 5(b). First we decompose given domain Ω into s subdomains Ω^p,

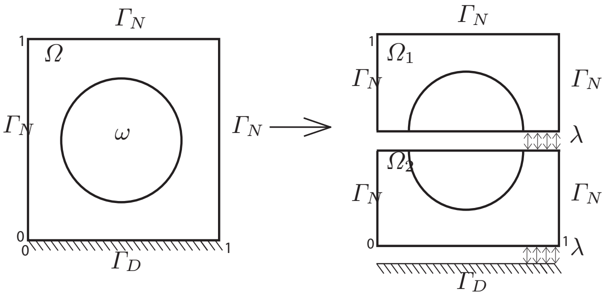


Figure 5. (a)Domain Ω(b)Decomposition of Ω

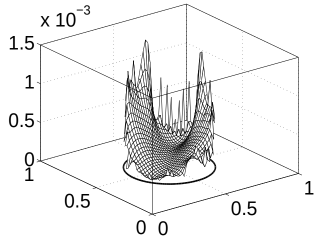


Figure 6. Differ. $|\hat{u}_h - \hat{u}|$ in ω

$p = 1, \ldots, s$. After finite element discretization of $\Omega = \Omega^1 \cup \cdots \cup \Omega^s$ we get the following quadratic programming problem

$$\min \frac{1}{2} u^T K u - u^T f \quad \text{subject to } Bu = c, \tag{4}$$

where $K = \mathrm{diag}(K_1, ..., K_s)$ denotes a symmetric positive semidefinite block-diagonal matrix of order $2n$, B denotes an $2m \times 2n$ full-row rank matrix, $f \in \mathbb{R}^{2n}$ and $c \in \mathbb{R}^{2m}$. The diagonal blocks K_p that correspond to the subdomains Ω^p are positive semidefinite sparse stiffness matrices with a-priori known kernels. The matrix B with its rows b_i and the vector c with the entries c_i enforce the prescribed displacements on the part of the boundary with imposed Dirichlet boundary conditions and the continuity of the displacements across the auxiliary interfaces. The continuity requires that $b_i u = c_i = 0$, where b_i are vectors of the order $2n$ with zero entries except 1 and -1 at appropriate positions. The matrix B is called gluing matrix.

The problem (4) is a quadratic problem, which can be hardly numerically computed, because the stiffness matrix K is usually ill-conditioned and singular. To eliminate this drawback we use its dual formulation in terms of Lagrange multipliers and we get

$$\min \frac{1}{2} \lambda^T F \lambda - \lambda^T d \quad \text{subject to } G\lambda = e, \tag{5}$$

where $F := BK^\dagger B^T$, $G := -R^T B^T$, $d := BK^\dagger f - c$, $e := -R^T f$. For solving this problem, we can also use orthogonal projectors as in previous approach. We define the orthogonal projector P onto the kernel of G as $P = I - G^T(GG^T)^{-1}G$ and the problem (5) is equivalent to the problem

$$\min \frac{1}{2} \lambda^T PFP\lambda - \lambda^T P(d - F\lambda_{Im}) \quad \text{subject to } G\lambda = 0, \tag{6}$$

where λ_{Im} is a particular solution of $G\lambda = e$ used to homogenize this equation. The problem (6) is solved by the projected conjugate gradient method with preconditioning (ProjCG) [5].

4.1 Numerical example

Let us define the linear elasticity problem

$$\left.\begin{aligned} -div\,\sigma(u) &= f \quad \text{in} \quad \omega, \\ u &= c \quad \text{on} \quad \partial\omega, \end{aligned}\right\} \tag{7}$$

where the righthand sides of (7) are $f = -div\,\sigma(\hat{u})$, $c = \hat{u}|_{\partial\omega} \in \mathbb{R}^2$, and $\hat{u}(x,y) = (0.1xy, 0.1xy)$, $(x,y) \in R^2$. Imposed Dirichlet boundary conditions on $\partial\omega$ are enforced by the Lagrange multipliers λ defined on $\gamma := \partial\omega$. The domain ω is defined as $\omega = \{(x,y) \in R^2 | (x-0.5)^2 + (y-0.5)^2 < 0.25^2\}$, and is embeded into the fictitious domain $\Omega = (0,1) \times (0,1)$ which is devided into two subdomains, see Fig 5(b). The corresponding relative error of an approximate solution $\hat{u}_h$ to the exact solution $\hat{u}$ is in the L_2 norm $E_{rel,(L_2(\omega))^2} = 1.6381 \cdot 10^{-2}$, see Fig. 6 for errors distribution. The goal is to extend the presented approach to 3D and general numbers of subdomains.

References

1. Dostal Z.: *Optimal quadratic programming algorithms: with applications to variational inequalities*, Springer: NY, 2009.
2. Glowinski, R.; Pan, Tsorn-Whay; Periaux, J.; *A fictitious domain method for Dirichlet problem and applications*, Computing, Volume 84, Numbers 1-2, 2008, 69-96.
3. Golub, G.H.; Van Loan, C.F.: *Matrix computation*, 3rd ed. The Johns Hopkins University Press, Baltimore 1996.
4. Haslinger, J.; Kozubek, T.; Kucera, R.; Peichl, G.; *Projected the Schur complement method for solving non-symmetric systems arising from a smooth fictitious domain approach.*. Lin. Algebra Appl., 14(2007), 713-739.
5. Kozubek, T. et al. *Total FETI domain decomposition method and its massively parallel implementation*, Computers and Structures, submitted
6. Kozubek, T.; Mocek, L.; *The use of the Fourier transform for solving linear elasticity problems*, AMMCS-2011 Conference, Waterloo, Canada, July 25 - 29, 2011, abstract accepted
7. Mocek, L.; *Fast solver based on Fourier transform for linear elasticity problem*, SNA'11 - Seminar on Numerical Analysis (2011), 83-87.

The Improvement of Projected SD for Quadratic Optimization with Separable Elliptic Constraints

Lukáš Pospíšil

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.pospisil@vsb.cz

Abstract. This article is a first step to deal with problem with recently developed algorithm MPGP (modified proportioning with gradient projections) for minimization of a quadratic function subject to separable elliptic constraints. In some cases the number of iterations of this algorithm is really big. In this paper we try to detect these cases and we make a proposal of improving it. But for good reason we get in simpler algorithm - Projected Steepest Descend method. Only two-dimensional problem is cosidered.

Keywords: Projected Steepest Descend, Quadratic programming, Elliptic Constraint, Quadratic Interpolation

1 Introduction

Problem of minimizing quadratic function subject to elliptic constraints arises in dual formulation of contact problems with given friction, i.e. the formulation in terms of normal and tangential contact stresses. For motivation see [6].

In this paper we solve only simple two-dimensional problem with convex quadratic function

$$\bar{x} \stackrel{\text{def}}{=} \arg\min_{x \in \Omega} \frac{1}{2}(Ax, x) - (b, x) \ , \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is a convex set defined by one elliptic constraint

$$\Omega \stackrel{\text{def}}{=} \{x \in \mathbb{R}^2 : x^T Bx \leq 1\}$$

with given symmetric and positive-definite matrices $A, B \in \mathbb{R}^{2,2}$ and the vector of right sides $b \in \mathbb{R}^2$.

2 Projected Steepest Descent method

We consider the simplest line-search optimizing algorithm - Projected Steepest Descend method. The algorithm is shown below.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 561–566.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Input: $A, b, x_{00}, B, \epsilon$

$k := 0$
$x_0 := P(x_{00})$
while $\|g_k^P\| \geq \epsilon$ **do**
  $x_+ := x_k - \alpha_k g_k$
  if $x_+ \in \Omega$ **then**
    $x_{k+1} := x_+$
  else
    $x_{k+1} := P(x_+)$
  end if
  $k := k + 1$
end while

Output: x_k, k

where

- $x_{00} \in \mathbb{R}^2$ is arbitrary initial aproximation,
- $P : \mathbb{R}^2 \to \Omega$ is projection onto Ω defined by

$$P(x) \stackrel{\text{def}}{=} \arg\min_{a \in \Omega} \|x - a\| \ ,$$

- $g_k = Ax_k - b$ is gradient of cost function in aproximation x_k,
- $g_k^P = P(x_k - \alpha_k g_k) - x_k$ is projected gradient,
- $\alpha_k = (g_k^T g_k)/(g_k^T A g_k)$ is coeficient of directional minimizing.

This algorithm can be found in e.g. [1] or [2].

3 Searching for the worst case

For our next observations we solve problem (1) with using particular values

$$A \stackrel{\text{def}}{=} \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}, \quad b \stackrel{\text{def}}{=} \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

and testing matrix of elliptic constraints

$$B_{\varphi,\delta} \stackrel{\text{def}}{=} R_\varphi^T B_\delta R_\varphi,$$

where

$$R_\varphi \stackrel{\text{def}}{=} \begin{bmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{bmatrix}, \quad B_\delta \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 \\ 0 & \delta \end{bmatrix}, \quad \varphi \in \langle 0, 2\pi), \delta \in \langle 1, \infty) \ .$$

Obviously, matrix $B_{\varphi,\delta}$ has spectrum $\sigma_B = \{1, \delta\}$, so we can control major and minor axes modifying parametr δ. The rotation of ellipse can by modified using parameter φ. Some examples are shown in Fig. 1.

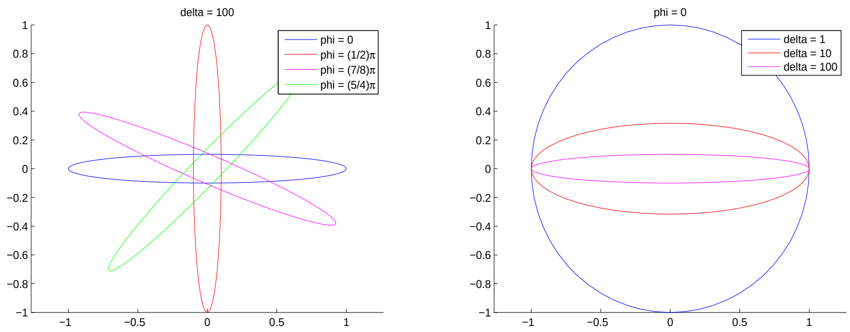


Fig. 1. Depencence of parameters in testing matrix $B_{\varphi,\delta}$

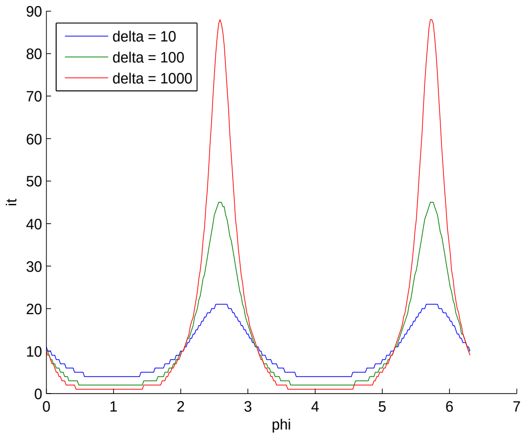


Fig. 2. Various choice of parameters in testing matrix $B_{\varphi,\delta}$

Now we study behaviour of algorithm using matrix $B_{\varphi,\delta}$ with constant $\delta \in \{10, 100, 1000\}$ and alternative $\varphi \in \langle 0, 2\pi)$. We demand accuracy of solution $\epsilon = 10^{-4}$. For out next observations, the number of iterations will be the most important. Our results are plotted in Fig. 2. It means that there exists a rotation of ellipse such tahat the number of iterations of Steepest Descend method is larger than in all other cases. This state we call "the worst case". In the next investigation, our aim is to decrease number of iterations in the worst case.
In left part of Fig. 3 is shown the progress of iterations in the worst case of the previous problem with choosing $B_{2.597,1000}$. In the right side is shown a reason why the number of iterations is so large - approximation in next iteration overjumps the solution. That means the coefficient α_k is too large and we have to use little more computation to estimate it properly.

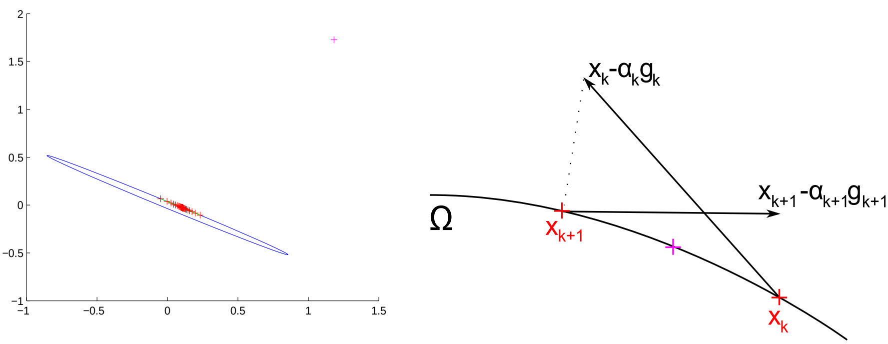


Fig. 3. Progress of iterations in the worst case

4 Quadratic interpolation

There is a question why α_k from Steepest Descend method is not the most sufficient coefficient. It is because it originally minimizes the quadratic function in direction g_k. But this is no more necessary. Let choose $x_0 = P([-1, 1]^T) \in \Omega$ (it is the left extreme point of ellipse) and plot function values of $f(x_0 + \alpha g_0)$ for various $\alpha \in \mathbb{R}_+$. We can assure that α_0 from SD really minimizes f in direction g_0. But we want to minimize f along boundary of Ω - in fact we want to choose α_k to minimize $f(P(x_0 + \alpha g_0))$. Let us plot function values of the last function for various α. Comparison of $f(x_0 + \alpha g_0)$ and $f(P(x_0 + \alpha g_0))$ can be found in Fig. 4.
To minimize $f(P(x_0 + \alpha g_0))$, we can use one of one-dimensional optimizing methods (e.g. Bisection). However, in computation of each approximation we will need to compute projection onto boundary of elliptic constraint.
There is a better idea - because $f(P(x_0+\alpha g_0))$ is quasi-quadratic, we can choose suitable three points, interpolate them with a quadratic function and this is

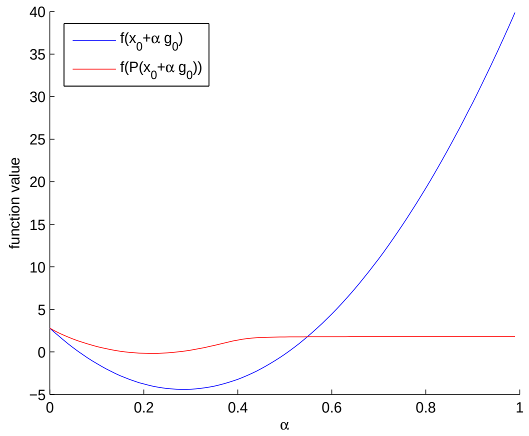


Fig. 4. Comparison of $f(x_0 + \alpha g_0)$ and $f(P(x_0 + \alpha g_0))$ for various α.

considered as an aproximation of original function $f(P(x_0 + \alpha g_0))$. Denote this approximation polynom by

$$f(P(x_0 + \alpha g_0)) \quad \approx \quad \Phi(\alpha) \stackrel{\text{def}}{=} \xi_1 \alpha^2 + \xi_2 \alpha + \xi_3 \ .$$

We can choose simply three points $0, \alpha/2, \alpha$. Then Φ has to satisfy interpolation property

$$\Phi(0) = f(x_0) \quad \wedge \quad \Phi(\frac{\alpha}{2}) = f(P(x_0 + \frac{\alpha}{2} g_0)) \quad \wedge \quad \Phi(\alpha) = f(P(x_0 + \alpha g_0)) \ ,$$

so coefficients ξ_1, ξ_2, ξ_3 can be computed as solution of system

$$\begin{aligned} \xi_1 0^2 + \xi_2 0 + \xi_3 &= f(x_0) \\ \xi_1 (\alpha/2)^2 + \xi_2 (\alpha/2) + \xi_3 &= f(P(x_0 + (\alpha/2) g_0)) \\ \xi_1 (\alpha)^2 + \xi_2 \alpha + \xi_3 &= f(P(x_0 + \alpha g_0)). \end{aligned}$$

Because $\Phi(\alpha)$ is one-dimensional convex quadratic function, we can simply derive the minimum as

$$\tilde{\alpha}_0 = -\frac{\xi_2}{2\xi_1} \ .$$

If we use this computation for every $x_k \in \partial\Omega$, we can rapidly increase a rate of convergence of algorithm as you can see in Fig. 5, where we made the same test as before but using a new coefficient. In this idea of quadratic interpolation I was inspired by [4]. For general information about interpolation and aproximation see e.g. [5].

5 Conclusion and further reseach

Obviously, this improvement with quadratic interpolation works for solving two-dimensional problems, but our goal is to solve much more comlicated problems,

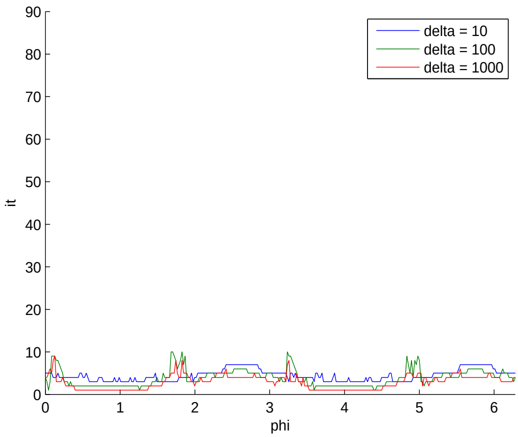


Fig. 5. Searching for the worst case using a new coefficient.

i.e. in higher dimensions, with combinations with unconstrained components or combinations with linear inequalities. For these problems we use active set method and modify the recently developed more robust MPGP [6].
Also the technical proof of convergence is missing and has to be abjected.

6 Acknowledgements

This research has been supported by the grant GA CR 201/07/0294 and the Ministry of Education of the Czech Republic No. MSM6198910027.

References

1. Pospíšil L.: *Minimizing Quadratic Function with Separated Quadratic Constraints.* Bachelor's thesis, VŠB-TU Ostrava (2008).
2. Došlý O.: *Základy konvexní analýzy a optimalizace v Rn.* MU, Brno (2005).
3. Dostál Z.: *Optimal Quadratic Programming Algorithms.* Springer Optimization and Its Applications, Vol.23 (2009).
4. Birgin E. G., Martínez J. M., Raydan M.: *Nonmonotone spectral projected gradient methods on convex set.* SIAM J. Optim., vol. 10, No. 4, p. 1196-1211.
5. Vondrák V., Pospíšil L.: *Numerické metody 1.* syllabus, project MI21, http://mi21.vsb.cz (2011).
6. Dostál Z., Kozubek T. : *An optimal algorithm with superrelaxation for minimization of a quadratic function subject to separable spherical constraints with applications.* Math. Program. Ser. A, DOI: 10.1007/s10107-011-0454-2, Springer (2011).

Using Radial-Basis Functions on Image Reconstruction

Aleš Ronovský[1] and Pavel Vlašánek[2]

[1]Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ales.ronovsky@vsb.cz
[2]Faculty of Science, Department of Informatics and Computers,
University of Ostrava, 30. dubna 22, 701 03 Ostrava
pavel.vlasanek@osu.cz

Abstract. There are lots of possibilities how to damage images or photos. Logos, scratches, noise and so on. This paper is focused on digital reconstruction of an old photograph scanned into digital image. This method of reconstruction uses RBF interpolation on damaged places in photography.

1 Image Reconstruction

Image reconstruction generally describes techniques how to get reconstructed image closest to original of damaged image or to anticipated results if we don't know original image. In principle we need to determine damaged pixels in image and then reconstruct those pixels using some reconstruction techniques.

One class of techniques of image reconstruction are methods using interpolation. Basic interpolation method is linear, which finds nearest values in x and y axis and take average. But in this paper we use more sophisticated method, interpolation method using Radial-Basis functions.

We use in this paper scan of an old photograph. On this scan are damages as scratches, cracks, stain of coffee and other mechanical damages. Our goal is to check function of our algorithm of reconstruction to these types of damages. First task was made a special mask to determine damaged pixels, i.e. pixels determined to reconstruction.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 567–572.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Radial-Basis Functions

We could write approximation using Radial-Basis functions as follows

$$s(x) = \sum_{i \in I} \lambda_i \varphi(\|x - i\|)\, x \in \mathbb{R}^d$$

where I is set of known points, $\varphi : [0, \infty) \rightarrow \mathbb{R}$ is one dimensional function and coefficients $(\lambda_i)_{i \in I}$ are real numbers [4]. We could use any norm, but as in many cases are we using Euklidean norm. Approximation, that we are searching for, is linear combination of patterns of given function. $x \rightarrow \varphi(\|x\|)$, which is circular symetric around its center. Points $(x_j)_{j=1}^n$ we call centers of interpolation. In addition, we call φ for given norm Radial-Basis function.

If the set of known points $I = (x_j)_{j=1}^n$ is definite, terms of interpolation gives us linear system

$$A\lambda = f,$$

where

$$A = (\varphi(\|x_j - x_i\|))_{i,j=1}^n \,,$$

$\lambda = (\lambda_j)_{j=1}^n$ a $f = (f_j)_{j=1}^n$.

Advantage of interpolation using Radial-Basis functions is, that to find good interpolation is needed just few known points. In many important cases of using RBF is sufficient to know just two different points. There are no other requirements for their distribution. The exception is the function TPS (thin plate spline), we define $\varphi(r) = r^2 logr$. Using this function, which we use in our algorithm often as well, we get a singular matrix A for a nontrivial set of centers. For example, when the known points will lie on the unit sphere around center. However, this cannot happen during reconstruction of 2D image, if we require more than four known points for the reconstruction of a pixel.

In addition, we can show that it is appropriate to add a polynomial of degree $m >= 1$ to the definition of the approximate function s. We get the equation

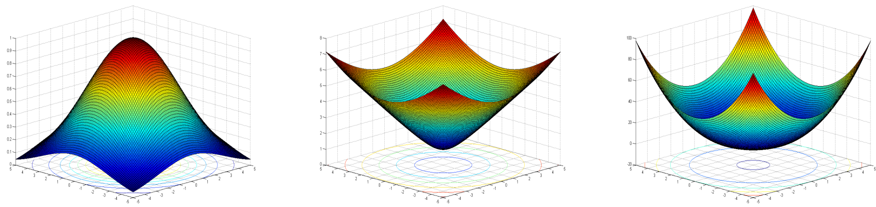
$$\sum_{k=1}^{n} \lambda_i \varphi(\|x_j - x_i\|) + P(x_j) = f_j, j = 1, 2, ..., n,$$

where x_i are known pixels and λ_i are coefficients of functions. We call polynomial P extension of system [5]. That leads to linear system [6]

$$\begin{vmatrix} A & P \\ P^t & 0 \end{vmatrix} \begin{vmatrix} \lambda \\ \gamma \end{vmatrix} = \begin{vmatrix} h \\ 0 \end{vmatrix},$$

$A_{i,j} = \varphi(r_{i,j}) = \varphi(\|x_i - x_j\|), i, j = 1, 2, ..., n, P$ is a vector of ones of length n for a constant expansion, or matrix of the coordinates of known pixels of size $n \times 3$ for linear extension and h is a vector of values of known pixels.

Selection of RBF greatly affect the results of interpolation. For various applications it is preferable to use different RBFs. E.g. Gaussian function $e^{-(\epsilon r)^2}$ is very sensitive to the choice of parameters [7], its use in practical tasks is small, but has some useful features for the theory. Other examples are the linear r, (MQ) multiquadric $\sqrt{(r^2+e^2)}$, TPS $r^2 logr$ and cubic r^3 (Fig. 1).



(a) (b) (c)

Fig. 1. (a) Gaussian function, where ϵ=0.25, (b) MQ with $\epsilon = 1$, (c) TPS.

3 Window Interpolation

For the image reconstructing we do not use global interpolation, but only local. For assume that the remote parts of image are not connected, we can reconstruct the corrupted pixel values only from values of known pixels in the neighborhood and thus greatly accelerate the progress of reconstruction. We use a floating window 5×5 pixels, where the center of the window is a pixel determined to reconstruction and from other undammaged or already reconstructed pixels we prepare a set of known pixels, which will be centers of interpolation. There are few other options which shape of the window we can use, but it has no significant iprovement than the window 5×5 (Fig. 2).

To avoid unwanted oscillations in the reconstruction, it is not wanted that known pixels near reconstructed pixel are arranged in a floating window on one side of the window and on the other side are all damaged [8]. It occurs in large contiguous damage. E.g. text, logos, etc. To prevent this, we add another pixel to the set of known pixels from the neighborhood, which we place across the damaged section. This pixel is placed into a sufficient distance so its contribution to the reconstructed pixel value was significantly smaller than the pixel values from the neighborhood, but not too far away to avoid unwanted oscillations of interpolation. If there is known pixel we use its value and if not, the value is calculated as the average value of the known pixels in the window. This will ensure a certain stability and will avoid unwanted oscillations of interpolation. This has significant impact on the quality of the resulting reconstruction.

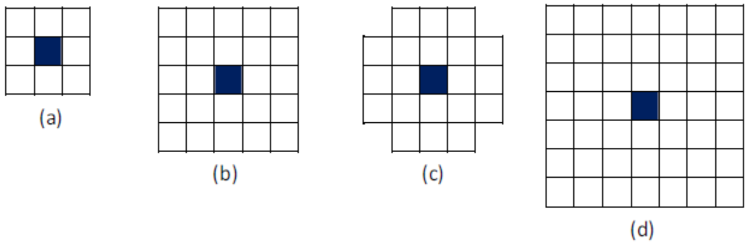


Fig. 2. (a) Window 3×3, (b) 5×5, (c) special shape of window, (d) 7×7.

4 Implementation

We have chosen the programming language C++ with SDL library [9] for our implementation. The program operates with BMP images in Microsoft windows platform.

We need damaged image for reconstruction and image which contains a mask defining bad pixels. The mask image contains black pixels representing damaged places and white ones which represents correct areas.

The program is working on CPU and assumes one computing unit and sequentional way of processing. The steps below describes way of work of our implementation:

```
while (damaged pixel exists)
{
  choose left to right walktrought on the whole image
    find damaged pixel
      if ((number of the known pixels around > c)
      && (processed pixel is not inside the hole))
        process
        insert new value to the image
        insert correction to the mask image

  choose up to down walktrought on the whole image
    find damaged pixel
      if ((number of the known pixels around > c)
      && (processed pixel is not inside the hole))
        process
        insert new value to the image
        insert correction to the mask image
}
```

We use a term hole for area which contains more damaged pixels together. In our approach we check if the side pixel of the processed pixels is damaged then we step to the end of the hole in the current line. This approach guarantees that the error should not be spread in the one direction and moreover guarantees steady of the error variance on the whole area of the holes.

5 Results

Error distribution depends on the way of the walkthrough. In our solution is error distributed to the center, which can be seen in the large holes. There are several types of typical damage. We can basically discern scratches, holes and regular shapes like text. These types of damage can be solved by different kind of a walkthrough. Concrete kind depends on the type of the damage and the main direction. Vertical damage is better reconstructed with up-down or down-up way and horizontal with left-right or right-left way.
On figures below are results of reconstruction of given photograph.

(*a*) (*b*) (*c*)

Fig. 3. (a) Part of original image, (b) with applied mask, (c) reconstruction.

Fig. 4. Reconstruction of selected parts.

6 Conclusion

Our implementation is able to reconstruct damaged image with usage of the Radial-Basis function. User has to deliver damaged image and mask which defines bad pixels. The program show results in view composed from original image, original image with applied mask, reconstructed image and error difference. For the article we have tested our implementation on the large old photo which contains typical kind of damaged areas as scratches and holes.

References

1. M. Bertalmio, G. Sapiro, C. Ballester and V. Caselles, Image inpainting, Computer Graphics, SIGGRAPH 2000, July 2000, 417–424.
2. ŽÁRA, Jiří, et al. Moderní počítačová grafika. 2. vyd. Brno : Computer Press, 2004. ISBN 80-251- 0454-0.
3. UHLÍŘ, Karel. Aplikace radiálních bázových funkcí v počítačové grafice a zpracování obrazu. Plzeň, 2007. 103 p. Dissertation thesis. Západočeská univerzita v Plzni.
4. Bradley John Charles Baxter. The interpolation theory of radial basis functions, 1992. 134 p. Dissertation thesis. Trinity College.
5. ZAPLETAL, Jiří. Aplikace Radiálních bázových funkcí. Plzeň, 2007. 91 p. Master's thesis. Západočeská univerzita v Plzni.
6. ZAPLETAL J., VANĚČEK P., SKALA V.: Influence of essential parameters on the rbf based image reconstruction. In Proceedings of SCCG 2008 (Slovakia, 2008), pp. 178–185
7. B. Fornberg, T.A. Driscoll, G. Wright, R. Charles: Observations on the behavior of radial basis functions near boundaries, Comput. Math. Appl., vol. 43. pp. 473-490, 2002.
8. Zapletal,J., Vanecek,P., Skala,V.: RBF-based Image Restoration Utilising Auxiliary Points, CGI 2009 proceedings, ACM ISBN 978-60558-687-8, pp.39-44, 2009.
9. http://www.libsdl.org

Wind Power Plant Analysis-Short Time Prediction of Electrical Power

Pavel Skalný

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavel.skalny@vsb.cz

Abstract. The theme of this paper is power prediction of wind power stations. This analysis was realized by two different methods. Discrete method, that on the basis of available data, estimates the interval in which the power, for given probability and time shift, will be found. So the method estimates future power at given period, if current performance is 0,200 (measured in kW), etc. Another method, that has been applied, is the method of exponential smoothing - one of the many time series methods by which future wind speed, that closely relates to electric power of wind power plant, was calculated. Main goal of this article is not to present a comprehensive system for power predicting, but to show some basic ways of procedure with similar problems.

1 Introduction

Wind power plants are not insignificant energy source. Unlike other types of power plants, wind farms are fundamentally dependent on meteorological conditions. The quality prediction of future power is, without any doubt, absolutely necessary. The prediction is of course based on meteorological data, which are usually predicted in hourly intervals. Even with accurate prediction of wind speed, there sre still significant fluctuations of values during that time. The goal of this paper is to quantify probability of wind farm during short time period and to Compare two different methods. One of compared methods was used in [1] the other is well known method of exponential smoothing which was many times applied for forecasting purposes. [2]-[3]. As a data source there was used a wind farm situated in Nový Jičín region between villages Veselí and Dobešov. Wind power plant consists of two wind turbines Vestas V90, 2x 2 MW rated power.. The measurements were made in 10 minutes (average values) intervals during 50 days (March 20 to May 9 of 2007). In total it is 7200 values of measured wind speed and power output for both of wind turbines. More information about WPP Veseli and other WPP analyses can be found in [4].

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 573–578.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Problem Formulation

Our goal is to calculate short time estimate of future power of wind farm. Provided that we know former wind speed and electrical power. The former information, are known in the long time period, so there is possible to make statistical analysis. The aim of the work isn't only show the possibility of short-time estimates of wind farm power, but also outlines the possibility of estimating future performance, without knowledge of weather forecasting.

3 Problem Solution

3.1 The discrete model

The function described below, which was designed in Matlab programme, can be used for computing the interval of wind power changes for probability p. The algorithm finds an interval [u, l] where 100 * p percent of the values of the given array is set. As a starting number for u and l, it is fitted to use a mean of array. This discrete model (More information about discrete modelling can be found in [5]) for predicting changes in performance was used mainly for the fact that the continuous probability distribution functions did not describe the issue ideally and for the fact that probability distribution of power varies markedly with time changes. An easy algorithm was used to find an array of values of power at given time shift. This algorithm found the new array of shifted values from interval, which was suitably chosen around former power. The interval was chosen to be smaller at low power values and to decrease with higher former power.

Analysis of estimates of future power discrete method

One of the article objectives is to realize if the discreet method as was used in the article „Wind Power Supply Stability Analysis"[1], can be used for prediction future power or if the method suits only for statistical analysis examined data. To give prominence to the problematic an experiment was realized. We used 90% of the data for discrete method analysis. On the remaining 10% of the data we found out if, is possible to predict future power. For each value (power) from tested data (10%) we found all similar values (Interval for acceptance was from 0.95*power-5 to 1.05*power+5). from these values we found all values after given time shift. On this knew array we applied the discrete method. The test was realized for probability p= 0.3, 0.6 and 0.9. In the table below test outputs are listed. The number before the slash indicates the number of successful estimates when future performance is in the predicted interval. The number after the slash indicates the total number of attempts.

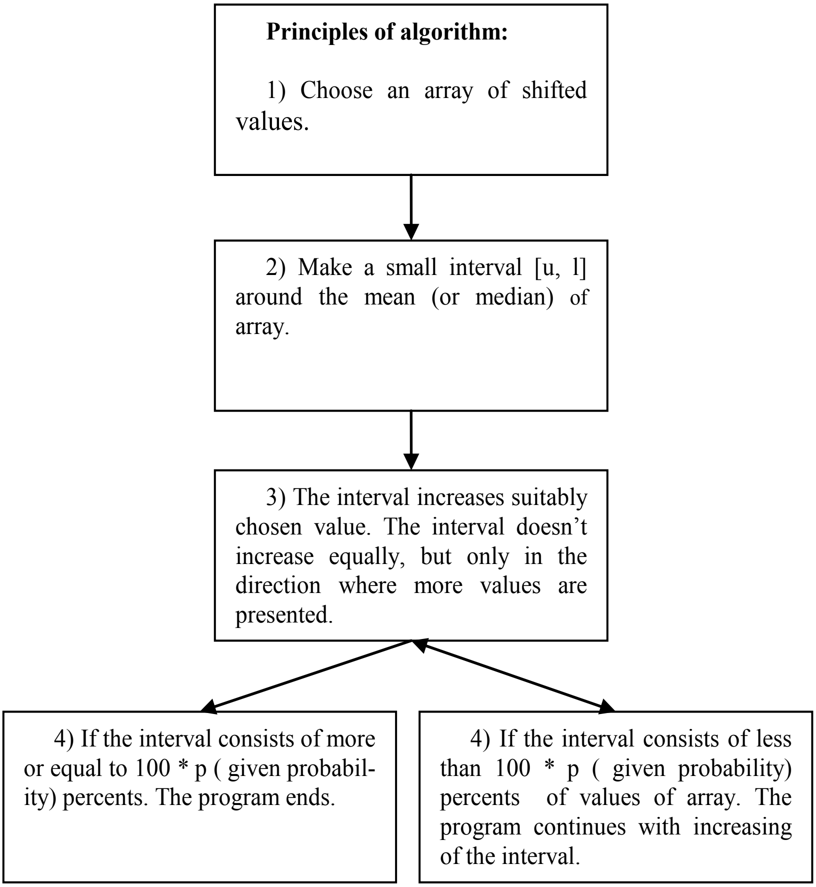


Fig. 1. Description of discrete method

10min		30min		60min	
P=0.3	243/784	P=0.3	214/782	P=0.3	189/779
P=0.6	427/784	P=0.6	394/782	P=0.6	387/779
P=0.9	593/784	P=0.9	567/782	P=0.9	562/779
90min		120min		180min	
P=0.3	178/776	P=0.3	174/773	P=0.3	215/767
P=0.6	356/776	P=0.6	371/773	P=0.6	389/767
P=0.9	545/776	P=0.9	553/773	P=0.9	542/767

Table 1.successful predictions using discrete method

As you can see in the chart above the proportion of successful attempts to all experiments is always lower than the given/set probability. And with increasing time the proportion of successful attempts as expected declines. In the chart below you can see average quantity (in KW) of estimated range for the time difference and the probability. Due to the maximum power of the equipment 2MW the average size of the estimated interval for meaningful prediction is too high. As shown in the attempt above discrete method is not suitable for predicting future power.

10min		30min		60min	
P=0.3	281	P=0.3	290	P=0.3	307
P=0.6	532	P=0.6	552	P=0.6	589
P=0.9	868	P=0.9	897	P=0.9	943
90min		120min		180min	
P=0.3	317	P=0.3	335	P=0.3	359
P=0.6	603	P=0.6	640	P=0.6	689
P=0.9	964	P=0.9	1002	P=0.9	1036

Table 2. Average quantity of estimated range

Method limitations

The program described above accounts only from statistical experience. It doesn't take into account the current trend of changes. Rather than predicting future it generates an interval of future power. To improve the prediction using the discrete method would certainly help to use more than one default output. Instead one value of power could be more effective to use sequence of powers. Risk of this approach is that is not certain that we mitt sufficiently large array of.

3.2 Exponential Smoothing

Other possibility how to predict the future power is to use appropriate time series.[6]-[7] As a time series we will understand a sequence of data, measured at uniform time interval

$$X = \{X_t : t \in T\} \quad (1)$$

where: T is a time we calculate with.

Simple Exponential Smoothing is a method which computes the forecasts of values and uses weighted average of all available previous data. The weights decline geometrically as they go back to the time.

$$s_1 = x_0 \quad (2)$$

$$s_t = \alpha\, x_{t-1} + (1-\alpha)\, s_{t-1},\ t=1,2,\ldots,n \quad (3)$$

where:

s_t: is the forecast for the value at the time period t

s_{t-1} is the forecast for the value at the time period t-1

x_{t-1} is the actual value of the series at the time period t-1

α is the smoothing constant ($0<\alpha<1$)

For multistep ahead forecast we use following formulas:

$$s_{t+l} = \alpha\, x_{t-1} + (1-\alpha)\, s_{t+l-1}, \quad (4)\ [6]$$

Method limitations

Unlike the discrete various method the exponential smoothing takes into account the current trend of changes. This fact helps to predict the future power or wind speed more precisely then the discrete method. Problem may arise if we would predict increasing series of high values (wind around 20 m/s or power around 2000kW) or decreasing series of low values (wind around 0 m/s or power around 0-20kW). In this cases time esponential smoothing doesn't take into account the fact, that probability of increasing high powers (or decreasing low powers) is smaller Fig(2). Time series takes into account only actual trend of sequence.

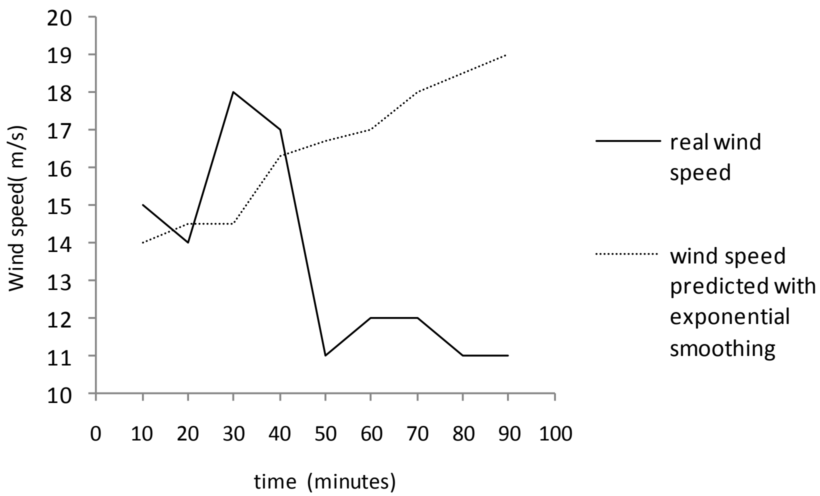


Fig.2 Strong wind error- Comparison between real wind and forecasts

3.3 Comparison

In this chapter Discrete and Exponential smoothing method's method will be compared.. Upper and lower limits were used so that the estimates for both methods were similarly large.

We use testing in Contingency Tables (see below) to find out, if the effectiveness of the forecast depends on chosen method. The test is realized for time prediction for 10, 30, 60 minutes. The null hypothesis is: "The effectiveness of the prediction does not depend on the chosen method" The alternative hypothesis:" The effectiveness of the prediction depends on the chosen method". We used chi-square test-the test of the independence in the table. In all three cases we reject the null hypothesis. the 99 % confidence level.

10 min.	S	F	30 min.	S	F	60. min.	S	F
d. m.	243	784		214	782		189	779
H.S	632	771		629	779		615	776

Table. 2. Contingency Tables for comparison discrete method with Simple exponential smoothing. S-Success, F-Fail

4 Conclusion

In this paper two methods for short time predicting of electrical power of wind farms were compared Using time series analysis is more effective than the discrete method for estimating future power. The discrete method is appropriate to quantify probability intervals during short time period, but it isn't very useful for specific prediction.

Acknowledgement: I would like to thank to Prof. Ing. Radim Briš CSc. for valuable comments and to Ing. Tadeusz Sikora Ph.D. for granting data source.

References:

[1] Sikora T, Skalný P. *Wind Power Supply Stability Analysis, to be printed in* EPE- Proceedings of 12th international scientific conference electric power engineering, vol 12, no 1 pp. 240-242, 2011

[2] Huitian Lu, Kolarik W. J, Susan S.Lu, *Real-Time performance Reliability Prediction,* IEEE Transactions on reliability, Vol 50, No. 4, December 2001, pp. 353-357.

[3] C. Nemes and F. Munteanu, *Development of reliability model forwind farm power generation,* Advances in Electrical and Computer Engineering, vol. 10, no. 2, pp. 24–29, 2010.

[4]S. Misak, L. Prokop, and J. Dvorsky, *Optimizing the mathematical model for prediction of energy production in wind power plants,* Przeglad Elektrotechniczny, vol. 87, no. 2, pp. 74–78, 2011.

[5] Meeker W., L.A. Escobar, *Statistical Methods for reliability data,* New York 1998.

[6] Lon-Mu Liu, *Time series analysis and forecasting,* Villa Park, 1996.

[7] V. Rogaliewitz , *Stochastické procesy (Analýza časových řad)*, Praha, 1993

Multilevel Solvers with Aggregations and K-Cycle

Vojtěch Sokol

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vojtech.sokol@vsb.cz

Abstract. The finite element analysis of properties of heterogeneous materials needs to deal with the ill-conditioning of the arising algebraic system caused by the heterogeneity. To overcome this difficulty we use multilevel preconditioner for CG method based on Schwarz-type method with inner K-cycle and appropriate coarse level construction.

Keywords: finite element method, Schwarz-type methods, multilevel methods, aggregation techniques, K-cycle

1 Introduction

In this paper, we present a multilevel preconditioner based on Schwarz-type method with coarse levels constructed by aggregation. The efficiency of the solver is examined on a model problem with strong heterogeneity. The model problem is discretized using finite element method. The arising algebraic system is ill-conditioned due to the strong heterogeneity present, therefore robust preconditioner with respect to heterogeneity is needed to efficiently solve the problem with CG (or other iterative) method. The preconditioner is further improved by the use of inner K-cycle on coarse levels.

2 Multilevel Iterative Methods

Multilevel iterative methods are know for their efficiency when solving regular boundary value problems. Their application to problems with material heterogeneity addresses the questions how to define the coarse level problem and how to transfer information between the levels to obtain an efficient multilevel method.

To be specific, let us consider the problem

$$Au = b, \quad u, b \in R^n \tag{1}$$

and the Algorithm 1. In the Algorithm 1, k denotes the index of level, where $k = L$ and $k = 0$ correspond to the finest and coarsest level respectively. Matrices A_k are usually obtained by the Galerkin formula $A_{k-1} = R_k A_k R_k^T$, $A_L = A$,

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 579–584.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

where R_k is a restriction matrix from the level k to a coarser level $k+1$ and R_k^T is a corresponding prolongation. In lines 2 and 6 ν_k steps of smoothing are performed by a smoother S. On the line 4 the algorithm can be used recursively when $k > 1$ and the exact solution is computed only when $k = 1$. This way we get multilevel method with $L+1$ levels. Further, one iteration of a multilevel algorithm will be used as a preconditioner.

Algorithm 1 One iteration of multilevel algorithm at level k

Input: A_k, b_k, u_k^i
Output: u_k^{i+1}
1: $u = u_k^i$
2: $u = S^{\nu_k}(A_k, b_k, u)$
3: $r_{k-1} = R_k\,(b_k - A_k u)$
4: $\tilde{u} = A_{k-1}^{-1} r_{k-1}$
5: $u = u + R_k^T \tilde{u}$
6: $u = S^{\nu_k}(A_k, b_k, u)$
7: $u_k^{i+1} = u$

In next subsections, the actual choice of the multilevel algorithm (section 2.1), restriction (section 2.2) and the K-cycle (section 2.3) will be addressed.

2.1 Multilevel Schwarz Preconditioner

The linear system arising from the finite element discretization is solved by the conjugate gradient method with multilevel additive Schwarz preconditioner. It uses decomposition of the computational domain Ω into overlapping subdomains Ω_i^δ and a coarse. This provides the following decomposition of the finite element space V_h:

$$\begin{aligned} V_h &= V_0 + V_1 + \ldots + V_k \\ V_i &= \left\{ v \in V_h, v \equiv 0 \text{ in } \Omega \setminus \Omega_i^\delta \right\}, \forall i \in \{1 \ldots k\}, \; V_0 = \mathrm{Range}(R_c^T) \end{aligned}$$

where $R_c = R_0$ is constructed by aggregation as described in subsection 2.2. If R_i are restrictions to spaces V_i then the two-level additive Schwarz preconditioner (B_{AS}) can be written as

$$B_{AS} = \sum_{i=0}^{k} R_i^T A_i^{-1} R_i,$$

where $A_i = R_i A R_i^T$ are FE matrices corresponding to problems on subdomains and on the coarse space. More details about Schwarz type preconditioners can be found e.g. in [4].

In the case of two-level Schwarz preconditioners, the system with matrix A_0 is solved directly, in the multilevel case the FE space V_0 is decomposed again and the operation $g = A_0^{-1} r$ is replaced by $g = B_{AS}^{(1)} r$, where $B_{AS}^{(l)}, l = 1$, corresponds to the Schwarz preconditioner on a coarser level. This procedure is repeated with

$l = 1, 2, \ldots$ until a desired size of the coarsest FE matrix is attained, which can be efficiently solved directly. The multilevel preconditioner on the finest level $B_{AS}^{(0)} = B_{AS}$ then forms a V-cycle multilevel method.

2.2 Aggregation Techniques

The coarse space and the arising restriction operator will be constructed by adaptive aggregation, which respects strong connections between degrees of freedom. To this end, we shall describe and test several techniques described in the literature, namely [1], [3], [2]. Note that all of the mentioned papers use a node-wise aggregation technique.

Here, we consider the aggregation technique for solving the problem (1). The aggregation divides the set of indices (nodes, degrees of freedom) $N = \{1, \ldots, n\}$ into disjoint subsets G_i of aggregates of unknowns, so that $N = \bigcup_{i=1}^{k} G_i$, $G_i \bigcap_{i \neq j} G_j = \emptyset$. Then the prolongation and restriction operators are defined by the boolean matrix R_c:

$$(R_c)_{i,j} = \begin{cases} 1 & \text{if} \quad j \in G_i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

We focus on aggregation techniques that exploit the information directly stored in the matrix A_h, these include algorithms by Vaněk et al. [3], by Scheichl and Vainikko [2] and by Notay [1].

The algorithm by Vaněk et al. starts with defining strongly-connected neighbourhood $S_i = S_i(\varepsilon)$ of a node i with threshold parameter ε:

$$S_i(\varepsilon) = \left\{ j \in N : |a_{ij}| \geq \varepsilon \sqrt{a_{ii} a_{jj}} \right\}, \tag{3}$$

and then separates nodes that are not strongly connected to any other nodes. Then main part of the algorithm can be briefly described as follows (details can be found in [3]):

step 1: form set of unaggregated nodes U, initially $U = N \setminus \{\text{isolated DOF's}\}$

step 2: for $j = 1, \ldots$ form initial aggregates G_j^0 from strong neighbourhoods, i.e. if $S_i \subset U$ then $(G_j^0 = S_i \;\&\; U = U \setminus G_j^0 \;\&\; j = j + 1)$

step 3: enlarge aggregates G_j^0 to G_i with respect to strong connection

step 4: process unaggregated nodes

The next aggregation technique is that of Scheichl and Vainikko [2]. The algorithm uses a strongly-connected graph r-neighborhood $S_i(r, \varepsilon)$ defined as follows:

$$S_i(r, \varepsilon) = S_i(r-1, \varepsilon) \cup S^+(r-1, \varepsilon))\,,\; S^+(r-1, \varepsilon)) = \bigcup_{j \in S_i(r-1,\varepsilon)} S_j(\varepsilon)\,, \tag{4}$$

$$S_i(1, \varepsilon) = S_i(\varepsilon) = \left\{ \{i\} \cup \left\{ j \in N : \left| \hat{A}_{ij} \right| \geq \varepsilon \max_{k \neq i} \left| \hat{A}_{ik} \right| \right\} \right\}, \tag{5}$$

where $\hat{A} = (\mathrm{diag} A_h)^{-\frac{1}{2}} A_h (\mathrm{diag} A_h)^{-\frac{1}{2}}$ is a scaled matrix and ε is a thresholding parameter for strong connections. The algorithm creates aggregates by finding a strongly-connected graph r-neighbourhood of a chosen seed node. To choose a good seed nodes an advancing front in the graph induced by nodes and edges of triangulation $\mathcal{T}_h$ is used.

The last aggregation algorithm by Notay [1] uses strong negative connectivity of nodes. For a node i, we define S_i as the set of all strongly negative connected nodes:

$$S_i(\varepsilon) = \left\{ j \in N : j \neq i, a_{ij} < -\varepsilon \max_{a_{ik}<0} |a_{ik}| \right\}, \tag{6}$$

where parameter ε is used as threshold for strong coupling. The sets $S_i(\varepsilon)$ are used to construct pairs of nodes that are most strongly negative connected. Recursively, the pairs can be aggregated into pairs of pairs (quadruples) etc.

2.3 K-Cycle

In Algorithm 1 the solution on the coarser level is approximated by a recursive use of the algorithm. However we can utilize a more sophisticated approach. The algorithm on the coarser level need not to be used directly but can be used as a preconditioner for a Krylov subspace method instead. In our case, the matrix on the finest level is a SPD matrix, the coarser matrices are generated by Galerkin formula, therefore we can use CG as a method of choice. More details about K-cycle its analysis can be found in [5].

3 Model Problem

The model problem on which we will test our preconditioner will be Darcy flow described by following equations:

$$\nabla \cdot v = f \, , \, v = -k \nabla u \quad \text{in} \quad \Omega = [0,2] \times [0,1] \tag{7}$$

with mixed boundary conditions: zero normal flux on the bottom and top sides and $u = 0,\ 1$ on the left and right vertical sides, respectively. The problem is discretized by linear triangular finite elements on a uniform grid.

The heterogeneity is represented by the permeability coefficient k. In our model problem, the coefficient is stochastically generated in such a way that $\log(k)$ has a normal distribution with the mean $\mu = 0$ and the variance σ^2.

4 Numerical Experiments

For the numerical simulations the domain $\Omega = [0,2] \times [0,1]$ was decomposed into stripes of width H_s and minimal overlap $\delta = h$. The permeability coefficient k was stochastically generated with the variance σ^2 set such that the coefficient jump $\frac{k_{max}}{k_{min}}$ hits the predefined value. The preconditioned CG are tested with the use of relative accuracy $\epsilon_{rel} = 10^{-6}$. The sizes of coarse spaces and the corresponding column fill-ins of coarse space matrices for various aggregations can be seen from Table 1. Table 2 shows numbers of CG iterations when number of inner Krylov iterations was set to zero (i.e. no K-cycle was used) and when number of inner Krylov iterations was set to five. All tables show mean values and standard deviations over 8 runs of the algorithm with different stochastically generated permeability field k. The reported results use aggregation parameters which produce good performance results and comparable numbers of iterations for all considered techniques.

	coarse space size			fill-in		
jump	c_1	c_2	c_3	c_1	c_2	c_3
10^0	8097	2025	510	6.3	6.8	6.7
10^4	8235	2077	532	6.6	6.7	6.6
10^8	9140	2532	720	6.4	6.6	6.3
10^{10}	9535	2775	822	6.4	6.6	6.3

(a) Notay, $\varepsilon = 0.1$, means

coarse space size			fill-in		
c_1	c_2	c_3	c_1	c_2	c_3
0.00	0.00	0.00	0.00	0.00	0.00
0.28	0.36	0.51	0.15	0.13	0.19
0.20	0.23	0.75	0.11	0.24	0.37
0.25	0.50	0.83	0.20	0.24	0.45

(b) Standard deviations (in %)

	coarse space size			fill-in		
jump	c_1	c_2	c_3	c_1	c_2	c_3
10^0	3655	242	18	6.7	6.4	4.4
10^4	8722	1555	260	6.3	6.4	5.0
10^8	12025	3595	812	6.1	6.4	5.6
10^{10}	12713	4211	1073	6.1	6.3	5.8

(c) Scheichl, $\varepsilon = 0.5, r = 2$, means

coarse space size			fill-in		
c_1	c_2	c_3	c_1	c_2	c_3
0.00	0.00	0.00	0.00	0.00	0.00
1.67	4.62	11.40	0.27	0.42	1.46
0.53	1.29	2.50	0.14	0.37	1.31
0.23	1.05	1.53	0.11	0.24	0.55

(d) Standard deviations (in %)

	coarse space size			fill-in		
jump	c_1	c_2	c_3	c_1	c_2	c_3
10^0	5471	638	85	6.9	6.7	6.2
10^4	8135	1469	209	6.3	6.7	6.3
10^8	10867	2809	500	6.2	6.7	6.4
10^{10}	11670	3345	672	6.2	6.7	6.4

(e) Vaněk, $\varepsilon = 0.16$, means

coarse space size			fill-in		
c_1	c_2	c_3	c_1	c_2	c_3
0.00	0.00	0.00	0.00	0.00	0.00
1.51	3.16	4.02	0.04	0.17	0.39
0.42	0.65	1.68	0.06	0.19	0.13
0.26	0.72	1.72	0.19	0.04	0.13

(f) Standard deviations (in %)

Table 1: Sizes of coarse spaces and fill-ins, 32385 DOFs, $H_s = 16h$.

5 Conclusions

In this paper multilevel Schwarz-type preconditioner with coarse space by aggregation and inner K-cycle was considered. The numerical results (Table 2) shows that the number of CG iterations does not increase or even slightly decreases with increasing heterogeneity. However it can be seen from Table 1 that price for this behaviour is increase of the coarse space size. From the same table it can be seen that all tested aggregation techniques can perform similarly (provided proper aggregation parameters are used).We can also see that when inner V-cycle is used (Table 2, left column) the number of iterations increases with

	mean			std. dev. (%)		
jump	4lvl	3lvl	2lvl	4lvl	3lvl	2lvl
10^0	55	39	23	0.00	0.00	0.00
10^4	64	48	29	2.02	1.89	1.22
10^8	65	44	23	4.03	4.23	3.29
10^{10}	61	38	21	6.56	6.81	6.23

(a) Notay, no K-cycle

mean			std. dev. (%)		
4lvl	3lvl	2lvl	4lvl	3lvl	2lvl
24	24	23	0.00	0.00	0.00
30	30	29	1.18	1.18	1.61
24	24	23	4.91	4.91	3.29
22	22	21	5.49	5.49	6.23

(b) 5 iterations of K-cycle

	mean			std. dev. (%)		
jump	4lvl	3lvl	2lvl	4lvl	3lvl	2lvl
10^0	64	53	29	0.00	0.00	0.00
10^4	70	53	27	1.61	1.41	1.32
10^8	64	38	20	4.26	4.98	4.38
10^{10}	58	31	19	5.97	9.79	2.81

(c) Scheichl, no K-cycle

mean			std. dev. (%)		
4lvl	3lvl	2lvl	4lvl	3lvl	2lvl
31	31	29	0.00	0.00	0.00
28	28	27	1.64	1.64	1.32
21	21	20	6.18	5.14	4.38
20	20	19	3.88	3.88	3.35

(d) 5 iterations of K-cycle

	mean			std. dev. (%)		
jump	4lvl	3lvl	2lvl	4lvl	3lvl	2lvl
10^0	63	54	31	0.00	0.00	0.00
10^4	68	51	27	2.92	1.47	1.98
10^8	65	39	21	4.05	4.55	4.27
10^{10}	60	34	18	5.59	5.75	6.46

(e) Vaněk, no K-cycle

mean			std. dev. (%)		
4lvl	3lvl	2lvl	4lvl	3lvl	2lvl
33	33	31	0.00	0.00	0.00
28	28	27	1.64	2.28	1.98
21	21	21	3.95	3.60	4.44
19	18	18	6.84	6.46	6.46

(f) 5 iterations of K-cycle

Table 2: CG iterations, 32385 DOFs, $H_s = 16h$.

increasing number of levels. The inner K-cycle can stabilize this dependence on the number of levels ((Table 2, right column). However the number of inner iterations has to be chosen carefully as it increases the computational complexity of the preconditioner. A thorough analysis of the computational complexity and a good choice of the inner K-cycle will be done in a near future.

References

1. Notay, Y.: Aggregation-Based Algebraic Multilevel Preconditioning. SIAM J. Matrix Anal. Appl. 27 (4), 998-1018 (2006)
2. Scheichl, R., Vainikko, E.: Additive Schwarz with Aggregation-Based Coarsening for Elliptic Problems with Highly Variable Coefficients. Computing 80(4), 319-343 (2007)
3. Vaněk, P., Mandel, J., Brezina, M.: Algebraic Multigrid by Smoothed Aggregation for Second and Fourth Order Elliptic Problems. Computing 56(3), 179-196 (1996)
4. Toselli, A., Widlund, O.: Domain Decomposition Methods - Algorithms and Theory. Springer, Berlin (2005)
5. Notay, Y., Vassilevski, P.: Recursive Krylov-based multigrid cycles. Numerical Linear Algebra with Applications 15(5), 473-487 (2008)

Object-oriented Distributed Matrix Data Structures Allowing Multi-GPU Computation

Martin Stachoň

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
martin.stachon@vsb.cz

Abstract. Distributed objects technology allows transparent parallelization of numerical linear algebra algorithms. We describe the implementation done in OOSol library using distributed objects technology CORBA. The object oriented design allows the same algorithms to be run on a single processor machine, as well on disjoint memory clusters or even using multiple GPU accelerators. Finally, we provide results of benchmarks done on several machines with GPU.

Keywords: numerical algorithms, distributed objects, GPU, CORBA, middleware

1 Introduction

Perhaps the most used programming model used today for developing parallel algorithms is message passing (usually using MPI) in the form of SPMD (Single Program Multiple Data). However, most linear algebra algorithms can be described in the terms of basic operations on objects (matrices). For such algorithms, explicit message passing and program branching lead to code which is more difficult to read as well as maintain. It appears we can instead use distributed data structures (mainly for matrices) using data parallelism (which is already known from languages like Fortran-KDP). This way, we can use the same main program, no matter it will be running on a uniprocessor, SMP, distributed memory cluster or Multi-GPGPU (General Purpose Graphical Processing Unit) cluster. The parallelism is executed by choosing appropriate class implementing abstract matrix interface, hiding the details from the programmer of an algorithm.

Furthermore, within the implementation of data-parallel classes, we can replace explicit message passing and SPMD model with distributed objects technology, such as CORBA, to hide the details of explicit messaging and marshalling of data from the programmer. Using distributed objects also fits well within overall object oriented design of an application. Instead of explicit program branching and messaging, parallelization is done by using remote object interfaces.

VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Implementation

2.1 OOSol Library

The implementation has been done is OOSol library, developed at Department of Applied Mathematics of Technical University in Ostrava. It aims to develop object oriented library of numerical solvers. The library is writen in C++ language. The library provides classes for vectors, matrices (sparse CRS, skyline, full) and iterative solvers (Conjugate Gradients, MPGRP, SMALBE) operating upon matrices and vectors. The library was sequential, using only one CPU. However, exploiting the object oriented desing and features of distributed objects technology allowed parallelization of it by providing new, data parallel data classes without the need to rewrite existing algorithms and data structures.

2.2 Block Matrices

At this moment, data parallelism is supported for matrices, as operations on matrices are most intensive for memory storage and computation power. However, operations on vectors could be implemented in the same way.

We decompose a matrix naturally to the form of block matrix :

$$A = \begin{bmatrix} B_{1,1} \dots B_{1,l} & \\ B_{2,1} & \\ \vdots \quad \ddots \vdots & \\ \quad\quad B_{k-1,l} & \\ B_{k,1} \dots B_{k,l} & \end{bmatrix}$$

. One block $B_{i,j}$ of a matrix shall be one remote matrix object within the memory space of specified CPU or GPU. Operations on a block matrix can be decomposed to operations on individual blocks, for example calculating matrix by vector dot product (DGEMV BLAS operation) can be expressed as

$$y = \beta \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \\ y_k \end{bmatrix} + \alpha \begin{bmatrix} B_{1,1} \dots B_{1,l} & \\ B_{2,1} & \\ \vdots \quad \ddots \vdots & \\ \quad\quad B_{k-1,l} & \\ B_{k,1} \dots B_{k,l} & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_l \end{bmatrix} = \begin{bmatrix} \beta y_1 + \alpha(B_{1,1}x_1 + \ldots + B_{1,l}x_l) \\ \beta y_2 + \alpha(B_{2,1}x_1 + \ldots + B_{2,l}x_l) \\ \\ \dots \\ \\ \beta y_k + \alpha(B_{k,1}x_1 + \ldots + B_{k,l}x_l) \end{bmatrix}$$

2.3 Distributed Objects Middleware

From the distributed objects technologies available today, we choose CORBA, because it is available in C++ language, in which OOSol library was written, and also provides good performance and portability. We chose the omniORB implementation, because it is free, supported and delivers good performance.

2.4 Worker-Manager Model

The implementation uses worker-manager model. There are two types of processes - *manager*, which is the the control program, contains main program code and algorithms, as well as distributed data structeres in the form of CORBA remote references. The other type of process is *worker*. Workers have only basic interface for creating matrices and performing operations on them, they have no knowledge of the main program's algorithm and perform only operations remotely invoked on them by manager (see figure 2.4). Therefore, workers are independent of the main program, unlike in the SPMD model, where all workers are running the same program code, branching in individual sections.

Usually, on a machine with n processors, n worker processes are launched.

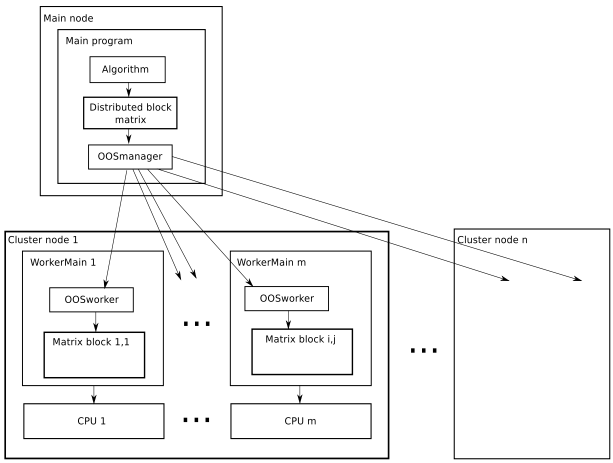


Fig. 1. Scheme of worker-manager model.

2.5 Individual blocks and GPU

As described above, matrices are decomposed into individual blocks. The master application accesses them by abstract remote matrix interface. Thus, different matrix implementation classes can be exchanged easily. OOSol already provided uniprocessor implementations for several matrix types, and we also added new classes for LAPACK optimized BLAS and also for GPU accelerated matrices. These were implemented using CUBLAS and CUSPARSE libraries, which are part of CUDA SDK and provide GPU accelerated matrix operations.

The use of CORBA for distributed matrices allows Multi-GPU computation, where one process is launched per GPU. The GPUs do not have to be in the same memory space.

3 Benchmarks

The implementation was benchmarked on several machines with NVidia GPUs. Unfortunately, the machines have only one or two GPUs. The machines were : Mamba (GeForce GTX 460), Tesla (2× Tesla C2050) and Hubert (Tesla C1060).

The benchmark problem is calculating the displacement of a 2-D membrane. Discretization of the problem using the Finite Elements Method leads to a system of linear equations with positive definite matrix, which can be solved effectively using the Conjugate Gradient method. By switching matrix classes at individual block, we benchmarked GPU vs. CPU implementations. By changing the discretization parameter h, we can control the size of the matrix.

3.1 Full matrix

Table 1 and figure 2 shows the computation time on test machines with increasing size of the problem. We can see that the speedup with single Tesla GPU is around 16 over single CPU for large enough matrices and 25 when using two Tesla GPUs, giving effeciency of 1.5 for dual-GPU computation.

Table 1. Calculation times for full matrices.

SIZE	CPU Tesla01	GPU Mamba	GPU Hubert	GPU Tesla01	2xGPU Tesla01
16	0.000517	0.001286	0.001737	0.002076	0.008566
100	0.000761	0.002535	0.003795	0.004049	0.012555
1600	0.349597	0.044811	0.07177	0.044383	0.057919
3600	3.11362	0.249982	0.347696	0.245637	0.245823
8100	24.1823	1.60169	1.90376	1.53134	1.09698
14400	102.028	-	6.56189	6.11234	4.04863
19600	275.813	-	13.7975	-	7.86928

3.2 Sparse matrix

Table 2 and figure 3 shows the computation time when using sparse matrices. We see that using GPU provides only 1.93 times relative speedup over CPU for sparse matrices. Multi-GPU could not be benchmarked because of technical issues at dual Tesla machine.

4 Conclusions

Distributed object technology has proved to be flexible and easy to use to turn sequential algorithms to scalable parallel ones. It also allows using GPU accelerators easily.

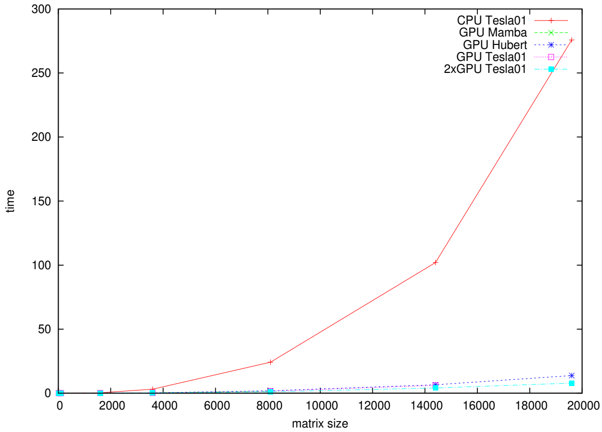


Fig. 2. Calculation time using full matrices for various machines depending on the size of the problem.

Table 2. Calculation times for sparse matrices.

SIZE	CPU Mamba	CPU Hubert	GPU Mamba	GPU Hubert
16	0.000517	0.000519	0.000861	0.000996
100	0.000547	0.000547	0.0013	0.001601
10000	0.031968	0.034001	0.030073	0.061988
250000	5.94499	10.8302	3.67078	7.88964
1000000	48.5665	93.7485	31.6323	65.4265
4000000	396.21	748.035	239.822	579.937
6250000	922.606	1852.93	476.869	1116.48

For accelerating Conjugate Gradients on GPU, we have achieved good speedup for full matrices, but only small for sparse matrices. This is caused by the fact, that the test problem was simple and the stiffness matrix was very sparse. When only a few off-diagonal elements are present, $O(n)$ operations has to be done to calculate the matrix-vector product, but the cost to transfer the vector to GPU memory and back is also $O(n)$. For full matrices, there is always $O(n^2)$ operations in computing the product.

Further work will focus on testing in environments with more GPUs and mixing matrix blocks stored on CPUs and GPUs (hybrid parallelization).

5 Acknowledgements

This work has been supported by internal grant SP/2010173.

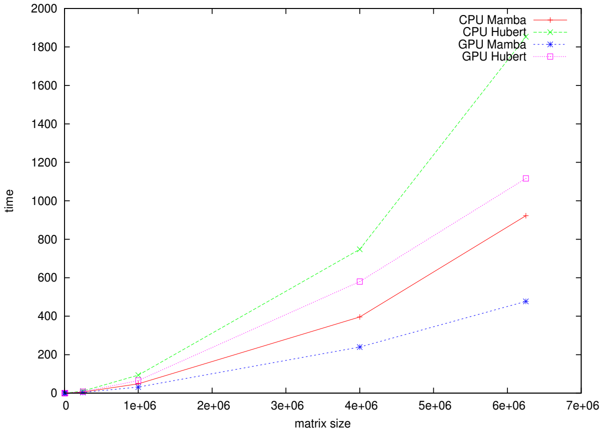


Fig. 3. Calculation time using sparse matrices for various machines depending on the size of the problem.

References

1. Vondrák V., Kozubek T., Kabelíková P., Horák D., Merta M., Stachoň M. et al. OOSol library http://am.vsb.cz/oosol/
2. Stachoň M. *Implementace distribuovaných datových struktur do knihovny OOSol.* (Diploma thesis), VŠB - Technická Univerzita Ostrava 2010.
3. NVIDIA Corporation, *CUDA CUBLAS library*. 2010.
4. NVIDIA Corporation, *CUDA CUSPARSE library*. 2010.

Monte Carlo as a Tool for a Dynamic Reliability Problem

Adam Zdráhala

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
adam.zdrahala@vsb.cz

Abstract. The main purpose of the paper is to illustrate and model, by means of proper examples, a complex dynamic system and to assess its reliability performance via direct Monte-Carlo simulations and bootstrap method. In principle, the simulation method is not a new method, but the paper shows that in connection with new computation technology which gives us a chance how a complex dynamic system could be effectively evaluated. One of the main problems in reliability assessment of complex dynamic systems is to take into account time dependencies of the system structure resulting from changes of its physical parameters. One way to solve the problem is the use of a Petri net approach. Results obtained by both exact analytical approach and the hybrid-stochastic Petri net approach will be confronted with the newly modified Monte Carlo approach.

Keywords. dynamic reliability, Monte Carlo, Petri nets, simulations, test-case, dynamic system, bootstrap

1 Introduction

The dynamic reliability approach takes into account changes (evolution) of the system structure (hardware). For instance, the dynamic reliability allows modelling a human operator (or an electronic control system) naturally. In these cases, the structure of the system is usually changed in order to keep the functionality and/or safety of the system. The evolution of the system can be modelled by modifications of values of so called process variables [6] (Pasquet et al. 1998).

It would be stressed here that more conventional techniques, which are based on Boolean modelling, such as Fault Trees and Event Trees, are not suitable for modelling of dynamic systems because of statistical dependency between values of physical parameters and state of components. Recently, Neural networks and Petri nets approaches were used as tools for reliability analysis of dynamic systems [6] (Pasquet,et al. 1998).

The structure of this paper is organized as follows. In Section 2, a newly modified Direct Monte Carlo Method (DMC) and bootstrap method are shortly introduced. To

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 591–596.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

emphasize the computation efficiency of the DMC method, well known benchmark has been deeply studied and its reliability has been quantified. In third section the results are presented both Monte Carlo method alone and in connection with bootstrap method.

2 Simulation approach

The DMC method may be interpreted in such a way that it converts a dynamic system to a deterministic system. We generate times of occurrence of all possible dynamic events from known probability distributions. This eliminate the system dynamics and the system can be evaluated from reliability point of view as a non-dynamical system. Repeating the simulation process to some extent and having good computation equipment (both software and computer technology) we are able to obtain results of high quality, that are comparable with both exact analytical results and referenced simulation results and in some cases even more accurate.

2.1 DMC procedure

The basic idea of the simulation process is described in [8] (Marseguera & Zio 1996). As mentioned above, unknown times of dynamic events are generated first of all, then all generated times are ordered. After that, minimum of all these times is found and the simulation process moves the system into it. We reconfigure system parameters given by new functional state of the system and evaluate reactions. New time evolution is then created taking into account minimum of these times and previously generated dynamic event times. We repeat this until system failure will occur or until mission time will be over.

Direct Monte Carlo method allows mutual interactions between deterministic events. Dynamic events influence deterministic changes but dynamic events cannot be interacting with deterministic changes.

The fact that the simulation process is very effective is based on merits of a high-performance programming language for technical computing MATLAB. Newly developed commands and object oriented programming procedures have been successfully used at the program generation. Simulation methodology in connection with high performance computing of dynamic systems has been shown to be viable within this paper.

2.2 Bootstrap method

B. Efron (1979) introduced the Bootstrap method. The method is based on very simple idea. We generate new samples from original sample by random. It means, each observation has a same probability to be a part of new sample. Each observation can occur more than once or not at all. It’s like drawing balls from a hat and then putting it back. We repeat this procedure many times (hundreds times). So we get hundreds "pseudo" samples. Bootstrap method is used for confidence intervals of failure probability in our paper.

2.3Bootstrap confidence interval

Recall the familiar one-sample confidence interval for the mean of a Normal population. This interval is based on the Normal sampling distribution of the sample mean and the formula for the standard error of the sample mean $\bar{x}$ and the formula $s/\sqrt{n}$ for the standard error of $\bar{x}$.

When a bootstrap distribution is approximately Normal and has small bias, we can use essentially the same recipe with the bootstrap standard error to get a confidence interval for any parameter.

Another approach is used when the normality assumption isn't satisfied. We generate enough samples then count our statistic for each sample and find $\alpha/2$ and $1-\alpha/2$ quantiles. We got $1-\alpha$ confidence interval for our statistic.

3 Banchmark system

This system was investigated in several papers [1], [2], [3], [4] (Dutuit et al. 1997, Chatelet et al. 1998, Beati & Caira 2004, Škňouřilová & Briš 2008). Consequently we are able to compare our results with other solutions. System is defined as follows:

The system consists of a tank containing a fluid whose level is controlled by means of three active components. Components 1 and 2 are two inlet pumps while component 3 is an outlet valve. All components work independently and each of them is either in On (active or stuck) or in an Off (active or stuck) state and assumed to provide the same rate of level changes: q = 1.5 m.h-1.

For each component, the transition rates from state 0, 1 (On, Off) to state 2, 3 (stuck On, stuck Off) are equal. They are 1/438 h^{-1}, 1/350 h^{-1} and 1/640 h^{-1} for component i = 1, 2, 3, respectively. Only one process variable considered, that is the liquid level h. At t = 0, the system is assumed to be in equilibrium state, i.e., the configuration of the components is (On, Off, On) and the liquid level is h_0 = 4 m. Control laws modify the state of the components to keep the liquid between two limits: h_0 - 1 m and h_0 + 1 m (Table 1).

Two possible failure events are considered: dry out [level ≤ (h_0-3) m] and overflow [level ≥ (h_0-3) m]. Another two liquid level are added that are $h_0 + 3 + \Delta h$ and $h_0 - 3 - \Delta h$, Δh is a small number. In our case Δh = 0.001. These two adition liquid levels help us enforce same properties of the system like is descripted (Fig. 1).

Table 1. System reaction on liquid passing thru important level limit

Passing thru level	Components configuration			Passing thru level	Components configuration		
	C1	C2	C3		C1	C2	C3
1	Dry out			2	On	On	Off
3	On	Off	On	4	Nothing		
5	On	Off	On	6	Off	Off	On
7	Overflow						

3.1 Results of the test-case

Results are demonstrated in Tables 2-3. Our DMC method is represented by averages generated from thirty repeats (Monte Carlo loops) with 90% confidence interval. Each MC loop results from one million simulations. The DMC method is compared with Petri net approach referenced from both [1] (Dutuit et al. 1997, denoted by SPN(1)) and [2] (Chatelet et al. 1998, denoted by SPN(2)).

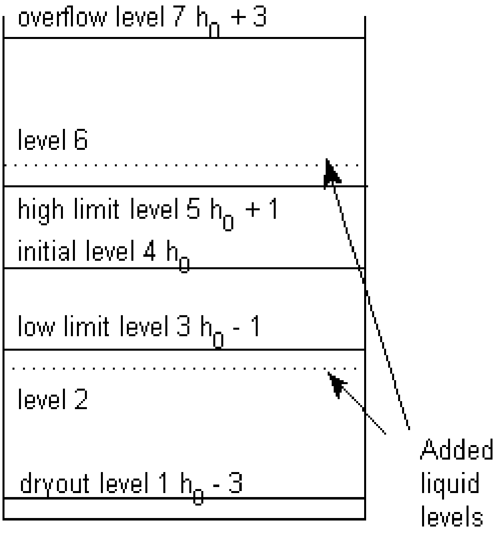


Fig. 1. Tank and important liquid level limits

Table 2. Dry out probability for basic tank system

t [h]	DMC average	conf.	SPN(1) av.	conf.	SPN(2) av.	conf.
200	0.0230	$4.730\ 10^{-5}$	0.0232	$0.247\ 10^{-2}$		
400	0.0656	$8.020\ 10^{-5}$	0.0666	$0.409\ 10^{-2}$	0.0667	$4.09\ 10^{-4}$
600	0.0932	$8.869\ 10^{-5}$	0.0932	$0.477\ 10^{-2}$	0.0955	$4.82\ 10^{-4}$
800	0.1078	$9.730\ 10^{-5}$	0.1075	$0.508\ 10^{-2}$	0.1103	$5.14\ 10^{-4}$
1000	0.1146	$9.937\ 10^{-5}$	0.1151	$0.523\ 10^{-2}$	0.1175	$5.28\ 10^{-4}$

Table 3. Overflow probability for standard tank system

t [h]	DMC average	conf.	SPN(1) av.	conf.	SPN(2) av.	conf.
200	0.1989	$1.395\ 10^{-4}$	0.2035	$0.660\ 10^{-2}$		
400	0.3584	$1.569\ 10^{-4}$	0.3691	$0.791\ 10^{-2}$	0.363	$7.88\ 10^{-4}$
600	0.4309	$1.674\ 10^{-4}$	0.4427	$0.815\ 10^{-2}$	0.437	$8.13\ 10^{-4}$
800	0.4638	$1.658\ 10^{-4}$	0.4767	$0.819\ 10^{-2}$	0.470	$8.19\ 10^{-4}$
1000	0.4784	$1.562\ 10^{-4}$	0.4917	$0.820\ 10^{-2}$	0.485	$8.20\ 10^{-4}$

As you can see. Monte Carlo method gives very good results. The results are given for one million Monte Carlo loops. It took five minutes of computation time. 90% confidence interval is shorter in comparison with another methods. But for counting confidence interval, we had to repeat Monte Carlo simulation 30 times. It means, 30 million Monte Carlo loops. It took one and half hour. It is little more then we want. We want to compute confidence intervals without dramatically increasing computation time.

That's why we used bootstrap method for confidence interva. We divided results of Monte Carlo loops into ten sets of the same size. Evaluated mean value for each set. This was our random sample. Then we generated ten thousand bootstrap samples from random sample and evaluated bootstrap confidence intervals. This procedure is very cheap for computation time. Computation time was increased just by 40 seconds. The results are presented in table 4, 5.

Table 4. Comparison classic confidence intervals and bootstrap confidence intervals for tank dry out

t [h]	DMC average	conf.	DMC av.	bootstrap conf.
200	0.0230	$4.730\ 10^{-5}$	0.0230	$0.247\ 10^{-3}$
400	0.0656	$8.020\ 10^{-5}$	0.0657	$0.3785\ 10^{-3}$
600	0.0932	$8.869\ 10^{-5}$	0.0934	$0.393\ 10^{-3}$
800	0.1078	$9.730\ 10^{-5}$	0.1079	$0.4525\ 10^{-3}$
1000	0.1146	$9.937\ 10^{-5}$	0.1147	$0.5405\ 10^{-3}$

Table 5. Comparison classic confidence intervals and bootstrap confidence intervals for tank over flow

t [h]	DMC average	conf.	DMC av.	bootstrap conf.
200	0.1989	$1.395\ 10^{-4}$	0.1995	$0.765\ 10^{-3}$
400	0.3584	$1.569\ 10^{-4}$	0.3586	$0.689\ 10^{-3}$
600	0.4309	$1.674\ 10^{-4}$	0.4310	$0.651\ 10^{-3}$
800	0.4638	$1.658\ 10^{-4}$	0.4642	$0.7285\ 10^{-3}$
1000	0.4784	$1.562\ 10^{-4}$	0.4788	$0.7735\ 10^{-3}$

4 Conclusion

We applied the Direct Monte Carlo approach for well known benchmark. Our results have been compared with other papers. The method seems to be accurate enough. When we want just a point estimator for the parameter. It's effective enough, too. But if we want interval estimator computation time is dramatically increased. So we used bootstrap method for computation interval estimator. It seems to be very effective method, because we got confidence intervals accurate enough with very small computation time increase. Monte Carlo method in connection with bootstrap method is very effective for both point estimators and interval estimators.

References

[1] Dutuit, Y., Chatelet E., Signoret, J.P. & Thomas, P. 1997. Dependability modelling and evaluation by using stochastic Perti nets: application to two test cases. *Reliab. Eng & System Safety* 55, 117–124.

[2] Chatelet, E., Chabot, J.L. & Dutuit, Y. 1998. Event representation in dynamic reliability analysis using stochastic Petri nets. *Proceedings of the Fifth InternationalWorkshop on Dynamic Reliability: Future Directions.*

[3] Beati, J. & Caira, M. 2004. Benchmark on Dynamic Reliability -An approach based on Dynamic Event Tree Analysis, Rome, December 13, *Workshop on dynamic reliability: results and lessons on a benchmark*. By the Italian Association of environmental, safety and reliability analysts, 3ASI.

[4] Škňouřilová, P. & Briš, R. 2008. Coloured Petri Nets and a dynamic reliability problem. *Journal of Risk and Reliability* 222, No.4, 463-720; ISSN 1748 006X.

[5] Škňouřilová, P. 2009. Dynamic reliability models based on Petri nets. *PhD Thesis*, Technical University of Ostrava.

[6] Pasquet S., Chatalet E., Padovani E., & Zio E. 1998. Use of Neural Networks to evaluate the RAMS` parameters of dynamic systems. *Proc. of International conference on safety and reliability ESREL'98, Trondheim, Norway*. Pg.723-728.

[7] Labeau, P.E. 2002. Pressurization test-cases for dynamic reliability; Workshop - The "test case" activity of the ISdF and its contribution to the development of dynamic reliability studies, *European Conference on Safety and Reliability ESREL 2002, Lyon*.

[8] Marseguera, M. & Zio, E. 1996. Monte Carlo approach to PSA for dynamic process systems. *Reliability Eng. & Syst. Safety* 52; 227-241.

Using Software-defined Radio Concept in Communication Systems to Analysis M-QAM Digital Modulation Technique

Mohamed Al-Wohaishi and Radek Martinek

Department of Measurement and Control, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{wohaishi, radek.martinek.st1}@vsb.cz

Abstract. The speed at which new digital communication systems are being developed is increasing dramatically. Growing quantity of transmitted information and growing quantity of users require development of modern communication systems based on new principles; existing systems are quickly becoming obsolete technologically. Digital radio and television broadcasting systems are used multistate digital modulation M-QAM for transmission of information. This paper presents real results, which illustrate the link between number of symbols in M-state QAM modulation, SNR and BER. Constellation diagrams of transmitted and received symbols (with superposition of noise) were presented in the article. Simple picture is transmitted through simulated radio channel to show the result of signal impairments. Experiments were done using software defined radio concept of communication system. Modular PXI HW platform was used in connection with graphically oriented development environment. This combination of modular HW and flexible SW components allows changing the communication protocol, modulation scheme, frequency bandwidth and other parameters in a very simple way by changing the software part of the system.

1 Introduction

Nowadays, modern devices often use many different technologies to provide customers with a wide array of functionality. This paper describes the analysis of digitally modulated signals in modern communication systems using software defined radio concept, in a digital communication system, digital information can be sent and received by subtly changing characteristics of a carrier sine wave. However, detecting changes in phase (PSK and QAM) is much more difficult and requires a synthetic instrumentation for analysis of the signals [1].

This approach is already state of the art. Modern radio receiver chipsets already count on embedded cores, where software implements many parts of the functionality.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 597–602.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

A software-based solution can be modified in the very late stages of the development process, and to some extent be reused across SDR platform generations. Certain standard software functions and libraries on the market are licensable as intellectual property (IP) modules or core development kits (CDK). Audio and video decoder software libraries are a typical example of such IP. Another advantage of these software solutions is the possibility for updates or bug fixes after delivery of a product, when a problem encountered in the field requires countermeasures [2].

2 Vector Signal Generator - NI PXI 5670

Along with the development environment LabVIEW, complemented by extended Modulation Toolkit library, the PXI module could be used to generate the required test signals for verifying the possibilities of digital transmission systems which use the new standards. RF vector signal generator, the NI PXI-5670, shown in figure 1, represents the generator of user-defined waveform (arbitrary waveform generator) working a resolution of 16 bits and sampling rate of 100MS / s (400MS/s in the interleaved mode) with a depth of memory up to 512 MB and the real bandwidth of 20 MHz Using a digital upconverter along with this module can generate signal in the range of 250 kHz to 2.7 GHz with random modulation scheme such as: AM, FM, PM, ASK, FSK, MSK, GMSK, PSK, QPSK, PAM, and QAM, see [3].

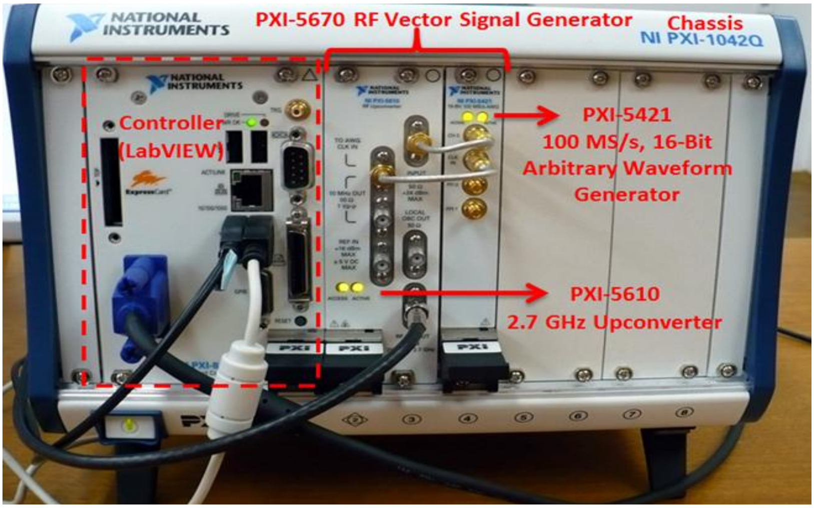


Fig. 1 Vector Signal Generator (NI PXI-5610 – RF Upconverter, NI PXI-5421 -100MS/s AWG)

3 Vector Signal Analyzer – NI PXI 5661

For the analysis of digitally modulated signals a NI PXI-5660 module was used, shown in figure 2. This module is a very compact solution (30% of normal weight and cubature of separate devices in this class), allowing very rapid measurement of digitally modulated signals in the range from 9 kHz to 2.7 GHz. With the real bandwidth of 20 MHz, but with possible flow of data 132 Mb/s over the PCI bus, this solution represents tremendous progress in the contrast of 1 MB throughput via GPIB interface for connection of single vector signal analyzers, see [4].

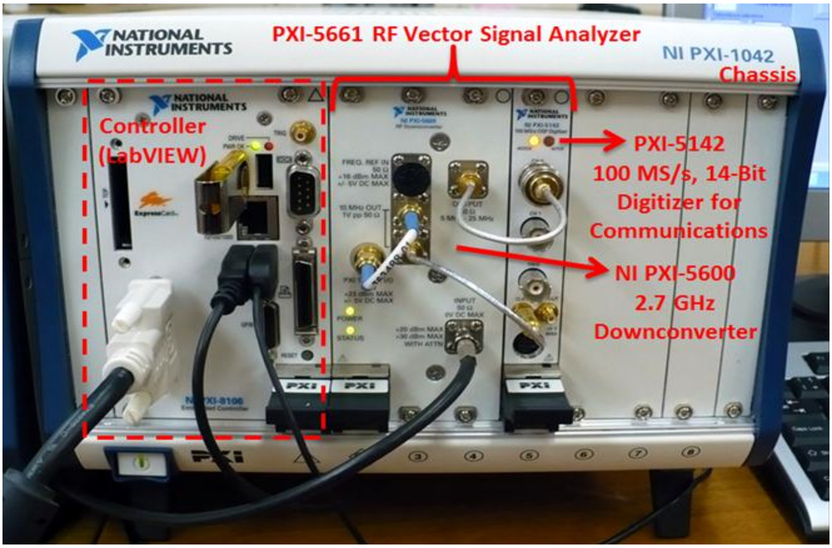


Fig. 2 Vector Signal Analyzer (NI PXI-5600 – RF Downconvertor, NI PXI-5142 -100MS/s OSP Digitizer)

Authors of this article focused on multistate M-QAM modulation schemes, which are widespread in digital TV broadcasting today.

Terrestrial DVB-T broadcasting uses QPSK, 16-QAM and 64-QAM modulation schemes, while terrestrial DVB-T2 broadcasting, which allows transmission of high definition picture format, uses 256-QAM modulation scheme.

4 Multistate M-QAM Modulation

We investigate on of the more widespread in digital TV broadcasting family of modulation schemes, QAM works by using M different combinations of voltage magnitude and phase to represent N bits, as described by the relationship $M = 2^N$. When N is an even integer, the constellation is regular with I and Q each representing 2^{N-1} bits. When N is an odd integer, the constellation is not necessarily symmetrical, and finding an optimal distribution of sample points is not straightforward, for which the signal can be generally given by:

$$s(t) = \sqrt{2E_s / T}\left[I(t)\cos\omega_0 t - Q(t)\sin\omega_0 t\right] \tag{1}$$

Where I (t) and Q (t) are the baseband I and Q waveforms, respectively

$$I(t) = C_0 \sum_n a_n^I \sqrt{T} \sigma(t - nT), \tag{2}$$

$$Q(t) = C_0 \sum_n a_n^Q \sqrt{T} \sigma(t - nT), \tag{3}$$

Here $\sigma(t)$ has unit energy and C_0 is chosen so that the normalizing condition is satisfied. This requires averaging over all the pairs (a_n^I, a_n^Q), weighting each by 1/M, the integral in is simply T, so [5]:

$$1/C_0 = \left[(1/M)\sum_{a^I, a^Q}\left[(a^I)^2 + (a^Q)^2\right]\right]^{1/2}. \tag{4}$$

The probability of error in symbol transmission is in M-QAM modulated transmission channel with AWGN noise and coherent demodulation, determined by the equation below, [6]:

$$P_{EMQAM} \cong 2\left(1 - \frac{1}{\sqrt{M}}\right) erfc\left(\sqrt{\frac{E_{\min}}{N_0}}\right), \tag{5}$$

Where *erfc* is complementary error function.

5 Results of simulations for tested M-QAM Modulations

Figure 3 shows the positions of symbols in constellation diagram for tested M-QAM modulated systems. Parts (a, c and e) show different M-QAM modulations without noise. Parts (b, d and f) show different M-QAM modulations with SNR 10 dB in transmission channel – the communication system couldn't work properly for higher number of states under this condition.

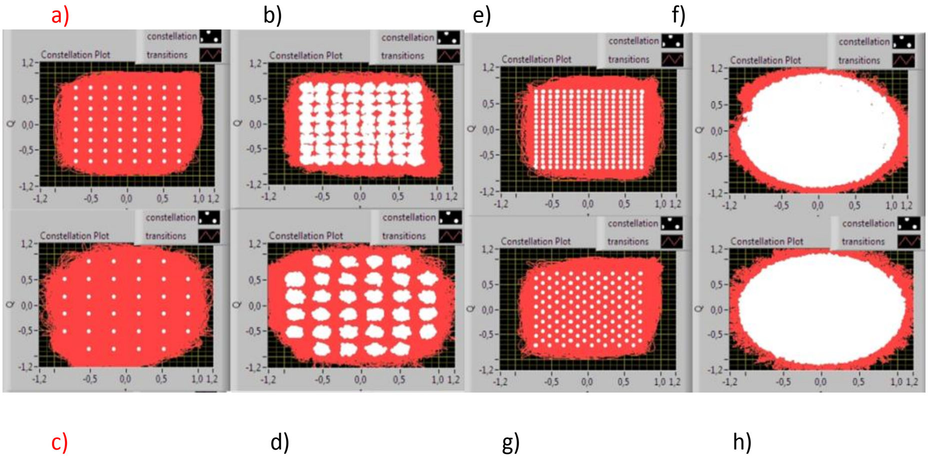


Fig. 3 Constellation diagrams for tested communication systems

The simulation of different M-QAM modulation shows that increasing of the state number, leads to an increase of transfer rate (transfer more bits per symbol). The downside however is that with the growing number of states BER increases at the same transmission power as a result of worse distribution of symbols in constellation diagram, as shown in figure 4.

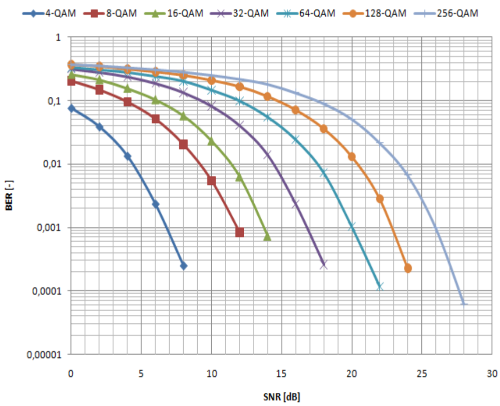


Fig. 4 Measured BER dependency on SNR

6 Conclusion

This article described the case study of software defined radio communication system implemented on National Instruments PXI modular HW platform using graphically oriented development environment LabVIEW.

The synthetic instrument for analysis of digitally M-QAM modulated signal was implemented on the same platform. Some basic advantages of SDR concept were demonstrated by SW configuration of basic parameters of the communication system.

Basic features of different types of digital signal modulation were demonstrated by results of measurements, which were executed on implemented synthetic instrument.

7 References

[1] Martinek, R.; Al-Wohaishi, M.; Zidek, J.: Software based flexible measuring systems for analysis of digitally modulated systems. RoEduNet 2010 9th,Sibiu, Romania 2010, ISSN 2068-1038, ISBN 978-1-4244-7335-9, On page(s):397.

[2] SCHLICHT, Michael; SPEITEL, Martin. EE Times Design [online]. 2001-01-17 [cit. 2011-01-25]. Software defined radio: defining the challenges. Dostupné z WWW: http://www.eetimes.com/design/microwave-rf-design/4212244/Software-defined-radio--defining-the-challenges?cid=NL_TestMeasurement&Ecosystem=test-and-measurement .

[3] National Instruments: RF Vector Signal Generator NI PXI-5671, Data Sheet, 2005 http://www.ni.com/pdf/products/us/20055170101dlr.pdf

[4] National Instruments: RF Vector Signal Analyzátor NI PXI-5660, Data Sheet, 2005 http://www.ni.com/pdf/products/us/4mi469-471.pdf

[5] I. L. W. Couch, (2007) “Digital and Analog Communications systems’’, 7th ed. New Jersey, USA: Pearson Prentice Hall.

[6] Ke-Lin Du, M. N. S- Swamy, (2010) “Wireless Communication Systems’’ Cambridge University Press, ISBN-13: 9780521114035.

Conducting Polymers – Polyaniline

Jiří Bocheza, Stanislav Hejduk, and Jan Vitásek

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba

Abstract. Conductive polymers are a very interesting group of organic polymers where polyaniline (PANI) is potentially the most interesting. The use of this polymer is offered in sensor applications, in integrated technology, etc. Because it isn´t too researched material, it was necessary to perform basic measurements. In this article are discussed the results of measurements of homogeneity PANI films, conductivity, polarization and reflectivity.

1 Introduction

Conducting polymers are a special group of organic polymers that are known for their conductivity compared to commonly known polymers. Between these polymers include e.g. polyacetylene, polyaniline (PANI), etc.

As a very promising material appears just polyaniline which is distinguished its conductivity (units S/cm), small thickness of films (50-200 nm), flexibility, ability response to change outer conditions (e.g. conductivity and color change depending on the pH value), its functionality (different possibilities of using) and cheap production.

There are many applications that could use some properties of PANI such as the production of simple flexible semiconductor integrated circuits, conductive inks for printers, sensor systems for monitoring of changes in chemical composition on the change in conductivity, sensor systems for monitoring of magnetic or electromagnetic fields, etc. [1].

The primary aim, before examining the application properties of PANI, was to verify and measure basic behavior of this material. The first measurement was measurements of spectral transmittance for determining the homogeneity of the thickness of layers and debugging the production process. Next step was measurement of conductivity and measurement of reflectance and polarization in dependence on the angle of the incident beam to the PANI film.

2 Polyaniline

History of polyaniline begins around 1840, when it was described by J. Fritzsche. This polymer was originally used as a natural dye. Interest in its conductive properties

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 603–608.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

increased up to the late seventies of the twentieth century and ever since takes place research in the field of conductive polymers and their various applications in practice.

One of the basic properties of PANI is that it occurs in two basic forms. The first form is polyaniline salt which has a conductivity of units S/cm and has a green color. The second form is polyaniline base which has a conductivity of 10^{-10} S/cm and has a blue color. Transfer between different forms is possible by changing the pH where the conductive form is obtained by bringing of PANI into the environment with a pH>7 and non-conductive form is obtained by bringing of PANI into the environment with a pH<7.

2.1 Production

Production of polyaniline is simple and cheap. They are used commonly accessible chemicals such as aniline hydrochloride (ANI HCl) and ammonium peroxodisulfate (APS). The whole reaction takes place in water environment where the individual chemicals in the form of powders mix separately in water (H_2O) in the required concentration [2]. Then the two solutions are mixed together and dipped into a solution of the subject which we want to cover by the PANI film. Reactions and application of PANI film will take about 10 minutes, when the solution thickens, turning blue and then green. Then the subject coated with PANI film is pulled out and rinsed with diluted hydrochloric acid (HCl) [2]. The remaining polymer can be isolated by filtration, dried and then processed as a conductive powder. Production efficiency is almost perfect and the thin film of PANI can be applied to a wide range of materials.

3 Measurement of basic properties of PANI

The first measurements on PANI films were focused on verifying of the basic behavior. It was the measurement of permeability, conductivity, reflectivity and polarization.

Measurements were performed on PANI films created on laboratory micro-glasses of a size 76x26x1. The films were created as described above at room temperature. Individual samples were created in provisional conditions at the Department of Telecommunications at VŠB - TU Ostrava and also with controlled procedure by chemists from the Center of nanotechnology at VŠB - TU Ostrava.

3.1 Spectral transmittance and homogeneity of the PANI films

Measurement of spectral transmittance was performed to determine the homogeneity of the created PANI films. We assume that the light transmittance of the material depends on thickness of the material. We determine on the basis of this assumption whether the thickness of PANI film is homogeneous or inhomogeneous.

The main aim of this measurement was to find a production process in collaboration with chemists from the Center of nanotechnology at VŠB - TU Ostrava, which would ensure the production of homogeneous PANI films.

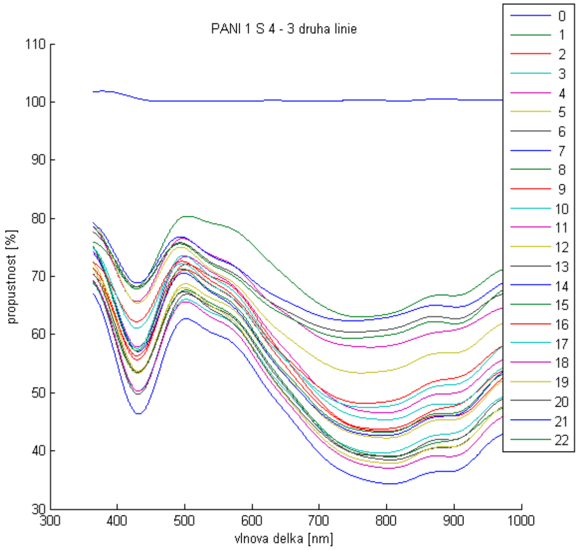


Fig. 1. Measurements of homogeneity of the first samples [3].

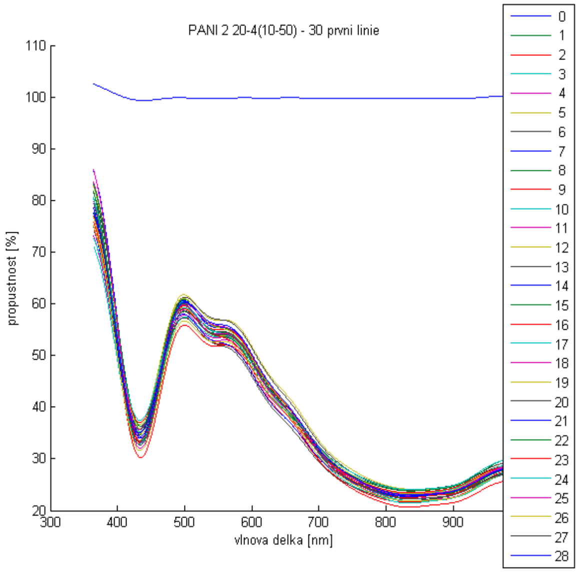


Fig. 2. Measurements of homogeneity after the debugging of the production process [3].

Measuring kit and the measured results were obtained by the student of VSB - TU Ostrava as a part of his diploma thesis [3]. Micro-glasses with films of PANI were inserted gradually between optical fiber. Individual points on the micro-glass were measured gradually using of stepper motors with step 2 mm in two or three lines. Figures 1 and 2 represent spectral characteristics of percentage transmittance in individual points on the micro-glass with PANI film. Figure 1 presents the results from one of the first produced samples. With the gradual improvement of the production process was achieved homogeneity of films with a difference <10% (Figure 2). This difference was influenced much of the outside interference, since measurements were not performed in the completely shaded room [3].

3.2 Measurements of conductivity

Conductivity is one of the most important properties of this polymer, and therefore it was necessary to determine what values the conductivity reaches and if it behaves consistently.

Orientation measurement of conductivity was performed on a sample of PANI, which was made in makeshift conditions in our department. Power supply was connected to a micro-glass with crocodile clips. Supply voltage was in the range 0-30V. For the more accurate results will be necessary to create a more sophisticated method of connection and measurement.

The measured conductivity of the sample with the PANI film is shown in Figure 3. The measured values shows, that conductivity increased slightly with increasing current and values of passing currents was in μA values.

We performed also a measurement of non-conductive PANI form. We found that for the voltage range 0-30V was conducted no current and thus we verified the non-conductivity of second PANI form.

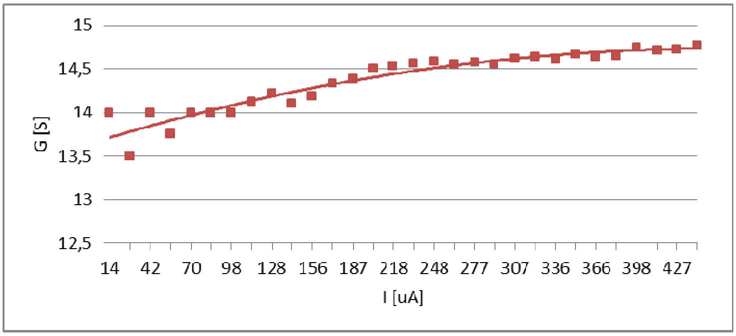


Fig. 3. The measured conductivity of the sample with a PANI film.

3.3 Measurement of reflectance and polarization

We measured the reflectance and polarization for different impact angles for linearly polarized laser beam with a wavelength λ = 650nm. We measured samples with PANI in the conductive and non-conductive form. We observed the course of the reflectance and polarization after reflection. Measurements were performed for the impact angles 23.8 ° - 69.5 °. Configuration for measurement and the measurement procedure are described in detail in the diploma thesis [3].

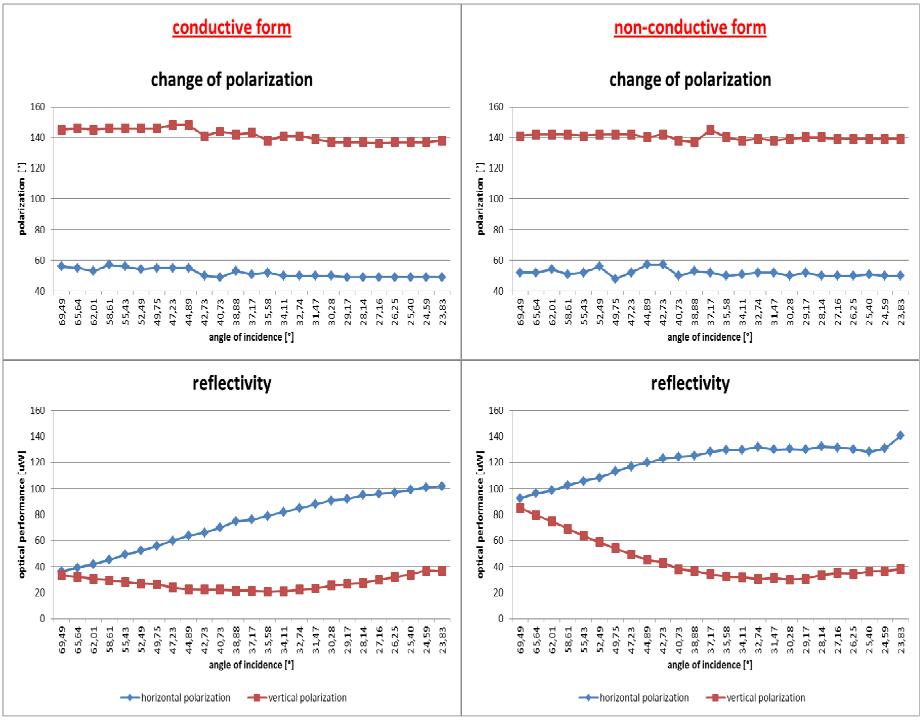


Fig. 4. Change of polarization and reflectivity depending on the impact angle and polarization of source.

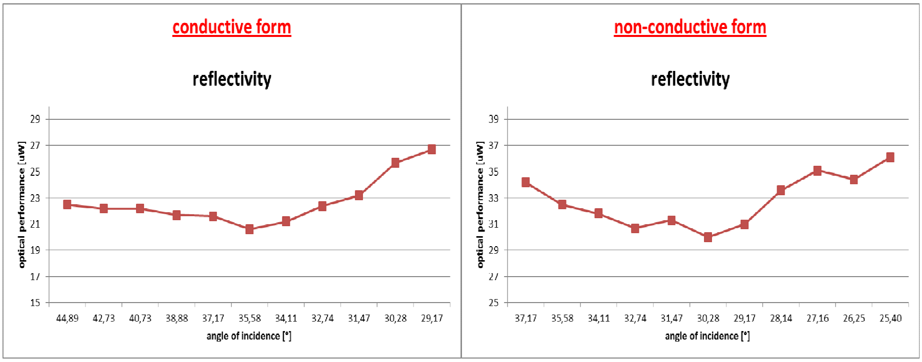


Fig. 5. Detailed view of minimum reflectance value for the conductive and non-conductive PANI form with vertical polarization source.

Measurements were performed for horizontal and vertical polarization of laser beam. Figure 4 shows the courses of changes of polarization and reflectance depending on the impact angle of the laser beam. The measurements were performed for conductive form of PANI (Fig. 4, left graphs) and non-conductive form of PANI (Fig. 4, right graphs).

Waveforms for the conductive and non-conductive form of PANI (Fig. 4, top graphs) show that polarization of the reflected beam is almost unchanged when the impact angle is changing and we can say that it behaves a constant manner. Variations on a waveform were probably caused by extraneous interference, because the measurement was performed in imperfectly darkened room.

The graphs in Figure 4 (down) show waveforms of reflectance. We can see that a higher reflectance is at wavelength λ = 650nm with non-conductive PANI form.

Interesting is shift of the minimum value of reflectance for vertically polarized source. Figure 5 shows a detailed look at this shift. It is evident that the minimum value of reflectance occurred for the conductive and non-conductive form of PANI at different impact angles of laser beam. The difference in impact angles is about 5°.

4 Conclusion

The article informs briefly about conductive polymers, particularly about potentially very promising polyaniline (PANI). We performed basic measurements, which characterize this conductive polymer closer, because it isn´t too researched material with respect to sensor technology.

The first measurements were aimed at measuring of homogeneity of PANI films on the basis of spectral transmittance, where it managed to debug the production process for the creation of homogeneous films, in collaboration with chemists from the Center nanotechnology VSB - TU Ostrava.

Next we performed orientation measurements of conductivity, where the results show that the conductivity slightly increases with increasing current and values of passing currents for the voltage range of 0-30V are in μA values.

The dependence of reflectance and polarization on the impact angle was the last measurement. The most interesting finding from this measurement is the shift of the minimum value of reflectance for vertical polarization source. The minimum reflectance values are obtained at different impact angles for the conductive and non-conductive form of PANI.

References

1. Prokeš, J., Nešpůrek, S., Stejskal, J.: Vodivé polymery: Inteligentní materiály pro nové století. *Vesmír*. January 2001, 80, pp. 35-38. ISSN 1214-4029.
2. Stejskal, J., Sapurina, I., Prokeš, J., Zemek, J.: In-situ polymerized polyaniline films. *Synthetic Metals*. 1999, 105, pp. 195-202. ISSN 0379-6779.
3. Svatuška, J.: Measurements of Static and Dynamic Charakteristics of Luminophore Materials. Ostrava, 2011. Diploma Thesis. VŠB – TU Ostrava.

Synthesis of Speechlike Signals for Masking Acoustic Information

Artem Ganiyev[1], Oleg Zelmanski[2], Anna Kubankova[3], and Petr Münster[3]

[1]Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
artem.ganiev.st@vsb.cz
[2]Information Security Department, Faculty of Telecommunications, Belarusian State University of Informatics and Radioelectronics, 6. Brovka Str., Minsk, 220013, Republic of Belarus
7650772@rambler.ru
[2]Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Purkynova 464/118, 612 00 Brno, Czech Republic
{shklya, petr.munster}@feec.vutbr.cz

Abstract. The structure of the "speechlike" signal generation system containing speaker verification module for selection the allophone database corresponding to the speaker for active acoustic information security by masking is developed. The voice activity detection (VAD) algorithm is developed and realized. Speech segmentation methods are analyzed. The speech segmentation algorithm is developed and realized. The article contains the description of the synthesizer of speechlike signals, voice activity detection module and speech segmentation module as well.

1 Introduction

Protection of information especially acoustic or speech information is the important task. Because speech is the fundamental way of communication and the main part of the private information is transferred by speech. There are two types of technical means of the information security – active and passive. The active means's task is to reduce speech intelligibility by increasing the noise level and the passive means's task is to reduce speech intelligibility by reducing the signal level. One of the active techniques of information protection is active masking.

Analysis of the modern masking systems shows that the greater part of them are based on the "white" or "colored" noise that is not reliable. There are some systems that utilize "speechlike" noise but they are not developed enough.

VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

So there is necessity to develop reliable and effective means for active acoustic information protection using “speechlike” signals built of the conversation to be masked.

The speechlike signal synthesizer which is related to active means of the information security is proposed. It consists of voice activity detection module, speaker verification module, speech segmentation module, speech classification module and speech synthesis module.

2 Speechlike signal synthesizer

The structure of the “speechlike” signal generation system is represented in the figure 1.

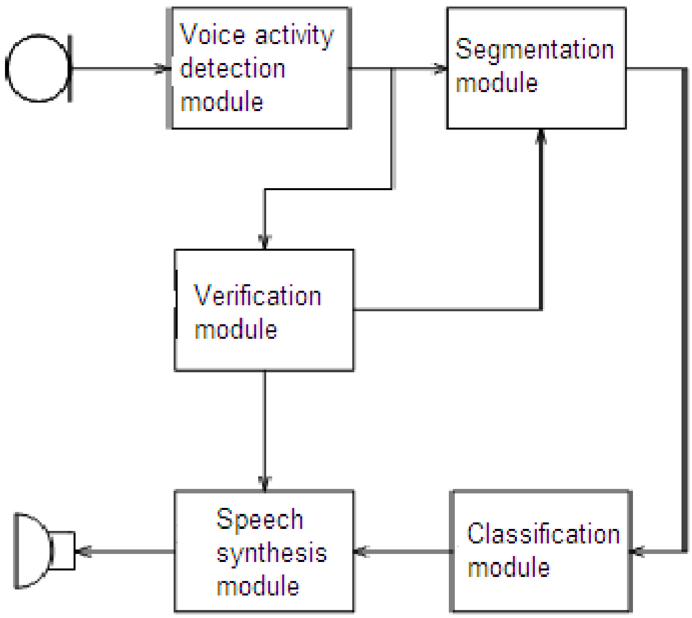


Fig. 1. Block diagram of the “speechlike” signal generator

The principal aim of the “speechlike” signal generation system is masking private conversation therefore the system must be automatically activated by this conversation. That is why it contains voice activity detector that starts the “speechlike” signal generator when the conversation begins. The process of separating conversational speech and silence is called the voice activity detection (VAD). VAD is an important enabling technology for a variety of speech-based applications including speech recognition, speech encoding, and hands-free telephony. The function of a VAD algorithm is to extract some measured features or quantities from the input signal and to compare these values with thresholds, usually extracted from the characteristics of the noise and speech signals. The voice activity detection algorithm [1] consists of four basic steps. At first input signal is framed by multiplication by windowing function. Hamming window can be used as windowing function. Then the RMS, zero-crossing rate and spectrum estimation [2] of the signal are calculated. After their values are compared with corresponding thresholds the decision is made [3, 4]. The structure of the voice activity detection module is represented in the figure 2.

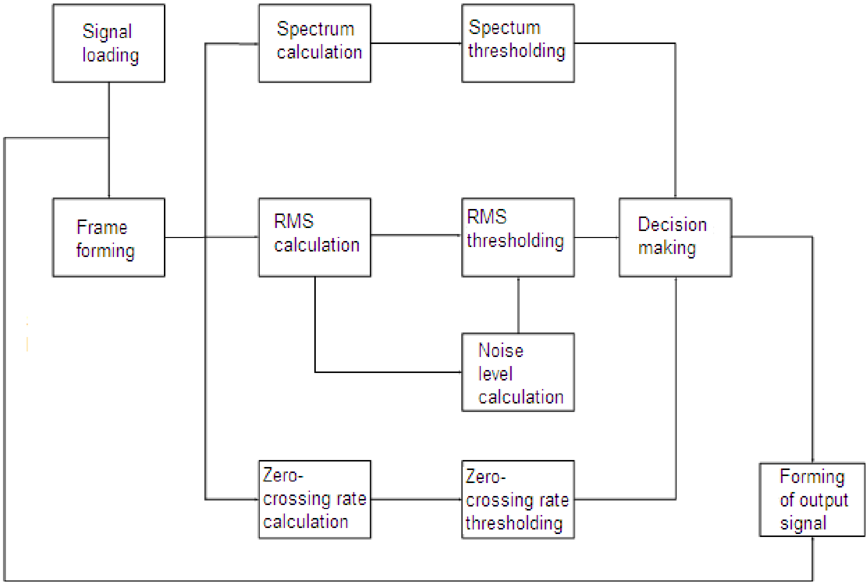


Fig. 2. Block diagram of the voice activity detecting module

Then a speaker can be identified by the speaker verification module [5]. If the speaker is identified the corresponding allophone database which contains phonetic units of the concrete speaker and was founded beforehand can be chosen. If the speaker cannot be identified the allophone database must be formed. It is executed in the speech segmentation and speech classification modules.

Speech segmentation is intended for separation of homogeneous oscillation parts of the signal which correspond to different types of phonemes [6]. The speech segmentation module is the fundamental element of speechlike signal generating systems for information security. The speech segmentation module's aim is to segment the speech that was extracted from the signal by voice detector. The goal of this stage is to divide the speech signal into phonemes. There are some different approaches to speech segmentation. Wavelet analysis, spectrogram and sonogram can be mentioned among them [6-8]. The speech segmentation module is based on the idea of signal spectrum variation analysis by the exploration of linear spectral frequencies of the signal. The structure of speech segmentation module is represented in the figure 3.

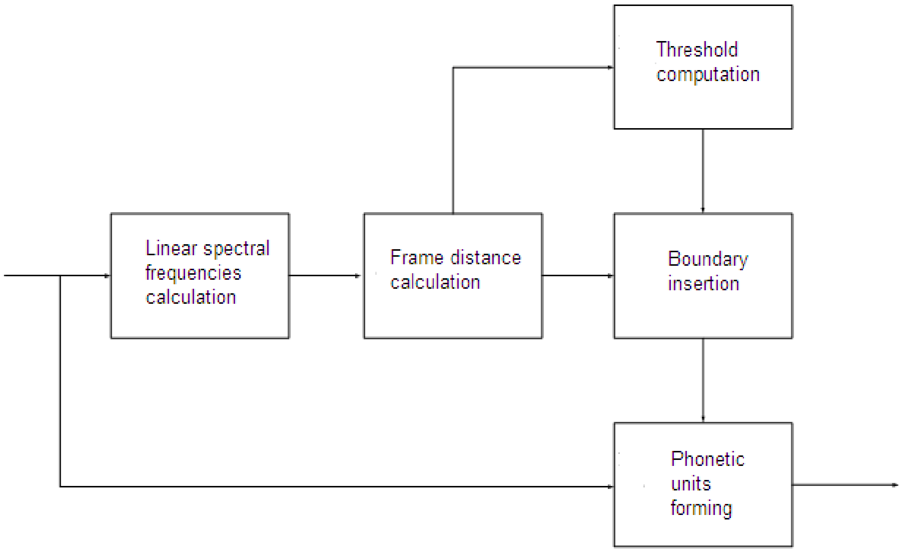


Fig. 3. Block diagram of the speech segmentation module

After segmentation phonemes should be classified and then the "speechlike" signal should be generated using these classified phonemes [9]. The speech generation is realized in the speech synthesis module. The pseudo text corresponding with statistical characteristics of a language is produced. Then it is sounded by the phonetic units from the allophone database of the concrete speaker.

3 Conclusion

The structure of the proposed "speechlike" signal generation system with biometric speaker verification for active masking private conversion is developed and considered. The voice activity detection algorithm is developed and released. It provides self-acting start of the system. The analysis of different speech segmentation methods is carried out. The speech segmentation algorithm is developed and released.

References

[1] Zelmanski, O. Voice activity detector on the basis of comparison of parametric features with their threshold values / Doklady BSUIR. – 2008. – № 7(37). – C. 72–76.

[2] Flanagan, J.L. Computer studies on parametric coding of speech spectra. / J.L. Flanagan, S.W. Christensen. – JASA, Vol. 62. – №2. – 1980. – P. 420–430.

[3] Detection of speech activity using feature model adaptation: пат. US № 6.993.481 B2, МПК G10L 151/20, G10L 15/08, G10L 15/06, G10L 15/04, publ. 7.11.2002.

[4] Voice activity detector : пат. 6,556,967 US, МПК7 G10L 11/02, G10L 11/00 / Douglas J. Nelson, David C. Smith, Jeffrey L. Townsend ; The United States of America as represented by the National Security Agency, Washington, DC; 12.03.1999; publ. 29.04.2003.

[5] Zelmanski, O. Text-dependent speaker verification system / Doklady BSUIR. – 2009. – № 4(42). – C. 37–41.

[6] Radova, V. Method for segmentation of voiced speech signals into pitch period segments / V. Radova, O. Silhan // NORSIG 2002 : proceedings of fifth IEEE Nordic Signal Processing Symposium, October / Norway, 2002. – P. 68–74.

[7] Kamakshi, P. Automatic segmentation of continuous speech using minimum phase group delay functions / P. Kamakshi, Nagarajan, M. Hema // Speech Communication. – 2004. – Vol. 42. – P. 429–446.

[8] Mermelstein, P. Automatic segmentation of speech into syllable units. / P. Mermelstein // J. Acoust. Soc. Amer., V. 58. – № 4. – 1975. – P. 880–883.

[9] Ali, A.M.A. Acoustic–phonetic features for the automatic classification of fricatives. / A.M.A. Ali, J.V. Spiegel. – J. Acoust. Soc. Am., V. 109. – № 5. – Pt. 1 – 2001. – P. 2217–2235.

Mobile FSO Remote Controlled Model

Stanislav Hejduk, Jiří Bocheza, and Jan Vitásek

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{stanislav.hejduk, jiri.bocheza, jan.vitasek}@vsb.cz

Abstract. In this paper we shortly brief about mobile free space optics concept, which is capable to maintain optical link not only for point-to-point connection. For this reason we built optically controlled car model. So we can easily simulate the mutual movement between transmitter and receiver. For optical communication we use two high power LED diodes at wavelength 850nm. Main objective was to build the model capable to show how the FSO network will behave in the constant position changes in various ambient lighting conditions.

1 Introduction

Optical wireless communications, known as free space optics (FSO) are the way to transmit data by using light. As a transmission medium can be used practically any transparent material that allows the existence of LOS (line of sight). Transmission parameters of different environments may differ, so for our purposes we will consider as a transmission medium only atmosphere.

Today, this technology is mostly used as point-to-point link between buildings in cases where can't be used radio transmission, and construction of a fixed link would be too expensive (heavily built-up area, etc…). Parameters of such connections are affected by atmospheric conditions (e.g. fog [1]).

In mobile FSO communications we don't expect communication over long distances, so we don't have to worry about fog. But we must deal with changes in ambient light that would cause interference in the form of noise [2], and with variations in the intensity of the communication channel due to reciprocal movement of the receiver and transmitter. These factors are accompanied by constant changes in the intensity of radiation incident on the photodetector.

In light of these facts, will increase demands on the receiver sensitivity and the transmitter optical power. This creates a problem in the form of higher energy consumption and reduces the mobility of device.

It was the reason why for the initial experiments was rebuilt radio controlled model of the car to FSO controlled model. This combination allows us to quickly and easily find the basic problems with movement in various environments.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 614–618.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Model Construction

As the basis of design, we used RC model of car, which was deprived of original electronics with the exception of motor for propulsion and servomotors for steering.

Model is now completely controlled by a newly designed electronics on the principle of FSO.

2.1 Transmitter

In order to control the servomotors installed in the model, we need to send signals compatible with their electronics, which means PWM modulation.

On the next figure we can see how the communication frame for single servo looks like.

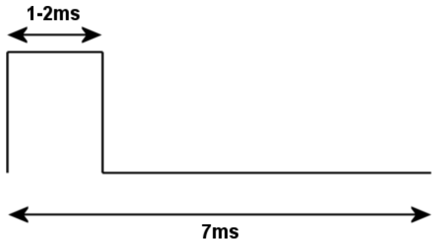


Fig. 1. Communication frame for PWM controlled servo.

Frame is 7ms long and servo position is controlled by pulse of width 1 to 2ms, voltage levels correspond to the TTL logic.

Because there are 2 servomotors, we must solve the problem how to transmit both signals simultaneously. Thanks to the shape of communication frame, we can move each other in time by 3,5 milliseconds. As a result we are able to easily transmit both signals in one optical channel. In order to allow the separation of the signals at the receiving side, we must choose different frequencies of carriers for each servomotor.

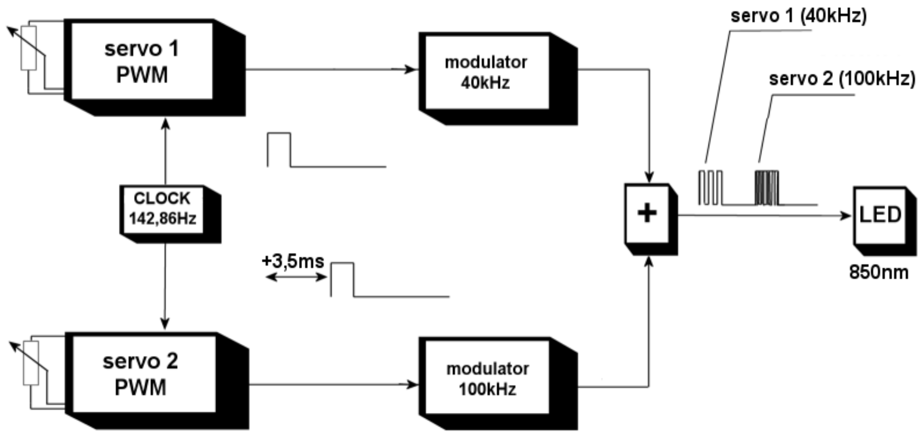


Fig. 2. Modulator block diagram.

2.2 Receiver

Main part of receiver is photodiode, which is connected in the reverse direction (photoconductive mode). This allows faster response to incident light. For useful signal separation are used two band-pass filters tuned to the frequencies of corresponding servomotors. This filtration is also useful for the removal of basic interferences from environment (fluorescent tubes, etc.). After basic amplification and filtration comes the automatic gain control (AGC), which ensures a constant output voltage even if optical power incident on the photodetector vary in time.

The system also provides a useful feature that in the case of weak signal reception causes a shortening of pulses. With this system, servomotors stop the model in bad signal reception and wait for the moment when the signal will be sufficient to control.

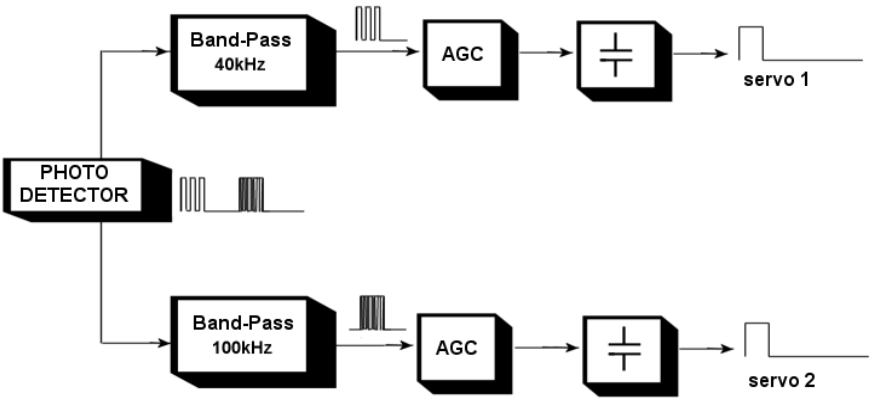


Fig. 3. Demodulator block diagram.

3 Results

In the current state the model is able to work in indoor areas at a distance of about 6 meters while respecting LOS. In the outdoors, especially in direct sunlight, the range is slightly smaller due to greater outside interference.

Light from LED disperses conically and at any cross section of this cone can be emitted power calculated as [3]:

$$\frac{SA_R}{SA_T + \pi/4\,(\theta R)^2} \tag{1}$$

Where SA_R is the surface area of receiver, SA_T is the surface area of the transmitter, θ is divergence angle and R is range in meters.

If we look at the measured values, we find that on the receiving photodiode must fall at least 4 microwatts of useful optical power to ensure correct functionality. This value is relatively high, but it’s necessary to maintain the usable signal to noise ratio. It is most likely caused by the used photodiode (BPW34) which is not equipped with filter, which means that photodetector is sensitive to wavelength area 400 – 1100nm.

On the next picture (Fig. 4.) we see the spectral analysis of the transmission LEDs in comparison with the spectrum of sunlight and fluorescent lamps. If we compare the individual waveforms, we see that at a wavelength around 850nm have sunlight only small values, so the interferences should be lower than in the visible spectrum.

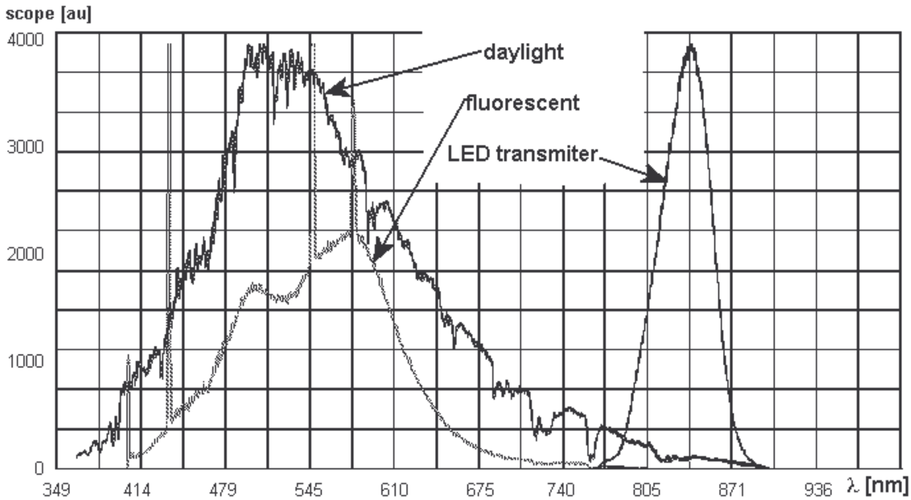


Fig. 4. Spectrum comparison of daylight, fluorescent lamp and LED transmitter.

4 Future Development

Aim is to adapt the system to cover the selected area without changing the direction of base stations while ensuring maximum mobility of mobile stations.

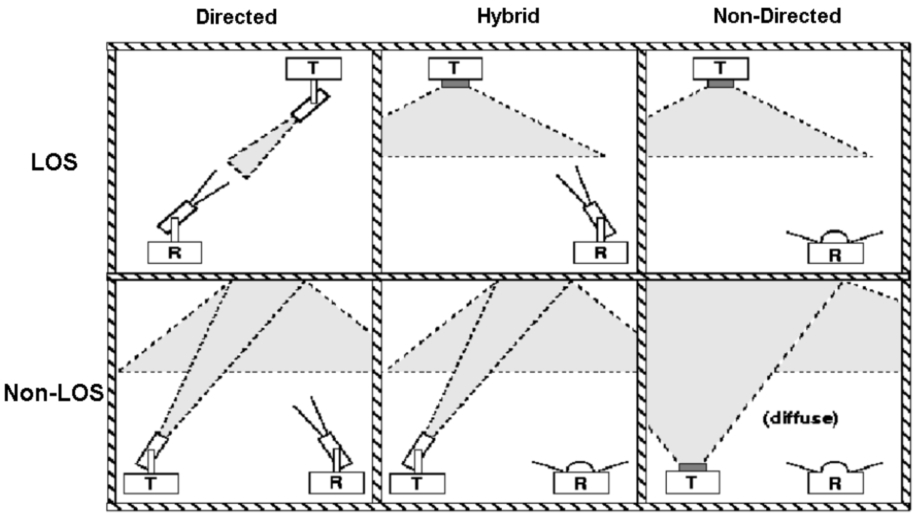


Fig. 5. Different optical coverage system realizations [4].

Basic configurations of communication systems are in the 5th figure (Fig. 5.) which shows differences in design for directional, diffuse and hybrid systems [4].

Further improvement is possible when using multi element receivers and transmitters [3], [5]. This provides much better coverage of surrounding area, because using of a single photodetector can't cover sufficiently wide area. This system however requires a sensitive management because at one time should not be too many elements operational. Otherwise there is a limitation of transmission speed due to multipath propagation. The same system applies also for multi-element transmitters. As a result we need a system that continuously selects the best receiver and transmitter to achieve the best system performance.

5 Conclusion

At the end of our basic experiments we have a function prototype for testing the coverage of mobile FSO network in selected areas. However, our system has currently only simplex data transmission capability.

Mobile FSO networks have great potential for future use in indoor environment, since they have compared with the radio networks several advantages:

- Precise definition of covered area.
- Unlicensed band.
- There is virtually no EMI, so those networks can be used in sensitive areas.

Of course there are also a few disadvantages.

- Higher energy consumption due to the large coverage area.
- Low bandwidth of power LEDs (only a few MHz).

References

1. Vitásek, J., Látal, J., Hejduk, S., Bocheza, J., Koudelka, P., Šiška, P., Skapa, J., Vašínek, V. Fog attenuation influence on Bit Error Ratio of Free Space Optics link. In *Knowledge in Telecommunication Technologies and Optics 2011 KTTO 2011*. 1st edition. Ostrava : VSB-Technical University of Ostrava, 2011
2. Hejduk, S., Vašínek, V., Bocheza, J., Vitásek, J., Látal, J., Ganiyev, A. Noise analysis of dark rooms and their evaluation. In *Knowledge in Telecommunication Technologies and Optics 2011 KTTO 2011*. 1st edition. Ostrava : VSB-Technical University of Ostrava, 2011
3. Jayasri Akella, Chang Liu, David Partyka, Murat Yuksel, Shivkumar Kalyanaraman, and Partha Dutta. Building Blocks for Mobile Free-Space-Optical Networks. Proceedings of IFIP/IEEE WOCN, Dubai, United Arab Emirates, March 2005
4. Adrian Mihaescu, Pascal Besnard: Indoor Wireless Optical Communications, 8th International Conference on Communications, Bucharest, Romania, 2010
5. Jaw-Luen Tang, Yao-Wen Chang, Design of an Omnidirectional Multibeam Transmitter for High-Speed Indoor Wireless Communications, EURASIP Journal on Wireless Communications and Networking, Volume 2010, Article ID 728468

IVAS System for The Indect Project

Lukas Kapicak and Pavel Nevlud

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{lukas.kapicak, pavel.nevlud}@vsb.cz

Abstract. The paper deals with multimedia support and remote control of Asterisk SW PBX. Main function of the Asterisk is connecting audio and video calls between two or multiple users. We developed technique for recording and playback of video files. Another our task in The INDECT project was remote control of Asterisk SW PBX. Remote control we developed over CLI interface and AMI interface. The IVAS system is using all of these interfaces to communicate to The INDECT portal.

Keywords: Indect, Asterisk, Video calls, CLI, AMI, SOAP.

1 Introduction

We are cooperating on the INDECT project [4] and we work on package 6. The main task of work package 6 is the INDECT portal. Our role is preparing multimedia functionality and interface for remote control of Asterisk through this portal. This task and project is called IVAS – Interactive Video Application System. IVAS is based on the open-source telephone exchange known as Asterisk.

Asterisk is software that can be installed on any personal computer, turning it into a communication server. Main functions of Asterisk include establishing audio and video calls, serving as a voicemail VoIP gateway etc. We extended Asterisk with recording and playback video files [1].

For developing of our remote control access to Asterisk through CLI and AMI interface we had chosen NetBeans. Netbeans is an open-source, JAVA based developing platform [8].

1.1 Testing configuration

We focused only on open-source software. Using of open-source software is one of the main conditions of the INDECT project. We used Asterisk 1.6 like telephone exchange [5]. For developing purposes we installed Netbeans in version 6.9.1 [8]. Hosting of our Web Services application [3] was provided by Glassfish server in

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 619–624.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

version 3.0.1. All of these applications needs JAVA runtime environment. Glassfish server could be situated on the same computer as Asterisk is or on a different computer in network. Communication works over TCP/IP connection so for local server we have to use local IP address.

For testing purpose we had chosen as end SIP client open source project Linphone. Linphone supports multiple operating systems, like Linux, Windows, MacOS, Android, Iphone. Linphone supports video with codecs: H263, H263-1998, MPEG4, theora and H264 with plugin x264.

2 Video recording and playback

Asterisk that enables recording and playback of video files has PULL and PUSH functions. The PULL function requires some end-user interaction, such as choosing a particular video. In this situation, the DTMF key input has to be applied to select a specific file. Mobile devices are equipped by web browser and through the web portal, users can establish video call to their devices.

The PUSH function, on the other hand, allows other users to choose a specific video file and subsequently routes these video files to a particular end-user or a group of the end-users. This function does not require any interaction by end users.

The app_mp4.c extension has to be installed into Asterisk. This extension consists of two components. First one is mp4play [2] that enables playing video files directly on end-users devices. The second is called mp4save and enables saving video files on a host´s disk or a different storage media.

2.1 Video recording

To record videos, we used smartphone with the Android operating system and the Ekiga softphone installed on a personal computer. To record video files, we can usually use front or rear facing cameras in mobile devices. Certain new HTC smartphones, however, are not equipped with a front-facing camera.

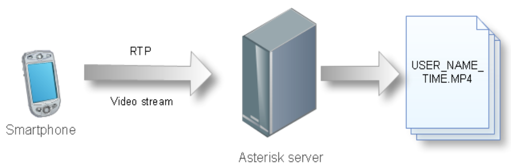


Fig. 1. Video recording

Fig. 8 depicts our testing configuration for video recording. The Asterisk server was the destination. The code below illustrates the `extension.conf` set up for an automatic recording of all video files.

```
exten => 784,1,Answer()
```

```
exten => 784,2,mp4save(/tmp/${CALLER(num)},D%d:M%m-
%H:%M:%S_%$}.mp4)
exten => 784,3,Hangup()
```

Recorded video files are tagged with caller's ID and time when the video call was established. The name of the video file will be used for choosing in database.

2.2 Video playback

Video playback can be initiated from both sides. Figure 2 shows its initialization by the user. The end-user chooses a particular video file through the DTMF key input. This technique is limited as regards the number of possible choices.

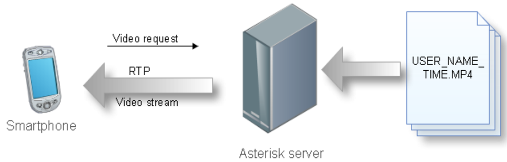


Fig. 2. Video playback

We show the `extension.conf` configuration file for video playback. This configuration file contains information about the number which the user will dial for the video playback from the server side. The name of the video file depends on the content of the saved video.

```
exten => 782,1,Answer()
exten => 782,2,mp4play(/tmp/video.mp4)
exten => 782,3,Hangup()
```

Fig. 3 shows the concept that we are working on at the moment. Selecting video files will be initiated through a website, and the requested video file will be chosen by a coordinator and PUSHed to the end-user device. All the user has to do is to answer the incoming call.

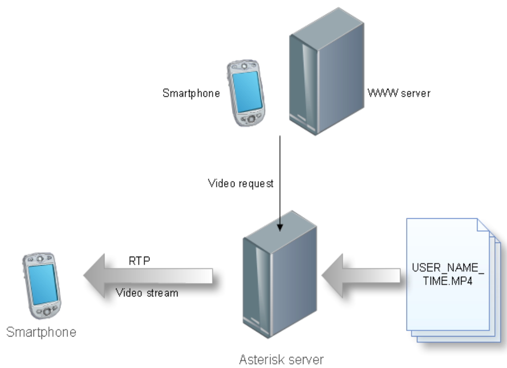


Fig. 3. Video playback third part initialization

2.3 Video playback initialized by web service

Web services server will be cooperating with the Asterisk and through SOAP [9] (link) protocol is web services server communicating with other web services servers or with web services clients. Asterisk has not have implemented SOAP interface so we have to develop techniques for cooperating Asterisk with web services servers and web services clients.

In figure 4 you can see architecture with web services server.

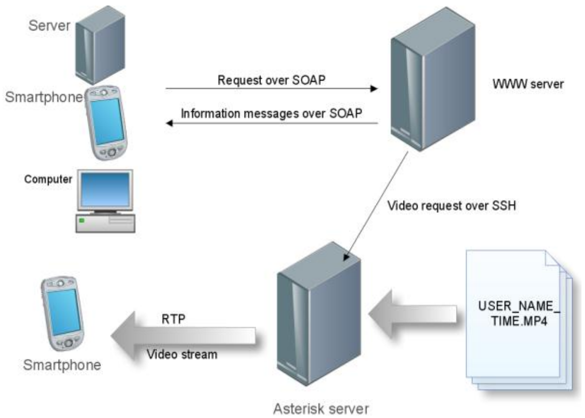


Fig. 4. Video playback with SOAP protocol

Using AMI (Asterisk Manager API) [6], we have implemented application that is able to connect to Asterisk server and issue various commands or read events over SSH. Application uses commands over CLI asterisk interface [7].

3 Remote control of Asterisk over SOAP protocol

Web service server can be connected to the Asterisk on local machine or on a remote computer. Information about connected users is performed if we fill the gap with the word “true”. After this we generated SOAP request message with specific code:

```
<S:Body>
    <ns2:GetOnLineUsers xmlns:ns2="http://WebServices.ws/">
        <getFullInfo>true</getFullInfo>
    </ns2:GetOnLineUsers>
</S:Body>
```

Response to this message consists of information about all connected users to the Asterisk. Structure of Response message is shown below:

```
        <ns2:GetOnLineUsersResponse
xmlns:ns2="http://WebServices.ws/">
                <return>
```

```
<currentIP>192.168.1.31:5060</currentIP>
                <online>true</online>
                <ping>1</ping>
                <userName>bob</userName>
            </return>
            <return>
<currentIP>192.168.1.34:47391</currentIP>
                <online>true</online>
                <ping>3</ping>
                <userName>dan</userName>
            </return>
        </ns2:GetOnLineUsersResponse>
    </S:Body>
```

Generating whole list of all users is provided with `GetAllUsers`. Action command is similar to command for generating list of online users. The difference is only in a tag. `GetOnLineUsers` which is substituted by `GetAllUsers`. Answer to this request is a list of all users. If users are online, it will generate response with IP address and ping.

3.1 Architecture of Remote control of Asterisk

In Figure 5 you can find architecture of our Web Services. User of web interface (Web Service requester) send request message to Web Services provider (with SOAP protocol [10]). Web service server analyzes this message and via AMI interface (on default TCP port 5038) send Action command to the Asterisk. Asterisk then perform this Action command and send Events message to the Web service provider and this server informs appropriate Web Service requester.

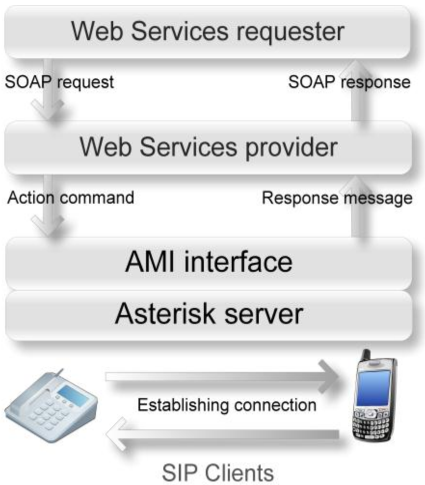


Fig. 5. Establishing connection between two SIP users.

4 Conclusion

Asterisk is communication server with support for switching voice and video calls through number of audio and video codecs. Asterisk with program mp4play records and plays video files saved on server storage directly in the end user equipment. From Asterisk log we even know if user received this video and then we have backwards control if user received video call from server.

Web Services in connection with Asterisk offers web management of Asterisk PBX. With this web management users are able to control Asterisk via SOAP messages. Our solution is primary designed for coordinators in office who can receive information about online users or they can connect two users or user with Asterisk. At the moment we have implemented 3 basic services but we are working on implementation of other services. Among these services belong streaming video on demand, streaming video from IP camera, connection of multiple users. These services will be offered by Asterisk and controlled by Web Services management.

References

1. Voznak, M., Kapicak, L., Zdralek, J., Nevlud, P., Rozhon, J.: Video files recording and playback with VoiceXML, 2011 10th WSEAS International Conference on EHAC'11 and ISPRA'11, 3rd WSEAS Int. Conf. on Nanotechnology, Nanotechnology'11, 6th WSEAS Int. Conf. on ICOAA'11, 2nd WSEAS Int.Conf. on IPLAFUN'11, pp. 350-354.
2. Kapicak, L., Zdralek, J., Nevlud, P.: Asterisk - video recording and playback, 2010 Knowledge in Telecommunication Technologies and Optics 2010, pp. 25-29.
3. Kapicak, L., Nevlud, P., Zdralek, J., Dubec, P., Plucar, J.: Remote Control of Asterisk via Web Services, TSP 2011, accepted to conference, presented will be in August in Budapest 2011.
4. INDECT Project http:/www.indect-project.eu
5. Meggelen, J., Smith, J., Madsen, L.: Asterisk: The Future of Telephony, O'Reilly Media, 2 edition, 2009, pp. 59 – 79.
6. Madsen, L., Bryant, R.: Asterisk Cookbook, O'Reilly Media, April 2011.
7. Voznak, M.: Voice over IP. VSB-Technical University of Ostrava: College book, 1st. ed., Ostrava, 2008.
8. Boudreau, T., Glick, J., Greene, S., Woehr, J., Spurlin, W.: Netbeans: The Definitive Guide, O´Reilly Media, 1. Edition, 2002, pp. 321-352.
9. Bean, J.: SOA and Web Services Interface Design: Principles, Techniques, and Standards, Morgan Kaufman, 2009, pp. 310 – 329.
10. Iverson, W.: Real World Web Services, O´Reilly Media, 1. Edition, 2004, pp. 101 - 126.

Communication Among Cars by Optical Free Space Networks

Petr Koudelka

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.koudelka@vsb.cz

Abstract. Safety of road users, reducing mortality on European roads, the growth of individual transport and reducing C02 emissions are some of the many questions that come to the fore in the issue of automotive industry. Currently, the only possible solution to these questions is the introduction of information-communication systems. Newly established organization dealing with this problem based their solutions on the utilization of wireless communications in the fields of radio waves. However, due to changes in the quality of lighting in cars through the implementation of LED technology, in conjunction with adaptive system of lighting, seem to be optical communication systems based on micro-cells networks as an interesting alternative.

Keywords: Automobile industry, PLED, V2V, C2C, V2V2I, LED.

1 Introduction

Over the last 30 years the automobile industry became the engine of many European countries, investments to the building factories and researching of new technologies for cars are increasing every year. With the increasing availability of buying a car and together with a dense network of expressways is increasing number of accidents caused by man. Accidents on roads and motorways in the European Union claiming 39 thousand lives in 2008. Czech Republic recorded a total of 1 076 deaths that year [1]. Insurance companies reckon the losses arising from accidents on the value of 160 billion euros, which is equivalent to 2% of EU GDP. By the year 2020 is expected to increase an individual transport of 32% [2]. Based on these alarming reports, the EU decided to significantly fund research and implementation of measures to reduce by half fatal road accidents by the year 2010, compared to the situation in 2001.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 625–631.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Intelligent cooperative systems V2V2I (Vehicle-to-Vehicle-to-Infrastructure)

Intelligent cooperative systems are based on a combination vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and traffic control center-to-infrastructure (TCC2I) showing promise of major improvements in both, efficiency of safe transport, as well as external security for all road users. The general scheme of system is presented in Figure 1 Intelligent cooperative systems increasing the time horizon, the quality and reliability of the information provided to driver on his / her surroundings, other vehicles and all other road users, allowing a better driving conditions that leads to greater security and more efficient and comfortable mobility. According to the communications consortium vehicle-vehicle (Car2Car Communications Consortia) V2V2I services are divided into three main groups [3]:

1st services related to security,
2nd services related to the optimal utilization of transport infrastructure,
3rd services related to other applications.

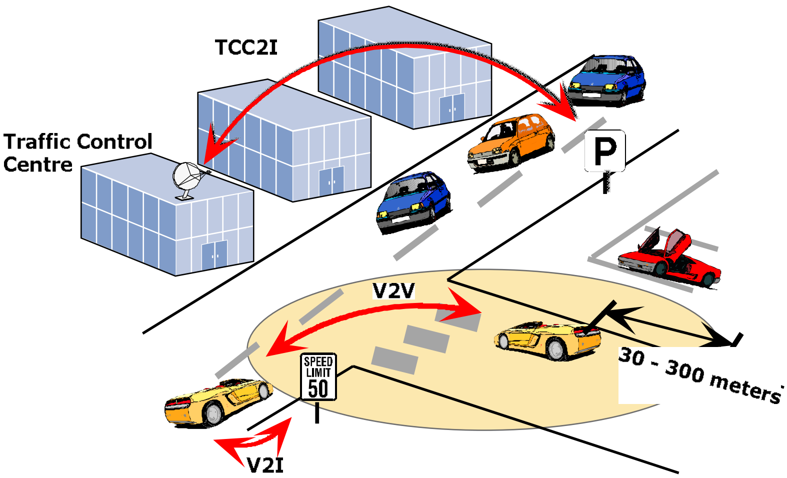


Fig. 1. Intelligent cooperative systems V2V2I

3 Communication Technologies of Intelligent cooperative system V2V2I

Organization dealing with this issue of intelligent cooperative system V2V2I built their solutions on the use of wireless communications in the fields of radio waves. In this area the European telecommunications standardization institute (ETSI) allocated the frequency band 5.9 GHz (the frequency region from 5.875 to 5.925 GHz). This band was derived from today available IEEE 802.11 communication system. This derived technology is called IEEE 802.11p (WAVE), described in the related documents as WLAN or CALM M5 and is divided into the following frequency bands:

- 10 MHz band from 5.885 to 5.895 GHz for network control and critical safety applications (control channel)
- 10 MHz band from 5.895 to 5.905 GHz for critical security applications, three 10 MHz bands from 5.875 to 5.885 GHz and from 5.905 to 5.925 GHz for applications to increase transport efficiency and safety-critical applications without time-consuming,
- two 10 MHz bands from 5.855 to 5.875 GHz for applications not corresponding to transmission security. Allocation of channels is shown in figure 2.

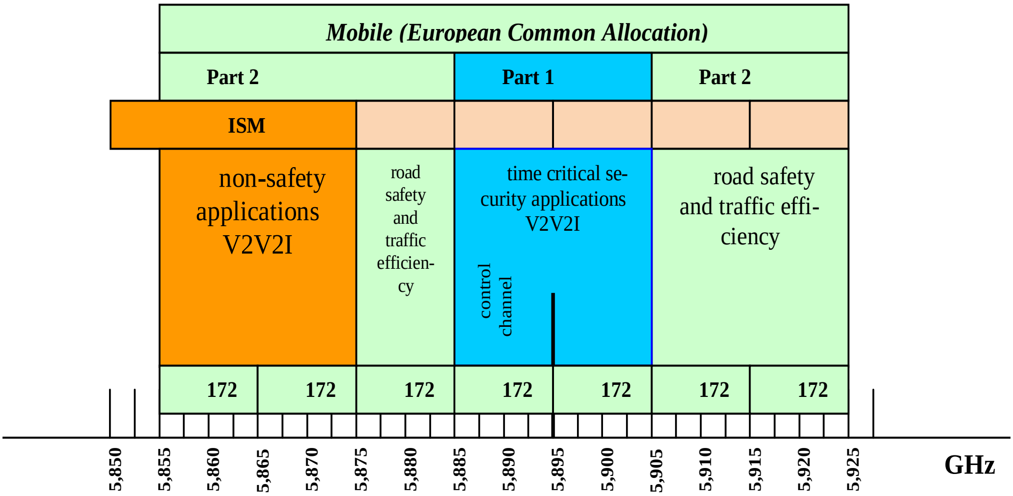


Fig. 2. Frequency bands allocated for V2V2I communication

Other communication technologies are based on the standardization called CALM (Continuous Air Interface for Long and Medium range). The subject of standards CALM is to provide a standardized set of wireless interface protocols and parameters for long and medium range, high-speed connection ITS using one or more media, with multipoint and network protocols and upper layer protocols. Defined communication technologies in addition to technology IEEE 802.11p are [4]:

- Cellular systems: GSM / HSCSD / GPRS (2/2.5G) and UMTS (3G),
- Wireless systems in 60 GHz band,
- Communication in the infrared optical spectrum,
- Wireless LAN (WLAN) at 5 GHz, IEEE 802.11a/b/g/n.

4 LED deployment in the automotive industry

As an alternative to radio waves based technologies, there are communication systems in the visible and infrared optical spectrum. A massive deployment of LEDs to the automotive industry speaks for the use of these systems. LEDs consume only about third of energy in providing the same level of lightning as traditional light sources. Energy conservation plays a big role in their deployment. In the evaluation by the U.S. Department of Energy 2003 was used information from the.

The LEDs are nowadays commonly encountered in luxury cars, they are used in the form of rear lights, but some car manufacturers used them as well in headlights applications - Figure 3, 4. Implementation of LED in products of the Japanese and European car manufacturers is around 40% of the light sources, while the massive deployment could be expected around 2011.

This year, the European Commission (EC) decided to enact a mandatory DRL (Daytime Running Lights) lights for new vehicles in all EU countries [5]. This type of lights should immediately emit light when starting the engine. Thanks to the rapid development of LED it is possible to assume that traditional bulbs completely disappear from new cars during 10 years. The advantage of LED over traditional light bulbs is clear. To achieve maximum output of LED are enough 3 ms, compared with traditional bulbs where the value is 200 ms. It is obvious advantage for stop lights when at speed of 40 km / h the time of 200 ms means saving 5 m stopping distance. Another advantage is the use of LED to transmit information without loading surroundings by electromagnetic smog and the risk of interference.

Fig. 3. LED based taillights

Fig. 4. Application of the DRL lights

4 LED deployment in the automotive industry

A major step in the field of road traffic safety is currently provision of communication between vehicles for the exchange of relevant information (V2V). An example might be when the vehicle A is moving and transmitting to moving vehicle B, that is behind the vehicle A, information such as the abrupt reduction of speed, speed, road conditions, information from ESP and steering angle, as shows Figure 5. This information can alert the driver of the vehicle B by acoustic or light signals or in other steps may be processed by independent information management functions. Such a temporarily created micro-cells optical networks could significantly eliminate the reaction time of man, which is around 1 s [6]. The estimated transmission rate for the transmission of important information is around the value of 400 kbps [7]. Measurements showed that they have bandwidth large enough to support data transmission in the range of Mbps.

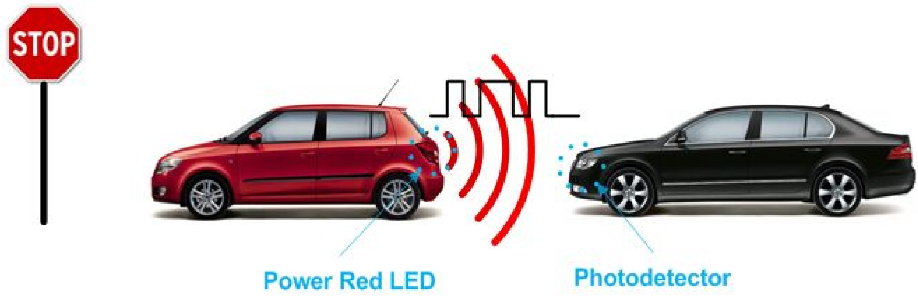


Fig. 5. Communications V2V in terms of the micro-cellular optical networks

In contrast with communications technologies such as IEEE 802.12p, which are able to communicate at distance from 500 m up to 800 m depending on the obstacles, require micro-cells optical networks always a direct visibility among transmitters. The worst conditions for these systems are also very sunny days, resulting in high levels of ambient lighting. This level is independent on the distance between vehicles. For optical communications the value of the light flux is reducing with the square of the distance and the received optical power fluctuates 5 decades up to 100 m distance by using a photodiode at the receiver with a spectral filter (FWHM = 50 mm). The receivers therefore require a wide dynamic range, high sensitivity and effective suppression of ambient light. In experiments it was found that such a temporarily formed network is capable of data transfer 1 Mbps over a distance up to 130 m. In order to verify power LEDs in laboratory conditions it was found that due to modulation is slightly reduced the value of light intensity as shown in figure 6.

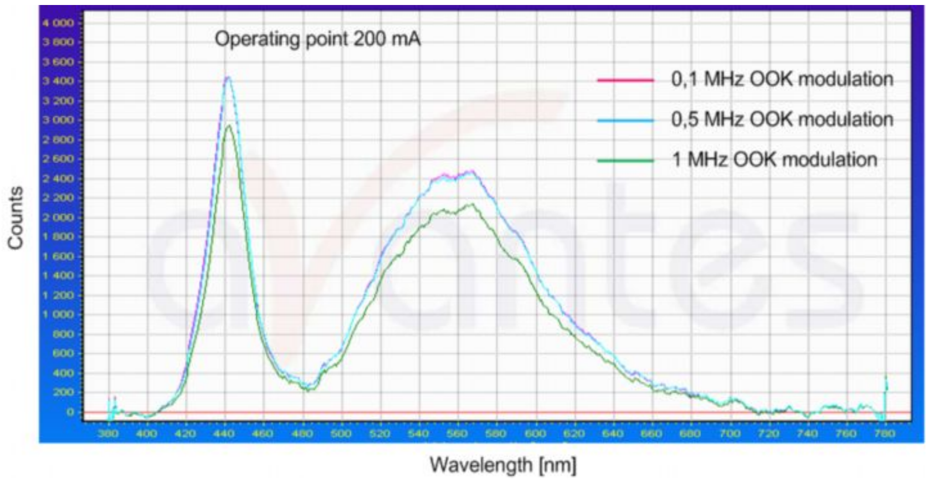


Fig. 6. Intensity change on power LED modulation value

5 Conclusion

Optical micro-cells communications are an alternative to radio technologies in the field of information exchange V2V. By organization GeoNet, which deals with the deployment of 802.11p technology to the automotive industry in the area of V2V and V2I, is clear that the radio system itself leads to congestion and interference of the free channels at high traffic density. Solution is offered in the form of power-controllable adaptive algorithms FPAV D-[8], however for unicast communication seems to be preferable micro-cells optical networks. Whole V2V2I system therefore should be hybrid. Dominant technology ensuring communication V2V and V2I would be standard IEEE 802.11p technology, but in situations like are long columns of cars, traffic jams or two close behind moving vehicles would be the primary information data from short-term generated vehicle-vehicle micro-cell optical network [9].

References

1. ETSC European Transport Safety Council official web site, <http://www.etsc.eu>
2. ENTRAC. Research framework, April 2006, <http://www.entrac.org>
3. C2C-CC Manifesto v. 1.1, August 2007, < http://www.car-2-car.org>
4. CEN – Evropský výbor pro normalizaci, INFSO G4/JJ D(2006) 701311, leden 2006, <http://www.silmos.cz>
5. LEDs Magazine, DRL-LED technology, September 2008, <http://www.ledsmagazine.com>
6. P. Reith and T. Eberz: Reduction of Stopping Distance Trough Chassis System Networking, Convergence 2002 Proceedings, Dietroit, 2002

7. T. Schaal and E. Zeeb: Optical Free Space Communication with LED Rear Lights, 5th International Symposium Progress in Automobile Lighting, Proc. PAL2003, pp.942-954, Darmstadt, 2003
8. A. Kovacs: Resource Sharing Principes for Vehicular Communications, EFKON Graz, Austria, 2008
9. GeoNET official web site, <http://www.geonet-project.eu>

Optical Power Measurement in Free Space Optical Link

Jan Látal and Tomáš David

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.latal@vsb.cz

Abstract. In our over-engineered world, the information is more valuable, the access to them is very important. Therefore it is necessary to develop new, faster and better ways of data transmission. One of the fast accesses is a Free Space Optical links, which are very secure against eavesdropping, but their disadvantage is dependence on atmospheric effects. Thos article summarizes some finding from measurement of optical power distribution in beam profile of Free Space Optical link, which is placed on the rector department and on the observatory.

Keywords: FSO, optical power, optical beam, FSO link.

1 Introduction

The telecommunication networks are still evolving and trying to meet the requirements of telecommunication services from private to commercial sector. This causes an unusual burden on telecommunications networks and their operators. Without alternative network technological systems, the overall effective data rate would be slowed. An alternative is a Free Space Optical link which is great at data rate and versatility. Free Space Optical link includes optical link between ground stations (buildings, poles), satellites and platforms in highs. The laser beam propagation in atmosphere is interesting for its variety of incident technical problems. The laser beam is impressed with atmospherically phenomena. These atmospherically phenomena are rain, snow, storm, and wind, whirlwind and atmospherically turbulences. This article primary aims for Gaussian beam description and measurement of optical power distribution in beam profile of Free Space Optical link, which is placed on the rector department and on the observatory.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 632–638.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 Laser beam

The wave character of light discharges idealization, in which the light is spatial concentrated and propagates without beam divergence. In Free Space Optics the Gaussian beam is frequently used, eventually top-hat beam or border beams.

3 Gaussian beam

The Gaussian beam is the most important beam of basic kind of beams for us, because just this kind of beam is emitted the most often from laser diode. The optical intensity distribution of Gaussian beam in cross plane comes up to Gaussian function, where the optical axis is also axis of symmetry. Optical power of Gaussian beam is concentrated in narrow cone. The beam width is is minimal in beam waist and grows up in both directions. The beam half-width *w(z)* is described by equation

$$w(z) = w_0 \cdot \sqrt{1 + \left(\frac{z}{z_0}\right)^2}$$

(1)

where w_0 is beam half-width in its the narrowest point, z is the coordinate of optical axis, z_0 is so-called Rayleigh distance. This is described as a distance in longitudinal direction of wave propagation from beam waist to point, where the cross plane is double. It deals

$$z_0 = \frac{\pi \cdot w_0^2}{\lambda}$$

(2)

,where λ is wavelength. The beam divergence θ is described by relation

$$\theta = \frac{\lambda}{\pi \cdot w_0}$$

(3)

For the radius of curvature of the beam *R(z)* deals

$$R(z) = z \cdot \left[1 + \left(\frac{z_0}{z}\right)^2\right]$$

(4)

4 Optical intensity of Gaussian beam

Optical intensity is a function of axial (axis direction) distance z and radial (radius direction) distance ρ. Radial distance is express as

$$\rho = \sqrt{\left(x^2 + y^2\right)}$$

(5)

Then the optical intensity of Gaussian beam is

$$I(\rho, z) = I_0 \cdot \left(\frac{w_0}{w(z)} \right)^2 \cdot e^{\left(-\frac{2 \cdot \rho^2}{w^2(z)} \right)}$$

(6)

Optical intensity maximum of Gaussian beam is in the middle of beam (axial distance *z* and radial distance *ρ* is equal 0). Gaussian beam radius *w(z)* grows up with increasing axial distance z. Optical intensity of Gaussian beam decreases with increasing axial distance *z* [1], [2].

5 Optical power of Gaussian beam

The total power carried by beam is determined by an integral of product of optical intensity and area of cross section of optical beam.

$$P = \int_0^\infty I(\rho, z) \cdot 2 \cdot \pi \cdot \rho \, d\rho \, . \qquad (7)$$

Where of is derived

$$P = \frac{1}{2} \cdot I_0 \cdot \left(\pi \cdot w_0^2 \right) \qquad (8)$$

The optical power of Gaussian beam is set as product of a half of maximal optical intensity and circle area with radius which is equal beam half-width.

8 Atmospheric transmission media

Free Space Optical Links are especially used in atmospheric transmission media but they are used indoor too. The atmospheric transmission media could be disadvantageous for Free Space Optical Link. In contrary conditions the data transmission could be troublesome, as far as impossible. Free Space Optical Links work in the troposphere. For the troposphere is a characteristic that water vapor condensates, many different influences generate there, which influence quality of the link. It could be rain, snow, whirlwind and fog. Sometimes the beam could be interrupt by birds. The troposphere is generally non-stationary and inhomogeneous media (it is supposed that these media is dielectric, linear, non-dispersive, and isotropic) a influence of these media on quality of transmission channel has an accidental character [4].

The main phenomena in atmospheric transmission media is [3]:

- extinction of optical intensity owing to turbulences in atmospherically media,

- extinction owing to dispersion and absorption on molecules and aerosols,
- fluctuation of optical intensity owing to turbulences in atmospherically media,
- fluctuation of optical intensity owing to fog, rain, snow,
- short interrupt of beam by birds.

9 Measuring installation for optical power measurement in Free Space Optical Link

For this task it was lent a measuring installation (fig. 1) from Brno University of Technology, which had to be set in working order.

This is a measuring installation produced by Festo Company with two axis with stepping motors. A measuring installation includes silicon detector, which measured optical power emitted by optical link from observatory, optical power meter, stepping motors with control units. In Matlab it was written software for control the stepping motors. The software enables to change measured wavelength, stepping rate, stepping size, time delay between steps etc. The communication between control unit and computer realized a serial bus RS-232.

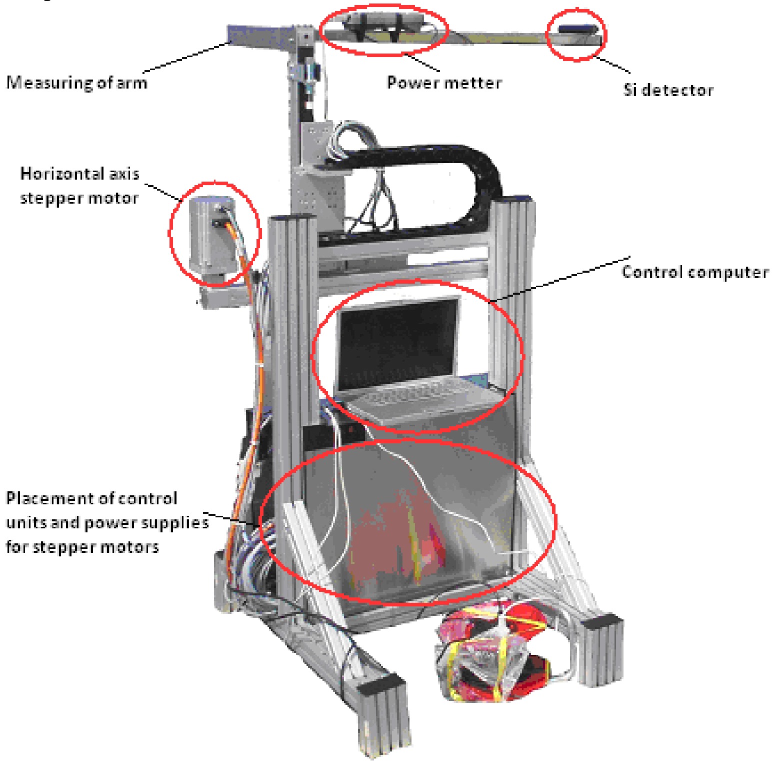


Fig. 1. Measuring installation

10 Optical power measurement in Free Space Optical Link

During the first measurements it was found that ambient sunlight very affected the measurements even if the optical power meter was shadowed in black pipe against direct sunlight. It was done several measurements at different wavelengths, at which the optical link works (830 - 860 nm), but the measurements were very affected by ambient light (fig. 2). Optical power values during a clear day were in range 40 - 60 μW. At night it was measured values around 1 μW.

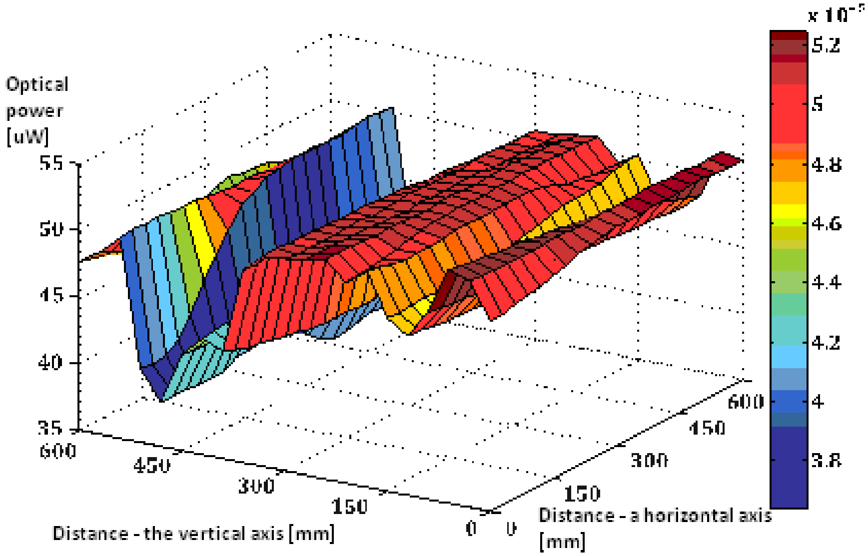


Fig. 2. Optical power measured 16.4.2011 at 14.00 hour [6]

Figure 4 was measured on 16. 4. 2011 at 14.00 hour, temperature 12,8°C, humidity 46%, northwest wind speed of 2.2 m/s, static pressure 1022,5 hPa. In the figure can be seen great optical power due to ambient sunlight. The waves in the figure are caused by ambient sunlight changes which are not so expressive in horizontal axis like in vertical axis. This effect comes up by fast fluctuation of optical power due to measurement method in raster.

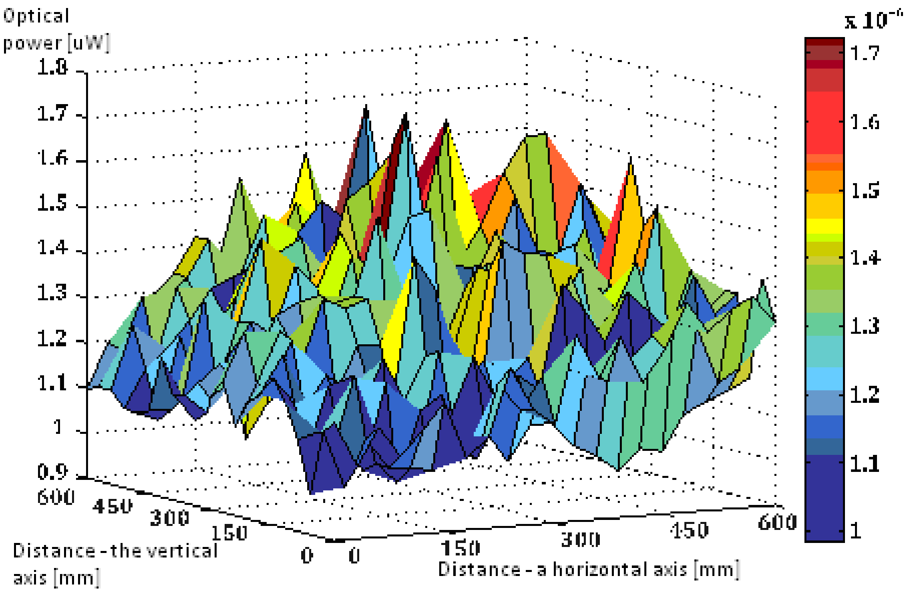


Fig. 3. Optical power measured 16.4.2011 at 22.00 hour [6]

In figure 3, which is from the same day like figure 2, but data were measured at 22.00 hour, is seen the fluctuation of optical power meter. The measurement conditions were following: set wavelength for optical power measurement lambda = 840 nm, air temperature 9,4°C, humidity 58%, northwest wind speed of 2.2 m/s, static pressure 1022,7 hPa. Received optical power is very low in comparison to measurement in day; its value is around 1 uW. The middle point of raster in figure 3 corresponds to the middle point of transmitted beam. In the figure 3 it is seen the rounded shape of beam with the greatest optical intensity in the middle point. Figure 3 displays many peaks with great intensity. This is caused by fluctuation of beam by transmission during optical media.

11 Conclusion

Measurements, which were done, showed that the shape of Gaussian beam was strong deformed by transmission during optical media. The beam was emitted from optical head placed on rector department building A, it was transmit to the optical head placed on observatory and it was transmit back to optical head placed on rector department building A. The deformation is caused by several phenomena, which are described above. In the future we would like to aim for optical sources testing - lasers, which are used for light emitting.

This work was supported in part by GAČR GA102/09/0550 - Optical beams study for atmospheric static and mobile communications a SP2011/47.

References

[1] KULHAVÝ, K. Root.cz [online]. 2002-11-05 [cit. 2011-02-18]. *Bezvláknová optika*. Available at WWW: <http://www.root.cz/clanky/bezvlaknova-optika/>. [web site].

[2] CÍSAŘ, D. *Vliv atmosférických turbulencí na intenzitní profil laserového svazku*. Brno, 2009. 42 s. Bachelor thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication. [academic thesis].

[3] BÁRTA, M. *Vliv atmosférických turbulencí na optický svazek*. Brno, 2009. 92 s. Diploma thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication. [academic thesis].

[4] LÁTAL, J; KOUDELKA, P; VITÁSEK, J. *Měření vlivu velikosti překážek na přenos optickým pojítkem*. Elektrorevue [online]. 6.6.2010, 2010/39, [cit. 2011-03-25]. Available at WWW: <http://elektrorevue.cz/cz/download/mereni-vlivu-velikosti-prekazek-na-prenos-optickym-pojitkem/>. [e-article].

[5] ŠÁLA, Jakub. *Bezdrátový optický spoj*. Brno, 2010. 70 s. Diploma thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication. [academic thesis].

[6] DAVID, T. *Testovací měření na profesionálním atmosférickém optickém spoji*. Ostrava, 2010. 97 s. Bakalářská práce. Vysoká škola báňská - Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky.

Embedded Solution of SIP Communication Server

Karel Tomala, Jiri Vychodil, and Jan Rozhon

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{karel.tomala, jiri.vychodil, jan.rozhon}@vsb.cz

Abstract. This article discusses the development and implementation of embedded solution of SIP communication server with an easy integration into the computer network based on open-source tools. This device is a PBX system for IP telephony, suitable for small and medium sized organizations. Hardware device, ie the software installed on optimized proprietary hardware, can be plugged into an existing network just before the operation. It also provides an interface for administrators and end users, futhermore by using a standard VoIP protocol the system is compatible with any device and service provider of SIP. The article describes the system design and the individual elements, which are contained in the solution.

Keywords: Embedded Solution, Monitoring, OpenWRT, Security, SIP.

1 Introduction

Many large institutions operates a small offices with tens or hundreds of employees. A common requirement is the full integration of these departments in the organization's environment (examples are libraries and branch offices). With our proposed solution, which has the working title BESIP (Bright Embedded Solution for IP Telephony), the integration can be achieved easily with the use of IP telephony and supporting network infrastructure. The device is designed as a price acceptable solution that outside support SIP (Session Initiation Protocol) IP telephony also includes support services such as ENUM [1], secure communication using SRTP and TLS [2], monitoring of call quality, tools for detection and elimination attacks, billing and clear configuration via a web interface. Since the whole kernel is designed and implemented as open-source system can be further extended with additional modules.

Systems for integration and coverage of field offices are now fairly widespread. But, few of them offer modularity on such levels as our proposed solution, moreover during development our aim was to develop create a device that will be much cheaper than similar commercial products.

The following chapters refer about scheme of the system in more details, individual modules and applications that were used in the implementation are mentioned and analysis of the used hardware is described too.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 639–644.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

2 System Scheme, Modules and Applications

As mentioned above, the BESIP system is a modular solution where each element consists of several applications which are supported by core (Fig. 1). Modules are divided according to the function which they perform at:

- Core
- Security Module
- Monitoring Module
- PBX Module
- Module of services

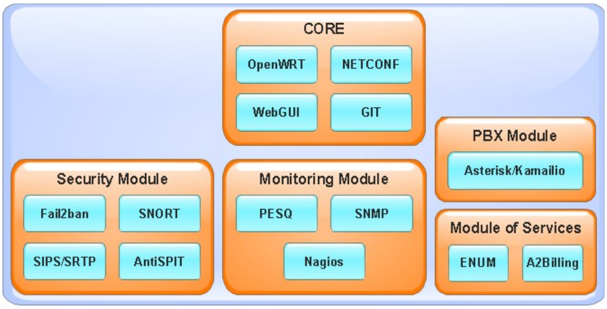


Fig. 1. BESIP Divided into Modules.

2.1 Core

The core of the system consists of the Linux distribution OpenWRT [3], which is directly designed for embedded devices and has very low demands on computing power. Instead of trying to create a single, static firmware, OpenWrt provides a fully writable filesystem with package management. This frees you from the application selection and configuration provided by the vendor and allows you to customize the device through the use of packages to suit any application.

To manage updates and revisions, we are using Git [4].Git is a free & open source, distributed version control system designed to handle everything from small to very large projects with speed and efficiency.

Another tool that is part of the core is the NETCONF protocol [5]. NETCONF provides mechanisms to install, manipulate, and delete the configuration of network devices. Its operations are realized on top of a simple RPC (Remote Procedure Call) layer. The NETCONF protocol uses an XML (Extensible Markup Language) based data encoding for the configuration data as well as the protocol messages. This in turn is realized on top of the transport protocol.

The last part of the Core Module is Web GUI support. This is done using Lighttpd [6], which has a small memory demands and is therefore suitable for embedded devices. Using the PHP interpreter also allows to generate AJAX and Java scripts, which provides better visibility and easy operation for end users.

2.2 Security Module

Security module is responsible for protecting the system itself against attacks from external subjects, as well as the analysis of these threats, and last but not least the module is responsible for signaling and media encryption. The protection of system against threats provides application Fail2ban. It is a tool which is able to block IP addresses in the firewall based on the logs scan. It can control a wide range of applications, depending on the set of configuration files. The regular expression is also included, in the configuration file, whereby Fail2ban detects suspicious activity. Blocking is subsequently performed through the rules of iptables.

Another way to protect the system against attacks directed at an IP telephony service is the implementation of Snort [7]. Snort is an open source network intrusion prevention and detection system (IPS / IDS), developed by Sourcefire. Combining the benefits of signature, protocol, and anomaly-based inspection, Snort is the most widely deployed IDS / IPS technology worldwide. It also enables flexible creation of rules by which the system then decides the next steps.

Security experts predict that SPIT (Spam over Internet Telephony) is a major threat in the future. The level of annoying factor is even greater than classical Spam. Imagine if we receive 5 Spam emails per day. We delete them and most of us have to take this as a daily routine. Now instead of emails we will receive call on the IP phone 5 times per day. Inconvenient thought. AntiSPIT is able to analyse and process input data from Call Detail Records (CDR’s) and consequently determine whether the used source will be inserted into blacklist [8]. CDR's are an integral part of every PBX and we decided to implement AntiSPIT into BESIP system.

The last element of the security module is the ability to encrypt calls using SRTP and TLS protocol. This security is ensured directly by communication server, in our case, the SIP PBX Asterisk in version 1.8.4.4 [9].

2.3 Monitoring Module

This module is able to monitor the speech quality for individual IP calls, as well as provide other monitoring of network devices in the network using SNMP [10] and Nagios tools [11]. In the case of monitoring the speech quality we are using our developed application for the Asterisk PBX, which determines the scale of speech quality – by using MOS (Mean Opinion Score) [12] (Fig. 2). This applications uses the PESQ (Perceptual Evaluation Speech Quality) [13] model. It is based on comparing the original signal with the degraded sample.

Fig. 2. Mean Opinion Score Scale.

Because even at the remote office can arise requirement for monitoring of network devices, BESIP offers several tools for analysis of network infrastructure and elements that are in it. The first tool is Nagios. Nagios is a popular open source computer system and network monitoring software application. It watches hosts and services, alerting users When things go wrong again and When They get better. Another is support for SNMP protocol, which was developed for the direct management of individual devices on the network and for communication between them.

2.4 PBX Module

This module is one of the most important of the entire system, as it contains the actual communication server for IP telephony calls. We have used a robust open source solution – Asterisk in the version 1.8.4.4, which offers a wide range of services with low processing power demands. All other modules are interconnected with Asterisk because of sending or retrieving data, Asterisk is also responsible for encrypting and comparing call quality using algorithm developed by us which is based on PESQ method.

Since BESIP system is still under development it was also considered the usage of a Kamailio [14] communication system. It is, however very difficult to configure it correctly, so the system will contain it only on the wishes of the end user.

2.5 Module of Services

Module of services contains tools for providing an additional services, such as billing or ENUM. End user can of course also define additional services that he/she needs, but above mentioned are part of the system by default.

Telephone Number Mapping (ENUM) is the process of unifying the telephone number system of the public switched telephone network with the Internet addressing and identification name spaces. Telephone numbers are systematically organized in the E.164 standard, while the Internet uses the Domain Name System for linking domain names to IP addresses and other resource information. Telephone number mapping systems provide facilities to determine applicable Internet communications servers responsible for servicing a given telephone number by simple lookups in the Domain Name System.

Another service that is supported by the BESIP System is A2Billing application. It is used for billing and monitoring operations in software PBX. In our case, the communications server can create a record of the call named CDR (Call Detail Record), but as such it is unable to process it. It cannot comfortably create, edit and process a huge number of participants. To eliminate these problems, Asterisk has a powerful mechanism for acquiring and storing data to and from databases. The MySQL database is used in BESIP system, from which then A2Billing analyzes the data and based on rules it charge it. Fig. 3 shows the connection of Asterisk PBX with A2Billing.

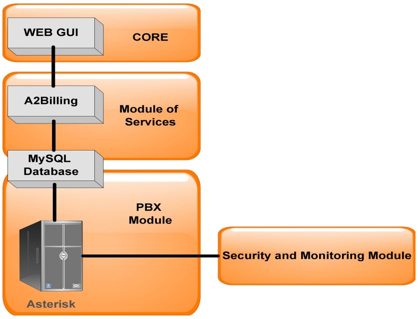


Fig. 3. Asterisk connection with A2Billing.

3 Hardware

Since the beginning of the development, the BESIP was planned as the most mobile, portable and especially low cost device. It was necessary to adapt to these conditions not only from the software point of view but also hardware on which the system will run. Outside of these conditions it also had to offer sufficient computing power for smooth operation of all modules, applications and participants. Another important factor was also low power consumption and reserve capacity for future services.

After a series of tests and analysis standard desktop PC with Atom processor was chosen. It consists of the following configuration:

Motherboard: Intel Packton D410PT
CPU: x86 Intel Atom D410 - 1,66 GHz
Memory: Kingston 1GB 667MHz CL 5
HDD: Kingston 16GB SSD Now S100
Case: Eurocase Mini ITX Wi-05

The price of this product is in good relative between price and performance. Furthermore, it is an widespread x86 architecture, which allows to install the well-known OS. Figure 5 shows the selected device.

Fig. 5. BESIP Hardware Solution.

4 Conclusion

We have developed and implemented a system with the working title BESIP, which allows easy integration of SIP IP telephony infrastructure to branch offices of large companies. The main advantages of this system are its low cost, high scalability and the possibility of adding or configuring a new or existing tools and the relatively small size and low power consumption.

Thanks to these features, the device will be deployed in real network traffic in our association, which will also be good opportunity to continue with actively development of the software product and where it will also be an intensive testing to debugging the solution.

References

1. Bradner, S., Conroy, L., Fujiwara, K.: The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM). RFC 6116, http://tools.ietf.org/html/rfc6116
2. Thermos, P., Takanen, A.: Securing VoIP Networks: Threats, Vulnerabilities, and Countermeasures. Addison-Wesley Professional; 1 edition, 2007, ISBN-13: 978-0321437341
3. Surhone, L. M., Tennoe, M. T.: OpenWrt. Betascript Publishing, 2011, ISBN-13: 978-6135271591
4. Loeliger, J.: Version Control with Git: Powerful Tools and Techniques for Collaborative Software Development. O'Reilly Media; 1 edition, 2009, ISBN-13: 978-0596520120
5. Enns, R.: NETCONF Configuration Protocol. RFC 474, http://tools.ietf.org/html/rfc4741
6. Bogus, A.: Lighttpd. Packt Publishing, 2008, ISBN-13: 978-1847192103
7. Voznak, M., Halas, M., Rezac, F.: Delay Variation Model for RTP Flows in Network with Priority Queueing. In Recent Researchers in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics - ISPRA '11 , pp. 344-349, Cambridge, UK, February 20-25, 2011, ISSN 1792-8133, ISBN 978-960-474-276-9.
8. Voznak, M., Rezac, F.: The implementation of SPAM over Internet telephony and a defence against this attack. 32nd International Conference TELECOMMUNICATIONS AND SIGNAL PROCESSING - 2009.Dunakiliti, Hungary, 26-27 8. 2009, ISBN 978-963-06-7716-5
9. Wintermezer, S., Bosch, S.: Practical Asterisk 1.4 and 1.6: From Beginner to Expert. Addison-Wesley Professional; 1 edition, 2009, ISBN-13: 978-0321525666
10. Case, J., Fedor, M., Schoffstall, M., Davin, J.: A Simple Network Management Protocol (SNMP). RFC 1157, http://www.faqs.org/rfcs/rfc1157.html
11. Barth, W.: Nagios: System and Network Monitoring. No Starch Press; Second Edition, 2008, ISBN-13: 978-1593271794
12. ITU-T: P.800 : Methods for subjective determination of transmission quality. ITU-T Recommendation, http://www.itu.int/rec/T-REC-P.800-199608-I/en
13. ITU-T: P.862 : Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs”, ITU-T Recommendation, http://www.itu.int/rec/T-REC-P.862/en
14. Goncalves, F. E.: Building Telephony Systems with OpenSIPS 1.6. Packt Publishing, 2010, ISBN-13: 978-1849510745

OptiSystem in E-learning Presentation

Jan Vitásek

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.vitasek@vsb.cz

Abstract. Nowadays there are several software which allow us to scan monitor and record mouse motion, writing text, objects moving etc. These software can be used to create detail service manual of applications. These software speed very up teachers work because each student can start the final video multiple times and he can better learn applications. In this article there is described Adobe® Captivate® 4 software.

1 Introduction

Software Adobe Captivate serves quick professional e-learning material creation with advanced interactivity [1]. This is an ideal instrument for interactive multimedia tutorials including testing and connection to Learning Management System (LMS). The special tools for simulations creation and all kind of software products controlling markedly extend the possibilities of this software in education. Due to publish in universal formats Adobe Flash and Adobe PDF the education materials are available to anybody in the whole world.

2 Adobe Captivate

It is very easy to create an impressive instructional presentation or work simulation in some application. Captivate automatically safe all actions which are carried out in recording mode on the screen. In editing mode it is possible further change, edit and add the other components into presentation or simulation [2]. Software includes.

- text caption, graphics, spoken comments and music,
- Flash animation and video,
- interactive elements (buttons, references, context help),
- knowledge tests and their evaluation.

After recording and editing it is possible to publish the presentation in Flash (swf) format and send it as email, safe in FTP server or use in web sides. The other alternative is safe in Windows executable file (exe).

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 645–651.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

We would like to stress that the class/style files and the template should not be manipulated and that the guidelines regarding font sizes and format should be adhered to. This is to ensure that the end product is as homogeneous as possible.

3 User interface

Captivate automatically records all actions which are done in an application on computer screen. There are three recording modes – automatic recording, full motion recording and manual recording.

3.1 Automatic recording mode

Adobe Captivate captures screenshots automatically and places them on separate slides. Mouse, keyboard, or system events are the common triggers for capturing screenshots. Automatic recording is the most commonly used recording method in Adobe Captivate [2]. Actions like drag-and-drop and mouse movements are automatically captured in the full motion recording mode. The recording is stopped by press End key.

3.2 Full motion recording mode

During full motion recording (FMR), the entire set of events is captured in real time as a video. Use FMR for movies that demonstrate complex procedures, like drawing or reshaping an object. FMR movies are also useful when demonstrating visual cues, such as the change in the shape of the pointer when it moves over certain objects. The recording is stopped by press End key [2].

3.3 Manual recording mode

The presentation can be constructed by manually taking screenshots during recording. The manual recording is used when it should be picked and chosen a few screenshots during the recording process. The procedure can get tedious for complicated procedures involving many steps. The recording is stopped by press End key [2].

In recording modes Captivate automatically generates captions or other effects like highlighting. The users can create, change or erase captions and other effects. Summary of features:

- automatic recording all actions which are done in some application including cursor motion and keyboard typing,
- intelligent full motion recording for video sequence recording,
- contemporary spoken comments recording,
- automatic creating captions to actions on the screen,
- set recording area.

4 Editing

After recording the list of all slides appears like filmstrip in Captivate editor. It is possible edit (add captions, sounds, videos, graphics etc.), move or erase each slide. Spoken comments and inserted sounds are editable in inbuilt sound editor. The great advantage is that it is not necessary record all slides again when some error appears or when slides sequence is incorrect [2].

- Transparent orientation in record filmstrip,
- unlimited possibilities of slides editing,
- add own captions, graphics, sounds, spoken comments, videos,
- timeline for exact synchronization of effects,
- inbuilt sound editor to edit sounds and spoken comments,
- full support Unicode for Multilanguage texts.

5 FSO link simulation in OptiSystem

OptiSystem is an application which was used for presentation. OptiSystem is an innovative optical communication system simulation package that designs, tests, and optimizes virtually any type of optical link from the physical layer to the transport layer according OSI model [3].

This application is open to our students, mainly to bachelor and master students. The students design and test some network therefore they need to meet OptiSystem and hereto there is used Captivate. The Captivate is used in lessons of Optoelectronics or Optical atmospheric communications.

5.1 Create new project

If Captivate 4 is run it appears dialog box where it is possible to choose open recent project, create new project, create template or read tutorials.

If it is chosen the Software simulation, it is opened new dialog box, where Screen area is chosen. There are two possibilities: select application or select screen area. After click on pop-up menu the all running programs appear to choice. Choice running program is better variant. At first the whole presentation should be made and then the size could be changed. The size change is nonreversible! Only copies should be resized.

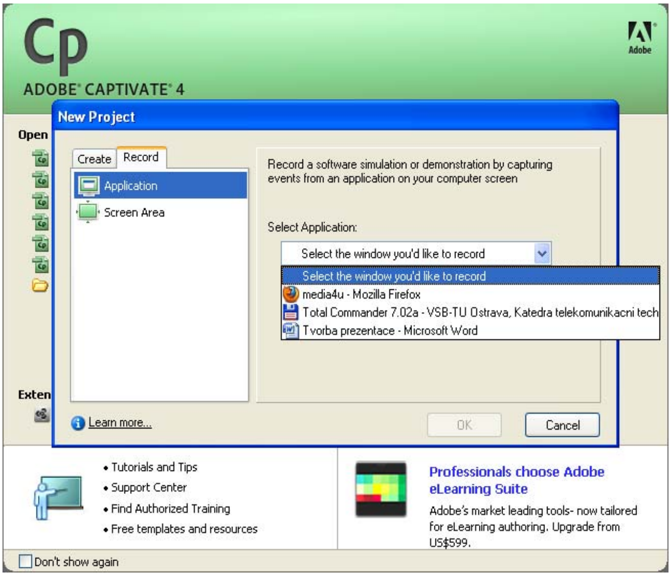


Fig. 1. Application choice for presentation

The choice is confirmed by press button OK. In left top corner there appear the recording icons. It is chosen recording mode. It is possible to change recording mode setting there. The recording is run by click on  button.

5.2 Recording and saving project

How it is described above, it is possible to choose:

- Automatic recording mode - Adobe Captivate captures screenshots automatically and places them on separate slides.
- Full Motion Recording, entire set of events is captured in real time as a video.
- Manual recording mode - manually taking screenshots during recording.

Now it follows self work in application which should be presented. If automatic recording mode was chosen, the user does not have to care for Captivate, slides are captured automatically. All edits, captions, revisions are made after as much as recording stop. The recording is stopped by press End button.

Now the whole presentation is made which has to be saved. The project name is written and project location is selected.

5.3 Project editing

Now it follows editing of created project. Single slides are subsequently ordered like on film strip. The slide sequence is simply changeable by catch the slide and moving at other place. The slides with mouse symbol contain move recording of mouse pointer and clicks on icons in applications. The slides with camera symbol contain video. In this concrete project slides with camera were done by icons moving from component library to layout. These slides include camera symbol and they are automatically made. All slides can be copied, erased or it is possible insert new slides etc.

Each action should be described by clear caption. On the left side there are icons to slides edit. It is possible to choose shapes of captions, change their size and position and edit the text which the caption includes.

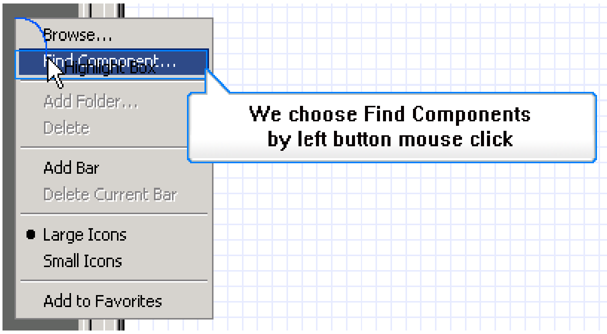


Fig. 2. Caption

The last very useful possibility is change of timing of single effects. It is very simple to change time duration of slide. This time could be shorten or extended. The caption start and caption end is optionally adjustable. Just it should be set time duration of caption and place it suitable in timeline of slide. It is possible to use more captions they could concur or overlap. Likewise it is possible to change time duration of typing, mouse moving, highlight box etc. It is possible to change all timelines of all slides. See next figure.

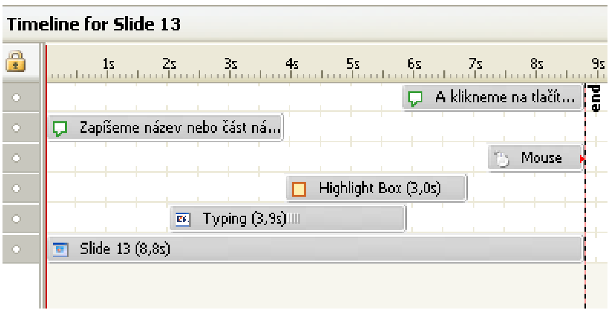


Fig. 3. Timeline

5.4 Resize

The resize of project allows to rescale into more suitable resolution. The resize is a nonreversible action therefore it is very useful to have a copy of original project. It allows to resize again.

5.5 Publishing

The final project has to be published it means click on Publish button. The type of file is chosen. The most frequently used choose is Flash for presentation in web browser [4] or executable file. As evidenced by figure 4 there are more possibilities as email, print in doc or docx file, upload to FTP server etc. The folder has to be chosen for published file.

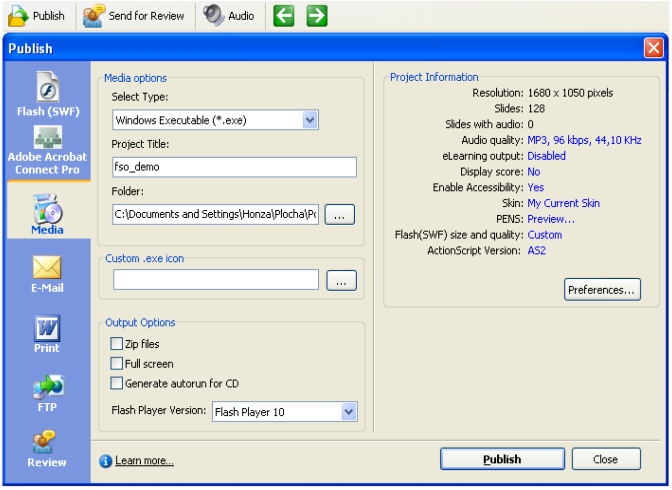


Fig. 4. Publishing

The Flash file runs at all computers which include some web browser which is practically each computer.

The final result you can see at figure 5. In the bottom there is a playback toolbar. It is possible to rewind the presentation, play, pause, go back one slide or go forward one slide or play 2x fast forward speed. There is displayed a timeline too.

Students can play the presentation more times, they can stop the presentation and ask the teacher or try it in OptiSystem. This is very useful and user-friendly.

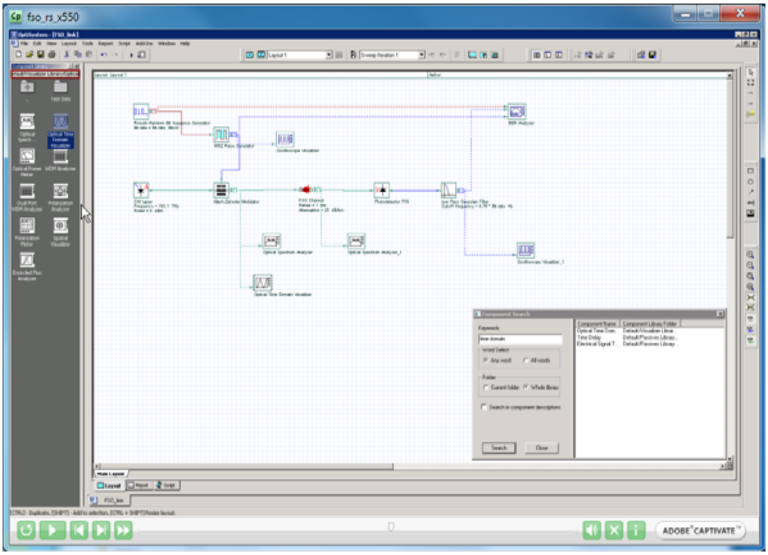


Fig. 5. Presentation play

6 Conclusion

Software Adobe® Captivate® 4 is very useful and user-friendly software for lessons which are aimed at application service. These lessons are very effective because it is possible to play more time the final presentation, students can replay the presentation at home, try everything again and aim at problem areas. In standard lesson the demonstration of some application is done at most onetime. This is insufficient to one's cost.

This presentation was used in Optical atmospheric communications lessons. The student's feedback was very positive.

This work was supported in part by an ESF project CZ.1.07/2.2.00/07.0351, GAČR GA102/09/0550 and SP2011/47.

References

[1] *Digital Media* [online]. 2011 [cit. 2011-02-28]. Adobe Captivate 5. Available from WWW: <http://www.digitalmedia.cz/produkty/adobe/captivate/>.

[2] *Help Adobe* [online]. 2009 [cit. 2011-02-28]. Using Adobe Captivate 4. Available from WWW: <http://help.adobe.com/en_US/Captivate/4.0/Using/captivate_4_help.pdf>.

[3] *OptiSystem : Component Library*. Kanada : OptiSystem, 2010. 1970 s.

[4] *Mozilla* [online]. 2011 [cit. 2011-03-10]. Doplňky Firefoxu. Available from WWW: <https://addons.mozilla.org/cs/firefox/browse/type:7>.

Self-Reproduction and Self-Assembly in Multi-Robot Systems

Michal Perdek

Silesian University in Opava, Bezručovo náměstí 13, 746 01 Opava, Czech Republic
michal.perdek@fpf.slu.cz

Abstract. This paper introduces the self-reproduction and the adaptation mechanism of multi-robot systems. The approach is derived from the evolution theory. During the replication multi-robot systems find lacking parts (modules) which are required to be connected into the already completed body-structure. Every additional connection in the replication process requires adaptation because body-structure has been changed and the adaptation tries to preserve certain behavior of these multi-robot systems. Approach utilizes multi-agent, evolution theory and also graph theory, which try to provide realization of the self-reproduction (self assembling) process.

1 Introduction

In the last few years, robotics has dealt with multi-robot systems which are inspired by natural organism. One single robot platform corresponds to one cell in nature, and multi-robots consist of many of these small cells (robots) that are able to aggregate to "multi-cellular" robot organism. Big advantage of these multi-robot systems is that they are able to perform tasks which an ordinary single robot does not achieve.

Basic idea in the multi-robotics is creating artificial constructs that are capable of autonomously construct their own copies using raw material taken from their environment. If we create such multi-robot system with enough material we can be able to create exponential number of these robots. Multi-robot systems are relatively new emerging field across several research areas such psychology, evolution, biology, engineering, cognitive science, neuroscience and many other fields. Interdisciplinary point of view is important for model design, since different types of self-reproduction may use different techniques and strategies to reproduce themselves. We can find various definition for formalization of the phenomenon of self-reproduction in interdisciplinary literature such as [1] or [2].

1.1 Multi-robot system

Multi-robot systems are assembled from many small autonomous robots (cells) that are controlled independently. In this paper we call these autonomous robots modules.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 652–657.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

Every single module has been equipped with transmitter, receiver and magnetic locks. Optional equipment is: cameras, sensors, wheels, servo motors, etc. Magnetic locks allow creating connection between modules (Fig. 1 shows an example of two multi-robot systems and their graph representation). This concept of connected modules and their connection can be regarded as multi-agent based system. Cooperation of agents is achieved by communication of the whole system. Modules can receive and transmit messages which can flow into actuation. Most solution use classical control approaches known from the control theory [3], but also bio-inspired and evolutionary concepts become more important [4]. We use decentralized approach [5], [6]; one module can control the whole system at one time. Because modules are able to do only simple "action", all modules contribute to make the whole system act as one entity, and behavior of this system has the form of emergence.

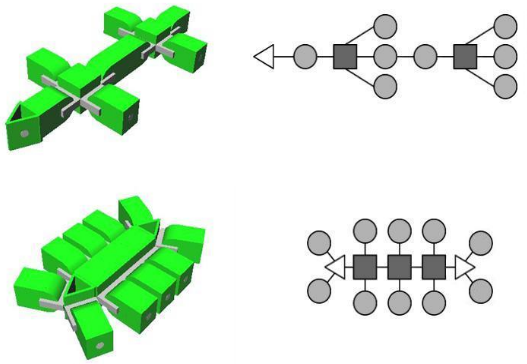


Fig. 1. Multi-robot systems and their body-structures representation

Modules are categorized into classes such as locomotive organs[1], main organs[2] or perception organs[3]. Each module knows its own type and other types. The same controller is used for the same type of modules, but with different behavior. In virtue of configuration, modules behavior is represent by chromosome that is described in Sec. 3. All modules in multi-robot system have the same chromosome that determines how the system will act. Chromosome consists of genes which determine specific behavior of specific modules.

2 Self-reproduction

Self-reproduction process starts with mapping of the body structure. Modules in our multi-robot system communicate with each other and exchange information about connection between other modules. This information is transformed to the tree (graph

---

[1] Locomotive organs - modules with servo motors, wheel and other types that enable movement

[2] Main organs – simple modules for connecting other modules through magnetic loch

[3] Perception organs – modules that have sensors or cameras

structure) that represents body structure (Fig. 2 shows an example of spider multi-robot systems and body-structure representation from different modules). These trees become the template of particular modules. Multi-robot system is capable of basic movement and tries searching new modules in environment. Parts found in environment are basically connected in the same way as the passage through the previously mentioned trees (body-structure representation) on free magnetic locks. Mapping process is triggered and body-structures are updated by creating new connection. Multi-robot system finds new modules and makes connections of modules until whole system contains two identical sub trees which correspond to the previously mentioned template. Self-reproduction process is based on growing and separating into identical subsystems by using edge cut method [7].

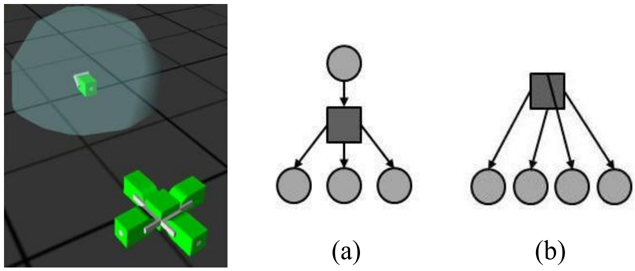


Fig. 2. Spider multi-robot system; (a) body-mapping from one of locomotion organs; (b) body-mapping from one main organ

3 Adaptation

In self-reproduction process we connect new modules which affect cooperative behavior to the system. We have to change behavior to preserve basic movement. The best behavior of the multi-robot system can be realized by taking advantage of using new module. Motion represents a classical controlling problem, which can be solved by genetic algorithm [8], [9], [10]. Module behavior is given by genome that is a part of our multi-robot chromosome. We focused on adaptation using genetic algorithm for better adaptation on changes in body-structure during the whole reproduction.

3.1 Genetics algorithm

Behavior of our multi-robot system was described by chromosome that was mentioned previously. This chromosome consists of genomes that determine particular activity for every single module. Genetic algorithm generates population that represents group of potential behaviors (chromosomes). Members of this population are evaluated by fitness function that represents member's quality. Better quality means better behavior. Algorithm uses genetic operators such as selection, mutation and crossover for creating new populations from the previous one [9], [10], [11]. Creating populations with using best individuals from previous generation and application genetic operators lead to better searching of state space then by using permutations. Genetic algorithms have intention to search better or optimal solution of our problem, but not necessary find best solution. Disadvantage of this approach is time and space complexity.

3.2 Genetic encoding and fitness function

Our spider multi-robot from Fig. 2 consists of two module types. Central module does not have movement parts and it is not involved into locomotion. But it forms connection among other parts. Behavior parameters of central module do not appear in chromosome. Chromosome consists of parts that are involved into locomotion directly with servo motors. Our chromosome on Fig. 3 contains four states (front, back, left, right). Every state consists of modules parameters involved into locomotion.

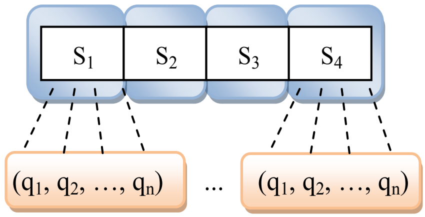


Fig. 3. Chromosome structure – S_i represents system states and q_i are parameters determine module behavior in specific state.

Spider multi-robot movement is based on oscillation, phase shift, frequency and amplitude. Behavior of specific module is given by function (1) where result determines angle for servo motor in specific time. Parameter s – phase shift lies between 0° - 360° and for our purpose it will be included into chromosome. Amplitude and frequency will not be included into chromosome. To simplify it, we use only one parameter for module in particular state.

$$\alpha = A \sin(2\pi Ft + (s\pi/180)) \quad \textbf{(1)}$$

If we involved all previously mentioned parameters, space state would be exponentially increased. For our multi-robot that means that we have only four parameters in one state which has 360^4 possibilities. New module from movement class adds one DOF to the whole system and space state for one movement will be increased on 360^n where n is number of movement parts. During self-reproduction chromosome will be compounded from the same states but the number of parameters in individual states will be increased.

Fitness function evaluates quality of individuals in population. In our case, this function judges movement quality. Parameters from individual in population were set for our multi-robot and then robot tries to move with these parameters for few seconds and after that it will be evaluated. Function derives benefits from the initial position and achieved position. Desired course will be expressed as a unitary vector from the initial position. After movement is done, achieved position will be used for calculation of the Euclidean distance D between these positions and deflection φ from desired course. Our fitness function will be compounded as (2)

$$F(i) = D - \varphi. \tag{2}$$

Adaptation on specific motion is performed in case that this motion does not fulfill satisfactory behavior. We do not adapt all motion in the same time, we use GA only for adaptation of one movement. If the multi-robot system needs to use a different movement, the adaptation starts again for this motion. During GA we create many individuals but we are interested only in the best solution in population that will be used for creating new generation, and the others will be thrown out. These thrown-out individuals may correspond with other courses and if the individual is sufficient or better than the currently used one, it will be replaced. This adjustment is able to bring better or less demanding searching in state space because new solution will be appearing also in other generation while GA will be searching specific solution.

3.3 Realization of adaptation

For our GA we use tournament method selection, one-point crossover and classical mutation. Parameters for realization GA for our self-reproducing process are shown in Table 1.

Table 1. Summary of simulations parameters, where n is number of modules directly involved into motion. Population size and number of generations will be changed during self-reproduction depending on number of motion modules in body-structure.

Setting	Value
Population size	$20n^2$
Number of generations	$20n^4$
Crossover probability	0,8
Mutation probability	0,03
Divergence	$5° = \pi/36$ rad
Distance	1,6

Number of motion modules is increased during self-reproduction and so time and space complexity increases, too. In our simulation that was created in simulation environment Webots, we successfully simulated self-reproduction with GA adaptation based on the principle that was mentioned above. During GA activity we evolve only those parts of chromosomes (genomes) that are directly responsible for particular movement in certain state. Changes in these genomes are updated into all modules in the system, where they replace the old part of chromosome. In this case spider-multi robot was object of interest and we created several simulations with this robot.

4 Conclusions and future work

In this paper we described self-reproducing approach that derives benefits from graph algorithm and multi-agent based systems. We focused on adaptation with using genetic algorithm for better adaptation on changes in body-structure during the whole reproduction. Proposed solution may provide better searching in state space with using unused individuals that are evaluated by fitness function. Realization of adaptation based on this concept described in paper proved reduction of time and space complexity. We simulated the whole process of self-reproduction and adaptation in simulation environment Webots.

Future work will include other techniques for adaptation and we will focus on emergent behavior that comes up from interaction among modules and interaction among modules and environment.

References

1. McMullin, B.: John von Neumann and the evolutionary growth of complexity: Looking backwards, looking forwards. In Bedau., M., A., (ed.): Artificial Life VII,MIT Press, Cambridge, MA., (2000) 261-264
2. Löfgren, L.: Relative explanations of systems. In Klir, G., (ed.): Trends in General System Theory. John Wiley and Sons (1972)
3. Wit, C. C., Bastin, G. and Siciliano, B.: Theory of Robot Control. Springer-Verlag, Berlin Heidelberg New York (1996)
4. Floreano, D. and Mattiussi, C.: Bio-inspired Artificial Intelligence: Theories, Methods, and Technolgies. MIT Press (2008)
5. Weiss, G.: Multiagent systems. A modern approach to distributed artificial intelligence. MIT Press, Cambridge, Mass (1999)
6. Kubík, A.: Intelligent Agents (in Czech). Computer Press, Brno (2004)
7. Perdek M.: Open-agent implementation in simulation environment Webots (in Czech). In Kelemen, J., Kvasnička, V., (eds): Kognice a umělý život X. Opava: Silesian university (2010) 275-280
8. Lal, S. P., Yamada, K. and Endo, S.: Emergent Motion Characteristics of Modular Robot through Genetic Algorithm. In Huang, Wunsch, Levine, Jo, (eds.): Advance Intelligent computing theories and applications: with aspects of artificial intelligence, Springer-Verlag, Berlin Heidelberg (2008) 225-234
9. Robinson, T. R.: Genetics for dummies. Wiley publishing, Hoboken, NJ (2005)
10. Koza, J.: Genetic Programming: On the Programming of Computers by Means of natural Selection, MIT Press, Cambridge (1992)
11. Nedjah, N. and Macedo Mourelle, L. (eds.): Evolvable machines: theory and practice. Springer-Verlag, Berlin Heidelberg New York (2005)

Registration of the object position changes in space using LabView environment

Paweł Michalski

Opole University of Technology, Faculty of Electrical Engineering, Automatic Control and Computer Science, Institute of Automatic Control and Computer Science
ul. Sosnkowskiego 31, 45-272 Opole, Poland
p.michalski@doktorant.po.opole.pl

Abstract. The paper presents a system for registering changes of object position in space, using a laser rangefinder, an HD camera, and precision biaxial motor. Advantages, limitations and the possibility use of the system are presented. The system software was made in a Labview environment.

1 Introduction

Inventory of large installations on the premises of industrial areas is a complicated and also time-consuming task. Currently, this is part where laser scanning techniques are used, which offers very good accuracy of measurement. Another advantage is the speed of the measurement. Documentation generated using a laser scanner offers a two-dimensional projections such as is conventionally inventory. In addition, however, the amount of information acquired as a result of laser scanning is disproportionately greater giving the operator the possibility to play with the installation of three-dimensional model with all the relevant details. With a properly prepared model installation, you can enrich it with additional information about the movements occurring in the installation. The operator will receive digital image of the object surface deformation, which is updated almost in real time. Thus prepared, the system can serve as the basis for the diagnosis of individual elements or entire systems

2 Hardware components of the system

For the purpose of the tracking system, made a set of markers that are identified and tracked by the developed device. The system allows tracking a single marker or entire sets of markers. What allows operator to track periodically several points during a single measurement cycle. The device consists of a precision biaxial step motors, which corresponds with the appropriate guidance camera and laser

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 658–663.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

rangefinder. As shown in Fig. 1 the HD camera is mounted on the pan and tilt motors. The camera is responsible for capturing an image and sends it to the corresponding application for the marker identification.

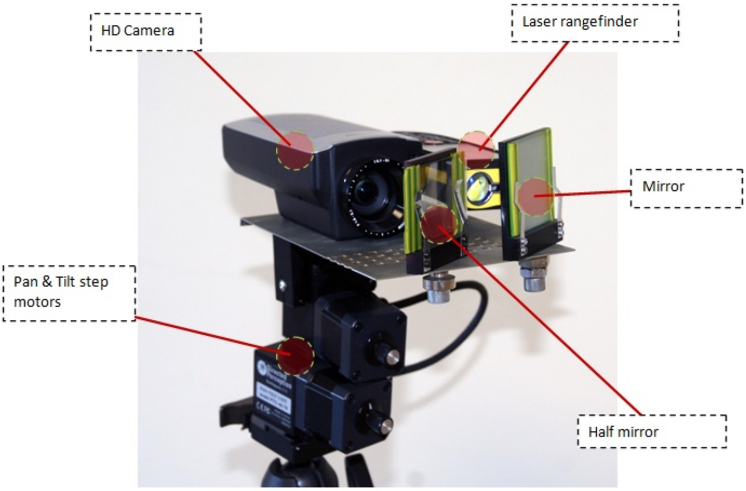


Fig. 1. Hardware components of the system

The camera has a 10 - fold optical zoom, which allows the identification of markers located at a distance of up to 50 m from the measuring system. The next element is laser rangefinder, which is performed to measure the distance markers. For each marker measurement is made three times with 0.5 seconds intervals to eliminate the environmental disturbance. The system consists of a set of mirrors to aim the laser beam in such a way that was exactly in the optical axis of the camera. When the system measures the distance to the marker and read the position of the stepper motors, the position of the marker in space can be determined. The position is calculated on the angles and the distance to the marker.

Fig. 2. Deployment pattern of markers.

The system requires the placement of one indicative marker of the installation, which is used to construct the coordinate system in which the other markers are oriented. The software is done periodically to follow the path designated by the operator during the first cycle. Operator using the software control panel guides the optical system on the markers. Then the search and the identification marker in the field of view are performed. When the marker is identified the system is centering the camera on the marker. After properly identifying marker the engine status, the ID value of the marker and the actually camera zoom is saved. Then the operator proceeds to the guidance system for the next marker. The measuring cycle can have from one marker to a virtually unlimited number of markers. Note, however, that the length of the exercise of individual cycles is dependent on the amount and method of deployment. The operator should identify markers in the order that the distance between them is as low as possible this will allow more efficient system operation. The example of the correct deployment of the markers is shown on Fig 2.

3 Software implementation in LabView

Designed software made in the LabView environment, based on image identification module IMAQ VISION. The application performs several stages; first step is to initialize all the devices and to establish communication with each of them. The communication system is quite extensive because of the software, each device communicates through other communication links such as TCP / IP, Bluetooth and RS-232. The application seeks to identify markers using edge detection modules included in the package National Instruments Vision.

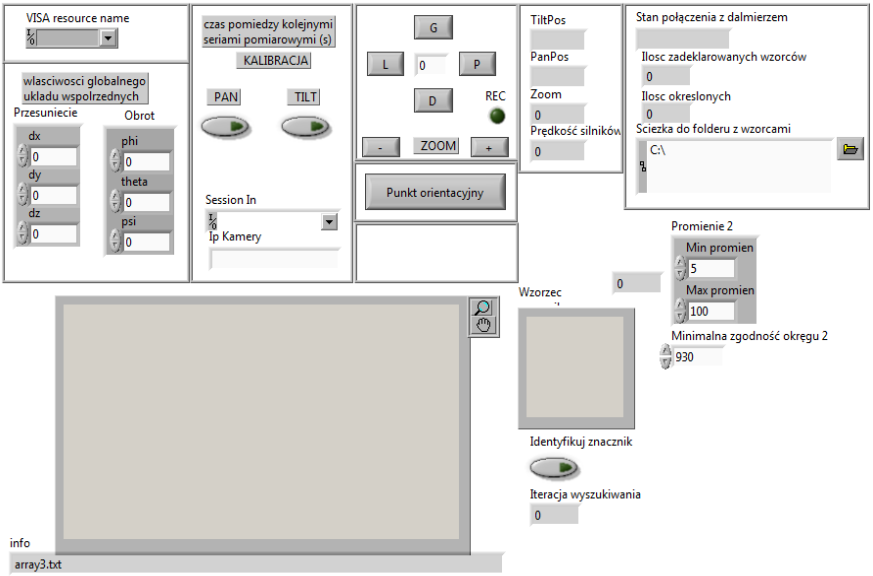


Fig. 3. Part of the control panel of the system

Data with the coordinates of identified markers are saved to a text file with the number of iterations needed to identify the marker and time at which the measurement was made. Such information can be easily stored in a database for later analysis. Data can also be automatically transferred to the engineering environments in order to update the model. Mapping model of the element movements is the most accurate when every element of the installation such as a pipe or boiler should have a few markers.

4 Summary

The developed system can be used everywhere where element movements beyond certain parameters can lead to damage of the object. The project was designed for piping, which under the influence of temperature and pressure is subjected to deformation. It will allow the early diagnosis of the installation and determination of those areas where deformations take dangerous limits. The use of the system can be much wider. An example of the application may be measuring the bridge elements motion, which under the influence of traffic starts to vibrate. Another application might be building heights where wind-loaded buildings deviate from its vertical axis. The system would allow an easy way to m whether these deviations are within the currently accepted standards.

References

1. 1. Zator S., Ryszczyk M,., Krawczyk M., Kwiatkowski J.: Intergracja baz danych diagnostycznych z rzeczywistym modelem przestrzennym, Transactions on Computer Applications in Electrical Engineering, XIII Scientific Conference Computer Applications in Electrical Engineering, Poznań April 1-16 2008, s.293-294
2. Głuch I., Krzyżanowski J: System diagnostyki geometrii obiektów energetycznych. Pomiary Automatyka Kontrola nr 9bis/2005 ss 166-169
3. Kabza Z., Zator S., Kwiatkowski L : Hybrydowe techniki pomiaru przemieszczeń instalacji technicznych, XII Międzynarodowa Konferencja Naukowo-Techniczna „Forum Energetyków” GRE2010, Zeszyty Naukowe Politechniki Opolskiej, Nr 336/2010, seria Elektryka z.64, s.133-134

A data partitioning problem in programming of massively parallel processors

Tomasz Przybyła

Opole University of Technology, Faculty of Electrical Engineering, Automatic Control and Computer Science, Institute of Automatic Control and Computer Science
ul. Sosnkowskiego 31, 45-272 Opole, Poland
tom.info.dok@gmail.pl

Abstract. The article presents the problem of data partitioning, which is crucial in the case of High-Performance Computing (HPC) in a massively parallel architecture. The OpenCL language is outlined, which is used for programming GPUs in order to realize the parallel computing problem. The calculations a performed on the Tesla C1060 and GeForce GT540M. The obtained results enable to property manage the task of grouping work-items into work-groups.

Introduction

The increasing computational performance of graphics cards in the confrontation with the traditional CPU has contributed to the search for new technologies enabling to use the potential of GPUs for tasks not related to the generation of graphics.

There are already many technologies that allow to program massively parallel processors; NVIDIA released CUDA architecture, Microsoft introduced the Direct Compute, ATI offered the AMD Stream technology. All of them are quite similar and are based on general-purpose computing on graphics processing units (GPGPU).

GPGPU computing makes of a graphics processing unit (GPU) for general purpose computing [1]. The model for GPU computing is to use a central processing unit and graphic processing unit together in a heterogeneous co-processing computing model.

OpenCL

OpenCL (Open Computing Language) is a low-level API (Application Programming Interface) that allows the use of GPGPU performance computing systems for

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 663–668.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

general-purpose computing. The development of OpenCL was initiated by Apple and developed by the Khronos Group [2].

OpenCL programming language is based on the ISO C99 specification, with the added support for vector types, vector operations, and qualifiers address space [2]. There is no support for function pointers and recursion. The language is based on the IEEE 754 (the numerical precision for all floating point operations) in order to ensure consistency of results across different platforms.

Implementation of parallel computing in OpenCL starts from defining the problem in the N-dimensional space [3]. In the case of the linear array of data, the problem can be represented in one-dimensional space, whereas the images will use the two-dimensional space and the spatial representations of data, such as 3D objects, with use the three-dimensional space.

Each independent component in the index space is called a work-item. Each work-item performs the same program kernel, but for different data [3]. OpenCL allows the grouping of work-items in the work-groups. All work-items belonging to the work-group are executed in a single device, giving it the ability to share local memory, and synchronization. Figure 2 shows the OpenCL memory model.

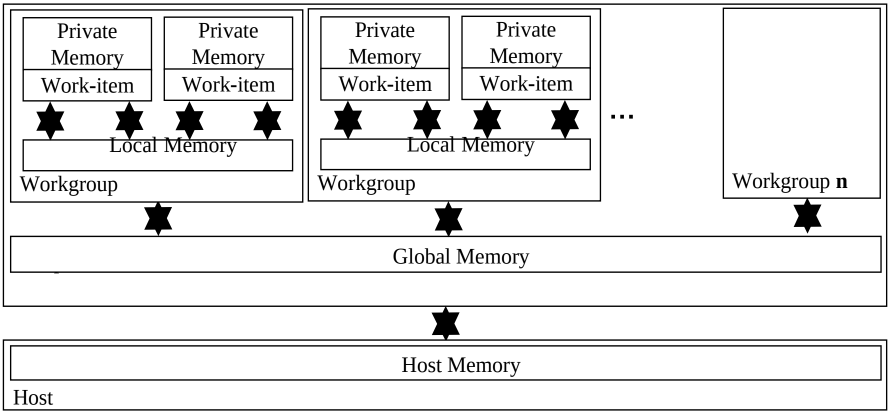


Fig. 1. OpenCL memory model

Heterogeneity of systems on which the calculations can be performed, carries with it some complications. There are systems that enable the creation of hundreds of work-groups containing thousands of work-items, but there are also systems that support individual work-items. A number of work-items used for the calculation depends on the specification of the algorithm and the architecture of the device. Information about the architecture provides the function clGetDeviceInfo [1], which allows the algorithm to better adapt to the computing device.

3 Data partitioning

If you use OpenCL for calculations on large data sets, an extremely important aspect is adequate data partitioning. It allows to distribute the processing load, reducing the execution time algorithm in heterogeneous systems.

Usually computing device contains hundreds of data processing elements. With a properly prepared code, we can control the amount of data used for the calculation of the processing element. For this purpose, it uses the function clEnqueueNDRangeKernel, that specifies the number of dimensions space, a number of work-items in each dimensions and the number of work-groups.

OpenCL allows the grouping of work-items. This operation can be performed directly or indirectly [3]. Both in the first and second cases, the programmer specifies the total number of work-items used in the calculations. The difference is that in the case of grouping the indirect method, OpenCL groups the work-items, while in the direct method, the programmer defines the size of the work-groups. An indirect method is selected by setting the parameter local_work_size to NULL [1].

Each work-group has a unique ID that uniquely defines the position of the group in the whole index space. Group ID is determined by the size of the work-group, which in turn is determined by a local_work_size parameter in the function clEnqueueNDRangeKernel.. Value of the local_work_size parameter determines the number of work-items included in the work-group in each dimension [2]. Work-group size cannot exceed the value returned by the function clGetKernelWorkGroupInfo ().

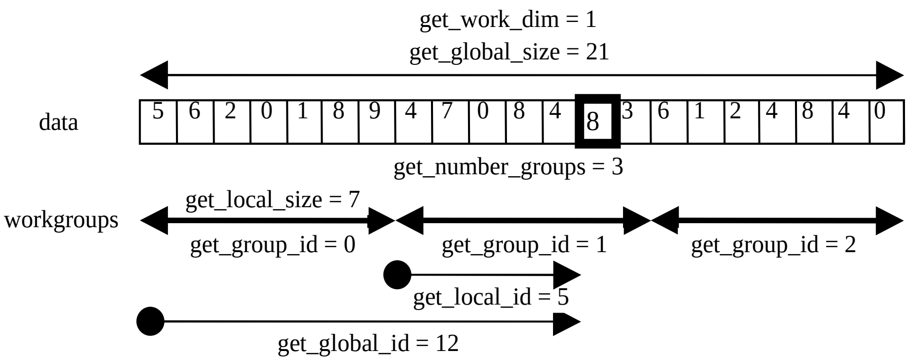


Fig. 3. Identifiers in the one-dimensional space

Besides the global ID, each work-item in the work-group has a local ID identifies its own position in the group. By accessing local ID, we can coordinate the work-item processing in the work-group.

4 GPU Computing

To examine the potential effects of work-group size on the execution time of the algorithm, a program is written for multiplication of two square matrices. The calculation was carried out on the Tesla C1060 and GeForce GT540M. Details of the systems are shown in Table 1.

Table 1. Comparison of computational resources

	NVIDIA Tesla C1060	GeForce GT540M
Max clock	1296 MHz	1344 MHz
Number of parallel compute cores	30	2
Number of processing elements	240	96
Max work group size	512	1024

Measurement of execution times of the algorithm are made for different sizes of work-groups and different sizes of the matrix. The results (in milliseconds) are presented in Table 2.

Table 2. Timing results for different sizes of work-groups

		Tesla C1060					GeForce GT540				
		Work-group size					Work-group size				
		2x2	4x4	8x8	16x16	Auto	4x4	8x8	16x16	32x32	Auto
Matrix size	512x512	148	77	37	20	18	258	66	25	35	23
	1024x1024	827	410	218	111	123	2130	534	213	285	268
	2048x2048	6293	2319	1190	776	794	17120	4274	1733	2285	2149
	4096x4096	46748	20089	8164	5671	5632	---	---	13941	18486	17218

The measurements clearly indicate that work-group size has a significant impact on the execution time of the algorithm. To effectively use the GPGPU performance computing systems the size of the work-group should be adapted to the specifications of your task. You should normally set the largest, supported by the computing device, the work-group size. Research has shown that automatic division to the work-groups not always gives the best results.

Summary

Appropriate grouping of work-items to the work-groups is one of many techniques that allow you to shorten the task execution time in the massively parallel architecture. Only the skillful use of the potential GPGPU systems and technologies OpenCL offers great opportunities for implementing the general-purpose computation.

References

1. Munshi A.: *The OpenCL Specification version 1.1*, Khronos OpenCL Working Group, 2010.
2. Kirk D., Wen-Mei Hwu W.: *Programming Massively Parallel Processors. A Hands-on Approach*, Morgan Kaufmann, 2010.
3. AMD: *Introducing to OpenCL Programming. Training Guide.* Advanced Micro Devices, 2010.

High resolution thermograms via distortion correction

Mirosław Lasar

Opole University of Technology, Faculty of Electrical Engineering, Automatic Control and Computer Science, Institute of Power Plants and Measuring Systems, ul. Prszkowska 76, 45-758 Opole, Poland
m.lasar@doktornat.po.opole.pl

Abstract. The article describes how to obtain high-resolution thermal images from a camera equipped with telephoto lens of resolution 16-times greater than the standard lens. Using the 320x240 pixels camera we can receive the 1.3 megapixel thermonogram in a similar way as for panorama images . The process of automated precise measurement of thermograms makes use of a biaxial turntable on which the camera is installed. The application is developed in the LabVIEW driver using CompactRIO Real-Time Controller. The complex thermograms can be analyzed by the software provided with the camera.

1 Introduction

An infrared camera used to record the temperature at a distance by detecting the infrared radiation emitted by the real objects [1]. It allows you to present measurements in a thermogram, with different color palettes (Figure 1). Unfortunately, when we are using the standard and wide-angle lens, on the thermogram appears a radial distortion. Then it is harder to fold thermograms in larger images. It manifests itself through the image distortion, visible especially in the most remote from the optical axis points.

© M. Krátký, J. Dvorský, P. Moravec (Eds.): WOFEX 2011, pp. 668–673.
VŠB – Technical University of Ostrava, FEECS, 2011, ISBN 978-80-248-2449-9.

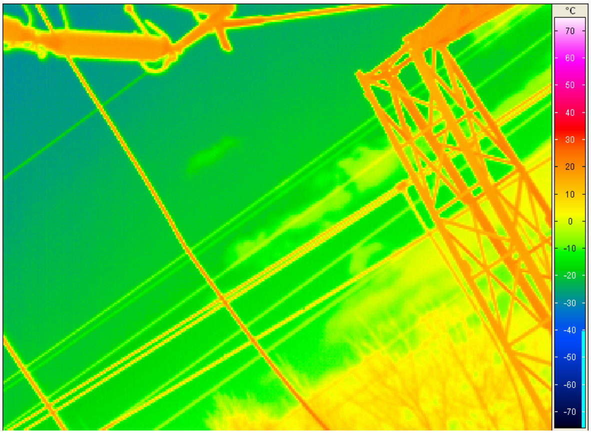


Fig. 1. Example of thermal image with a resolution of 320x240 pixels

Depending on the shape of the image distortions which distinguishes between two types of distortion (barrel and pillow).

Geometric distortion are correcting using three methods [2]:

- Interpolation - consisting in drawing up the table with the value assigned to a given value of distortion radial radius. When correcting a specific point interpolates an array of two values, the range in which the test pixel.

- Tabular - in which it is assumed that in small parts of the picture distortion value is constant. Correction in this case is to draw from the table field distortion values, which are within the coordinates of specific point.

- Polynomial - the most accurate of the above., Calculated using formulas new coordinates every point on the image.

2 Methods to correct distortion of thermograms

In order to notice and correct the distortions that occur in lenses used a resistance wire, which is heated to a temperature significantly higher than ambient (about 60°C). For the purposes of visualization changed display range thermogram temperatures, so that the background has adopted the color black. (Fig. 2). The measurements were repeated several times by rotating the camera horizontally. Then measurements were repeated with each turning axes, resulting in the thermogram were horizontal lines. Following the submission of thermograms obtained with each grid.

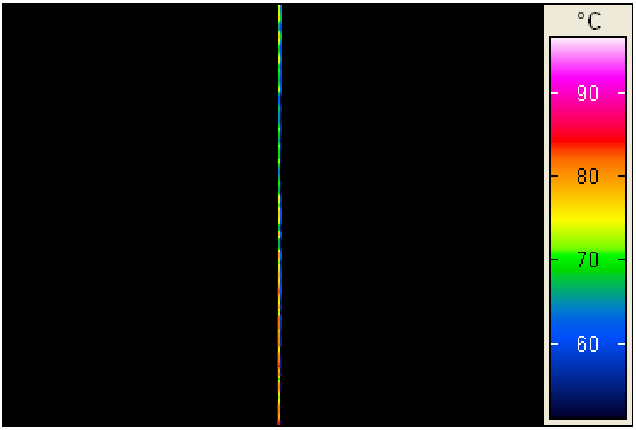


Fig. 2. Marked chosen measured points

Correction was to shift the image pixels in the resulting image, which was the visual effect of straightening the grid lines. Moving Pixels also leaves behind artifacts, which resulted in the blanks. To fill these side effects uses interpolating distortion correction, known from computer graphics. On the basis of pixels adjacent to pixel creates a new pixel is created, one of several possible methods of interpolation. Correction algorithm uses the following mathematical equations [2]:

$$\begin{aligned} x_u &= x_0 + (x_d - x_0)(1 + k_1 r_d^2 + k_2 r_d^4 + \\ y_u &= y_0 + (y_d - y_0)(1 + k_1 r_d^2 + k_2 r_d^4 + \end{aligned} \qquad (1)$$

where:

x_u, y_u – point coordinates after distortion correction
x_0, y_0 – the point of best symmetry
x_d, y_d – pixel coordinates before the correction
$k_1, k_2, \ldots$ – radial distortion coefficients defining
r_d – the radius vector

The most accurate method of distortion correction methods is polynomial, but it was not used, because this would create ambiguous situations (eg at the interface between two different objects) that can lead to deterioration of results. Thermograms joining algorithm in the mesh distortion was to change the ambient temperature at -273.15°C and the lower display threshold was set to 50°C, resulting in the thermogram is a black background. The next stage was compared with successive thermograms temperature at the same point and replaced with higher values.

Thermograms for submission algorithm consisted of a sequence of iterative combining temperature data files. Initially, the data files are combined in the lines. The idea of submitting poems thermograms are shown in Figure 3 Then, the columns joined together in the same way. Camera can take pictures on the tab, which can facilitate the manual consolidation of individual thermograms together.

24,7	28,9	29,8		32,3	28,9	38,4		24,7	28,9	29,8	32,3	28,9	38,4
25,8	25,3	26,3		33,4	31,5	33,5		25,8	25,3	26,3	33,4	31,5	33,5
26,2	26,3	26,3		32,5	35,6	33,3		26,2	26,3	26,3	32,5	35,6	33,3
26,3	26,9	27,1	+	36,6	35,6	38,9	=	26,3	26,9	27,1	36,6	35,6	38,9
26,3	26,7	26,8		38,9	37,9	33,2		26,3	26,7	26,8	38,9	37,9	33,2
27,1	27,2	27,3		33,5	30	37,5		27,1	27,2	27,3	33,5	30	37,5

Fig. 3. The idea of submitting thermograms

For testing was used infrared camera VarioCam Head with a resolution of matrix 320x240 pixels. The control system consisted of a precision turntable biaxial consisting of modules NR360SP9 / M (Fig. 6), which was controlled by RTC the CompactRIO controller with the NI 9512 modules and 9022, and P70530 (Fig. 5). Communication between the controller and the computer were made via Ethernet, and data were transmitted from the camera through a FireWire port.

Infrared camera mounted on a tripod with a swivel and a heater wire shown in Figure 4 Measurements were carried out for the three lenses: wide angle (64^ox50^o), standard-draft (36^ox24^o) and telephoto (8^ox6^o).

Fig. 4. Measuring equipment

Fig. 5. A cRIO set with NI 9512 and 9022 modules, and the step motor controller P70530

.

Fig. 6. Two-axis turntable consisting of two modules NR360SP9/M

For the control measurement system designed and built an application in LabVIEW, which control panel shows the Figure 7. It allows you to control the turntable in manual and automatic, and registration of infrared images.

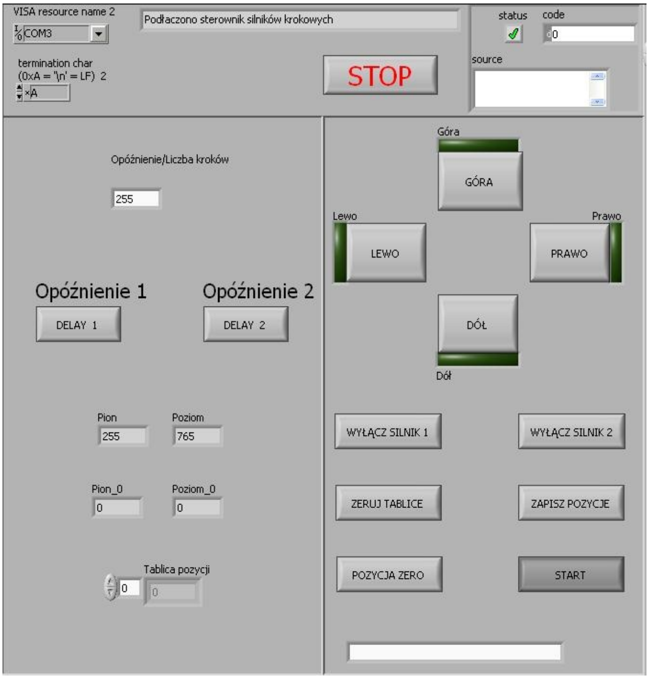


Fig. 7. Application to create a sequence of measurements

3 Measurements

It was noted that the greatest distortions occur in the wide-angle lens. Figure 8 shows the resulting grid before and after correction, consisting of 44 thermograms. As can be seen in the lens barrel distortion that occurs, which is very visible. For the standard lens distortion were not as large, but still occurred and required correction. The best results were obtained when mounted telephoto. The effect of deposit of 37 thermograms are presented in the form of a grid in Figure 9. For this lens is not visually apparent geometric distortion.

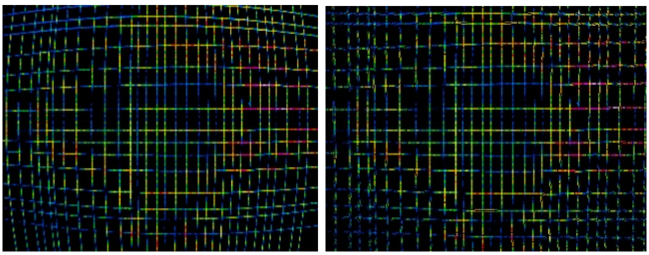


Fig. 8. Picture of the grid obtained from the thermograms for the wide-angle lens before (left) and after correction (right).

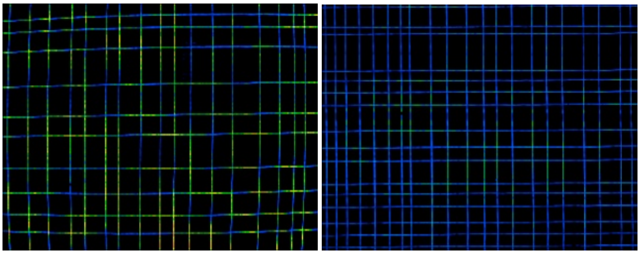


Fig. 9. Picture of the grid obtained from the thermograms for the standard lens (left) and telephoto (right) before correction.

Using the thermograms, not only for temperature measurements, but such a balancing energy flow through building partitions, or to map the three-dimensional models [3] to correct the errors of temperature measurement, for lenses with a focal length 12.5 mm and 25 mm should be applied the algorithm geometric distortion correction. Figure 10 shows the photo section of the node installation for central heating and submit six thermograms obtained in a thermal image 1920x240 pixels resolution.

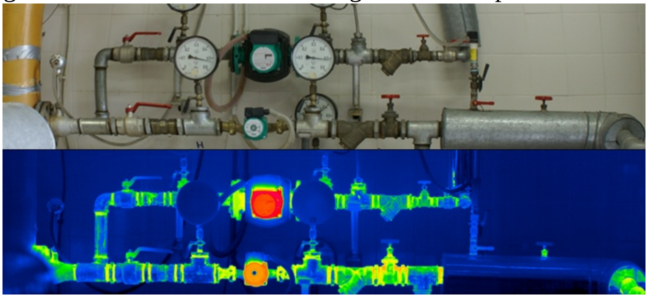


Fig. 10. Photo of piece installation (up) and the effect of submission 6 thermograms (down)

Complex and corrected thermogram can be submitted together with photographs (such functions have infrared cameras that support programs), as shown in Figure 11.

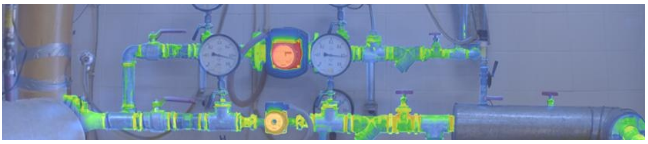


Fig. 11. Photo installation imposed thermogram

The resulting high-resolution thermograms were also used, among others for applying a texture (thermogram) of the cloud of points obtained from laser scanner. The effects of the imposition of the scan thermogram are shown in Figure 12. Dimensional model of the surface temperature distribution can be used both as initial conditions and data for comparison with the results of finite element calculations.

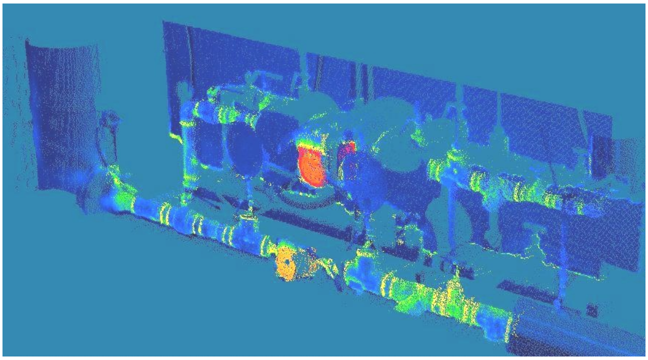


Fig. 12. 3D model with thermogram

Built measuring system allows you to make high-resolution thermal images of objects, which are followed by slow temperature changes. The time required to execute a complex series of thermographic images of the 16 is about 15 seconds.

References

1. Minkina W.: Pomiary termowizyjne – przyrządy i metody, Wydawnictwo Politechniki Częstochowskiej, Częstochowa
2. Podlasiak P. :Wyznaczanie parametrów dystorsji kamery na podstawie zdjęć obiektów naturalnych i korekcja zdjęć" Archiwum Fotogrametrii, Kartografii i Teledetekcji Vol. 14, Warszawa 2004
3. Zator S., Ryszczyk M.: Korekta pomiaru termowizyjnego w oparciu o trójwymiarowy model obiektu, Przegląd Elektrotechniczny, nr 12/2008, s. 323-325

Author Index

Editors:	Michal Krátký, Jiří Dvorský, Pavel Moravec Dept. of Computer Science VŠB – Technical University of Ostrava
Title:	WOFEX 2011
Author:	composite authors
Place, year, edition:	Ostrava, 2011, 1st
Page count:	693
Edit:	VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava-Poruba Czech Republic
Print:	Tiskárna Tiskservis gen. Sochora 1764, 708 00 Ostrava-Poruba Czech Republic
Impression:	200

Not for sale.

ISBN 978-80-248-2449-9