UNIVERSITY OF ŽILINA

TRANSCOM 2011

9-th EUROPEAN CONFERENCE
OF YOUNG RESEARCH AND SCIENTIFIC WORKERS

PROCEEDINGS

SECTION 3

INFORMATION AND COMMUNICATION TECHNOLOGIES

ŽILINA June 27 - 29, 2011
SLOVAK REPUBLIC

UNIVERSITY OF ŽILINA

TRANSCOM 2011

9-th EUROPEAN CONFERENCE
OF YOUNG RESEARCH AND SCIENTIFIC WORKERS

under the auspices of

Ing. Eugen Jurzyca
Minister of Education, Science, Research and Sport of the Slovak Republic

&

Prof. Ing. Tatiana Čorejová, PhD.
Rector of the University of Žilina

SECTION 3

INFORMATION AND COMMUNICATION TECHNOLOGIES

ŽILINA June 27 - 29, 2011
SLOVAK REPUBLIC

Edited by Eva Gavláková, Peter Brída

ISBN: 978-80-554-0372-4

TRANSCOM 2011
9-th European conference of young research and scientific workers

TRANSCOM 2011, the 9-th international conference of young European researchers, scientists and educators, aims to establish and expand international contacts and co-operation. The main purpose of the conference is to provide young scientists with an encouraging and stimulating environment in which they present results of their research to the scientific community. TRANSCOM has been organised regularly every other year since 1995. Between 160 and 400 young researchers and scientists participate regularly in the event. The conference is organised for postgraduate students and young research workers up to the age of 35 and their tutors. Young workers are expected to present the results they had achieved.

The conference is organised by the University of Žilina. It is the university with about 13 000 graduate and postgraduate students. The university offers Bachelor, Master and PhD programmes in the fields of transport, telecommunications, forensic engineering, management operations, information systems, in mechanical, civil, electrical, special engineering and in social sciences.

SECTIONS AND SCIENTIFIC COMMITTEE

1. TRANSPORT AND COMMUNICATIONS TECHNOLOGY.

Scientific committee: Černý J. (CZ), Drozdziel P. (PL), Dydkowski G. (PL), Gašparík J. (SK), Havel K. (SK), Janáček J. (SK), Jánošíková Ľ. (SK), Kampf R. (CZ), Kavička A. (CZ), Kazda A. (SK), Novák A. (SK), Palúch S. (SK), Peško Š. (SK), Rievaj V. (SK), Šulgan M. (SK), Volek J. (CZ), Žarnay M. (SK), Žarnay P. (SK)

2. ECONOMICS AND MANAGEMENT.

Scientific committee: Bartošová V. (SK), Blašková M. (SK), Borkowski S. (PL), Březinová O. (CZ), Ďurišová M. (SK), Glebocki K. (PL), Graźulis V. (LT), Hittmár Š. (SK), Hrnčiar M. (SK), Kucharčíková A. (SK), Lyakin A. (RUS), Rostášová M. (SK), Rybakov F. (RUS), Seemann P. (SK), Strenitzerová M. (SK), Tomová A. (SK), Veretennikova B.O (RUS)

3. INFORMATION AND COMMUNICATION TECHNOLOGIES.

Scientific committee: Dado M. (SK), Diviš Z. (CZ), Drozdová M. (SK), Hudec R. (SK), Huotari J, (FIN), Keil R. (DE), Klimo M. (SK), Kolev P. (BG), Kotsopoulos S. (GR), Koudelka O. (A), Kováčiková T. (SK), Madleňák R. (SK), Matiaško K. (SK), Ranc D. (FR), Spalek J. (SK), Vaculík J. (SK), Vaculík M. (SK), Vrček N. (CR), Wieser V. (SK), Zábovský M. (SK)

4. ELECTRIC POWER SYSTEMS. ELECTRICAL AND ELECTRONIC ENGINEERING.

Scientific committee: Altus J. (SK), Blažek V. (DE), Brandstetter P. (ČR), Capolino G. A. (FR), Consoli A. (IT), Čápová K. (SK), Dobrucký B. (SK), Janoušek L. (SK), Luft M. (PL), Rusek S. (ČR), Szkutnik, J. (PL), Špánik P. (SK), Vittek J. (SK)

5. MATERIAL ENGINEERING. MECHANICAL ENGINEERING TECHNOLOGIES.

Scientific committee: Adamczak S. (PL), Bokůvka O. (SK), Dzimko M. (SK), Guagliano M. (I), Kunz L. (CZ), Meško J. (SK), Neslušan M. (SK), Nicoletto G. (I), Palček P. (SK), Skočovský P. (SK), Takács J. (H)

6. MACHINES AND EQUIPMENTS. APPLIED MECHANICS.

Scientific committee: Dekýš V. (SK), Gerlici J. (SK), Chudzikiewicz A. (PL), Jandačka J. (SK), Kalinčák D. (SK), Malenovský E. (CZ), Medvecký Š. (SK), Nemček M. (CZ), Sága M. (SK), Sitarz M. (PL), Szava I. (RO), Zapoměl J. (CZ), Žmindák M. (SK)

7. CIVIL ENGINEERING.

Scientific committee: Bujňák J. (SK), Ferrero A. M. (IT), Garbuz A. (UA), Horváth F. (HU), Ižvolt L. (SK), Melcer J. (SK), Petkova R. (BG), Plášek O. (CZ), Malachova A. (RU), Ungureanu V. (RO)

8. SOCIAL SCIENCES.

Scientific committee: Banáry B. (SK), Cabanová V. (SK), Grecmanová H. (CZ), Hádková M. (CZ), Kráľová Z. (SK), Larry Fast (USA), Lengyelfalusy T. (SK)

9. SECURITY ENGINEERING. FORENSIC ENGINEERING.

Scientific committee: Artamonov S. V. (RU), Burg H. (DE), Dudáček A. (CZ), Horák R. (CZ), Kasanický G. (SK), Klučka J. (SK), Leitner B. (SK), Navrátil L. (CZ), Podbregar I. (SLO), Poledňák P. (SK), Šimák L. (SK), Štofko S. (SK)

ORGANIZING COMMITTEE

CHAIRPERSONS Čelko Ján, Bokůvka Otakar

EXECUTIVE SECRETARY Vráblová Helena

MEMBERS Bača Tomáš, Bačová Katarína, Baďurová Silvia, Belan Juraj, Bomba Lukáš, Brída Peter, Brumerčík František, Gavláková Eva, Hampl Marek, Harušinec Jozef, Horváth Peter, Hrbček Jozef, Jošt Jozef, Kittel Ladislav, Kolla Eduard, Koniar Dušan, Kopas Peter, Lanďák Milan, Lieskovský Anton, Mendrošová Katarína, Mišiaková Kvetoslava, Mokryš Michal, Mrvová Miroslava, Mužíková Karolína, Pácha Matěj, Peterková Andrea, Pilát Peter, Pitor Ján, Raždík Ján, Smetana Milan, Spuchľáková Erika, Šípek Michal, Šramová Veronika, Tengler Jiří, Tkáčová Gabriela, Vaško Alan, Vaško Milan, Vlček Jozef, Vrábel Ján, Vráblová Lucia, Závodská Anna, Zvaríková Katarína, Žiačková Vladimíra

SECTION 3 INFORMATION AND COMMUNICATION TECHNOLOGIES

REVIEWERS:

Benčo Miroslav
Brída Peter
Čáp Ivo
Dado Milan
Drozdová Matilda
Dúha Ján
Fabuš Juraj
Hrbček Jozef
Hudec Róbert
Janech Ján
Janota Aleš
Jarina Roman
Káčik Daniel
Klimo Martin
Kováčiková Tatiana
Kremeňová Iveta
Madleňák Radovan
Matiaško Karol
Muzikářová Ľudmila
Paralič Martin
Spalek Juraj
Šimák Vojtech
Švadlenka Libor
Tomová Anna
Vaculík Juraj
Vaculík Martin
Wieser Vladimír
Zábovský Michal
Ždánsky Juraj

Note:

Authors are responsible for language contents of their papers

CONTENTS

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Financial Time Series Modelling with GARCH Models

[*]Ján Bábel
[*]University of Žilina,Faculty of Management Science and Informatics,Department of Macro and Microeconomics, Univerzitná 8215/1, 010 26 Žilina, Slovakia{jan.babel}@fri.uniza.sk

Abstract: This paper deals with heteroscedasticity and its presence in financial time series. By examining numerous time series it tries to find an answer to question: „How many financial time series are heteroscedastic? " If time series is heteroscedastic, the assumption for using GARCH models is fulfilled. Ii would be interesting to know the number of financial time series for which the use of GARCH model is adequate.

Keywords: heteroscedastic, GARCH, ARCH test, currency cross rate

1. Introduction

GARCH models are designed to capture certain characteristics that are commonly associated with financial time series:

- fat tails
- volatility clustering
- persistence
- mean-reversion
- leverage effect

The first model that provides a systematic framework for volatility modeling is the ARCH model of Engle (1982). Bollerslev (1986) proposes a useful extension of Engle's ARCH model known as the generalized ARCH (GARCH) model.

$$\begin{aligned} r_t &= \mu + a_t, \quad a_t \sim N(0, \sigma_t) \\ a_t &= \sigma_t v_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \end{aligned} \qquad (1)$$

where r_t is a log return series, a_t is the innovation in time t, v_t is a sequence of iid random variables with mean equals to 0 and variance equals to 1. Variables α_i a β_i are called ARCH and GARCH parameters and are computed by maximizing the likelihood function (MLE). σ_t^2 represents the estimated volatility of time series[1].

2. Detection of Heteroscedasticity

Heteoscedasticity can be usually found in majority of financial time series, but before using GARCH models, we need to test the residuals of time series for the presence of heteroscedasticity. If there is a heteroscedasticity, model with conditional heteroscedasticity, GARCH model, can be

[1] [9] TSAY, RUEY S.: Analysis of Financial Time Series, page 114

used. If there is no indication of heteroscedasticity, assumption of using GARCH models is violated.

For this purpose, I choose to examine currency cross rates and commodities (gold, silver and platinum) offering for free by Oanda currency site. The period under examination is 1st of June 2009 to 13th of October 2010, together 500 daily observations. 50 national currencies and commodities were chosen, which gives together 50 choose 2 combinations. We obtain 1225 financial time series, specifically currency cross rates and spot prizes of gold, silver and platinum.

Before estimating a full ARCH-GARCH model for a financial time series, we need to test for the presence of ARCH effects in the residuals. If there are no ARCH effects in the residuals, then the GARCH model is unnecessary and misspecified.

Since an ARCH model can be written as an AR model in terms of squared residuals as

$$\varepsilon_t^2 = a_0 + a_1\varepsilon_{t-1}^2 + \dots + a_p\varepsilon_{t-p}^2 + u_t \qquad (2)$$

a simple Lagrange Multiplier (LM) test for ARCH effects can be constructed based on the auxiliary regression (3). Under the null hypothesis that there are no ARCH effects: $a_1 = a_2 = \dots = a_p = 0$, the test statistic

$$LM = T.R^2 \sim \chi^2(p) \qquad (3)$$

Where T is the sample size and R^2 is computed from regression (3) using estimated residuals.[2]

According to ARCH LM test with number of lags equals to 12, from overall 1225 financial time series, 220 were considered as homoscedastics (18%), which is a surprising fact according to theoretical assumptions of financial time series.

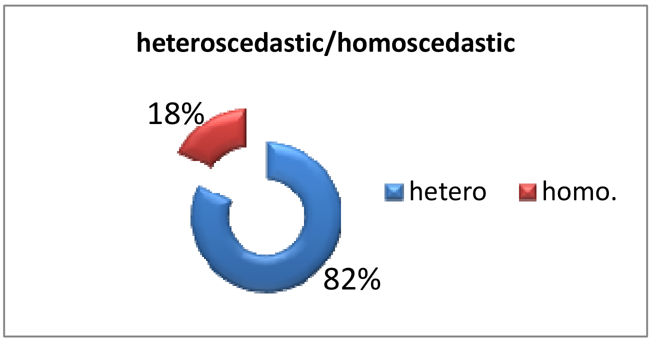


Fig. 1. Comparison of hetero. and homo. Series.

The analysis was made within R environment, using packages quantmod and tseries for data downloading. Detailed frequency table of every single currency and its presence as homoscedastic, is shown in figure 2. The most frequent homoscedastic time series in combination, with other currencies or commodities, were platinum (33x) and gold (32x) .

[2] Modelling Financial Time Series with S-PLUS, Zivot, Wang (2005)

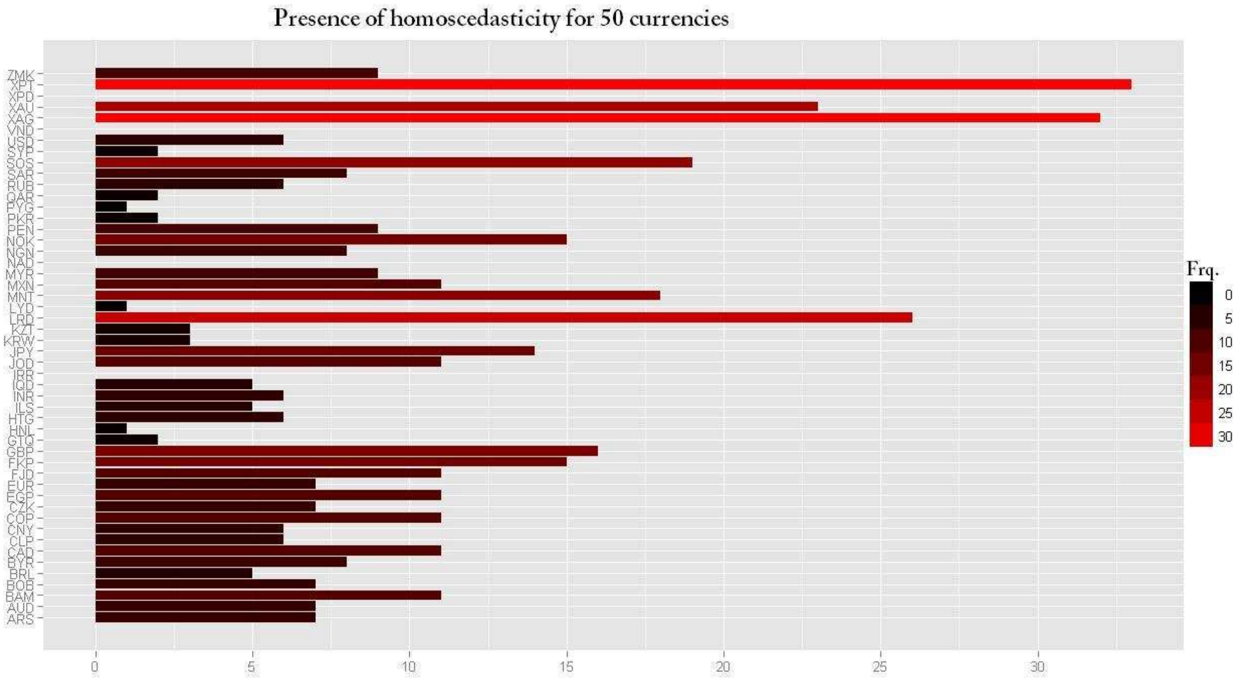


Fig. 2. Presence of homoscedasticity for chosen 50 currencies according to ArchLM test

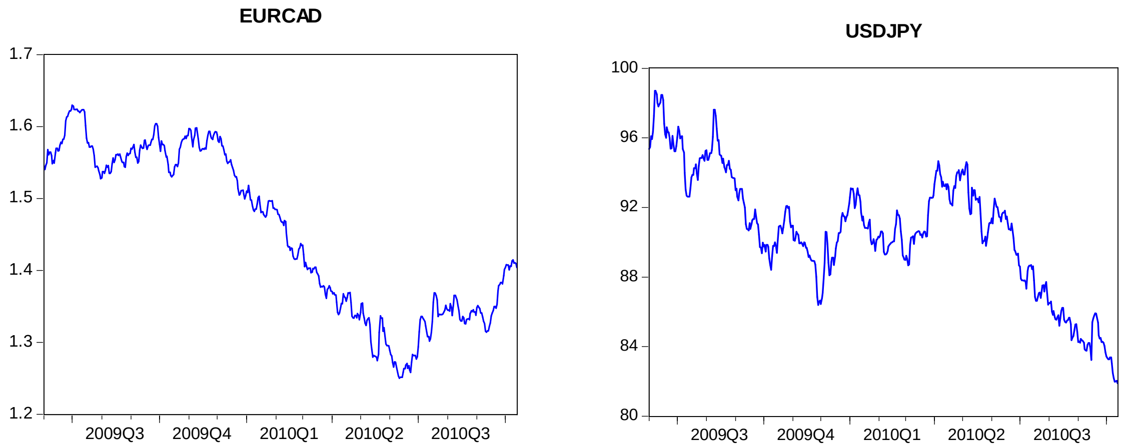


Fig. 3. Cross currency rates: Euro against Canadian dollars and American dollar against Japnesse Yen

Then I examined 2 time series, one from each group and build GARCH models. EUR/CAD represents heteroscedastic time series, USD/JPY homoscedastic. Theoretical assumption is, that GARCH models should work better with heteroscedastic time series and reach lower RMSE than linear ARMA models. Both time series are shown in figure 3.

USD/JPY	*Aproximation*	*Prediction*	*Aproximation*	*Prediction*
Error statistics/models	*ARMA(1,0)*	*ARMA(1,0)*	*ARMA(1,0)+GARCH(1,1) + t distribution*	*ARMA(1,0)+GARCH(1,1) + t distribution*
Root Mean Squared Error(RMSE)	0.430612	0.431572	0.430640	0.435503
Mean Absolute Error (MAE)	0.304027	0.264251	0.303666	0.271930

Tab. 1. Results for homoscedastic time series USD/JPY

EUR/CAD	*Aproximation*	*Prediction*	*Aproximation*	*Prediction*
Error statistics/models	*ARMA(1,0)*	*ARMA(1,0)*	*ARMA(1,0)+GARCH(1,1)+ GED fixed 1.7*	*ARMA(1,0)+GARCH(1,1) + GED fixed 1.7*
Root Mean Squared Error(RMSE)	0.006870	0.005692	0.006870	0.005588
Mean Absolute Error (MAE)	0.004918	0.004245	0.004907	0.004170

Tab. 2. Results for heteroscedastic time series EUR/CAD

3. Conclusion

From results listed in Table 1 and 2 follow that GARCH extension of ARMA models is suitable when heteroscedastic time series is examined (EUR/CAD example). The joint estimation of ARMA+GARCH models beats ARMA model primarily in prediction part. These findings correspond with theoretical assumptions of these models. However, surprising is the fact, that almost one fifth of all currency cross rates are homoscedastics.

Acknowledgement

This work was supported by Slovak grant foundation under the grant VEGA No. 1/0667/10.

References

[1] BOLLERSLEV, T. *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics 31: 307–327, 1986.

[2] CARMONA, RENE A.. *Statistical Analysis of Financial Data in S-Plus*, Springer, ISBN: 978-0-387-20286-0, 2004.

[3] ENGLE, R.: GARCH 101. *The Use of ARCH/GARCH Models in Applied Econometrics*, Journal of Economic Perspectives—Volume 15, Number 4, pages 157–168, 2001.

[4] MARČEK, D., MARČEK, M. *Analýza, modelovanie a prognózovanie časových radov s aplikáciami v ekonomike*, ES ŽU Žilina, ISBN:80-7100-870-2001.

[5] MARČEK, M. – PANČÍKOVÁ, L. – MARČEK, D. *Ekonometria a soft computing.* Žilina: EDIS – vydavateľstvo ŽU, 2008.

[6] TSAY, RUEY S. *Analysis of Financial Time Series, Second Edition*, John Wiley & Sons, Inc., ISBN-13 978-0-471-69074-0, 2005.

[7] ZIVOT, E., WANG, J. *Modelling Financial Time Series with S-PLUS®*, Springer, 2005, ISBN: 978-0-387-27965-7

[8] OANDA, *The currency site (2008) http://www.oanda.com/*

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Detection of Abnormalities in ECG Using MATLAB

*Branko Babušiak, *Michal Gála
*University of Žilina, Faculty of Electrical Engineering, Department of Electromagnetic and Biomedical Engineering, Univerzitná 1, 01026 Žilina, Slovakia, {branko.babusiak, michal.gala}@fel.uniza.sk

Abstract. The neural networks have many ways of usage in technical field. They have been applied successfully to speech recognition, image analysis and adaptive control, in order to construct software agents or autonomous robots. The electrocardiography (ECG) signal is periodical and therefore it is quite predictable. In this paper it is described usage of neural network for ECG signal prediction. Successful prediction of ECG is used for abnormalities detection – artifact, extrasystole, apnea, etc. An automatic abnormalities detection system was created for offline and online purpose. Functionality of this system is also described in this paper.

Keywords: ECG, neural network, signal prediction, detection of abnormality.

1. Introduction

The neural network is computational model based on the features abstraction of biological neural systems. Neural networks are an efficient, pervasive, and powerful means of computation. The creation of the neural networks was inspired by the study of the human brain. Neural networks are pattern classifiers. They do not store "knowledge" in a memory bank. The information is distributed throughout the network and is stored in the form of weighted connections. The most valuable characteristics of neural networks are adaptability and tolerance to noisy data. In general, they are well suited for applications that involve classification of input (e.g., digital image, natural language, and speech processing). Neural networks are not appropriate for problems that require precise, unary answers, such as solving mathematical problems [1][2][3][5].

The first main difference between neural network and computer program is that neural networks are robust against the errors. If computer program contains an error, the system will collapse. Errors or divergences in neural networks do not lead to collapse in the system. The second difference is that neural networks perform their tasks in parallel (increases speed of computation). The parallel processing of signal is typical feature of the brain neural networks [1][2][3][5].

The basic element of neural network is a neuron. The structure of the neuron is in the figure 1.

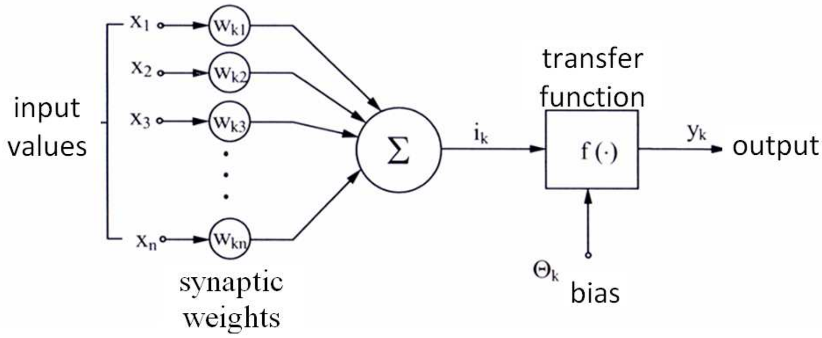


Fig. 1. Basic model of the neuron.

An electrocardiogram (ECG) is a graphic product from an electrocardiograph which records the electrical activity of the heart over time. An ECG displays the voltage between pairs of the electrodes, and the muscle activity that they measure, from different directions. This display indicates the overall rhythm of the heart, and weaknesses in different parts of the heart muscle. It is the best way to measure and diagnose abnormal rhythms of the heart [7][8].

The heart activity repeats periodically; therefore, in the record there are recognized intervals which are similar and repeated. Each interval consists of a P wave, a QRS complex and a T wave (fig. 2).

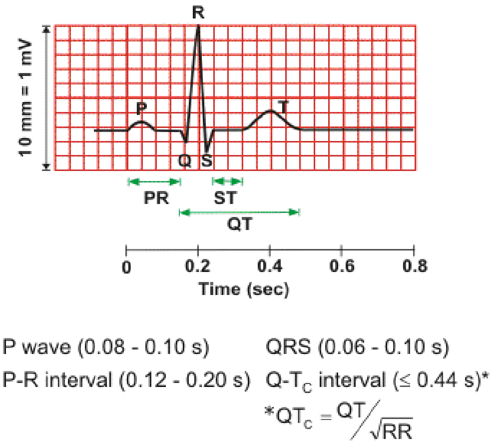


Fig. 2. Schematic representation of normal ECG curve.

2. Offline prediction of ECG

Let us consider that "healthy and calm" ECG record consists of periodically repeated intervals. Then it is possible to predict the following time sample of the digitalised ECG record if a few previous samples are known. Neural network is used for estimation of the next time sample. Realization of this task is performed in the MATLAB environment. The neural network will be designed and simulated over the real ECG signal. A short interval of this signal is displayed in the figure 3. The signal is ten seconds long and its sampling rate is 100 Hz.

At first it is needed to extract suitable training signal from the whole ECG signal. As it is seen in figure 3, signal has fluctuating (sinusoidal) progress due to breathing process. Fluctuating progress is best seen on R peaks. For that reason the training signal (set) has to be a part of the whole signal which includes at least one period of mentioned sine wave. A training signal (red part of figure 3) consists of 620 samples.

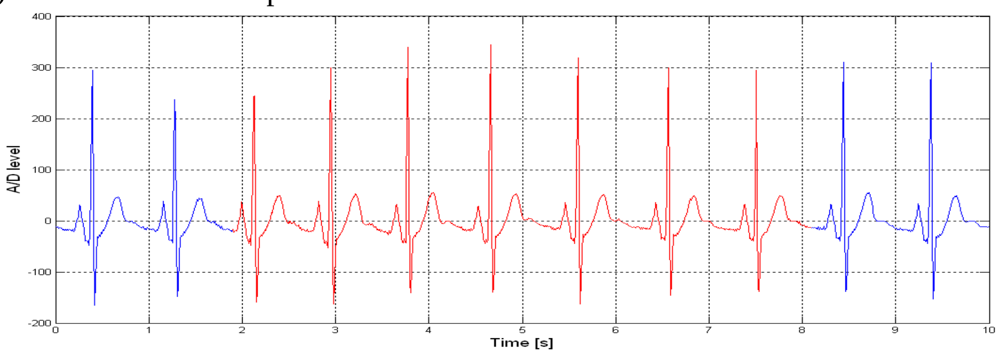


Fig. 3. ECG signal (length: 10 sec.; sampling rate: 100 Hz). Training set for neural network is marked by red colour.

Now, let us try to create neural network which will be able to estimate next sample if previous five samples are known. Two different types of neural network – the linear layer with one neuron and multi-layered backpropagation neural network are used to solve this problem in comparison purpose.

The linear layer with one neuron is shown in the figure 4. Transfer function is linear. It is very important to create correct matrix of inputs **P** and vector of expected outputs **T**. Each column of the matrix **P** stands for five consecutive samples (dimension **P** is [5x620]). **T** is a row vector with 620 expected outputs (output = following sample after five samples in **P**).

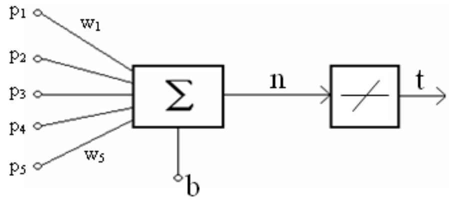


Fig. 4. Linear neuron.

Training process takes approximately 30 milliseconds (all time data are measured on mobile Intel Pentium Dual-Core CPU T4300 2.1 GHZ). In verification process of trained network it was used the part of signal outside of training set range. Test signal is also 10 second long (1000 samples). Test signal (blue) and the result from neural network (red) are shown in the figure 5. Prediction of new 1000 samples takes 15 milliseconds on average.

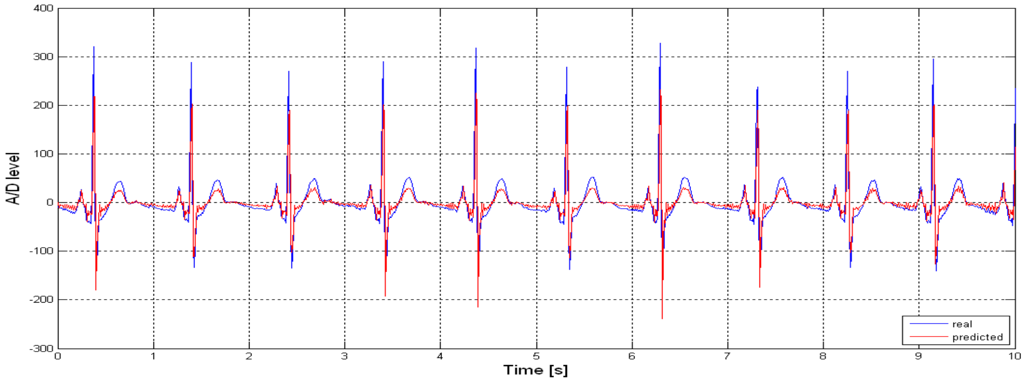


Fig. 5. Prediction of the ECG by using simple linear neuron.

In the second case it is created the backpropagation neural network with three layers (figure 6). The network has three neurons in the first layer, two neurons in second (hidden) layer and one neuron in the third layer. It is used log-sigmoid transfer function (logsig) in each layer. The learning rate was set on value 0.05 and tolerable error is 5.10^{-5}. For computation of weights *w* and biases *b* it was used trainlm algorithm which is based on Levenberg-Marquardt optimalization. Designed network was applied to the same signal as in the previous case. The result is shown in the figure 7.

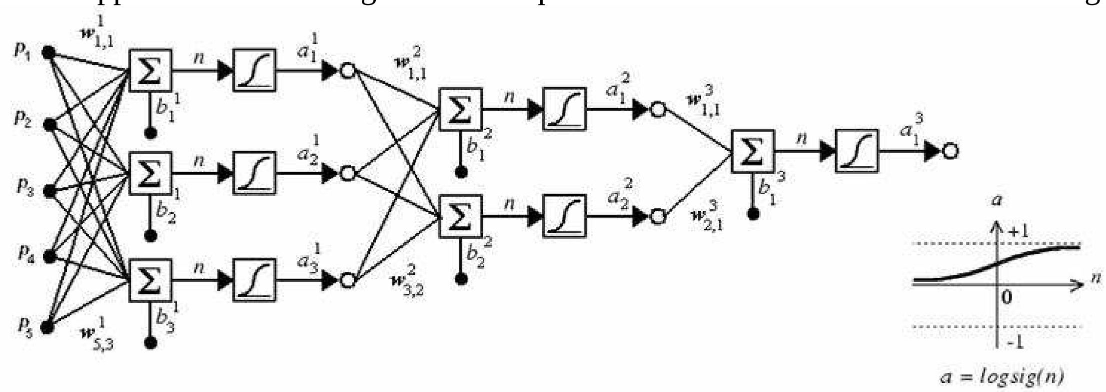


Fig. 6. Three-layered neural network. Transfer function on the right.

Duration of training process vary and depends on number of iteration. For example, it takes 3.1 seconds for 68 iterations and 2.2 seconds for 49 iterations. Prediction of new 1000 samples takes 25 milliseconds on average.

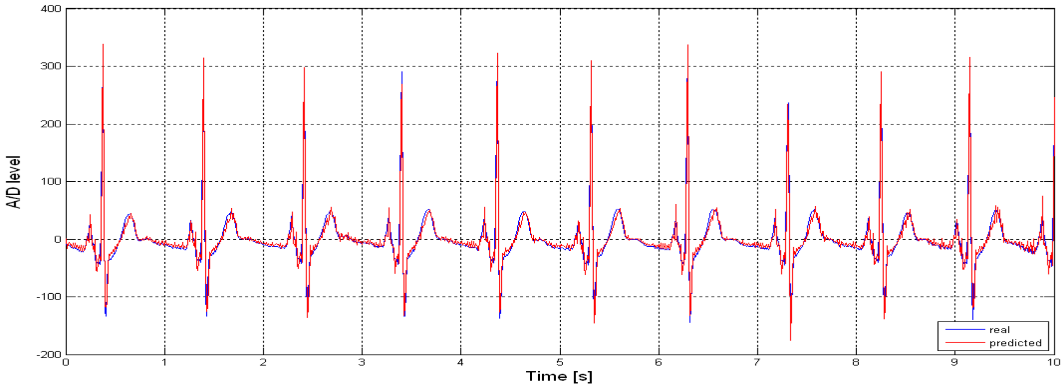


Fig. 7. Prediction of the ECG by using three-layered neural network.

The three-layered neural network achieves better precision in predicted signal in comparison with simple neuron. But it takes longer time for prediction on the other side. In the figure 8 it is displayed difference between real and predicted signal for both cases. The difference is computed as:

$$d(k) = \left| r(k) - p(k) \right|, \quad (1)$$

where *d* is difference signal; *r* is real signal; *p* is predicted signal and *k* is sample order.

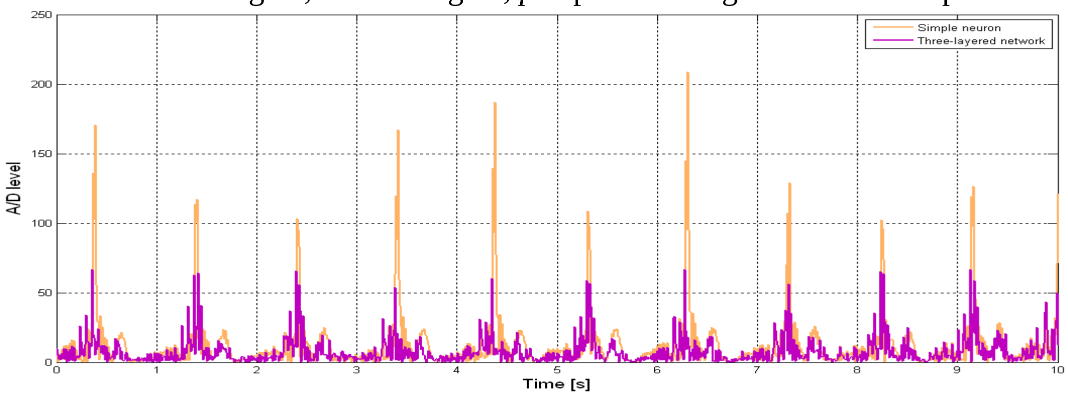


Fig. 8. Difference between real and predicted signal.

3. Detection of abnormalities

The three-layered neural network is used for the abnormality detection system because of previous result. More layers do not lead to better prediction of ECG. Automatic detection system is very useful in evaluation of long-time ECG records. Long-time ECG (e.g. all-day monitoring) contains many abnormalities (extrasystoles, movement artefacts, temporary bad electrode contact, etc.). Designed system helps cardiologist to focus only on parts which are suspicious for any kind of abnormality.

System was tested on long-time record (7 hours and 48 minutes). For example, in the figure 9 it is shown 8 seconds long part of signal with presence of extrasystole (blue signal; extrasystole at 22nd second). Predicted signal is shown in figure 9 (red). In the figure 10 a difference between real and predicted signal is displayed. Big value of difference stands for extrasystole in the ECG. If boundary for abnormality is set to value 0.15 (fig. 10 horizontal line) then system is able to detect any unexpected change in the ECG. In order to eliminate detected abnormalities which are caused by movements (muscular activity) it is appropriate to use low pass filter to reduce higher frequencies. Detected abnormalities are then marked in the record. System is time-saving because cardiologist can focus only on marked regions in long-time ECG.

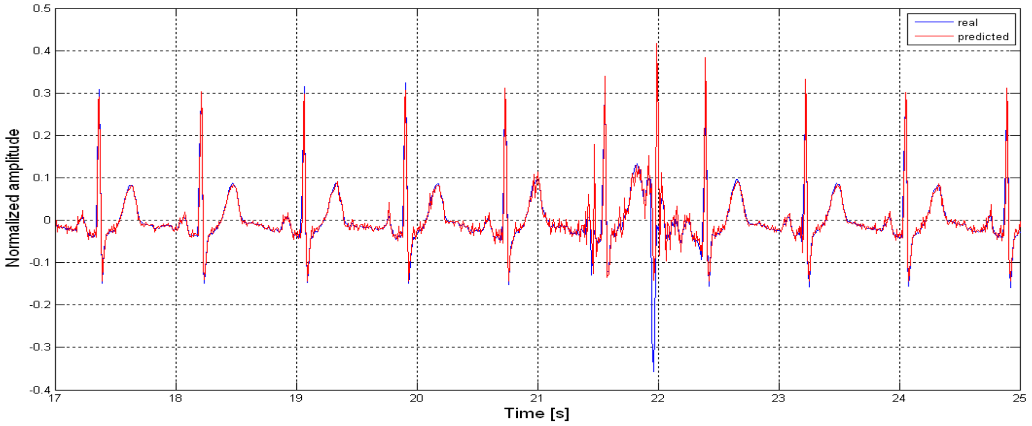


Fig. 9. Signal with presence of extrasystole.

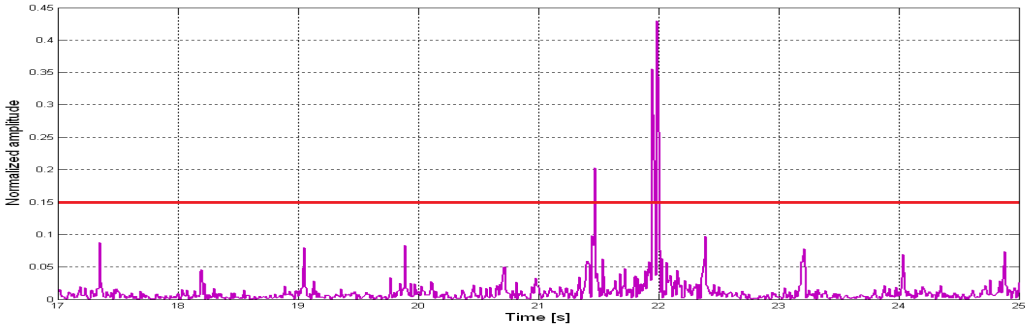


Fig. 10. Difference between real and predicted signal. Horizontal line represents boundary for abnormality.

4. Online abnormality detection

Online prediction is dedicated for cardiomonitors which communicate with PC station via Ethernet (fig. 11 A). For this purpose it was created C# application ECGgenerator which simulates cardiomonitor and it transmits ECG signal to the network - intranet, internet (fig. 11 B). The application can run on arbitrary PC connected to the network. ECGgenerator loads real ECG record from the specified binary file. Each sample of signal is transmitted one hundred times in a second. It corresponds to sample rate of 100 Hz.

Prediction is realized for every half second (50 samples). Right after prediction difference between real and predicted signal is computed. Therefore prediction and evaluation of difference is a half second delayed. If difference exceeds boundary of abnormality, system will warn user by beep sound, warning message on the screen etc. Moreover ECG included with positions of abnormalities is stored in to hard drive. While prediction and evaluation are made next samples from cardiomonitor (ECGgenerator) are still received in the background without any delay. Processing of each 50 samples takes about 15 milliseconds.

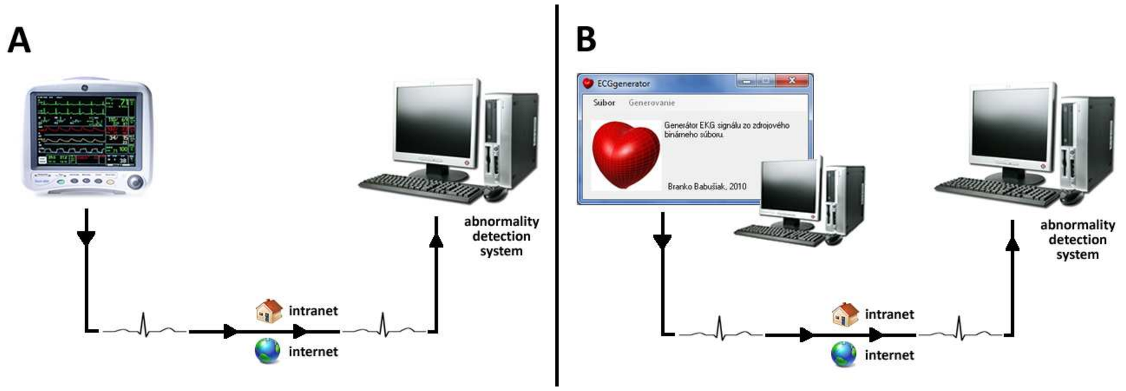


Fig. 11. Online abnormality detection system connected via Ethernet.

5. Conclusion

In this contribution it was presented using of neural network in the ECG signal prediction. It was created abnormality detection system, which is not finalized yet. Designed system can be used for detecting abnormalities in either offline or online applications. Offline detection system considerably helps cardiologist to evaluate long-time (a few hours) ECG record. Online detection system provides long-distance monitoring of ECG thanks to internet connection.

Authors still work on improvements of proposed detection system. They would like to change training set and retrain neural network in dependence on changing heart beat frequency. For this purpose QRS detection algorithm will be implemented into the detection system. Detection of abnormalities in online application will be accelerated by using CUDA technology – parallel computing on Graphical Processing Unit by NVIDIA.

References

[1] SINČÁK, P., ANDREJKOVÁ, G.: *Neurónové siete I.* (inžiniersky prístup), ELFA Press, Košice, 1996.
[2] SINČÁK, P., ANDREJKOVÁ, G.: *Neurónové siete II.* (inžiniersky prístup), ELFA Press, Košice, 1996.
[3] VONDRÁK, I.: *Umělá inteligence a neurónové sítě*. Katedra informatiky, FEI, VŠB – TU Ostrava, 2002.
[4] MATLAB Neural Network Toolbox: http://www.mathworks.com/access/helpdesk/help/toolbox/nnet
[5] Wikipedia, the free encyclopedia: http://www.wikipedia.org
[6] Neural network: http://gseacademic.harvard.edu/~elsheish
[7] MOHYLOVÁ, J., KRAJČA, V.: *Zpracování signálů v lékařství*, Žilinská Univerzita v Žiline, 2004.
[8] JAVORKA, K.: *Lekárska fyziológia*. Osveta, Martin, 2001.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Data Dissemination in VANET

[*]Tomas Baca , [**]Anton Lieskovsky
[*]University of Zilina, Faculty of Management Science and Informatics, Department of software technologies, Univerzitna 8215/1, 010 26 Zilina, Slovakia, {tomas.baca }@kst.uniza.sk
[**] University of Zilina, Faculty of Management Science and Informatics, Department of informatics, Univerzitna 8215/1, 010 26 Zilina, Slovakia, {anton.lieskovsky}@fri.uniza.sk

Abstract. In this paper we will look at issues of vehicle communication in VANET. The main objective is to present an algorithm for distributing data in VANET. Distribution of data in VANET is essential for vehicle communication. In addressing this issue it is necessary to face the problems that are caused by environmental influences or lack of resources.

Keywords: VANET, Data dissemination, Vehicle, Replication.

1. Introduction

The idea of vehicle communication in VANET promises many benefits of intelligent transport systems, e.g. reduced number of accidents, mitigation of accident consequences, better use of road and resources such as time, fuel and space for new innovative applications. However, before the widespread deployment of vehicle communication, it is necessary to solve many problems and challenges in this area. Vehicle communication does not only create virtual information space, but must secure a firm definition of rules for its use. This is the only way to create stable, safe and acceptable system usable for ITS (intelligent transportation system).

This article examines the issue of data distribution in VANET. Successful distribution of data is a prerequisite for successful functioning of VANET. In the paper we firstly introduce the importance of vehicle communication, and then we look at design of messages that serve for communication in VANET. The main objective is to present the algorithm of data distribution and the results of its use

2. Why Vehicular Communication?

In the early automobile industry, road infrastructure and various types of cars were regarded as autonomous systems. Mutual influence and growth (higher number of cars, larger and more sophisticated transport infrastructure) of these systems required their connection with traffic signs. The whole system was again considered as autonomous.

Since (today) the number of accidents, congestion, number of cars and size of road infrastructure are increasing we can consider the current autonomous system to be insufficient. The next stage of the automobile industry (from the point of view of intelligent transport systems development), is to find a way to control the whole system. This idea tends towards the concept of fully automated cars that in future will be capable of "self-service" (drive by itself). We can see the beginnings of this new system in the automation of some of the usual driving manoeuvres. The best known applications of this automation include: self parking, automatic speed adaptation ... Another important step towards improving traffic condition will be the integration of vehicle communication via VANET (Vehicular ad-hoc network). It is possible to achieve a collaborative approach in VANET by sharing information that individual cars can identify from the environment.

Examples of utilizing the benefits of automobile communications are spreading Warnings (messages on accidents or traffic restrictions), and the like. The big advantage of such cooperation lies in the immediate car to car communication and informing on real-time local information.

Vehicle communication brings new functionality which implementation will help in the progress of research on intelligent transport systems (ITS).

3. Related work on data dissemination

Research on the data distribution is often presented by term data replication. Replication techniques are based on different principles.

Technique based on access frequency see [1] divides the data on "hot data items" and "cold data items. This division is made on the basis of query (requests) frequency from mobile nodes on the source data.

Another way to spread data is their replication based on the date, as suggested in [2]. Data replication in VANET is also possible to implement in terms of database systems. In [3] the VANET philosophy is presented as distributed database system with architecture adapted for mobile networks.

Efficient data distribution is a challenge for the design of adaptive algorithms that ensure the distribution of information in VANET with minimal cost on resources that VANET provides.

4. Description of simulation

In our experiments we have simulated a system of early warning for road users of impending danger on the road. The problem was simplified to the simulation of cars driving in road network illustrated by the picture. Cars were capable of wireless communication and therefore were forming VANET network.

Drivers of cars were facing danger in form of puddles on the road. Driving through a puddle can be considered dangerous. However, if the driver had been warned in advance, he could pass through it slowly and safely. Thus we equiped the cars in the simulation with capability to detect puddles when they were driving through them and they were also capable of autonomous warning message transmission.

In the simulation cars were driving in urban environment such as displayed at the picture. Roads (grey) were bordered by buildings (green) which were forming obstacles for wireless communication. Cars (yellow) were driving in speeds ranging from 10 ms-1 to 20 ms-1. They were entering the road network through “gateways” (light green at the bottom of the map) and later on leaving it again through the gateways. The whole road network has spanned at the area 4500m long and 3000m wide totaling approximately 25500m of roads.

For our experiments we were changing parameters of dissemination algorithm described in next section. For every tested combination of values 25 simulation runs were evaluated. Every simulation run was 50000 simulation seconds long.

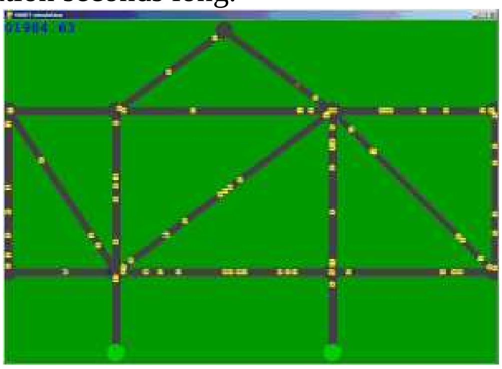


Fig. 1. The sample of simulation software.

5. Design of algorithms for data dissemination

The main idea of the algorithm is to keep the information within certain geographical area till the information is not outdated. Messages are being sent as broadcasts and received by cars within communication range. A car that received such a message will rebroadcast it if it is not further from the original source of the message than certain distance (controlled by parameter DISTANCE) and also if the information is not outdated (is not older than value of the parameter KEEPALIVE). The algorithm is event driven.

Algorithm can be controlled by several parameters:

- RANGE = maximal distance,
- LIFETIME = how long is information considered valid,
- DEFFERTIME = time interval between retransmissions.

When car is driving through a puddle:

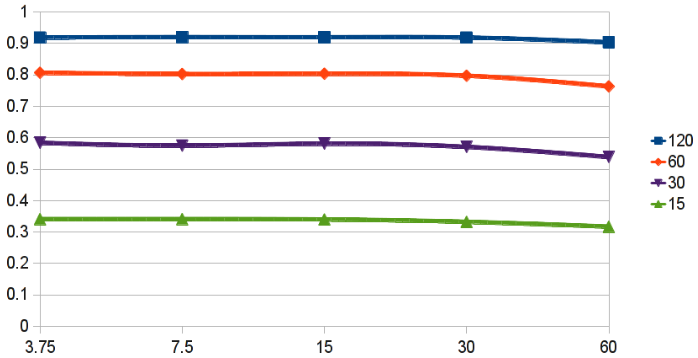
1. Send a message containing:
   - POSITION = geographical coordinates of puddle,
   - EVENTTIME = time when the car detected that it is in the puddle,
   - SENDER = identification of sender of the message.
2. When car received a message
   1. Update information about puddles on the road, prefer most recent data based on message.EVENTTIME
   2. Wait for DEFFERTIME
   3. If current time - EVENTTIME < LIFETIME and distance between current position and POSITION < RANGE and the message contains most recent information available then rebroadcast the message else stop.
   4. go to step 2

6. Results

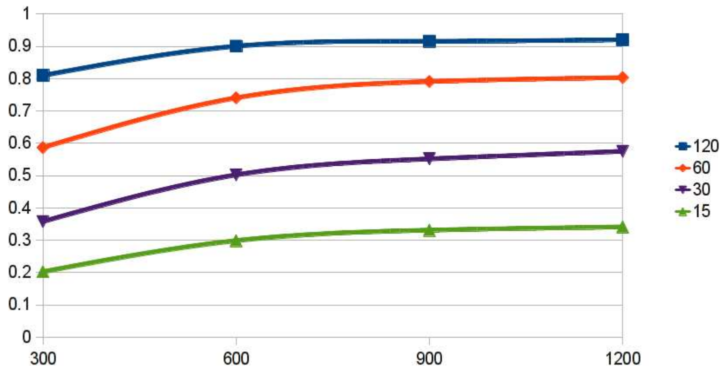
In our simulation, parameters RANGE and LIFETIME were bounded together. LIFETIME was set to RANGE/10 because that was the maximum time for car to stay within the range anyway (speed of cars was ranging from 10 to 20 ms-1). In addition to parameters of routing algorithm we were also experimenting with density of road traffic, measured only approximately as average number of cars in the road network.

We were evaluating the algorithm by Warned Drivers Ratio (WDR). That is how many times a driver had been warned when driving through a puddle in proportion to all times a driver was driving through a puddle.

Our results are visualised at graphs 1 and 2. At both graphs there are results for various densities of traffic. At graph 1 one can observe influence of DEFFERTIME (X-axis) on WDR (Y-axis) when RANGE was fixed at 1200 m. At graph 2 one can observe influence of RANGE (X-axis) on WDR when DEFFERTIME was fixed at 7.5 s.



Graph 1: Influnece of DEFFERTIME, when RANGE = 1200



Graph 2: Influence of RANGE, when DEFFERTIME = 7.5 s

Graphs 1 and 2 does not include all the results we obtained from our simulations. We experimented with larger set of parameter values. However, influence tendencies of parameters are clear.

From the graph 1 one can see that higher value of DEFFERTIME causes worse value of WDR. However the impact is only slightly noticable. From graph 2 one can observe noticable improvement of WDR with rising value of RANGE. It does have significant impact especially with lower values.

However density of traffic proved to be very important. It has very strong influence on the resulting value of WDR as one can see from both graph 1 and 2. One can observe significant gaps between lines displaying results for different densities of traffic.

7. Conclusion

We have proposed an algorithm for distribution of safety information in VANET and tested it in computer simulation. The algorithm has proved to be able to give good results, value of WDR over 90%, with parameters properly set and with road traffic dense enough.

As stated at the beginning of the article the idea is to introduce control of the whole system because of road traffic getting more and more dense. Thus we can allege that with the proposed algorithm we are on good way in accomplishing the goal and we can continue in augmenting it for our further research.

References

[1] T. Hara, Y.H.Loh, S.Nishio, "Data Replication Methods Based on the Stability of Radio Links in Ad Hoc Networks", Journal of the Information Processing Society of Japan, Vol.44, No.9, Sept. 2003, pp.2308-2319.

[2] B.Xu, O.Wolfson, S.Chamberlain, Y.Yesha, "Adaptive Lazy Replication in Unreliable Broadcast Networks", Conference on Extending Database Technology,2000.

[3] Janech, Ján: Distributed Database Systems in the Dynamic Networks Environment. In: International Journal on Information Technologies and Security. Sofia : Union of Scientists in Bulgaria, 2010. n◦ 1. ISSN 1313-8251.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Speech Recognition with Fuzzy Flip-flops

*Štefan Baďura, *Stanislav Foltán
*University of Zilina, Faculty of Management Science and Informatics, Department of InfoCom Networks, Univerzitná 8215/1, 010 26 Žilina, Slovakia, {stefan.badura, stanislav.foltan}@fri.uniza.sk

Abstract. In this paper we present a new model of neural network. Our design is a special kind of network used in tasks of speech recognition. Basic hierarchical structure is proposed with all necessary stages for its creation. Fuzzy flip-flop element is taken as basic building unit (node). Lindenmayer's system (L-system) is used for connecting nodes into neurons. Evolution rules, genetic programming, GPU computing are utilized to improve performance and final results. For general evaluation of proposed methods, the whole model will be evaluated in task of bimodal speech recognition.

Keywords: Fuzzy flip-flop, speech recognition, neural network, neuron.

1. Introduction

The topic of this work is the speech recognition, utilizing network of fuzzy flip-flops. The construction unit of the network, the fuzzy flip-flop (can be understood as a neural component of a neural network), and it was designed at the Department of Communication Networks [1]. Unlike L. T. Koczy, who introduces a network of JK flip-flops [2] as an alternative to a neural network, this is based on classic RS flip-flop. Significant change is the utilization of probabilistic algebra instead of Boolean algebra. Another significant change is the addition of new inputs (reward, punishment, limit and memory). These changes serve to influence firing properties of neurons. Now we can affect power and duration (speed of spontaneous forgetting) of excitement. The idea of fuzzy flip-flops is main motivation of our work and our system design, but it is more than just a basic element. The whole system consists of many stages; they are briefly described in following sections.

The second section describes a basic design of our system. The 3rd section outlines some features as they were defined in [1]. The 4th section describes more complicated structures as neurons, layers and population are introduced. In last section are briefly described inputs for our problem (video sequences and audio records) and it concludes this paper and proposes some future steps.

2. Network design

As it was mentioned above, the main problem and motivation that has to be solved is a proposal of network design. It can be separated into different stages. At first basic element – construction unit is designed, which provides some important abilities for our purposes (as it was described in introduction). For more details see [3]. Next, these basic elements are connected into more sophisticated structures, in our case we are talking about neurons. Next stage describes mutual interconnection of constructed neurons into a layer. Other stages consist of enlarging whole network into layered structure. In final stages the whole system grows into population. For more details see Figure 1. In this paper we briefly describe all stages. Our model is intended to be used especially in tasks of speech recognition, therefore dynamic properties of inputs need to be considered. In our case the dynamic properties are learned in network structure.

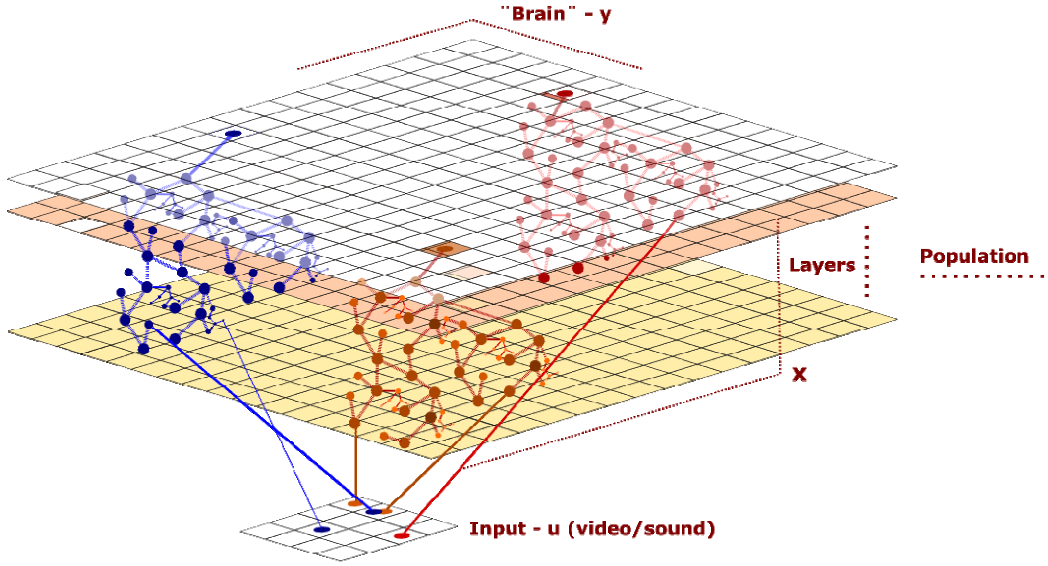


Fig. 1: System design. It consists of more populations, each population represents one network. Network is constructed from layers. In layers are located neurons, those are constructed from basic flip-flop elements.

3. Basic element

As the basic element we understand fuzzy flip-flop which provides suitable properties for our model. These properties could be: one basic element must support its neighboring element (principles of Hebbian influence). Next, this element must be able to provide some possibilities of manipulation – make a relationship between elements stronger or weaker. The important property is also to remember state of node (suitable time period). The Figure 2 shows example of designed basic element and the same figure on the right shows mutual interconnection of two such elements, this can be considered as basic element in our model.

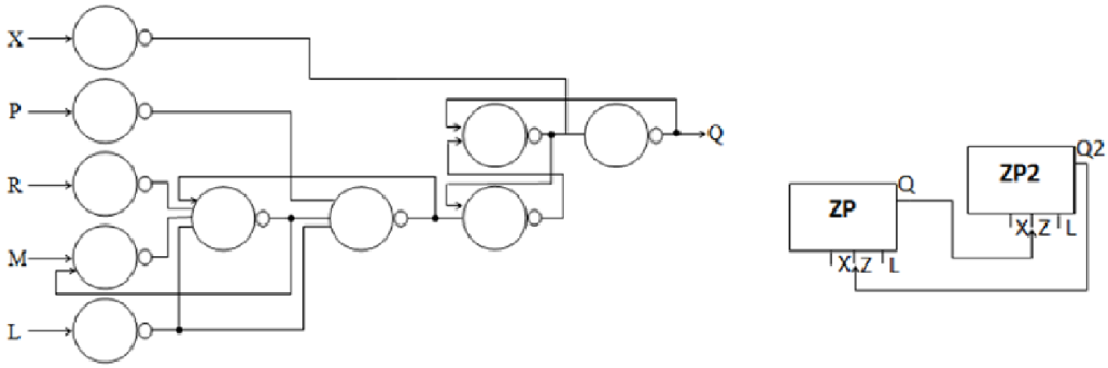
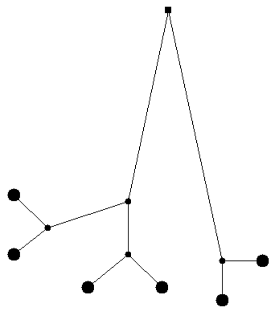


Fig. 2: The left image shows basic flip-flop. The right image shows interconnection of two such flip-flops into one basic element. X – Input into flip-flop; P,R,M,L – Parameters used for manipulation of flip-flop (punishment, reward, memory, limit); Q- Output

4. Neuron

Design of an initial network is inspired by complex patterns in nature (growth of plants, shell design, etc.), which can be described by a fractal structure. Lindenmayer's system (L-system) is used for prescription of such structures (prescription of a neuron and then the network will consist from the number of same neurons). L-system is defined as G = (V, A, P), where V is a set of symbols (alphabet), A is a string of symbols from V defining the initial state of the system (axiom), and P is a set of rules defining the way variables can be replaced with combinations of constants and other variables.



```
NDendrit{ ; Juraj Boron
 Angle 8
 Axiom FFX
 X=F[Z+FX]-F[Z-X]+X
 Z=@.8FZ
}
```

Fig. 3: The example of dendritic structure with initial axiom FFX, pair of rules X, Z and after two iterations, the final structure is generated.

For a later network structure adjustment we use genetic algorithms. The question raised now is that what options for adjusting does genetics offer? There is variety of options:

Genetics can be applied for the very basic element (neural component) for the modification of L-system prescription (rules generating L-system, or final string generating the fractal structure) or for direct modification of the generated network, etc.

5. Layer building

In previous section we described principles of building neuron structure. Each such structure has a root node, a set of these nodes creates "output layer" in meaning of Kohonnen layer or human brain cognitive layer. For neurons there are rules of mutual connection. E.g. a basic connection rule is based on Euclidean distance in given layer. If two nodes of different neurons (or the same neuron) are closer than any given distance, they are interconnected each other (see Figure. 4).

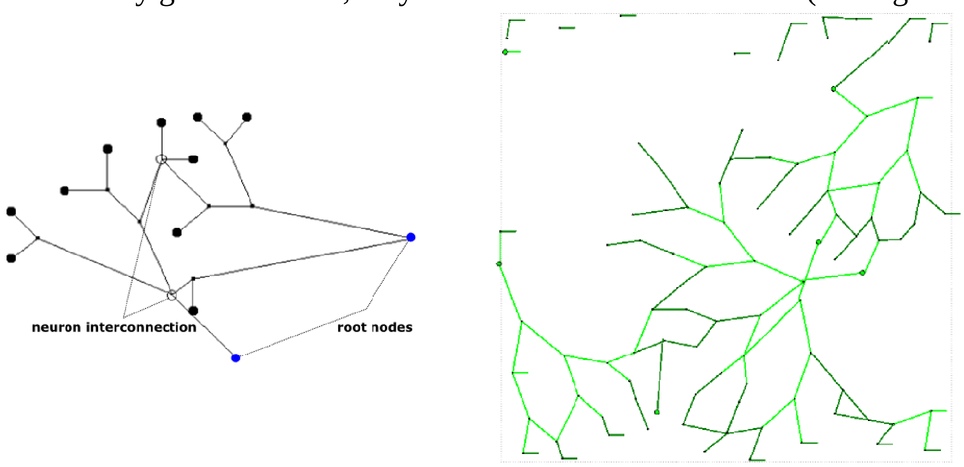


Fig. 4. The image on the left shows basic principle described above. The image on the right is an example from our demo application.

Some nodes from one neuron can be located in different layer – layered structure is generated. Nodes in these layers can be interconnected each other – the same principle as was described at the beginning of this section can be used. For example also all inputs nodes create an input layer, but this layer has not structure design.

6. Output evaluation

For a network evaluation, we have proposed the objective function that takes power of network excitation and the location of excited parts as the input values. The idea behind this process copies monitoring of human brain activity with EEG (Electroencephalography), so it can be said that we are trying to simulate a neural system of human. We believe that final network should be able to recognize any kind of input. Given that the department is experienced in speech processing, we decided to test our network at the speech recognition problem. Potential advantage of our approach to speech recognition over current approaches is that we do not need dictionary of known expressions in external memory. All words are stored/coded in main network structure. In association with this fact, we also believe that our system would be much faster in recognition than already known systems.

7. Conclusion

An input for our network will be spectrum of speech or features extracted from images. Not necessarily all input values will be counted. This will be matter of network structure. If we consider network as one layer of a system, then input spectrum will form another layer. Both layers will have coordinate system. Only unconnected inputs of neurons from the layer of network will be connected to inputs. In this connection, unconnected input with coordinates [x,y] will take value from layer with inputs from coordinates [x,y] (values x, y modulo width, height of a predefined window in spectrum), this is one example of providing inputs for network. The network behavior can be affected also by different strategy at presentation of inputs. It is goal of experiments to find out what effect brings multiple submission of the same input, if the output will be different from the case, when the network will get same input just once, etc.

Last but not least is the question of effective realization of experiments. Given the expected size of the network, we expect utilization of high performance computing for recalculation of network state.

System speed-up could be reached also by utilization of GPGPU at genetic programming part of solution. For this purpose we would like to use computation on graphics processing units, which is relatively lowcost solution (GPUs by nVidia with CUDA architecture).

References

[1] KLIMO, M., BORON, J. *Temporary properties of RS fuzzy flip-flops.* In FSTA 2010 (10th International Conference on fuzzy set theory and applications), 2010.

[2] KÓCZY L. T., LOVASSY R., FUZZY FLIP-FLOPS AND NEURAL NETS, IEEE INTERNATIONAL FUZZY SYSTEMS CONFERENCE, FUZZ-IEEE 2007, 1-6, ISBN: 1-4244-1209-9, 2007.

[3] KLIMO, M., BORON, J. Dynamické vlastnosti pravdepodobných fuzzy klopných obvodov. In ITAT 2009, ISBN: 978-80-970179-1-0, 2010.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Potential of RFID Middleware and EPCIS

*Filip Beneš
*VŠB-Technical University of Ostrava, Institute of Economics and Control Systems, International RFID Laboratory, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, {filip.benes}@vsb.cz

Abstract. Nowadays, we have been experiencing stronger and stronger penetration of RFID technology into our daily lives. Commercial networks and manufacturing corporations are both facing the question how to implement this technology not to stay behind the competing business subjects. For successful setting of RFID both properly chosen hardware (readers, aerials and tags) and operating software applications are necessary. These applications connect hardware with information systems of corporations and form so called middleware. Corporations thus deal with the problem which middleware to choose for their purposes. One of the solutions which have been handled in our International RFID Laboratory at Technical University of Ostrava is the WebSphere Sensor Events. This system meets international standards and provides middleware infrastructure for creation and administration of sensor-based solutions in business sphere. The aim of this article is to make the reader familiar with features and potentials of similar systems. Simultaneously, I want to present the idea of similar system usage in car industry.

Keywords: RFID, Tag, EPCIS, Middleware, RTLS, ePedigree, traceability

1. What is RFID

RFID is an abbreviation for radio-frequency identification. Any object of our real life, no matter if it should be people, animals or things, can be identified. An object intended for identification is marked with so called "tag". Information from the object marked this way can be then obtained using radium waves.

1.1. RFID Tag

From hardware point of view, RFID tag consists of microchip and aerial encapsulated in paper, plastic material or ceramics. Energy for communication is received from carrier wave sent by the aerial of the reading device.

For some applications, also so called active tags are used which have an additional own source of electrical energy and it is not necessary to energize them with the reader. Information about the object is called "Electronic product code (EPC)" and is stored on the microchip, while the aerial takes care of communication.

1.2. Reading device

Main purpose of RFID reading device is to communicate with tags. Load data from them and send it further into information systems. The majority of reading devices have their reader and aerial part separated for the purpose of modularity, by connection of several aerials and creation of various configurations for ensuring proper reading in particular environment. Some reading devices are compact, for example the handhelds. Recently, there are also RFID extensions for mobile phones and PDA.

1.3. Difference from barcodes

RFID tag broadens features of barcodes in such way that its direct visibility is not necessary. It offers dramatically bigger reading distance. It allows loading of more elements at once and has much bigger capacity for storing information. Among its outstanding features there is also a possibility to add active elements for various quantity monitoring, such as temperature, humidity, movement and rotation in all directions.

1.4. RFID Middleware

Up to now, I was writing about the hardware part of RFID technology. For making use of information from tags it is however necessary to have it transfered to information systems. After the first manipulation (which takes place in the reading device), another process comes in the higher layers of RFID system. Communication method of reader is described by protocols. With regard to the fact that in complex RFID system various types of readers can be present supporting various protocols, it is necessary to make use of the tool which is called Middleware. In this case, Middleware serves for gathering, filtration and consequential delegation of data from RFID readers to other components of system.

1.5. EPCIS

Electronic Product Code Information Services (EPCIS) is a standard designed for exchange of information between business partners in their logistic network based on RFID technology. Together with information systems, RFID is the major technological means enabling identification, gathering and exchange of information about products across all logistic network. EPCIS is crucial for product traceability. Traceability could be defined as an ability to gain and verify information about tracked item from views of its history, current status and location of the item in the logistic network.

The basic building stone of EPCIS system is the event-driven data processing. Event is to be understood as some real event related to the RFID identification. Event must be presented in the form which is comprehensible to software components. From system point of view, it is carried out with XML document and its metadata. From these events, we can understand the link to particular business processes. Events may be related to identification of one or more objects, identified with the EPC, aggregation of objects on pallet, business transaction or a mere inventory check of particular object type.

The event includes, among others, also the time stamp, indicating the exact time of occurrence, list of EPC codes associated with the event, identification of the event relation to business transaction and reading location in which the event was recorded.

Activities of International RFID Laboratory also include the analysis of particular EPCIS solution possibilities. In the EPCIS field, there are open source solutions, such as EPCIS Fosstrack, we commonly work with and use for demonstrative purposes. Due to frequent communication with industrial companies we must focus also on the potential of more sophisticated solutions, such as BizTalk Server from Microsoft or WebSphere Sensor Events from IBM.

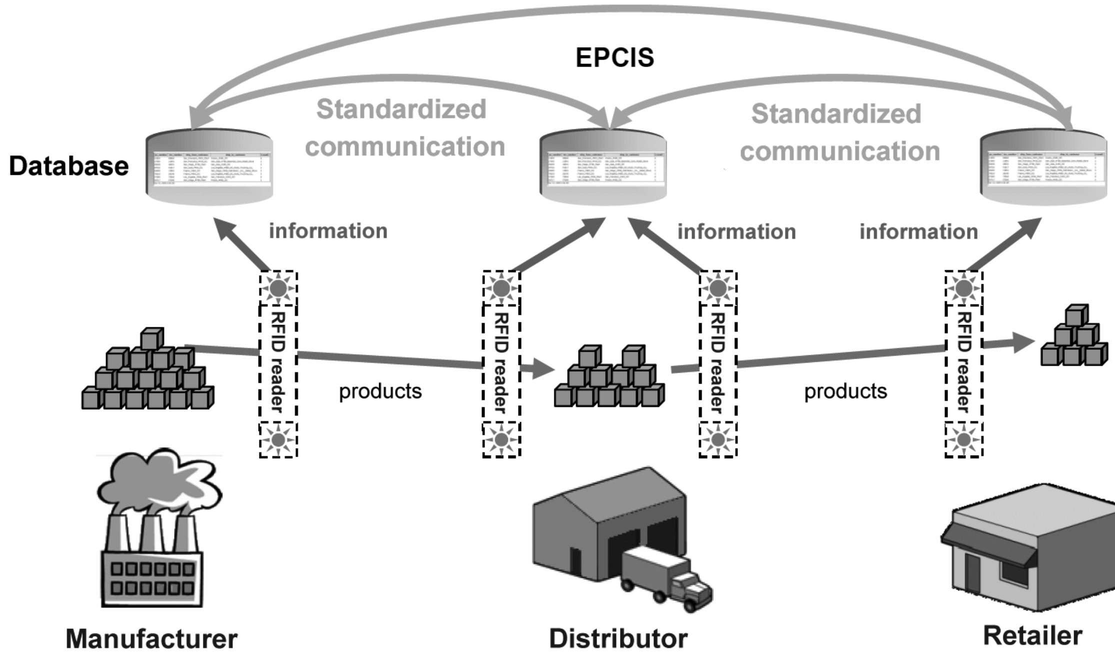


Fig. 1. Diagram of EPCIS based communication [4]

2. WebSphere Sensor Events

The advantage of this system lies in its openness towards many sensors. The sources of data for this platform include, among others, sensors of barcodes, GPS and RFID tags, both active and passive. System enables setting of RFID architecture and its interconnection to physical readers. This platform provides those who dont have RFID hardware at their disposal with simulation of reading devices and tags. InfoSphere Traceability Server, a full-featured EPCIS system and a robust database at once, also belongs to the group of similar programs.

WebSphere Sensor Events contains several model examples of usage concerning the issues of Container Tracking, Track and Trace of Goods, ePedigree and Asset Visibility Client.

2.1. Container and goods Tracking

Container Tracking is used in case there are returnable wrappings, containers or other returnable materials present in the logistic network. Using this component demonstrates their traceability from the producer warehouse and later return to their owner.

The Track and Trace of goods use case provides a base to implement the serialization and tracking of goods through manufacturing, shipping, and receiving. The goods is stored on the pallet equipped with the RFID tag. This tag contains the EPC code linked with the EPCIS database in which all information received during transport through reading gates of logistic network participants is stored.

2.2. ePedigree

Thanks to InfoSphere Traceability Server, the manufacturer inserts commissioning and shipping information - which is important for the distributor - to the system. In case the distributor also has the EPCIS system, he can add more information and the final vendor (and thus also the customer) can reach the complete pedigree information. Usage of ePedigree is related in the model case especially with health care, in particular with distribution of pharmaceuticals. It helps to exactly track down where the particular medicine comes from and if by chance it could be a counterfeit.

2.3. Asset Visibility Client

Another domain where the RFID tags may be applied are so called RTLS systems - locators working in real time. The active tags and reader device networks are used here. Similarly to the GPS positioning, the triangulation and trilateration is used for detection of exact location of selected element. Generally, the more dense network of reading locations we create in the transmitting range of active tag, the more accurate the information about its exact location we receive. The advantage in comparison to the GPS is especially the fact that we can cover also the inner premises of roofed buildings where the GPS signal doesn’t reach. Asset Visibility Client enables visualization of building ground plans and then, in real time, recording of particular element movements in given area. Enables a quick response to emergency situations or security breaches.

3. Usage for monitoring of cars in parking places

One of the ideas I’ve been working on recently is usage of RTLS systems for needs of local car industry. The basic concept lies in usage of reading locations network and RFID tags for covering of huge parking places and monitoring of vehicle movements in these parking places. This system would be able to exactly determine where the particular car is located at a given moment, it would be able to find out who parked the vehicle, it would store all important data about the vehicle and made it easy to track it down.

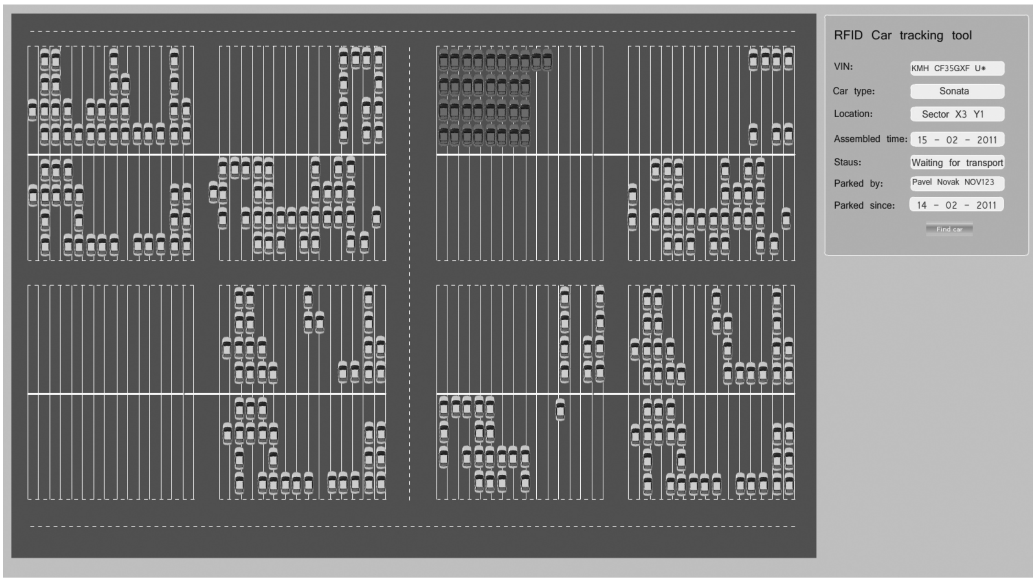


Fig. 2. Deisgn of car tracing tool environment

4. Conclusion

This article aimed for making readers familiar with RFID middleware and EPCIS potential and its usage across the whole spectrum of business processes. It also brings in the idea of using similar system in car industry for localization of vehicles in parking places.

References

[1] LAHIRI, SANDIP. *RFID Sourcebook*. IBM Press, 2006

[2] HUNT, D., PUGLIA, A., PUGLIA, M. *RFID: A Guide to Radio Frequency Identification*. John Wiley & Sons, 2007

[3] *InfoSphere Traceability Server Deployment Guide version 2.0*. IBM Redbooks, 2008

[4] SAMEK, M. *IBM application of Electronic Product Code Information Service*. 2010

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Attractive Virtual Educational Portal for Schoolchildren of Primary Schools Based on Data Mining

* Jakub Berthoty, *Vitaly Levashenko, *Antónia Kľučková, *Igor Janic, *Michal Kavecký, *Dušan Dragula, **Bera Bučková
* University of Žilina, Department of Informatics, Žilina, Slovakia, 10026, Vitaly.Levashenko@fri.uniza.sk
** Primary School, Školská 49, Žilina-Zavodie 010 04

Abstract. This paper is describing the methodological and application aspects of elaboration of Attractive Virtual Educational Portal for Primary School Schoolchildren with Data Mining Support.

Keywords: Education, Data Mining, databases.

1. Introduction

In the ongoing age of informatization, you can hardly find a household which would not include a computer or an internet connection. Even the youngest members of the family, primary school schoolchildren have access to this technology, which is positive on one side, because today, it is important to know the basics of working with a computer and this skill is considered essential. On the other side, most of the schoolchildren consider computer and internet just a gaming instrument due to the vast amount of games available freely on the internet (www.onlinehry.sk, www.superhry.cz, www.zahraj.sk) or games bought by their parents. Most of these games have absolutely no educational value, and in many cases, playing them leads to a gaming addiction. Furthermore, this draws their attention away from education or relaxation and has a negative impact on their character [1]. Even the teachers from the elementary school in Žilina - Závodie, whom we have cooperated with on modules development for two years now, complained to us about the problem they are encountering every day. Their pupils do not do their homework and therefore it would be good to support them in solving routine tasks.

Therefore we decided to design and implement a project of attractive virtual educational portal that will provide schoolchildren with possibility to exercise routine tasks playfully and independently from school duties. This portal should be used mainly at home. But if teachers are interested, they can incorporate it into educational process in school environment, the choice is theirs.

So the goal of the project is to draw children to solving of routine tasks and to do it via playful form. Our project will also be unique in the way that apart from the educational module itself it will incorporate also the module for datamining.

2. Data Mining and Education

While solving a large number of exercises, an enormous amount of data can fill the databases, as these data represent the answers of students to the set exercises. However, these data are usually not used because normal educational portal do not contain instruments for student answer analysis, even if these data could point out interesting and potentially surprising dependencies. This information could be used not only in pedagogy, but also in the educational process itself. Intelligent educational systems, which can accommodate the educational process to the skills of individual students thanks to analyzing his correct and erroneous answers and setting a strategy for education of this individual, can serve as an example. If we want to create such an educational system, the effective instruments for its creation would certainly include DataMining methods. DataMining, meaning deep data analysis, is a process of extraction of data patterns from large databases and searching for hidden data dependencies. DataMining methods include decision trees, decision rules, clustering, classification etc. [2-5]. Most of the DataMining methods are supplemented by a mathematical apparatus for handling of erratic data, with the use of multivalued and fuzzy logic.

The Data Mining algorithms will also provide means for handling of noise data which are basically anomalies within the data sets with erroneous or missing values [6,7]. This will make the Data Mining tools much more resistant to errors, and will improve overall performance in the area of data analysis.

To facilitate the process of data classification (clustering), the Portal Data Mining layer (see next chapter) will contain a Fuzzy C-Means Clustering module [8-11], which will be able to create a desired number of clusters from any available data, while using multiple types of distance measures (Euclidean, Chebyshev, weighted etc.) [12].

3. Attractive virtual educational portal

The paper aims to create an attractive virtual educational Portal for the schoolchildren of primary schools with the realization of mechanism for obtaining useful information to support intelligent learning process. Realization of virtual educational Portal is designed as a software solution in the form of 4 basic layers (see Fig. 1).

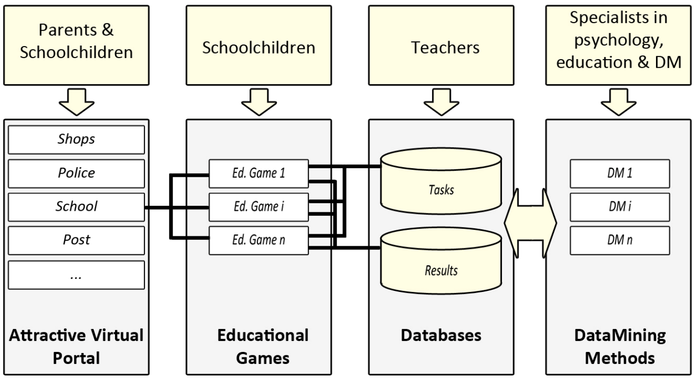


Fig. 1: Attractive Virtual Educational Portal.

The first layer consists of an interactive medium, serving for the purpose of communication between the students in the form of a game. It will be a virtual club, which will allow the students to register for free using their general information (name, surname, age, school, class, place of residence ...). The registration results in a student game profile, which is, at the beginning, identical for every member of the virtual club. Through the games, the student can improve his profile. The game environment contains surrounding objects of the real world: school, rail station,... For a correctly finished task, the student receives credits, which can be used in these objects to gain something and improve his profile. The schoolchild might not even notice that he is being educated while playing a game. While playing, his motivation to be the best beside with his/her peers, makes him wants to improve his profile further and with it, improve his knowledge.

As a result, this layer can be used to mimic real life, and encourage personal responsibility for their player profile.

The second layer is created by educational modules, realized in the form of games. The tasks solved during the games are corresponding to the basic school subjects. Their goal is to provide new information for the students and improve existing knowledge, all through playing the games. For completing the tasks, the students are assigned credits. The concept of the project supports adding new modules (even after the project is finished).

The first two layers are mainly used by students and their parents. Each module represents the application in the form of a game as some type of scenario. These applications adhere the following conditions: (a) open and universal structure, and (b) independence from the topic of education. The first condition ensures the progressive development of the education Portal by developing new modules, created after the completion of the project. The second condition allows the independence of each module slot on the particular subject matter and teaching assignments.

An important factor in the implementation of the first two layers of the Portal is the consulting help of teachers and psychologists who have the required knowledge about the education of schoolchildren of a primary school.

The third layer - database, integrates 2 databases. The first database contains the tasks to be solved by the students in a unified and formalized form. The second database is a database of reactions from the students, so it contains their answers. Its role is to store the answers for future analysis with usage of the effective methods on intelligent analysis. Collected data allow for the statistics of responses, which will conduct a source for the next layer [13, 14].

The fourth layer contains methodologies of intelligent data analysis derived from the learning of schoolchildren. The base of the analysis will consist of the original method of data mining developed by the authors of the project. These methods make it possible to identify a set of interesting dependencies, useful for the education of schoolchildren.

The paper foresees the creation of several methodologies directed at increasing efficiency of education process. In connection with this project, we demonstrate the possibility and usefulness of this promising procedure for the example analysis of the basic triplet (schoolchild - teacher - the role):

a) analysis of success - schoolchild is to designate a limited number of key questions of the test, allowing for quick but reliable and objective way of schoolchild know-ledge evaluation. As a result of finding the key is-sues is the possibility of shortening the review-time of knowledge and preparation of task and the organization of mini-tests in lessons to achieve a high degree of credibility and objectivity of the evaluation;

b) analysis of teaching skills - by tracking the adoption of the learning topics. As a result, the teacher can determine the topics that the schoolchildren are failing at and find them difficult to embrace with a view to their subsequent amendment and revision;

c) analysis of the content - of tasks expect the research of the results testing own test questions. This allows for example to identify issues of mutual correlation between results, which will increase the adequacy and effectiveness of test assignments.

To achieve the objectives of this layer is necessary to solve the following tasks in the field of decision support, based on the analysis of fuzzy data using a priori information:

- formulation and research of optimization criteria decision-making rules for different types of tasks with the help of apparatus much-valuable logic and fuzzy logic
- synthesis of new algorithms for knowledge discovery in databases and calculation of reliability-based systems for use in decision support
- development of practical recommendations in the form of specific methodologies and software for using the proposed algorithms for solving practical problems associated with analysis of test results of schoolchildren (in collaboration with the primary school Žilina - Závodie).

For the promoters of the project, it is very important to obtain the results from other Slovak schools and universities and to provide access to the Portal and to the obtained results.

To complete the description of the use of presented software solution, it is necessary to define 3 groups of users to work with presented software product.

The first group consists of parents and schoolchildren who have immediate access to the first two layers of the Portal.

The second group is formed by teachers, working mainly with the third layer. Their interaction with the Portal will be controllable via an intuitive assistance forms, allowing the process to ensure complementarily and editing tasks. Teacher must not necessarily expert in the field of informatics. Completion of other modules in the process of their preparation will be through the Portal administrator - such as teacher of informatics.

The third group consists of experts in the 4th layer, who will use it to process the knowledge acquired on the basis of information stored in the databases. This may serve as a useful point for the establishment of relations between experts and teachers. The experts group also contains teachers who are interested in pursuing research or education, pedagogy and specialists in artificial intelligence and decision support.

Important contributions of the project are publications whose authors are project promoters and teachers. The basic theme of publications: a) information generated by the Portal and mathematical apparatus, that form its basis, b) teaching and methodological results of processing the data obtained in the process of functioning of the Portal, which will be useful for specialists in pedagogy and child psychology.

4. Conclusion

The authors expect the effect of using the paper's results in two ways:

1) Usage of the proposed educational Portal will provide an effective system of education for schoolchildren, bringing positive social and educational effects. Preparing schoolchildren on the first stage of primary school will lay a solid foundation for their future education, which is one of the current tasks of the Slovak society. Access to the results obtained using established software will be useful for specialists (teachers and psychologists) in understanding the schoolchildren and can serve as input for further research in other fields related to education and data analysis.

2) Development of new algorithms for data processing is the basis for a future application for intelligent analysis of input data. This will ensure conditions for a competitive production.

References

[1] GREGUSSOVÁ, M., KOVÁČIKOVÁ, D. Sú naše deti vo virtuálnom prostredí v bezpečí? *Bratislava : Výskumný ústav detskej psychológie a patopsychológie*, http:// www.zodpovedne.sk/ kapitola4.php?cl=bezpecne_na_internete, 2008.

[2] Mitchell T., Machine Learning, *McGraw-Hill*, 1997.

[3] Witten I., Frank E., Data Mining: Practical Machine Learning Tools and Techniques (Second Ed.), *Morgan Kaufmann*, 525 p. ISBN 0-12-088407-0, 2005.

[4] Paralič J., Objavovanie znalostí v databázach, *Technická univerzita v Košiciach*, ISBN 80-89066-60-7, 2003.

[5] Berka P., Dobývání znalostí z databází, *Academia, Prague*, 368p., ISBN: 80-200-1062-9.3, 2003.

[6] O. Maimon, A. Kandel and M. Last, "*Information-Theoretic Fuzzy Approach to Data Reliability and Data Mining*", Fuzzy Sets and Systems, Vol. 117, No.2, pp.183-194, 2001.

[7] O. Maimon, A. Kandel, and M. Last, "*Information-Theoretic Fuzzy Approach to Knowledge Discovery in Databases*", in Advances in Soft Computing - Engineering Design and Manufacturing Editors: R. Roy, T. Furuhashi, and P.K. Chawdhry.), Springer-Verlag, pp. 315-326, 1999.

[8] Gan G., Ma C., Wu J., Data Clustering Theory, Algorithms, and Applications V2, Society for Industrial and Applied Mathematics Philadelphia, Pennsylvania, American Statistical Association, 488 p., 2007.

[9] Xu R., Wunsch D.C., Clustering, Institute of Electrical and Electronics Engineers, Wiley, 364 p., 2009.

[10] Berkhin P., Survey of Clustering Data Mining Techniques, Accrue Software, Inc., 56 p., 2002.

[11] Lee H.M., Chen C.M., Chen J.M., Jou Y.L., An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy, IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 7 p., 2001.

[12] Pedrycz W., Knowledge-Based Clustering – From Data to Information Granules, Department of Electrical and Computer Engineering, University of Alberta, 337 p., 2005.

[13] Martincová, P.: *Operačné systémy*. 2. vyd., ISBN 80-7100-413-8, 1997.

[14] Matiaško, K.: *Databázové systémy*. 1. vyd. ISBN 80-7100-968-7, 2002.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Estimation of Cardiovascular Patient Risk with a Bayesian Network

*Jan Bohacik[1], **Darryl N. Davis
*University of Žilina, Faculty of Management Science and Informatics, Department of Informatics, Univerzitná 8215/1, 01026 Žilina, Slovakia, {Jan.Bohacik}@fri.uniza.sk
**University of Hull, Faculty of Science, Department of Computer Science, Cottingham Road, HU6 7RX, United Kingdom, {d.n.davis, J.Bohacik}@hull.ac.uk

Abstract. Cardiovascular decision-making support experiences increasing research interest of scientists. Ongoing collaborations between clinicians and computer scientists are looking at the application of data mining techniques to the area of individual patient diagnosis, based on clinical records. An investigation of a Bayesian network learnt according to a generated decision tree with cardiovascular data for estimation of patient risk in cardiovascular domains is presented. Promising experimental results are also provided.

Keywords: classification, cardiology, Bayesian networks, medical data mining.

1. Introduction

A major challenge facing healthcare organizations (hospitals, medical centers) is provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are cost-effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Healthcare organizations must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems [5].

The research reported in this paper considers assessing the risk of individual patients. For assessing the risk of a cardiovascular patient, computational classification models can be used. Classification models are typically used in data mining which is one of the steps of the more general knowledge discovery in databases [3]. As a database, a collected cardiovascular dataset is used by us. Knowledge discovery in the cardiovascular dataset provides knowledge and tools of use for prediction of a cardiovascular patient's risk and its improvement.

The paper is organized as follows. In Section 2, the used cardiovascular dataset and the used clinical model are described. Our designed Bayesian network as a classification model is described in Section 3. Section 4 contains the details of our experiments. Section 5 concludes this paper.

2. Cardiovascular dataset

As a dataset, a group of 839 instances (cardiovascular patients) classified into two levels of risk and described by 17 attributes $\boldsymbol{A}$ as queries about patients' symptoms, medical history, clinical findings and results of physiological measurements is used. Instances are derived from clinical data collected at the Hull site (498 instances) and at the Dundee site (341 instances). This data is noisy, contains many null values and is problematic. It was transformed into the used dataset according to [2]. Describing attributes $\boldsymbol{A}$ are defined as $\boldsymbol{A} = \{ A_1; \dots; A_k; \dots; A_{17} \}$ = {*Age*; *Sex*; *Heart disease*; *Diabetes*; *Stroke*; *Side*; *Respiratory*; *Renal failure*; *ASA*; *Hypertension symptom*; *ECG*; *Duration*; *Blood loss*; *Shunt*; *PATCH*; *Coronary artery bypass surgery*; *Consultant*}. Most describing attributes are categorical with the exception of numerical *Age*, *Duration* and *Blood loss*. If A_k is a

[1] Ján Boháčik with all Slovak diacritics.

categorical attribute $A_k = \{a_{k,1}; \ldots; a_{k,l}; \ldots; a_{k,l_k}\}$ where $a_{k,1}; \ldots; a_{k,l}; \ldots; a_{k,l_k}$ are possible categorical values. $A_2 = \{a_{2,1}; a_{2,2}\} = \{female; male\}$, $A_3 = \{a_{3,1}; a_{3,2}\} = \{no, yes\}$, $A_4 = \{a_{4,1}; a_{4,2}\} = \{no; yes\}$, $A_5 = \{a_{5,1}; a_{5,2}\} = \{no; yes\}$, $A_6 = \{a_{6,1}; a_{6,2}\} = \{left, right\}$, $A_7 = \{a_{7,1}; a_{7,2}; a_{7,3}; a_{7,4}\} = \{normal; mildCOAD; modCOAD; severeCOAD\}$, $A_8 = \{a_{8,1}; a_{8,2}\} = \{no; yes\}$, $A_9 = \{a_{9,1}; a_{9,2}; a_{9,3}; a_{9,4}\} = \{one, two, three, four\}$, $A_{10} = \{a_{10,1}; a_{10,2}\} = \{no; yes\}$, $A_{11} = \{a_{11,1}; a_{11,2}; a_{11,3}; a_{11,4}; a_{11,5}; a_{11,6}; a_{11,7}\} = \{normal; qWaves; sTWaves; aFib60to90; aFibLT90; fiveEctopic; other\}$, $A_{14} = \{a_{14,1}; a_{14,2}\} = \{no; yes\}$, $A_{15} = \{a_{15,1}; a_{15,2}; a_{15,3}; a_{15,4}; a_{15,5}; a_{15,6}; a_{15,7}\} = \{armVein; legVein; otherVein; dacron; no; ptfe; stent\}$, $A_{16} = \{a_{16,1}; a_{16,2}\} = \{no; yes\}$, $A_{17} = \{a_{17,1}; a_{17,2}; a_{17,3}; a_{17,4}; a_{17,5}\}$. Class attribute C (= *Risk*) is used to classify instances into two possible categorical values c_1 and c_2 meaning risk levels (*low* and *high*, respectively). It is denoted by $C = Risk = \{c_1; c_2\} = \{low; high\}$. The values of class attribute C are generated according to the following heuristic clinical model [2]: an instance (cardiovascular patient) is classified into *high* if the patient's death or severe cardiovascular event (e.g. stroke, myocardial relapse or cardiovascular arrest) appears within 30 days after an operation.

3. Bayesian network based on a generated C.45 decision tree

Our Bayesian network is learnt on the basis of a decision tree. The decision tree is generated according to the C4.5 algorithm [6] used on collected data for all 17 describing attributes A_k, k=1, …, 17, class attribute C and 839 instances. The generated decision tree is in Fig. 1. It consists of the root node and the inner nodes (expressed as ellipses) associated with attributes, branches (expressed as lines) associated with possible values of attributes and leaf nodes (expressed as rectangles) associated with risk levels $c_j \in C$.

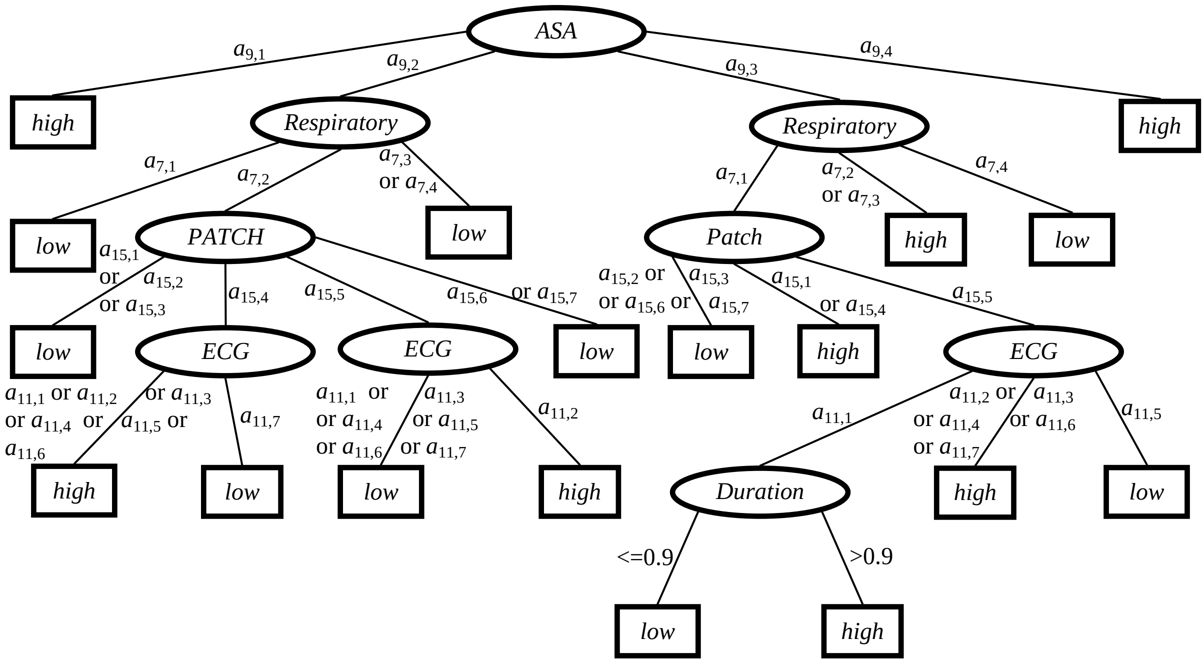


Fig. 1. Generated C4.5 decision tree on the basis of instance values for A_k, k = 1, 2, …, 17, and C.

The attributes associated with the root node and the inner nodes in Fig.1 are used as nodes in our Bayesian network in Fig. 2. Basically, a Bayesian network is a graph with arcs connecting nodes and no directed cycles (i.e., a directed acyclic graph), whose nodes represent random variables and whose arcs represent direct dependencies. Each node has a conditional probability table (CPT), which, for each combination of values of the parents, gives the conditional probability of each of its values. If there is a branch from an attribute to another attribute in Fig. 1, there is an arc from the attribute to the other attribute in our Bayesian network.

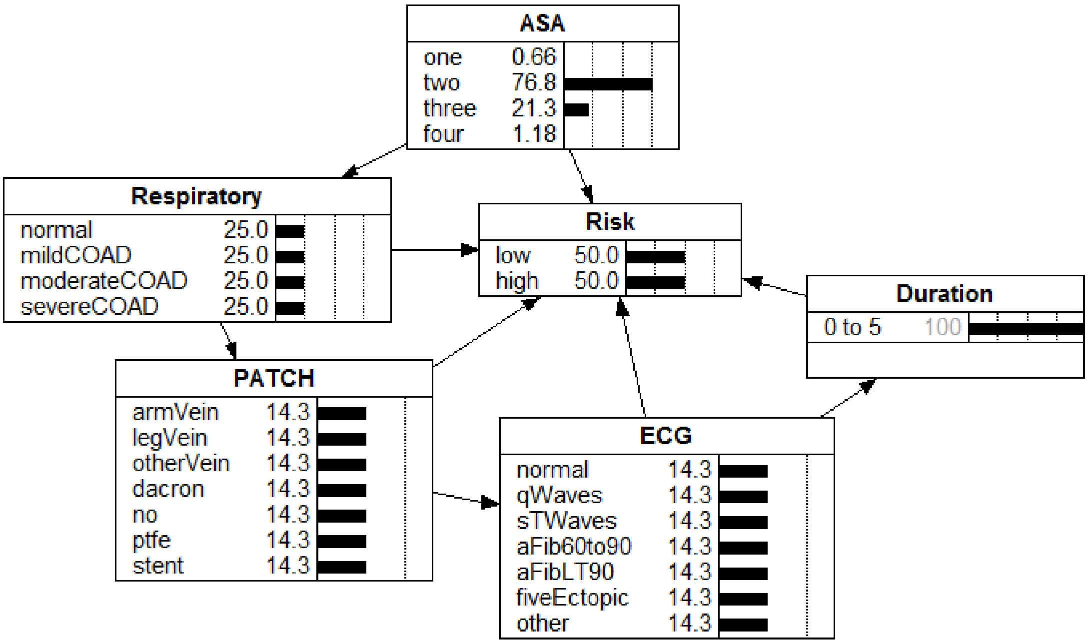


Fig. 2. Proposed Bayesian network.

The CPTs in our Bayesian network are learnt on the basis of data for attributes A_7, A_9, A_{11}, A_{12}, A_{15}, class attribute C and 839 instances with software library Netica^{TM} [4].

4. Experimental results

The main purpose of the experimental study is to compare the performance of our designed Bayesian network with other classification models on our cardiovascular dataset. Experiments were carried out with our Java software tool which is being developed with the intention of its integration into the medical decision-making support system of the BraveHealth system. The core algorithms are implemented in external libraries: Netica^{TM} [4] and Weka [6]. The performance of the particular classification models is measured with sensitivity = tp/(tp + fn), specificity = tn/(tn + fp), positive predictive value = tp/(tp + fp), negative predictive value = tn/(tn + fn) and accuracy = (tp + fn)/(tp + fp + fn + tn) where tp is true positive, fp is false positive, fn is false negative, tn is true negative, 'C is *low*' is negative and 'C is *high*' is positive.

Method	**SEN (%)**	**SPEC (%)**	**PPV (%)**	**NPV (%)**	**ACC (%)**
TreeBayesNet	77.78	96.63	80.33	96.09	93.80
Bayes	7.94	97.48	35.71	85.70	84.03
C4.5	4.76	98.60	37.50	85.42	84.51
NNge	15.08	90.18	21.35	85.73	78.90
MLP	15.08	89.62	20.43	85.66	78.43

Tab. 1. Experimental results.

The results of our experiments are given in Tab. 1 where TreeBayesNet denotes our Bayesian network described in Section 4 and implemented with Netica^{TM} and Weka, Bayes denotes a Bayesian network classifier implemented in Weka as class BayesNet, C4.5 is a decision tree classifier implemented in Weka as class J48, NNge is a nearest neighbor classifier using non-tested generalized exemplars [1] implemented in Weka as class NNge, and MLP is a neural network classifier using multilayer perceptron implemented in Weka as class MultilayerPerception. SEN is sensitivity, SPEC is specificity, PPV is positive predictive value, NPV is negative predictive value and ACC is accuracy. Since these classification models are considered to be used as a part of the medical decision-making support system of the BraveHealth system, they should avoid cases when high risk patients are labeled low risk and so sensitivity should be maximized. Our proposed classification model TreeBayesNet learnt from a generated C4.5 decision tree gives the highest sensitivity 77.78% of all classification models. Its accuracy with 93.80% is also the highest.

5. Conclusions

A classification model based on a Bayesian network learnt from a generated decision tree is proposed. It is employed together with several other essentially different classification models on data which is collected about cardiovascular patients. Our model gives considerably better results, especially with its avoidance of labeling high risk patients as low risk patients. However, further investigation, including a simultaneous use of more classification models, continues so that even fewer high risk patients are labeled as low risk ones. The aim of our study was to investigate/develop issues and software capable of being integrated into the BraveHealth system which will provide a patient centric approach for an integrated, adaptive, context aware remote diagnosis and management of cardiovascular diseases.

Acknowledgement

This work is funded by the European Commission's 7th Framework Program: BRAVEHEALTH FP7-ICT-2009-4, Objective ICT-2009.5.1: Personal Health Systems: a) Minimally invasive systems and ICT-enabled artificial organs: a1) Cardiovascular diseases.

References

[1] BRENT, M.: *Instance-Based Learning: Nearest Neighbour With Generalization*. Master's thesis. Waikato:University of Waikato, 1995.

[2] DAVIS, D. N., NGUYEN, T. T.: Chapter IX: Generating and verifying risk prediction models using data mining: A case study from cardiovascular medicine. *Data Mining and Medical Knowledge Management: Cases and Applications*. 1st ed. :IGI Global Inc., ISBN10: 1605662186, 2009.

[3] FAYYAD, U., PIATETSKY-SHAPIRO, G., SMYTH, P.: From data mining to knowledge discovery in databases. *AI Magazine*, Vol. 17, No. 3, pp. 37-54, ISSN: 0738-4602, 1996.

[4] NORSYS SOFTWARE CORP.: *Netica™ Application* [http://www.norsys.com/netica.html].

[5] PALANIAPPAN, S., AWANG, R.: Intelligent heart diseases prediction system using data mining techniques. *Int. Journal of Computer Science and Network Security*, Vol. 8, No. 8, pp. 343-350, ISSN: 1738-7906, 2008.

[6] WITTEN, I. H., FRANK, E., HALL, M. A.: *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd ed. :Morgan Kaufmann, ISBN: 978-0-12-374856-0, 2011.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Improvement of Telecommunication Offer with the Use of FTTH Networks

[*]Maciej Czaplewski
[*]University of Szczecin, Faculty of Management and Economics of Services,
Department of Economics and Organization of Telecommunication,
ul. Cukrowa 8, 71-004 Szczecin, Poland, maciej.czaplewski@wzieu.pl

Abstract. The article describes the importance of development of FTTH (Fiber-to-the-Home) networks for companies functioning on the telecommunication markets. It mainly focuses on the advantages of FTTH emphasizing the possibilities of offering new services for end customers by companies which will build up their own FTTH network and in consequence eventually transform it into competitive advantage.

Keywords: FTTH networks, telecommunication.

1. Introduction

The Markets of Telecommunication Services are becoming increasingly competitive. On the one hand, it is the result of progressive liberalization and globalization processes of these markets. On the other hand, this state is a corollary of the technical progress and technological innovation occurring in the area of information transmission and communication.

Both of these trends lead to an increase in the number of operators offering services in the telecommunications markets, among which, apart from traditional telecommunications operators, there are also Internet providers, cable television operators and sometimes also entities from other sectors (such as energy or rail transport).

This situation leads to an even further increase of competition in the particular Telecommunication Services Market, which intensifies and extends to the whole market of information and communication, due to the expansion of service offerings by all the major providers of this market and the increasing overlap of these offers between these providers, i.e.:

- telecom operators
- Internet providers,
- cable TV operators.

Those above mentioned operators using:

- technological progress and the new information and communication technologies as well as
- already held technical potential and expert knowledge,

increasingly pursue a, so called, Triple Play strategy, and offer services that contain:

- telephone services,
- data services,
- digital TV services.

This, in consequence, leads to a blurring of distinctions between offers of telecom, Internet and cable television operators, and subsequently, to further competition increase in the market of information and communication, including the Telecommunication Services Market.

The new market situation forces companies operating in market of information and communication, including the Telecommunication Services Market to find ways to distinguish themselves from the growing number of competitors.

Finding rational ways to distinguish themselves from the competition requires the identification of needs reported by increasingly demanding customers, and subsequently, the adaptation, as far as possible, to those needs. Business practice indicates the growing consumer demand for modern services, which requires the provision of networks and technologies to transmit and process large amounts of data. Keeping in mind that fact, operators that wish to differentiate themselves from competitors should consider the creation of broadband networks with high data flow that reaches directly to the homes of users.

Such opportunities are primarily offered by FTTH (Fiber-to-the-Home) networks.

2. Characteristic of FTTH technology

FTTH is an solution from FTTx group basing on optical networks with high data rates, in which one can distinguish:

- point-to-point networks;
- Gigabit Passive Optical Network (GPON);
- Active Ethernet. [1]

Most FTTH networks are created in a point-point relationship between central object and the operator's apartment. FTTH Council defines Fiber-to-the-home as a telecommunications architecture in which a communications path is provided over optical fiber cables extending from the telecommunications operator’s switching equipment to (at least) the boundary of the home living space or business office space. [2] It is stated that for the smooth functioning of these networks, their length from a central point should not exceed 20km. [3]

Regardless of the type of the used solution, one should state it clear that FTTH optical networks have significant value, reflected by a high bit rate and a high speed of data transmission. These networks allow the transmission of 18 Tbit / s, which means almost 1000-fold higher speeds than those offered by the regular brass network. [4]

Parameters offered by FTTH technology means clear quantitative increase in transmission capacity in relation to the previously used technology of data transmission. Such high bit rate allows the provision of services, which couldn’t be offered with the use of other technologies.

The relationship between the potential of various technologies used in data transmission (in Mbit/s) and the requirements in this area occurring in offering various ICT services, in a synthetic manner, presents Figure 1.

The information presented in Figure 1 show a clear advantage of FTTH technology capacity over the previously developed information and communication technologies. The figure shows also the ability to offer, via FTTH network, high-definition television services, including three-dimensional television programs (3D digital television). From a business standpoint, this means the possibility to offer, by operators using FTTH technology, services that are not offered by competitors, who do not have this technology. Subsequently this can transform into market advantage over these competitors.

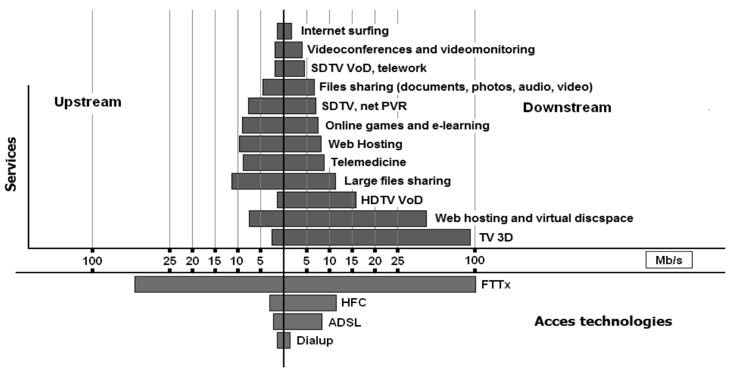


Fig. 1. The relationship between the potential use of technology and data requirements in this area occurring in the provision of ICT services

Source: H. Babis et. al., *Scenariusze rozwoju FTTH ze szczególnym uwzględnieniem możliwości oferowania nowych usług*, Grudzień 2008, p. 106.

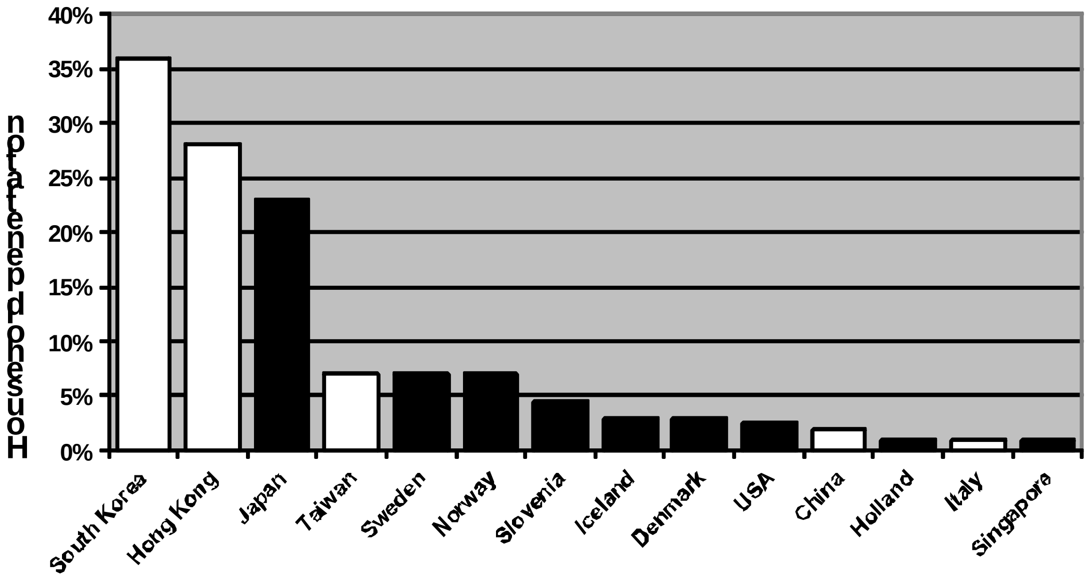
3. Main areas of FTTH technology utilization with a special emphasis on the situation in Poland

Level of FTTH technology utilization in business practice is created in particular by:

- concern about services available only with the use of this technology, such as next-generation digital television services and three-dimensional TV services (3D DTV);
- financial strength of market of information and communication operators operating in particular countries and the estimation of their capacity to recover expenditures on FTTH within an acceptable period;
- politics of individual countries regarding the development of the information society, as well as the financial strength of these countries, which eventually helps to support the high costs associated with the deployment of FTTH networks.

Characteristic for highly industrialized countries of East Asia developed technology and ICT sectors, high level of awareness about these technologies and high level of skills to use them as well as engaging these countries in the creation of modern communication networks mean, that these countries have the highest penetration of buildings or households connected directly with fiber-optic network.

List of countries in which, in the half of the year 2008, at least 1% of the buildings or dwellings were directly connected to optical network shows Figure 2



■ Economies with the majority of FTTH networks
□ Economies with the majority of FTTB+LAN networks

Fig. 2. Economies with the highest penetration of Fiber-to-the-Home (FTTH) or Fiber-to-the-building (FTTB)

Source: *With Robust Growth in Fiber to the Home Subscribers, Asia-Pacific Continues to Lead in FTTH Market Penetration.* FTTH Global Ranking (23-July-2008). http://www.ftthcouncil.eu-/documents/press_release/July%2008-%20FTTH%20Global%20Rankings%20FINAL.pdf.

However, it should be mentioned that countries and territories, which according to the Figure 2 show the highest penetrations, i.e. South Korea and Hong Kong as well as, occupying the fourth place, Taiwan are mostly using FTTB technology, supported by LAN (Local Area Network). This means that the fiber network is not brought directly to the apartment, but only to the building. Within a particular building the individual dwellings might be connected with the use of other, than fiber, technology. In consequence the maximal bit rate of LAN networks inside such buildings might be lower.

According to the classic use of FTTH technology (being considered by the paper), as shown in Figure 2, Japan is the leader and the remaining places are occupied by Sweden, Norway and Slovenia.

In the rankings, showing the number of homes connected to fiber lines, and the number of subscribers, who have direct access to fiber, Poland is on a distant place.

In December 2010, Poland had only 25,500 subscribers using FTTH providing access to the Internet at 100-120Mb/s speed or more. [5] Polish position compared to other selected European countries by number of FTTH subscribers in the year 2010 presents figure 3, while the Polish position in comparison with other countries from Central and Eastern Europe presents figure 4.

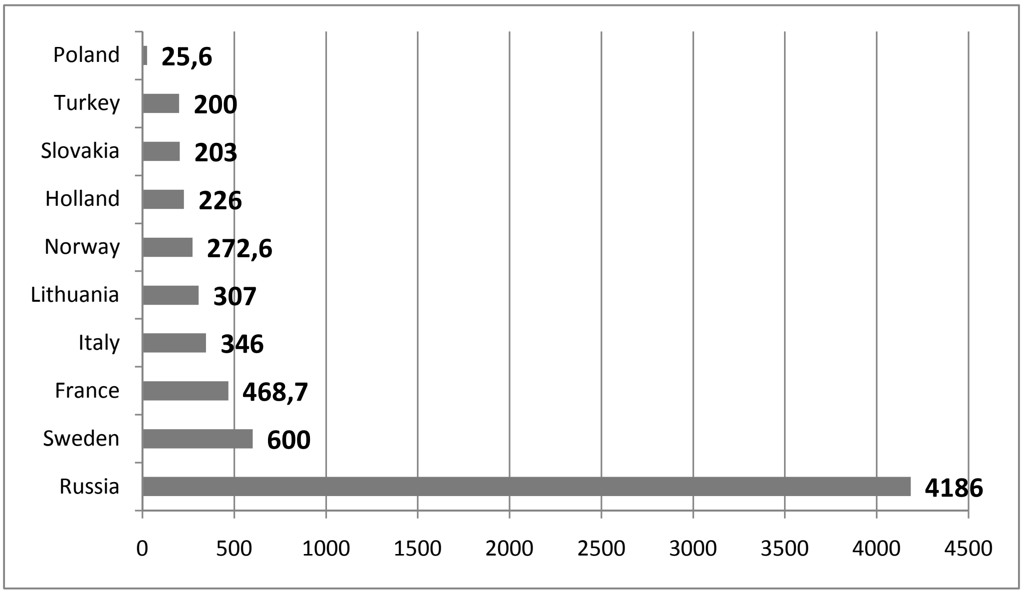
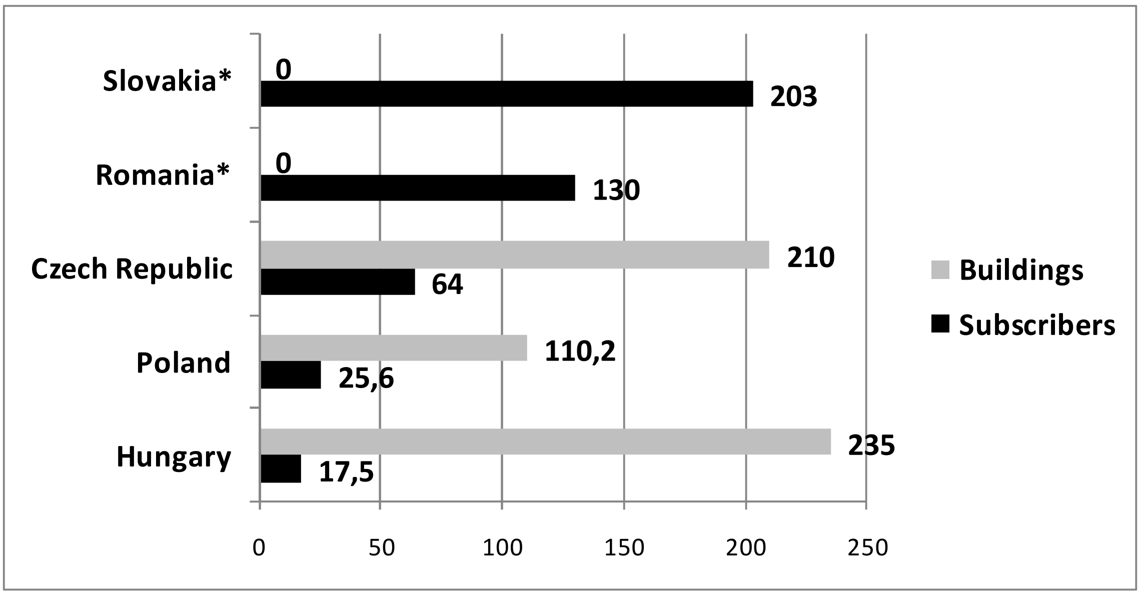


Fig. 3. Subscribers using FTTH in Poland and selected European countries in 2010 (in thousands).

Source: FTTH Council Europe after: “Rzeczpospolita” from 24.02.2011.



* No datas are available for Slovakia and Romania

Fig. 4. The number of subscribers and number of buildings having direct access to fiber (FTTH) in Poland compared to other countries of Central and Eastern Europe (in thousands).

Source: FTTH Council Europe after: “Rzeczpospolita” from 24.02.2011.

One indicates that the factor inhibiting investment in fiber is the desire to make the maximum use of existing ADSL networks, basing on copper cables. Therefore, the Polish national operator Telekomunikacja Polska invests sparingly in fiber technology. In Poland, the investment in fiber-optic cables are carried mostly by small operators, such as:

- INEA - operator from the city of Poznan,
- Telefonia DIALOG – a company operating mainly in Lower Silesia (Dolny Śląsk);
- Multimedia – a cable TV operator, locally investing in fiber, e.g. in such cities like Jaslo and Mielec.

The literature stresses that an unfavorable factor for the rapid development of FTTH networks are primarily the high cost of building this network. According to the “Comtext” magazine, cost of connecting an FTTH line to one dwelling ranges from 1,500 to 3,000 CHF (Swiss Francs). For this reason, an intensive development of FTTH networks is observed mainly in large cities. Examples of cities from the EU intensively deploying FTTH network, taking into account the main sources of funding for the construction and the degree of openness of access of the network are shown in table 1.

	Cologne	**Vienna**	**Paris**	**Hauts-de-Seine**	**Stockholm**	**Amsterdam**
Amount of connections	200 thousand. FTTH	1 million FTTH	1 million FTTH	FTTH	Dark fiber (unused fibers)	40 thousand FTTH
Municipal financial participation	250 millions	-	75-100 millions	Dup to 70 millions of subsidy	100 millions	6 millions
Open network	No	Aim - Yes	Yes?	-	Yes	Yes

Tab. 1. Examples of cities from the EU building FTTH networks, the source of funding and the degree of openness of these networks

Source: *An overview of Fiber*, 3rd November 2007, s. 4.

The information shown in the Table 1 indicate that the deployment of FTTH networks, in virtually every case, is financially supported by the city. Sometimes, however, one indicates a lack of coordinated approach in FTTH network deployment within particular cities, which is especially a result of fierce competition between telecom operators, cable operators and power plants.

Such situation may entail the risk of deployment - especially in cities - parallel FTTH networks, which would lead to an even further increase of costs of building this network. Moreover, another relevant issue is to provide a universal access to FTTH technology and not to limit it only to large municipal areas.

This indicates that for the prevention of possible irrational investment in FTTH technology as well as for the increase of eventual investor’s concern to provide universal access to this technology, some regulatory settlements may be helpful, including most of all the following:

- restrictions regarding the construction of parallel networks;
- to guarantee the use of the network without discrimination of anyone;
- establish an appropriate level of prices for network access, which allows the investors to achieve a satisfactory return on investments. [4]

4. Conclusion

The construction of FTTH networks is connected with high expenditures, but its advantages over the previously used technologies create a possibility for companies and even countries (in case it is used in a wide range), that will use FTTH a competitive advantage. The importance and status of this future-oriented technology, shows primarily the fact, that currently there is no technical alternative capable of sending such large amounts of data, and the demand for such transfers is permanently growing.

This indicates the need for development of FTTH networks while respecting the requirement to ensure the highest rationality of the process of its expansion.

Taking into account the high investment costs of FTTH networks construction and a growing and forward-looking role of this technology, one should consider the creation of collaborative associations gathering investors interested in building fiber optic infrastructure in a shared and coordinated way. Such investment would base on building a fiber-optic cable with multiple fibers, subsequently divided by the individual investors (e.g. according to the participation). The advantage of such solutions would be the:

- reduction of common expenses connected with the construction of FTTH networks,
- stimulation of innovative competition through presentation by individual fiber owners, their own network access offer as well as leased lines offers.

References

[1] SCHWEITZER, H. Presentation materials on www.hedresel.de/downloads/1presentaionh.schweizer.pdf, p. 12.

[2] FTTH Concil – *Definition of terms*, http://s.ftthcouncil.org/files/FTTH_definitions.pdf, p. 1.

[3] NOWAK, D., MURPHY, J., *FTTH: The overview of existing Technologies*, http://www.csi.ucd.ie/Staff/jmurphy/publications/-921.pdf.

[4] *Dank Glassfasernetz gestärkt aus der Krise*, “Comtext” 2009/05, p. 1.

[5] “Rzeczpospolita“ from 24.02.11.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

MAC Protocols for Directional Antennas Applications in Ad hoc Networks

[*]Tomáš Čakan, [*]Štefan Pollák, [*]Vladimír Wieser
[*]University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitná 1, 01026 Žilina, Slovakia, {tomas.cakan, stefan.pollak, vladimir.wieser}@fel.uniza.sk

Abstract. This paper introduces directional antennas application for Medium Access Control (MAC) protocols in wireless Ad hoc networks. Wireless Ad hoc network typically assumes the using of omnidirectional antennas at all nodes. By this way, two nodes communicate by using a given channel and all other nodes in the neighborhood are staying silent. With directional antennas, two pairs of nodes located in each other´s neighborhood may potentially communicate simultaneously, depending on the directions of transmission. Using directional antennas has many benefits compared with omni-directional ones: increasing spatial reuse, coverage range and network capacity. However, directional transmissions increase the hidden terminal problem, the problem of deafness and the problem of determination of neighbors' location. In this paper, we propose a review of MAC protocols for using directional antennas in Ad hoc networks.

Keywords: Medium Access Control, Directional antenna, Omni-directional antenna, RTS, CTS.

1. Introduction

Ad hoc network can be set up or deployed anywhere and anytime because it poses very simple infrastructure setup and none or minimal central administration. These networks are mainly used by community users such as military, researchers, business, students, and emergency services. In typical Ad hoc network nodes use omni-directional antennas. The medium in a wireless network is, by nature, a shared resource where a sender normally uses omni-directional broadcast mode to transmit a message for its intended destination. However, it is also possible to use directional antennas for transmission and reception in these networks. Using directional antennas some advantages, e.g. the increasing of spatial reuse, coverage range, network capacity and reducing interference by directing beamforms towards a requested direction. The using of directional antennas requires a new approach in the design of MAC protocols to fully exploit these advantages.

In this paper, we give overview of MAC (Medium Access Control) protocols for using directional antennas in Ad hoc networks [1].

1.1. Omni-directional and Directional antenna

While the omni-directional antenna is capable of sending and receiving signals from all directions (360 degrees), the directional antenna only from a specified direction, with typically greater gain than the omni-directional antenna. The direction in which the main lobe should point for a given transmission is specified by the MAC protocol. In fig. 1, the communication range using the omni-directional or directional antenna is compared [5].

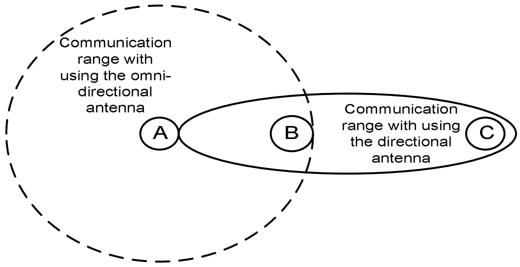


Fig. 1. Communication range using the omni-directional (B) and directional antenna (C)

2. RTS/CTS Mechanism in IEEE 802.11 MAC Protocol

IEEE 802.11 MAC protocol for omnidirectional antennas uses RTS (Request-to-Send) and CTS (Clear-to-Send) control messages. In this protocol, any node that wishes to transmit data must send a RTS packet before it can start the transmission. Node broadcasts a RTS packet for its intended receiver. If a receiver receives the RTS packet successfully, it replies to a CTS packet, and receiver can start transmitting a data packet. When a receiver successfully receives the data packet, then immediately sends an ACK (Acknowledgement) to node, which started a transmission. RTS and CTS packets obtain the proposed duration of data transmission, so all nodes within the radio range of transmitter and receiver must wait for duration of data transmission before they can transmit themselves. This reservation for data transmission is used for collisions prevention.

This characteristic of RTS/CTS mechanism overcomes the hidden terminal problems in wireless LAN environments. However, using this mechanism waste a large portion of the network capacity by reserving the wireless medium over a wide area [1].

2.1 Directional MAC (D-MAC) protocol

This protocol utilizes a directional antenna for sending RTS packets in a specific direction and omni-directional antenna for sending CTS packets [2]. The physical location must be obtained by using the global positioning system (GPS). Let’s assume the situation, where node B wants to transmit a data packet to node C (fig. 2). Assume that no other transmitting is in progress. Node B is starting to send a directional RTS (DRTS) packet, including the physical location information of B (in the direction of a node C). Whereby, node A doesn’t receive the DRTS from node B (even though node A is within B transmission range). If node C receives DRTS packet from node B, then node C sends an omnidirectional CTS (OCTS) packet. In this packet the information about location of node C and location of a node B are included. After the successful exchange of DRTS and OCTS packets, data packet is sent from node B to C using directional antenna. When node C receives the data packet, then sends an ACK message to node B using a directional antenna.

Now it's possible to communicate from node D to node E. Directional antenna of node D is blocked towards to node C, because node D received OCTS packet (exchange packet from node B to node C). The blocked antenna on node D is different from the antenna, which is directed toward to node E. If node D get information that data transmission to node E would not interfere with on-going data transfer from node B to C, then sends the DRTS packet to node E.

This modified MAC protocol improves performance by allowing simultaneous transmissions by using directional antennas.

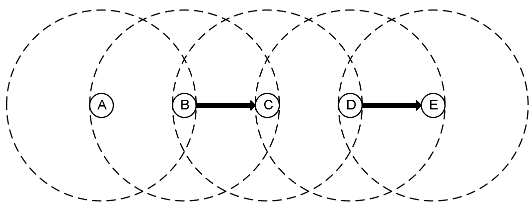


Fig. 2. Example of Directional MAC Scheme (D-MAC)

2.2 D-MAC protocol without location information

This MAC protocol uses directional antennas where the nodes do not have location information about the other nodes. The protocol uses RTS/CTS packets for identifying source and destination nodes. We assume that node listens ongoing transmissions on all elements of directional antennas [7]. Node which wishes to communicate to any neighbor, first sends omnidirectional RTS packet addressed to destination node. In fig. 3, node A sends RTS packet to node B transmitted on all antennas because node A does not know the position of node B. If node B receives RTS packet correctly, then reply by CTS packet again on all antennas. Node B notes the direction from which is receiving RTS packet by antenna which receives RTS packet in maximum power (antenna 2 in the figure). Similarly A estimates the direction of node B when receiving CTS packet. If the RTS/CTS handshake is performed successfully, node A begins transmitting the data packet (antenna 4). All neighbors of node A and B that hear the RTS/CTS dialog, use this information to prevent interfering with the ongoing data transmission.

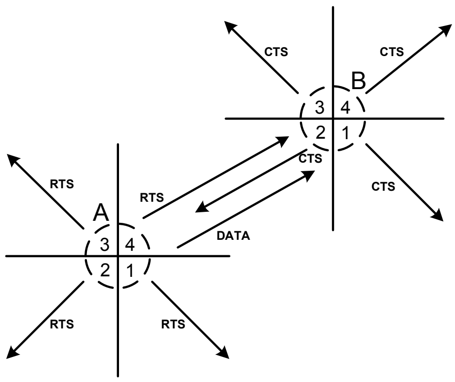


Fig. 3. Example of using directional RTS/CTS handshake in D-MAC Scheme

2.3 Multiple Beam Antenna Array MAC (MBAA-MAC) protocol

In this protocol, after successfully exchanging RTS/CTS packets, transmitter instantly does not send the DATA packet as in a classic IEEE 802.11 scheme. It waits for a time period called additional control gap (ACG) [3, 4]. This time period is inserted between the RTS/CTS and DATA packets. ACG time is used for neighbor nodes, in the vicinity of the sender and the receiver, as a chance to exchange their own control packets and schedule concurrent data transmission. The MBAA procotol can be used to obtain better spatial reuse by scheduling two or more concurrent transmissions.

For example, we assume two transmissions, one from node A to B and the other one from node C to D. Node A starts transmission by sending RTS packet (on all antenna elements) to node B. The RTS packet includes information about scheduled start times of A´s DATA packet and B´s ACK packet. Then node B responds with a CTS packet to node A including some scheduled timing information (on all antenna elements). The node B notes information about the direction of node A (using an element which receives maximum power of the RTS packet). The timing information is required such, that a node in the neighborhood of node A and node B can estimate whether or not it can receive a data packet from some other node simultaneously, while node A is transmitting data to node B.

2.4 Circular Directional RTS

This protocol is based on an innovative scheme of RTS packets. In this scheme, RTS packets are transmitted directional consecutively in a circular way and thus scan all the area around the transmitter [6]. In the fig. 3, it is assumed that transmitter A starts transmitting RTS packet in a dedicated direction (beam 4 in fig. 3). Transmitter sends RTS packet in all the area around the transmitter (sending RTS packet using all directional beam M of antenna). The RTS packet contains information about duration of RTS/CTS handshake. This information is transmitted around the transmitter using the circular RTS packet and neighbors are informed about the intended transmission.

The neighbors are using a simple algorithm to compute the direction of transmitter or receiver (neighbors are aware of the intended RTS/CTS handshake). If RTS packet is received correctly, then the receiving node transmits CTS packet towards the direction of the transmitter. When the transmitter completes the circular transmission of RTS packet, then listens to the omnidirectional medium to receive CTS packet. If CTS packet is received correctly, then the transmitter continues with transmission of the Data packet and ACK packet (these packets are transmitted directional). Sending RTS, CTS, DATA and ACK packets directionally, increases coverage area, compared with the omnidirectional transmission. This protocol doesn´t need any information about location of nodes.

3. Conclusion

In this work, the overview of MAC protocols for Ad hoc networks using directional antennas is proposed. The traditional IEEE 802.11 MAC protocol is using omnidirectional RTS/CTS handshake, which can waste bandwidth by reserving the wireless medium over a large area. By using directional MAC protocols and directional antennas the increasing spatial reuse, coverage range and network capacity is reachable. These MAC protocols are using a variation of the RTS/CTS handshake for discovering a position of the source and destination nodes or the position is discovered by using GPS. In summary, to find the best directional MAC protocol for Ad hoc networks with application of directional antenna to further improve quality of service parameters requires another research.

Acknowledgements

This paper was supported by the Scientific Grant Agency VEGA in the project No. 1/0336/10.

References

[1] BANDYOPADHYAY, S., ROY, S., UEDA, T. *Enhancing the Performance of Ad Hoc Wireless Networks with Smart Antennas.* ISBN 0-8493-5081-6, Boca Raton: Taylor & Francis Group, 2006.

[2] NASIPURI, A., YOU, J. YE., HIROMOTO, E. R. *A MAC Protocol for Mobile Ad Hoc Networks Using Directional Antennas.* Wireless Communications and Networking Conference, pages 1214 – 1219, vol.3, September 2000.

[3] KO, YOUNG-BAE, SHANKARKUMAR, V., VAIDYA, H, N. *Medium Access Control Protocols Using Directional Antennas in Ad Hoc Networks.* IEEE INFOCOM 2000, March 2000.

[4] VERMA, R., PRAKASH, A., VERMA, P.K., TYAGI, N., TRIPATHI, R. *A Novel MAC Protocol for MANETs using Smart Antenna System.* Power, Control and Embedded Systems (ICPCES) 2010, December 2010.

[5] ELBATT, T., ANDERSON, T., RYU, B. *Performance Evaluation of Multiple Access Protocols for Ad hoc Networks Using Directional Antennas.* Wireless Communications and Networking, 2003, March 2003.

[6] KORAKIS, T., JAKLLARI, G., TASSIULAS, L. *A MAC protocol for full exploitation of Directional Antennas in Ad-hoc Wireless Networks.* MobiHoc '03 Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, 2003.

[7] CHOUDHURY, R. R., YANG, X., RAMANATHAN, R., VAIDYA, H. N. *Using Directional Antennas for Medium Access Control in Ad Hoc Networks.* MOBICOM´02, September 2002.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

ARP Cache Poisoning

[*]Veronika Ďurčeková, [*]Ivan Dolnák
[*]University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitná 2, 010 26 Žilina, Slovakia, {veronika.durcekova, ivan.dolnak}@fel.uniza.sk

Abstract. The Address Resolution Protocol (ARP) is used by computers to map network IP addresses to physical MAC addresses. However it is known that ARP messages can be misused and ARP cache can be maliciously poisoned. This is a serious problem on the Local Area Network (LAN) because ARP cache poisoning can lead to Man-in-the-middle attack (MITM) or to Denial of Service (DoS). This paper is aimed at ARP cache poisoning and ARP spoofing attacks where an intruder can impersonate another legitimate host to gain access to sensitive data this way. The problem and technique of ARP spoofing is clarified in this paper and some selected solutions to prevent computer networks from ARP spoofing are proposed. One of the possible solutions was tested in real conditions and results are presented at the end of this article.

Keywords: Address Resolution Protocol. ARP Cache Poisoning. ARP Spoofing. Network Security. Arpwatch.

1. Introduction

The Address Resolution Protocol (ARP) is sometimes discussed in terms of position in the ISO OSI model. ARP was not developed based on the design principles and strict encapsulation hierarchy of this model and therefore such discussions still create conflicts about the exact operating layer within this model. Most often ARP is placed into the data link layer but since it requires the definitions of network addresses of the network layer, it is not unusual to find it referenced at that layer. [11] In this paper we refer to the ARP as exclusively network layer protocol.

The Address Resolution Protocol is responsible for the resolution of known logical IP address into a corresponding physical MAC address. This is done when a host wants to communicate with another host within or outside the LAN and does not know the physical address of the receiving side, respectively the destination physical address. [1] The process of resolving destination physical address consist of several steps beginning with the broadcast ARP request sent on the connected network segment asking for the MAC address of the host with known IP address. The host with the requested IP address than sends a unicast ARP reply message to the originator with the pair of IP and MAC address. [2]

One well known hacking technique is ARP Cache poisoning, which serves to forge ARP request or ARP reply. It is a technique used to sniff data, modify the traffic or in special case stop the through coming traffic and execute Denial of Service Attack (DoS). [2] This type of attack can be executed on every network, which uses ARP protocol.

There is a term and kind of attack – the ARP spoofing - that is sometimes classified as ARP Cache poisoning. In our opinion ARP Cache poisoning and ARP spoofing are two different computer attacks. The ARP Cache poisoning relates to ARP spoofing which is the background for the first mentioned attack. ARP spoofing is about the malicious host, who clones the MAC address, which could be simply eliminated by, for example switch port security.

2. Background and Problem Definition

The basic principle of ARP Cache poisoning is to send fake ARP frame to the LAN. Launching ARP Cache poisoning attack can be simply done by updating an existing ARP entry or inserting new forged entry in the ARP cache for a target host. The goal of an attacker is to associate his or her MAC address with the IP address of another host, mostly the default gateway in LAN. Any traffic meant for the gateway is then mistakenly sent to the attacker instead. And this is the point, when the attacker could either choose to forward the traffic to the correct default gateway without changes or modify the data before forwarding it. This attack performance could be also identified as Man-in-the-middle attack (MITM), which is also well known in hacker community. The attacker then can listen to the traffic between legitimate hosts. Another option is that an attacker could also launch a Denial of Service (DoS) attack against a victim by associating a nonexistent MAC address to the IP address of the victim's default gateway so that every packet that the victim sends is in a matter of fact sent to the wrong MAC address. So MITM and DoS attacks are two main purposes of executing ARP Cache poisoning. What should be said is that this attack can be launched from a compromised host, or an attacker's computer that is connected directly onto the same LAN segment. [3, 4]

In the current communication networks, ARP Cache poisoning is a popular method to perform a MITM attack inasmuch as eavesdropping belongs to the most potential attacks. This means that an attacker may be able to trick one or both hosts into thinking that the attacker's MAC address is the address of the other host, for example router, SIP server, DNS server, and so on. And the attacker is able to sniff all the traffic, while forwarding it on to the intended host. The flaw which makes the ARP Cache poisoning possible to execute is that some operating systems accept and replace an entry in their ARP cache regardless of whether or not they have sent an ARP request before.

3. Detection and Defence

The process of detecting ARP Cache poisoning and ARP spoofing attacks in LAN network based on Ethernet is relatively simple. To accomplish this, very common solution is to implement server machine with appropriate software, or to use functions implemented in network switches onto intended network segment. The detection and prevention of ARP attacks could be realized for example as in [5], where there are used some plug-ins to join a special ARP detection module to Intrusion Detection System like Snort. Another process is presented in [6], where is offered an architecture for detecting ARP attacks. It discusses a new practical technique performed by adding special external hardware element to the LAN network to work as a sniffer. External elements are combined and create architecture for practical implementation in production network.

Because of sharing the same LAN segment by end user computers and IP phones, defence mechanisms use to be more complicated as detecting methods. Solution of this problem could be:

1. using (manual) detection provided by network administrator(s) – for example the use of open source software running on server in intended network segment
2. using simple and effective solution – also known as ARP inspection, common function of managed LAN switches.

3.1. Open source software for attacks detection and defence

There are plenty of open-source software solutions, which try to help avoiding possible breaking attempts. For example, first solution which could be mentioned is **ArpON**, also known as Arp handler inspection. This package can detect and block all ARP poisoning and spoofing attacks on switched or hubbed LANs with static ARP inspection or dynamic ARP inspection approach.

For detecting suspicious ARP traffic could be used very favourite solution, which is Arpwatch. **Arpwatch** listens for ARP replies on a network and sends to network administrator a notification about the ARP entry change via an email. The notification informs network administrator about the possibility that in connected LAN network exists MAC address cloning, see Fig. 2. This software tool is quite famous and widely available but it depends on the network administrator, who should be able to differentiate between conventional traffic and ARP Cache poisoning attacks. It depends also on the ability of the network administrator to take fast and appropriate arrangements. [3]

There is available the GUI-driven software **XArp** with the use of Windows operating system. It performs ARP packet inspection on the chosen network interface based on configurable inspection filters and active verification modules.

There is also one simple solution for less challenging ARP spoofing attack, which is the use of static MAC addresses. However this does not scale on large networks, where the static mapping has to be set for each host.

3.2. ARP inspection in LAN switches

LAN switches from vendors like Cisco or Linksys provide function for ARP Inspection - detection of ARP attacks. This is also known as Dynamic ARP Inspection (DAI). It determines the validity of an ARP frame based on the MAC/IP address bindings stored in a DHCP Snooping database. To prevent ARP Cache poisoning or spoofing, a switch must ensure that only valid ARP requests and responses are relayed. To ensure that only valid ARP requests and responses are relayed, DAI takes the following actions. [8] First, DAI forwards ARP packets received on a trusted interface without any checks and it intercepts all ARP packets on untrusted ports. Second, DAI verifies whether each intercepted packet has a valid MAC/IP address binding before forwarding those packets which can update the local ARP cache. Last but not least, DAI drops, logs, or drops and logs ARP packets with invalid MAC/IP address bindings. As a result, all access switch ports should be configured as untrusted and all switch ports connected to other switches (trunks) should be configured as trusted. All ARP packets traversing the network from an upstream distribution or core switch could bypass the security check requiring no further validation. The model of using DAI could be seen on Fig.1.

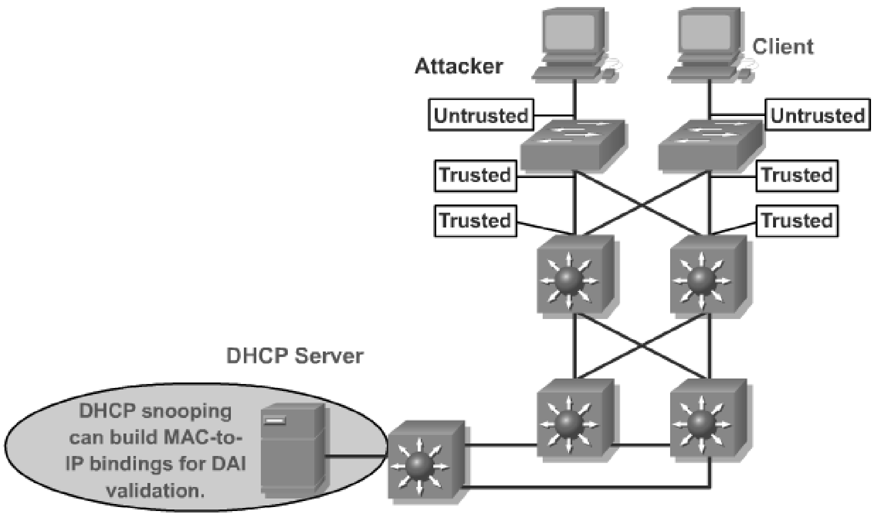


Fig.1. Dynamic ARP inspection (Source: Cisco, 2007).

3.3. ARP inspection based on open source software

From the above mentioned open source software solutions was successfully tested and implemented Arpwatch package, running on GNU/Linux operating system. This software was used in LAN segment, where servers, host computers and IP phones coexist without any security mechanisms implemented. We executed several ARP spoofing and ARP Cache poisoning attacks using Linux BackTrack 4 R2 distribution oriented on security penetration testing and we focused on security violation in LAN. All attacks were relatively simple, because there were used software packages available in distribution and the goal was to proof the concept of possible ARP attacks. On Fig. 2 is presented example of Arpwatch output, an e-mail message with “flip-flop” subject sent to the network administrator. Every e-mail like this one could mean presence of ARP Cache poisoning in LAN network segment. This figure illustrates result sent via e-mail in case of existence 2 MAC addresses to one IP address.

```
From: arpwatch@server.fel.uniza.sk (Arpwatch)
To: root@server.fel.uniza.sk
Subject: flip flop
Status: O

hostname: <unknown>
ip address: 10.0.0.243
ethernet address: 0:1a:37:24:4e:4a
ethernet vendor: <unknown>
old ethernet address: 0:c0:9f:54:83:f
old ethernet vendor: QUANTA COMPUTER, INC.
timestamp: Wednesday, March 01, 2011 12:44:19 +0100
previous timestamp: Wednesday, March 01, 2011 11:43:08 +0100
delta: 1 hour
```

Fig. 2. Output from Arpwatch open source package.

4. Conclusion

ARP poisoning is popular and common technique to perform a MITM attack in LANs, in which eavesdropping is simply one of the possible potential impacts. Results of testing proofed the existence of the possibility to use this kind of attack to eavesdrop running data communication, phone calls, etc. The test also proofed the necessity of securing LAN switches and server machines.

References

[1] XING, W., ZHAO, Y., LI, T. *Research on the defense against ARP Spoofing Attacks based on Winpcap.* Education Technology and Computer Science, Wuhan, 2010.

[2] PUANGPRONPITAG, S., MASUSAI, N. *An Efficient and Feasible Solution to ARP Spoof Problem.* Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Pattaya, Chonburi, 2009.

[3] ABAD, C. L., BONILLA, R.I. *An Analysis on the Schemes for Detecting and Preventing ARP Cache Poisoning Attacks.* Distributed Computing Systems Workshop, Toronto, 2007.

[4] AL-HEMAIRY, M., AMIN, S., TRABELSI, Z. *Towards More Sophisticated ARP Spoofing Detection/Prevention Systems in LAN Networks.* Dubai, 2009.

[5] XIANGNING, H., ZHIPING, J., XINLI, T. *The detection and prevention for ARP Spoofing based on Snort.* Computer Application and System Modeling, Taiyuan, 2010.

[6] DESSOUKY, M.M., ELKILANY, W., ALFISHAWY, N. *A Hardware Approach for detecting the ARP Attack.* Informatics and Systems, Cairo, 2010.

[7] ENDLER, D., COLLIER, M. 2007. *Hacking Exposed VoIP – Voice Over IP Security Secrets & Solutions.* New York : McGraw-Hill, 2007. ISBN 978-0-07-226364-0

[8] CISCO. 2007. *CCNP: Building Multilayer Switched Networks - 5.0.* [online]. San Jose : Cisco Systems, Inc., 2007-01-09, [cit. 2011-03-03]. Available on Internet: <http://www.cisco.com/web/learning/netacad/index.html>

[9] VOIPSA. 2005. *VoIP Security and Privacy Threat Taxonomy.* [online]. Voice over IP Security Alliance, 2005-10-24, [cit. 2011-03-03]. Available on Internet: <http://www.voipsa.org/Activities/VOIPSA_Threat_Taxonomy _0.1.pdf>

[10] WIKIPEDIA. 2009. *ARP spoofing.* [online]. Wikipedia : The Free Encyclopedia, 2009-03-07, [cit. 2011-03-13]. Available on Internet: <http://en.wikipedia.org/wiki/ARP_poisoning>

[11] WIKIPEDIA. 2011. *Address Resolution Protocol.* [online]. Wikipedia : The Free Encyclopedia, 2011-03-03, [cit. 2011-03-13]. Available on Internet: <http://en.wikipedia.org/wiki/Address_Resolution_Protocol>

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Analysis of Data at the Car Color Recognition

*Stanislav Foltán
*University of Žilina, Faculty of Management Science and Informatics, Department of Information Networks, Univerzitná 2, 01026 Žilina, Slovakia, {foltans}@gmail.com

Abstract. In this paper, I present an application of computer vision for car color recognition. Images used as the inputs for the application were taken in a real world environment, so aspects, like light or weather conditions cause problems in this task. My approach is a simplified version of the current advanced approaches. The main value of this paper is illustration of the results that are to be expected with such an approach. I have defined parts of the car body where I picked color samples. By making a histogram of sampled colors, by means of neural network, a dominant color is recognized.

Keywords: car color recognition, computer vision, GRB, HSV, image processing.

1. Introduction

In this work, I try to enable a computer to recognize color of an automobile captured in an image. Although many papers were published about computer vision, very few have looked at color vision in a real world environment. A review of techniques and algorithms for outdoor machine vision can be found [1]. The main problems introduced for outdoor color vision are the surface reflection and illumination at various climate conditions and periods throughout the day. For elimination of these aspects, authors have tried to estimate reflection parameters [2] or [3]. Another approach that I have examined used a CIE daylight model and its upgrades [4]. Paper [5] also presents the results of CIE daylight model in autonomous vehicle systems. The results address a large number of false positives, and authors propose using of robust models for color prediction. Robust models are not what I am interested in this work. The main goal is to add new functionality for an already established system for car brand recognition [6] with the lowest costs. This system would be able to recognize the basic and thereby, the most used car colors. To achieve the goal, I proposed sampling the input image, sorting the samples into color intervals of HSV color space (CIE color space could be used as well, like in [7]), and finally, by means of feed-forward neural network making assumptions about the color of the car. While at the beginning there is a question whether it is even possible to create a functional application without using robust methods, the conclusion of this work will produce an answer.

2. Problem solution

2.1. Problem analysis and preprocessing

The aim is to find out a method for car color recognition. To test the approach, I chose an entrance to a parking lot, where a camera is strategically placed. This camera is connected to an automatic system which produces snapshots in real time when a vehicle pulls in front of the barrier. These snapshots are used as the inputs for my application. Fig. 1 and Fig. 2 (left) illustrate the example of camera snapshot. Raw images from CCTV camera (AXIS 120A) are in JPEG format in resolution 640 x 480px.

The first step at recognition is to find a vehicle whose color is going to be recognized. This is a potential problem because it is not easy to extract foreground (to separate a vehicle) from a non-static background. Our background contains trees moving in the wind, walking pedestrians, etc., so

I need a special process for this particular use to find out which part of the image is important for later analysis and which is not. I solved the problem of car detection by localization of license plate [6] and later by approximate identification of the bumper and bonnet areas (Fig. 2 left). These are the car body parts which reflect the true car color with the highest probability. Fig. 1 illustrates the share of each part area of the total area of the image.

Other and even more critical problems that are experienced are illumination and surface reflectance (Fig. 1 right). These problems are typical for color vision in an outdoor environment.

Fig. 1. Images of the distinctive areas shown on a vehicle. The picture on the left shows an ordinary car, with the bonnet at about 30% of the car area. The picture on the right illustrates problem of reflection. Here the color orange represents the true/original color of the car (about 5% of the area) and yellow covers the remaining bonnet area and depicts the reflective, not original color of the car (about 10% of the area).

2.2. Calibration of colors

The main prerequisite for color recognition is the need to know the type of sampled color. For this experiment, I identified ten colors (see Tab. 1). That means I have to determine whether the sampled color is one of these, or if it is another color. This could be a problem while the input images are saved in form using an RGB color model. The problem is to assign ranges of RGB coordinates for each color. For this reason, I have decided to use an HSV color model (HSV stands for hue (range from 0° to 360°), saturation (range 0 to 100%) and value (range 0 to 100%)). The advantages in this color model are that the primary information about color is contained in the hue coordinate. For the calibration of color ranges, several photographic images and scales were utilized.

Color	Range
White	V>95
Silver	V<70,95)
Gray	V<35,70)
Black	V<35
Red	H<0,15) and H>340
Green	H<75,170)
Blue	H<170,270)
Yellow	H<45,75)
Orange	H<15,45)
Magenta	H(270,340>

Tab. 1. Values of HSV coordinates found out as a result of color calibration.

2.3. Image analysis

Once I am able to distinguish between color samples, I can proceed in determination of the primary vehicle color. Input for this step is the original image and knowledge of license plate positioning. The area above the plate, part of bonnet, appears to be the best place for taking samples. Another good candidate is the front bumper. Fig. 2 illustrates the sample area used in this research. For data reduction, not all pixels in the area of interest are sampled. Outputs of the analysis are two histograms of colors, each for one sampling area.

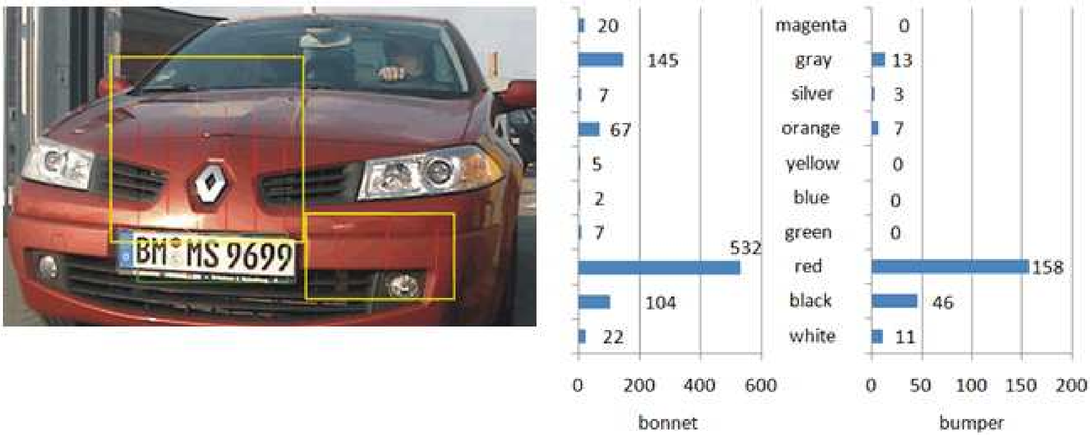


Fig. 2. The image on the left illustrates the concept of sampling. In yellow rectangles the bonnet and additional bumper area are depicted. Histograms for the bonnet and the bumper of this car are on the right. However from these two histograms the dominant color is clearly visible, this is not always true.

2.4. Color estimation based on the image analysis

The main color estimation is based on histograms like the ones in Fig. 2 right. Every result is not as obvious as in this example and some problems will occur in its use. If I take samples from parts which are not the car color, for example a grille, a number of other colors increases (in the case of a grille, usually black). If there is a reflection, a number of white or blue pixels increase. The main task is to use a method which can handle these obstacles. I have made experiments with two methods:

1. Estimation by means of decision tree. This method is described in detail in [8]. It uses probability of occurrence of each color in the picture to make the decisions between possible results.

	Numbers
Number of images	202
Number of negative results	22 (10.89%)
Number of negative results (A)	9 (4.46%)

Tab. 2. The results of color recognition by means of decision tree. (A) Number of negative results if mishmashes caused by illumination and reflection are tolerated.

2. Estimation by means of neural network (NN). The background of NN is generally well known, many papers were published, i.e. [9]. To address the question of whether the NN is able to recognize color from histograms better than if decision tree is used, I decided to do the following test. I used a simple feed-forward NN with 20 inputs, one hidden layer with 17 neurons, 4 outputs, sigmoid activation function, a backpropagation as a teaching method and a set of 202 input images as a training set. After each round of teaching, I tested the neural network with the same set of images. The results of this test are depicted in Fig. 3.

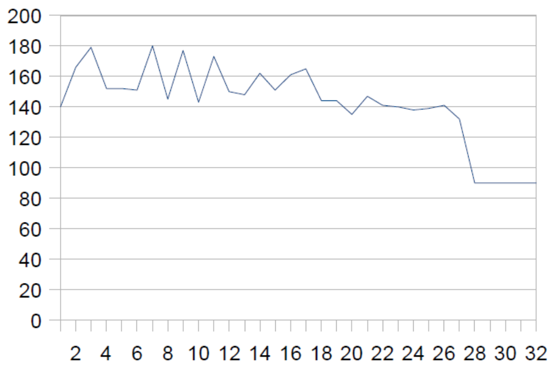


Fig. 3. Neural network learning process. The horizontal axis represents repetition of the learning process and the vertical axis represents the number of undistinguished inputs from total number of 202 inputs.

3. Conclusion

The objective of this work was to test possibilities of car color recognition in a real world environment. For this purpose I acquired a set of images from CCTV camera placed at the entrance of a parking lot. Real world environment means that the images contained in this set were taken in different light conditions, from different angles and ranges.

To achieve this objective, the input images were first pre-processed, which means that the parts of the car that were identified as the car color were selected. The image pixels from these car parts were sampled and sorted into histograms. Then two approaches for the histogram analysis were used. The first approach used predefined rules for recognition and the second one tested the odds of neural network.

From the results shown in Tab. 2, it is visible that the first method used for recognition brings relatively positive results. The main factors of negative results were heavy illumination and reflection (this can be partially eliminated if there were constant light conditions in a covered area, for example, using this method in a final application deployed in a closed parking garages). The most problematic issue is recognition errors in determining between black/blue, silver/blue and white/silver. If I take into account errors created by a change in lightning conditions which I have no control over, the results of experiments were 95.54% success on the dataset of 202 images.

In comparison, the results of determination based on frequency of occurrence, the success rate of the neural network method is much lower (55.5%). In this test, the errors were derived mostly from the category "absolutely wrong", not like the first method, where errors were caused by light conditions, making the results inaccurate.

In my opinion, the poor performance of the neural network is that the data sets could appear to be mutually exclusive for a neural network. As the idea of elimination was explained in [8], the information about the presence of many black pixels should have a different weight if many other pixels are red and another weight if they are blue. Here is the weakness in using the neural network approach in this color vision application and therefore, I would decide to utilize the first approach in the final application.

References

[1] SAHRAGARD, N., RAMLI, A.R. *A Review on Algorithms and Techniques for Outdoor Machine Vission*. In European Journal of Scientific Research, EuroJournals Publishing, Inc. 2009.

[2] TOMINAGA, S. *Estimation of Reflection Parameters from a Color Image*. In Computer Vision — ACCV'98, Lecture Notes in Computer Science 1999..

[3] MALONEY, L.T., WANDELL, B.A. *Color Constancy: a method for recovering surface spectral reflectance*. In Optical.Society of.America. J. Opt. Soc. Am, 1986.

[4] BULUSWAR, S.D., DRAPER, B.A. *Color models for out-door machine vision*. In Computer Vision and Image Understanding, 2002.

[5] BULUSWAR, S.D., DRAPER, B.A. *Color machine vision for autonomous vehicles*. In Int. J. Eng. Appl. Artif. Intell. 1998.

[6] FOLTAN, S., BADURA, S. *Robust car brand recognition from camera image*. MENDEL 2010, 16th International Conference on Soft Computing, 2010.

[7] TOTH, Š. *Using traffic sign recognition in map applications*, In Symposium GIS Ostrava, 2011.

[8] FOLTÁN, S. *Car color recognition from CCTV camera image*, In Theoretical and applied aspects of cybernetics 2011, Kyiv, Ukraine , 2011.

[9] BISHOP, CH. M. *Neural Networks for Pattern Recognition,* Oxford University Press, 1995.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Retrieval 2D Images Using Shape Descriptor

[*]Miroslav Frič, [*]Patrik Kamencay
[*]University of Zilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitna 1, 01026 Zilina, Slovakia, {Miroslav.Fric, Patrik.Kamencay}@fel.uniza.sk

Abstract. This paper deals about retrieval images based on the content. We briefly discuss problems current systems based on the annotation and introduce retrieval images based on the shape features. We present semantic gap and we shortly brief principles of Fourier descriptors. Simply experiments were done on base Fourier descriptors. In this case the combination two experiments show that extraction of one low level feature is not sufficient. However, retrieval by shape is still considered to be a more intrinsically difficult task compared to retrieval based on other visual features. In addition, the problem of shape retrieval becomes more complex when extracted objects are corrupted by occlusion or noise as a result of the image segmentation process. We propose extract more features from the image and with combining complex segmentation could be improve retrieval process.

Keywords: Fourier descriptors, shape, feature extraction

1. Introduction

In this last decade a drastic increase in the size of image databases has been realized. Multimedia databases deal with text, audio, video and image data which could provide us with enormous amount of information and which has affected our life style for the better. As illumination, Flick – very popular image hosting web, was upload more than two billion images from users between years 2004-2007 and volume of image still increasing [2]. Retrieval techniques are text based or content based [1-2]. However, text based annotation have a few disadvantages:

1. The set of words describing images is not fixed and annotation cannot be automatic, because understanding of the image is still difficult problem.
2. Annotation is not always accurately describes the content of the image.
3. It requires enormous effort if we would like to describe all images manually.
4. Every user describe each image differently.
5. Existing things and moments cannot be describing by text.

2. Semantic gap

Many text-based search engines are available on the World Wide Web, and they are among the most visited sites. Every day is adding much multimedia content on the popular web side. For searching information on the web site was develop text based search engines for instance *www.google.com*, *www.ask.ccom*, *www.yahoo.com* etc. The identification searched information give us a relevance result however, not relevance result for visual content. If we talk about image retrieval we using low level features as color, texture and shape for retrieval images. However, these low level features we can be extracted from the objects. They called low-level features because most of them are extracted directly from digital representations of objects in the database and have little or nothing to do with human perception. The all people perceive all things at the images with semantic based on high level features. If we can retrieval images we have to use low level features for instance we divide image based on their dominant color and showing similar image with chosen color. The dominant representations for the people are building, people, and animals at the images. This difference between low level features (color, shape, texture) and high level human semantic perception is called semantic gap. Reducing gap between low level features and high level features is still challenging for researcher from all over the world. Reducing the semantic gap and effective retrieval images by their visual content we can by developing a sophistically algorithm [3-4].

2.1. Fourier descriptors (FDs)

In order to make model shape and data shapes comparable, the shape representation must be invariant to translation, rotation and scale. It is difficult to achieve under spatial domain and we must involve large amount of computation. Using FDs, the problem can be solved. Rotation invariant of the FDs is achieved by ignoring the phase information and by taking only magnitude values of the FDs [3]. Shape boundary is set of coordinates (x_i,y_i), i = 1,2,3…N which are extracted out in the preprocessing stage by contour tracking technique. The centroid distance function is expressed by distance of the boundary points from the centroid (x_c,y_c) of the shape

$$r = ([x_i - x_c]^2 + [y_i - y_c]^2)^{1/2},\ i = 1,2,3,...,N\ , \tag{1}$$

where x_c,y_c are averages of x coordinates and y coordinates respectively. Due to subtraction of centroid (this centroid represents the position of the shape) from boundary coordinates, the centroid distance representation is invariant to shape translations. In order to apply Fourier transform, all the shapes in database are normalized to the same number of boundary points.

Fourier transform of ri, i = 1,2,3…N-1 is then given by

$$un = \frac{1}{|\,N}\sum_{i=0}^{N-1} r_i \exp\left(\frac{-j2\pi ni}{N}\right), n = 0,1,...,N-1\,. \tag{2}$$

The coefficients un, n = 0, 1,…, N-1, are called Fourier descriptors (FD) of the shape, denoted as FDn, n = 0,1,…N-1. The FDs acquired in this way is translation invariant due to the translation invariance of centroid distance. To achieve rotation invariance, phase information of the FDs are ignored and only the magnitudes $|FD_n|$ are used [6-9]. Scale invariance is achieved by dividing the magnitudes by the DC component, i.e., $|FD_0|$. Since centroid distance is a real value function, only half of the FDs are needed to index the shapes [8]. Finally, the following feature vectors are used as the Fourier descriptors to index the shape.

$$f = \left[\frac{|FD_1|}{|FD_0|}, \frac{|FD_2|}{|FD_0|}, ..., \frac{|FD_{N/2}|}{|FD_0|}\right]. \tag{3}$$

The similarity measure of the query shape and a target shape in the database is simply the Euclidean distance between the query and the target shape feature vectors.

3. Retrieval images based on the shape features

This simply experiment was carried out on the database (DB) by showing on the Fig 1. The input DB was chosen very carefully, with the simply shape. In this case the bottle was chosen. It contained over 16 images and extracted the dominant object in these images, the bottle. Images were first segmented based on Graph segmentation method. The input parameter “s” (threshold) set on 0.8 and parameter “k” (number of clusters) set on 300 were used. Best result was achieved for this input segmentation parameter. In next step, all segments were described by Fourier shape descriptor in size of 12 numbers for features vector. Consequently, every image were described by features vector of size n*12 where “n” is number of segments in image. Features vector was gained by comparing with reference features vector. Reference features vector was gained from the binary images of the bottle. For imaging, bottle was expressed by closed curve of binary bottle.

Fig.1. Input images with the bottle object

Difference between referenced features vector and vector from the real DB is minimum value. The DB contains similar types of images. Moreover, quality of segmentation had a substantial impact of the retrievable images. The precision retrieval images based on this method was 83, 6%. Similar experiment was made on similar DB. Similar object, which was chosen, has contained a shape of the bottle or very similar shapes (battery spray, etc.)

Fig.2. Reference image for gaining reference shape features

With using a same method on the DB was gained different results. Precision retrieval images was dropped to 43, 3%. Reason, why the results are different (distance between shape features), is that shape of the bottle is very similar as shape of the spray. The chosen segmentation method is not suitable, because segmentated objects were not completely. In many cases the chosen segmentation method does not work properly and retrieval process could be corrupted. Shape features were gained and in many cases we are not able to say, if the object of interest was the bottle or something else (spray, battery, etc). In addition, the problem of shape retrieval becomes more complex when extracted objects are corrupted by occlusion or noise as a result of the image segmentation process. The improving the accuracy of retrieval can be done with using a low level features as colour, texture or their combination. With extraction a many features of the object should be improved a precision of retrieval.

4. Conclusion

Retrieval images based on the shape is challenge for everyone on the area of image processing or image recognition. Many proposed method by extraction a shape features has been realized. The accuracy of retrieval system or method is important condition on the field of the CBIR systems. In this case was selected simply DBs in order to show that shape is dominant features for the people but not for the computer. In addition, the problem of shape retrieval becomes more complex when extracted objects are corrupted by occlusion or noise as a result of the image segmentation process. We propose gaining more features from the images as color or textures features. It will be suitable to use color descriptors, texture descriptor for improving a precision of retrieval images. On the other hand, it is useful using semantic descriptors for instance SIF

T (Scale-invariant feature transform) descriptor for gaining more semantic in the retrieval process.

Acknowledgement

This paper has been supported by the VEGA grant agency no. 1/0570/10, “Algorithms research for automatic analysis of multimedia data”.

References

[1] SAGARMAY, D., YANCHUN, Z. *An overview of content-based image retrieval techniques*. Advanced Information Networking and Applications, 2004. AINA 2004. 18th International Conference on, vol.1, no., pp. 59-64, Vol.1, 2004.

[2] RAFIEE, G.; DLAY, S.S., Woo, W.L. *A review of content-based image retrieval*. Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on, vol., no., pp.775-779, 21-23 July 2010.

[3] ZHANG, D., LU, G. *A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures.* Journal of Visual Communication and Image Representation, No. 14 (1). (2003), pp. 41-60.

[4] ZHOU, X.S., HUANG, T. S. *CBIR: From low-level features to high-level semantics*. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 3974 (April 2000), pp. 426-431.

[5] BACH, J. R., FULLER, C., GUPTA, A., HAMPAPUR, A., HOROWITZ, SHU, C. J. *An Open Framework for Image Management.* in SPIE Conference, On Storage and Retrieval for Image and Video Databases San Jose, CA, 1996, pp. 76-87.

[6] DESELAERS, T., KEZSERS, D., NEY, H. *Features for Image Retrieval: An Experimental Comparison* Information Retrieval 2008, Vol. 11. Issue 2. Springer. pp. 77-107.

[7] LUX, M.: *MPEG-7 photo annotation and retrieval*. Proceedings of the seventeen ACM international conferences on Multimedia, pp. 925-926, 2009, Beijing, China.

[8] SMEATON, A. F., OVER, KRAAIJ, P. W. *Evaluation campaigns and trecvid.* Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (New York, NY, USA, 2006), ACM Press, pp. 321-330.

[9] ZHANG, D., LU, G. *Review of shape representation and description techniques*, Pattern Recognition, vol. 37, pp. 1-19, 2004.

[10] BENCO, M., HUDEC, R. *The advanced image segmentation techniques for broadly useful retrieval in large image database*. NSSS IX, Tatranske Zruby, Slovak Republic, pp. 40-44, May 2006, ISBN 978-80-8040-344-7.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Supply Chain Based on the RFID Technology

*Petra Fuchsíková,**Vladimír Kebo
*VŠB-Technical University of Ostrava, 17. listopadu, Ostrava - Poruba, 708 33, Czech Republic, {petra.fuchsikov}@vsb.cza
**VŠB-Technical University of Ostrava, 17. listopadu, Ostrava - Poruba, 708 33, Czech Republic, {vladimir.kebo}@vsb.cz

Abstract. The article puts mind to the RFID based supply chain, especially in domain of shipping containers, which is more often discussed problem of the supply chain area. Article points out different types of shipping container problems, such as management of containers, hygienic compliance and container entirety, also provides a possible solution, based on sometimes controversial RFID technology. Article provides a short introduction to the RFID technology and describes a laboratory demonstration of the RFID based supply chain. Traceability is considered as a one of the most important added value of the RFID technology, article points out that it is possible to achieve traceability independently on the technology. Article summarize added value of the RFID technology and points out the main problem of its massive expansion.

Keywords: RFID, supply chain, shipping containers, traceability, barcodes, ROI, laboratory demonstration

1. Introduction

Actual situation in many companies, which are using shipping containers, is a usage of barcodes for its identification, and that happen in better cases. In a production process, especially food production process, it is very common to use many types of shipping containers for storing and transportation of semi-finished products as a part of technological process. Unfortunately for companies, many problems with containers occur, for example problem with availability of a specific container types. It is difficult to determine where exactly and in what state the specific containers are. We could prevent this by implementing traceability solution for shipping containers, based on the RFID technology. With the traceability solution we can cover whole lifecycle of shipping containers in a company.

Fig. 1. Shipping conteiner types

2. RFID technology

The RFID acronym stands for the Radio Frequency Identification. The basis of the technology is radio waves. Radio waves are used for wire-less and contact-less reading of data, which are stored in RFID chip. RFID technology consists of four main parts: RFID tag, Reader, Antenna and Middleware.

2.1. RFID tag

Micro-chip, data carrier, connected to an *antenna*, which can have a different shapes and which is used for communication with a *reader*. We differentiate three categories of the RFID tags: active, passive and semi-passive (sometimes semi-active) tags. The main difference is simple, passive tag doesn't have a battery and it has to be charged by an antenna signal to transmit data. It transforms pulses to power supply and broadcast a response. On the other side active tag receives and broadcast data by itself, it doesn't need extra charge.

2.2. Reader

Electronic device, which is providing communication with RFI tags. It also reads the information, which is stored in the RFID tag. We can have readers in different shapes, as mobile devices or as fixed reading gates, for different usage.

2.3. Antenna

In most cases there are more concurrently working antennas plugged to the reader. They should be switched consequently, so that each antenna would behave to reader in the same way. In most cases are antennas circularly polarized for the reason to read as many RFID tags as possible.

2.4. Middleware

Specific software used for collecting and filtering data for enterprise systems. It is part of the technology where collected data are processed and transported to valid information system formats.

3. Shipping containers in RFID based supply chain

One of the biggest problems in companies with area extensive production is a lifecycle of shipping containers. We can gain information about actual state or movement of shipping containers by using of the RFID technology for its identification. By using of automatic identification, we can determine where exactly and in what amount shipping containers are, in the real time. We can also determine if shipping containers are ready to use or just react on instant to its actual absence.

If we consider food production process, it is very important to comply with strict hygienic regulations, which are applied to shipping containers. It is very common that shipping containers, which are processed by washer installation, are not flagged. After a time It is very difficult to say if the specific container was washed or not. Compliance with hygienic regulations is very important and it is another big problem for food production supply chain. By using the RFID technology, we can flag shipping containers and note timestamp of the finished washing process to the information systems. Then we will be able to determine which shipping containers are ready to use in production process and which are not. It is also obvious that the RFID tag, which is build in the shipping container, has to be resistant to the washer installation, because temperature is growing to 100°C.

Another problem of shipping containers is its entirety, especially when the container itself is a part of production process on production line. If we want to guarantee quality and smooth flow of the production process, we have to prevent that damaged containers or its broken parts will stuck the production line, also we have to verify that container, which is used during production process is sufficient for the specific production process. In case that company has big amount of shipping containers from different suppliers, also from different time periods, it is very difficult to say how often was each container used or even what is its full lifetime. RFID technology can provide precise information of the shipping container lifecycle, so production companies can easily beware usage of shipping containers over expiration period.

3.1. Demonstration of RFID based supply chain

Team from ILAB RFID from VŠB-Technical University of Ostrava created model of supply chain based on the RFID technology. Simulation of production environment was done on the model. Whole model consist of two parts, which represent two stocks, stock A and stock B as can be seen on the picture. Both stock entrances are equipped with RFID gates, readers and antennas. Roller conveyer is used as a transportation platform. Simulation demonstrate shipping process from stock A and receiving process in stock B. Data from RFID readers are transferred through middleware to the enterprise information systems, where the business intelligence take a place.

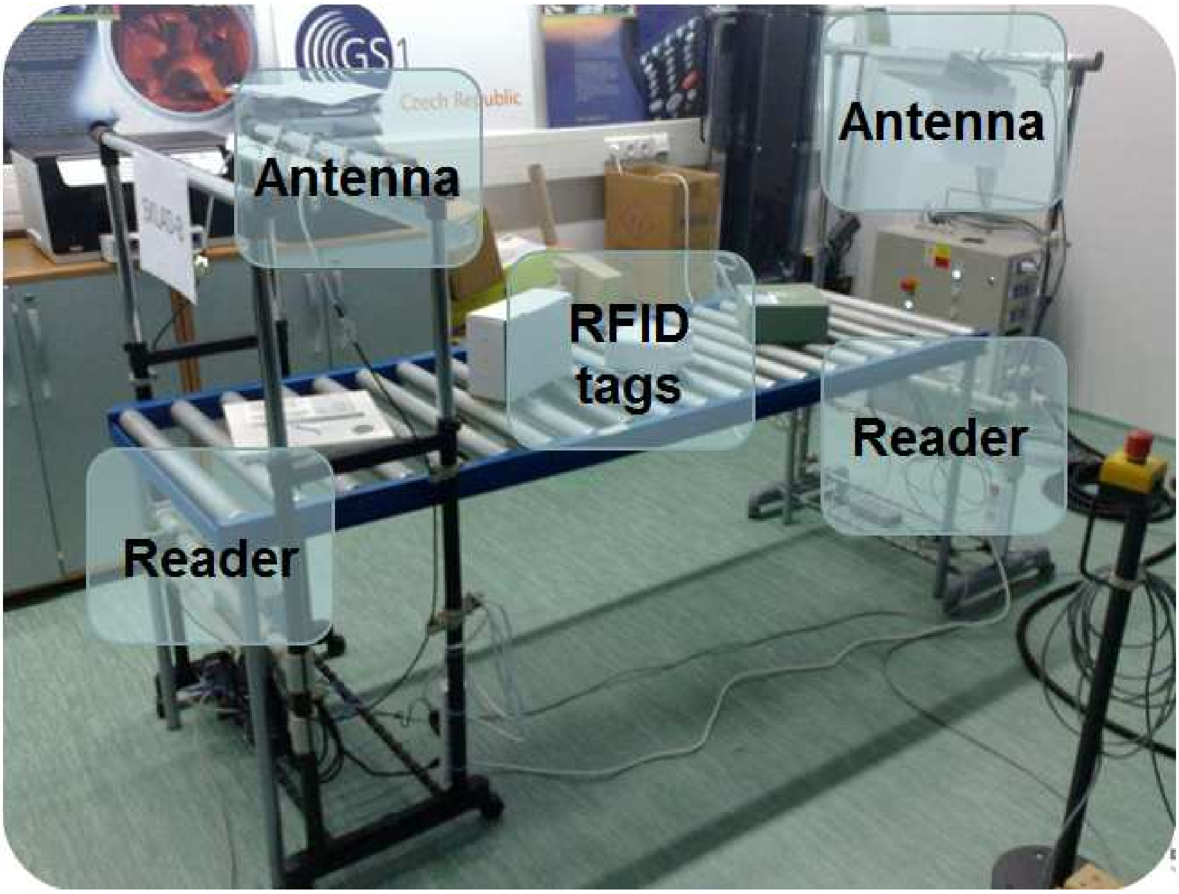


Fig. 2. Simulation of the RFID supply chain

4. Traceability

Traceability is defined as an ability to uniquely trace specific object and its lifecycle in process chain. Traceability is enabled through collection and evaluation of traced object data during its lifecycle.

Let’s introduce a model case of traceability usage. Consider simple supply chain, where products are transported. Supply chain consists of three parts: producer, distributor and retailer. From a product traceability point of view, we are interested in data such as *when*, *where* and *what* happened with a product. We collect data from different parts of the supply chain and subsequently we will gain data about whole product lifecycle. Very important is also data evaluation in the context of supply chain and its transfer to information systems.

If we will filter our interest on specific part of the supply chain such as food production part and choose shipping container as a main traceability object, then we can trace a container in different parts of production process and even in distribution process.

The basis of traceability is unique identification of an object, which is in RFID technology provided by EPC code. The RFID technology is not a condition of traceability achievement and you can easily use barcodes instead. However RFID technology has its added value such as effective automatization and reliability of reading process.

Complex RFID traceability solution can also trace a human factor during production process. It is possible to determine *which* employee done *what* activity and *when*. Production process can became more transparent and responsibility of employees can be seen.

5. Added value of the RFID technology

Implementation of the RFID technology to company processes is often conditioned by high initial cost investments. Question if and how to guarantee that the initial cost investments will return is answered by ROI analysis (Return-of-investments). ROI analysis will provide implementation possibilities and conditions and companies can then decide which types of solution to choose, depending on actual needs and costs.

Added values of RFID technology are summarized in next few points:

- Effective supply chain automatization and simplification
- Identification fraud prevention
- Higher endurance of RFID tags(temperatures, water etc.), compared to barcodes
- Effective and multiple tag readings, without direct visibility
- Reduction of employee costs
- Effective management of stocks, shipping containers etc., based on traceability solution
- The opposite of the added values is still a cost of the RFID technology, which is still the main decision maker.

6. Conclusion

It is very important to realize that the RFID technology is only a technical tool, which could or could not provide added value for companies. It is quite obvious that the barcode technology can provide sufficient and cheaper solution alternative in many ways, but the RFID technology is attracting attention of the companies more often.

Added value of the RFID technology is obvious, but costs are still too high for massive implementation. Positive aspect is that the costs are getting lower and lower.

One of the most discussed problems in supply chain is shipping container domain. It is also one of the positive examples of the RFID technology usage. RFID technology is very effectively solving the problem with effective manipulation and traceability of the containers. It is not providing only a modern and more expensive alternative for barcodes, but it is actually solving real problem.

References

[1] GLOVER, B., BHATT, H. *RFID Essentials.* First Edition. O'Reilly Media. 2006. 288p. ISBN: 978-0-596-00944-1

[2] MORADPOUR, S., BHUPTANI, M. *RFID Field Guide: Deploying Radio Frequency Identification Systems*. First Edition. Prentice Hall. 2005. 264p. ISBN: 978-0131853553

[3] CIHLÁŘOVÁ, Pavla. Úvod do EPC Global Network [online]. Prague : [s.n.], 2006 [cit. 2010-05-07]. Available on WWW: <http://www.rfid-epc.cz/download/prezen/RFIDWorkingGroup-EPCglobalNet.pdf>.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Wireless Sensor Network Positioning Based on MDS Method

*František Gaborik
*University of Zilina, Department of Telecommunications and Multimedia, FEE, Univerzitna 1,
010 26 Zilina, Slovakia, frantisek.gaborik@centrum.sk

Abstract. This document is focused on basic principles of localization in wireless sensor networks. Positioning is an important aspect in the field of wireless sensor network. The interest in wireless sensor networks localization grows further with advances in the wireless communication methods and sensing techniques and the consequent proliferation of all wireless sensor networks applications. MDS (Multi-Dimensional Scaling) methods play an important role in this field. This paper describes principle of MDS and investigates impact of various parameters on the positioning accuracy, e.g. the number of reference nodes or radio range of the nodes.

Keywords: localization, wireless sensor networks, MDS - multidimensional scaling.

1. Introduction

A Wireless Sensor Network (WSN) can be formed by hundreds of nodes which have limitations of memory, energy, and processing capacity. In this type of networks, one of the main problems is to get the localization of nodes.

The challenges in this hierarchy is: detecting the relevant quantities, monitoring and collecting the data, assessing and evaluating the information, formulating meaningful user displays, and performing decision-making and alarm functions are enormous. The information needed by smart environments is provided by distributed WSN, which are responsible for sensing as well as for the first stages of the processing hierarchy. The importance of sensor networks is highlighted by the number of recent funding initiatives. The complex focus on WSN where we can see connectivity between data acquisition network and data distribution network is shown in the following figure.

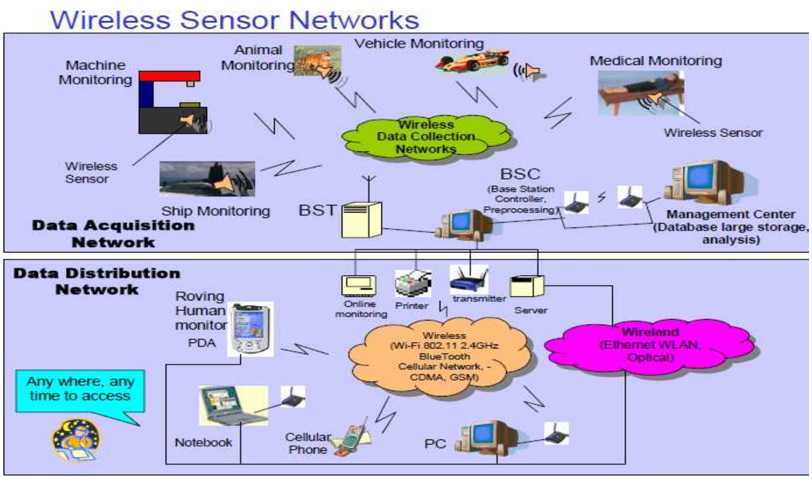


Fig. 1. Description the complexity of wireless sensor networks [3].

2. Positioning in Wireless Sensor networks

Location systems in WSNs use algorithms based on connectivity (range-free) and on methods with distance measurement considering complex techniques as: Angle of Arrival, Time of Arrival and Time Difference of Arrival [8]. Localization range-free algorithms using techniques based on connectivity require a large amount of reference nodes and a high node density because the propagation model for the radiofrequency signal is not perfectly spherical. The algorithms with distance measurement are very sensitive to errors during the measurements and require that the nodes are not collinear, consequently their application is difficult.

The power of use WSN is in the ability to deploy large codes of tiny nodes that assemble and configure themselves. Advantage is also adaptation mechanisms which can respond to changes in network topologies or can cause the network to shift between drastically different modes of operation.

Positioning process generally consists of three components:

- identification and data exchange,
- measurement and data acquisition,
- computation to derive location.

Simplified positioning process can be described: initiation of the process is based on the positioning request. The request is originated by location oriented application. The next step of the positioning process is to collect (measure) data that are used for position estimation. Calculating algorithm defines mobile device position estimation. The location information is forwarded to the target unit (e.g., localization server), that is handling and evaluating this information.

This paper is focused on MDS positioning. Concretely, we are interested in MDS-MAP method.

2.1. Localization method MDS-MAP

MDS is a method for visualizing dissimilarity data. There is existing central node - server in network, where multidimensional scaling MDS-MAP localization method used centralized access to localization points in wireless sensor networks. In this case when we don't know the latitude and longitude of a set of cities, we only know information about their distances.

Advantage of using MDS-MAP method is technical undemanding for nodes in networks and favourable price of nodes. Of course this method has disadvantages – complicated system of mathematics operations, precision of localization, less successful localizations nodes by weak signals. Using of these methods is special for hybrid localization methods.

The typical form of multidimensional scaling is to create a configuration of points in one, two, or three dimensions, whose inter-point distances are "close" to the original dissimilarities. The different variants of MDS use different criteria to define "close". These points represent the set of objects, and so a plot of the points can be used as a visual representation of their dissimilarities. Recently, MDS has been successfully applied to the problem of node localization in wireless sensor networks. [1]

Method MDS-MAP consists of these three steps:

- server compute shortest distances between nodes in the field of consideration,
- system is using classical multidimensional scaling to the distance matrix – paths between pairs of nodes,
- given adequate result nodes, matrix of results are mapped to their absolute results through a linear transformation.

Classical MDS requires to know the distance between every pair of nodes. The shortest path distance between two remote nodes can be provided by an estimate of the true Euclidean distance. This estimate is fine when the networks are dense or uniform, but is not good for very irregular ones. When the estimation is off, the result of classical MDS is not good [6].

For nodes localization can be used couple of types MDS methods. They are definite by similarity data – it can be qualitative (MDS is non-metric) or quantitative qualitative (MDS is metric).

Next classify of MDS model depends on numbers of similarity matrices, where is difference between MDS deterministic and MDS probabilistic. In any case, point of MDS is to find a configuration of nodes in network – multidimensional space some transformation to define distances between points in network. After determination of distances and numbers of matrices can transform relative maps to absolute maps through linear transformation and set position data of finding points.

For estimate of shortest distances between points in field can be used measurement of RSS - Received Signal Strength or ToA - Time of Arrival. In next step apply MDS to matrix of distances. Fig. 2 represents status matrix of distances based on MDS. The result of using MDS is relative map, where each node or point has some position Fig. 2, which is definite in matrix. This map is in different position then absolute map. All position data is by means of linear transformation transformed into absolute position data. In case of transformation in 2D space there is necessary know about minimum 3 points in our transformation area. On the basis of absolute map we will set geographic position of all nodes in the field.

	1	2	3	4	5	6	7	8	9	10
1	0	36,767	36,234	76,601	103,38	45,238	105,65	76,891	88,552	59,486
2	36,767	0	43,19	39,834	82,642	47,297	68,879	60,018	90,612	61,546
3	36,234	43,19	0	44,169	67,172	9,2453	71,846	40,657	52,371	23,276
4	76,601	39,834	44,169	0	42,808	39,858	29,046	20,184	54,084	40,884
5	103,38	82,642	67,172	42,808	0	58,144	31,96	26,719	28,172	43,896
6	45,238	47,297	9,2453	39,858	58,144	0	64,128	32,939	43,314	14,248
7	105,65	68,879	71,846	29,046	31,96	64,128	0	31,189	60,133	58,223
8	76,891	60,018	40,657	20,184	26,719	32,939	31,189	0	33,9	27,034
9	88,552	90,612	52,371	54,084	28,172	43,314	60,133	33,9	0	29,095
10	59,486	61,546	23,276	40,884	43,896	14,248	58,223	27,034	29,095	0

Fig. 2. Matrix of distances using for creating relative map [7].

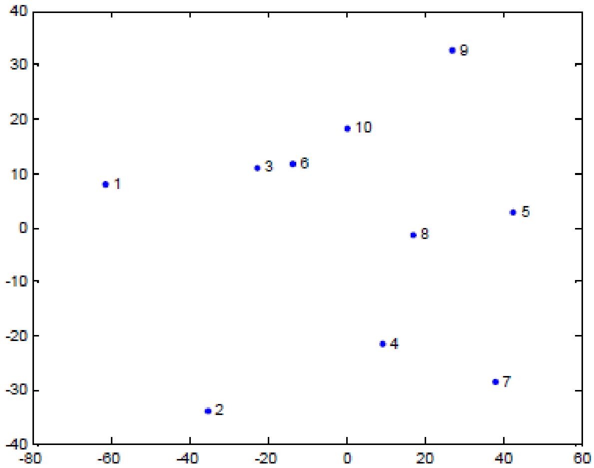


Fig. 3. Relative map based on MDS [7].

2.2. Simulation model

Simulation model takes into consideration a network of RNs and network of BN. Signals from particular nodes are independent to each other and all nodes in the model are deployed with omni-directional antenna.

Let $[x_i; y_i]^T$ i = 1,2…,5 are coordinates of RNs and $[x_r; y_r]^T$ i = 1,2…,30 are coordinates of BNs. Positions of particular nodes were generated by uniform distribution on the area 100 x 100 m. The results are based on 1000 independent runs. Radio channel is modeled as AWGN (Additive White Gaussian Noise) channel, i.e. it consists of two parts: path loss and white Gaussian noise [xxx]. The impact of channel properties is not investigated in these experiments, SNR = 9 dB is used in all experiments.

The positioning accuracy is compared by means of Root Mean Square Error *RMSE* and relative RMSE - Δ. They can be calculated as follows:

$$\Delta = \frac{RMSE}{R} = \frac{\sqrt{(x_r - x_{est})^2 + (y_r - y_{est})^2}}{R}.100\,[\%]. \qquad (1)$$

where $[x_r; y_r]$ are coordinates of true location; $[x_{est}; y_{est}]$ are coordinates of estimated location and R is radio range.

3. Simulation Results

This chapter analyzes simulation results obtained by MDS-MAP positioning method. Impact of following parameters on positioning accuracy is investigated:

- the number of reference nodes in the network,
- the radio range of particular nodes.

In the first experiment, an influence of the number of reference nodes in the sensor network on positioning accuracy was investigated. Radio range R of all nodes is 40 m.

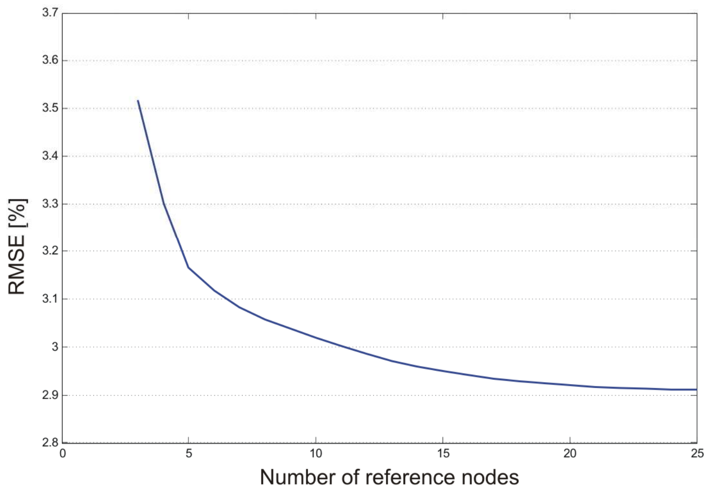


Fig. 4. RMSE vs. No. of RN.

According Fig. 4, ascending value of RNs involved in observed area means decreased RMSE. It causes the fact that the density of RNs is higher in the area. This fact is usable in wireless sensor networks where the big amount of RNs is assumed.

In the following simulation, impact of the radio range of reference nodes in the sensor network on positioning error was investigated. The number of all nodes was 30 and RNs was 5 in this case.

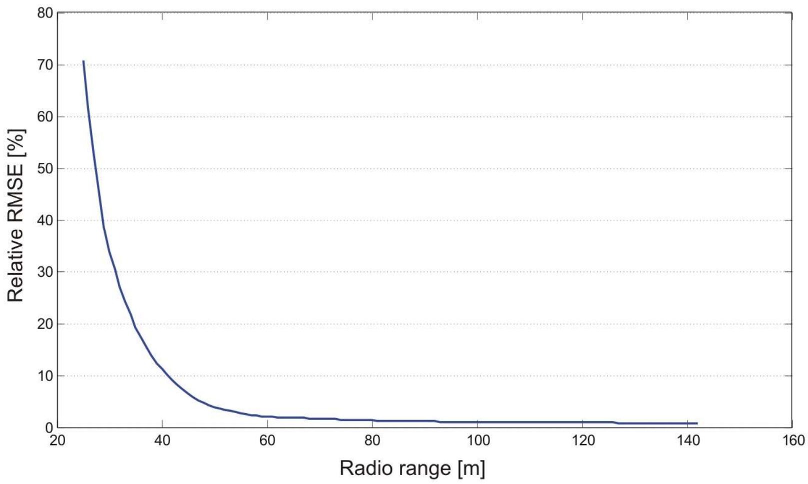


Fig. 5. Relative RMSE vs. R - radio range.

According Fig. 6, ascending value of radio range of involved nodes in observed area means decreased relative RMSE calculated by (1).

4. Conclusion

We discussed MDS-MAP methods for the location determination in wireless sensor networks. We analyzed the impact of following parameters, i.e. the number of all present reference nodes in the observed area and their radio range. The mentioned parameters were tested by means of extensive simulations.

According to the results, ascending value of RNs involved in observed area caused decreased RMSE. This fact can be utilized in huge wireless sensor networks. Ascending value of radio range of involved nodes in observed area means decreased relative RMSE.

Acknowledgement

This work has been supported by the Slovak VEGA grant agency, Project No. 1/0392/10 “The research of mobile nodes in wireless sensor networks”.

References

[1] GUOQIANG, M.,BARIS, F. Localization Algorithms and Strategies for Wireless Sensor Networks, University of Sydney, Australia 2009

[2] DOHERTY, L. Algorithms for Position and Data Recovery in Wireless Sensor Networks by, Submitted to the Department of Electrical Engineering and Computer Sciences, University of California.

[3] COOK, D.J., DAS, S.K., WILEY, J. *Wireless Sensor Networks* - ed., New York, 2004.

[4] Brida P., DUHA J., KRASNOVSKY M. *On the Accuracy of Weighted Proximity Based Localization in Wireless Sensor Networks*, in Proc. Personal Wireless Communications 2007, IFIP, 2007, Prague, Czech Republic, 2007

[5] BRIDA P., MACHAJ J., DUHA J., A Novel Optimizing Algorithm for DV based Positioning Methods in ad hoc Networks, ELEKTRONIKA IR ELEKTROTECHNIKA

[6] Dept YI SHANG, *Improved MDS-Based Localization*,. of Computer Science University of Missouri-Columbia, 2004

[7] DOBRUCKY, R. Localization in mobile wireless ad hoc networks, 2009, Diploma work

[8] MENDOZA, L. M., HUGO, V., ZARATE, S., DIAZ, A. P. An efficient algorithm for localization in wireless sensor networks based on internal array of nodes within cells. In ICPP Workshops, pages 405–412, 2005.

University of Žilina, Faculty of Electrical Engineering
Department of Telecommunications and Multimedia
Univerzitná 8215/1, 010 26 Žilina

- LABORATORY OF DIGITAL COMMUNICATIONS
- LABORATORY OF DIGITAL VIDEO PROCESSING
- LABORATORY OF ANALOG CIRCUITS AND SYSTEMS
- LABORATORY OF AUDIO SIGNAL PROCESSING
- LABORATORY OF OPTICAL COMMUNICATIONS
- LABORATORY OF MOBILE COMMUNICATIONS
- LABORATORY OF MULTIMEDIA TECHNOLOGIES

About Department

The Department of Telecommunications was founded in the year 1967. In first years, the activity of the Department was concentrated on circuits' theory and signals, digital and impulse techniques, transmission systems and switching systems, telecommunication networks and their reliability. Close connection between department and telecommunication practice was in mentioned time. Department activities were focused on modern trends in communication technologies. Several new laboratories were built. The education gradually increases in domain of software based subjects. Focus of department activity was extended to multimedia content research and creation. That was declared by modification of title department from January 1, 2008. At the present, Department of Telecommunications and Multimedia covers wide scope related to communication technologies and processes in education and research. Professional activities are connected to specialized laboratories. In the field of telecommunication technologies, attention is focused on problematic of communication networks, access technologies, convergence of network technologies with main activities oriented on quality of media services. In term of fixed networks, department has significant activities in the field of research and development technologies for broadband all optical networks. This field is closely associated with research activities at the Department of Physics. Wireless technologies are oriented on mobile and satellite communications, positioning systems as well as to DVB-x. The field of digital signal processing has significant research activities, especially from semantic analyses and annotation of audio and video signal perspective. Relatively new field at this department are multimedia technologies. In this case, main orientation is focused on technological part and also on part of multimedia content creation. It is represented by image composition, stylistics and work with multimedia data. The main target of this area is support of future multimedia services.
From the number of students' point of view, the department is one of the biggest departments at the Faculty of Electrical Engineering.

Contact:

University of Žilina
Faculty of Electrical Engineering
Department of Telecommunications and Multimedia
Univerzitná 8215/1 010 26 Žilina, Slovakia
tel. :+421 41 513 2201
fax. :+421 41 513 1520
e-mail : kt@fel.uniza.sk
web: www.kt.uniza.sk

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

MPEG-7 Audio Toolbox for MATLAB

*Bruno Gardlo, *Michal Kuba
*University of Transport, Faculty of Electrical Engineering, Department of Multimedia, Univerzitná 2, 01026 Žilina, Slovakia, {Bruno.Gardlo, Michal. Kuba}@fel.uniza.sk

Abstract. Audio content classification can be very helpful in various research fields, such as audio processing, audio-video or audio-visual quality estimation. Recently, there have been published many classification methods, using various audio features which translates temporal and frequency audio characteristics of the signal to either lower dimension, better computer-understandable format, or simply more practical domain for further signal processing. The most often used or standardized methods use audio features such as Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPC) or MPEG-7 audio features. Here presented, we will describe our tool for computing the aforementioned MPEG-7 audio features using the well know MATLAB environment.

Keywords: Audio Content Classification, MPEG-7, Matlab.

1. Introduction

At the time of writing this paper, there have been many tools for computing MPEG-7 features available for download. Further investigating and testing of these tools led to a decision, to program our own script, which will better suite our demands for processing and storing the extracted data from audio databases.

The MPEG-7 standard establishes the way the data should be stored and it prescribes to store it in XML compliant format. This often leads to usage of data mining algorithms for reading XML files and storing the data in other format needed for the classification method itself. This scenario is not often very useful for fast processing of the signal, since transcribing the MPEG-7 data files to format suitable for research-specific environment only extend the computational time.

Another problem is that MPEG-7 standard does not describe the specific algorithms for computing all the MPEG-7 audio features. This often leads to slightly different results, e.g. errors in estimation of the fundamental frequency could yield to differences in audio fundamental frequency based MPEG-7 features.

The paper is organized as following. Section 1 describes motivation for using MATLAB as our programming environment, Section 2 describes selection of MPEG-7 audio descriptors and at the end of the paper are described the conclusions.

1.1. Matlab Environment and Audio Signal Processing

The MATLAB environment is very efficient for computing audio properties, since at the time there are a lot of very well optimized toolboxes for speech processing. One can utilize the FFT functions, implemented windowing functions or tools for reading the audio files. Also the scripting characteristics of this program enable the user to rapidly develop scripts suitable for solving difficult and time consuming problems.

In MPEG-7 standard, many different audio features come as a function, which depends on other, or previously calculated MPEG-7 feature. Some Matlab toolboxes use different function for computing each of the audio features. Our scripting method on the other hand takes advantage of pre-computed values from former steps. This can significantly speed up computing process, since there is no need to compute some routines (e.g. Fast Fourier Transform) in every function all over again. Of course this lead to disadvantage that all MPEG-7 features must be computed for given audio files at one time. But since there is often demand to work with parameterized audio databases, not the audio files itself; this can be also considered as a great advantage.

2. MPEG-7 Audio Features

Describing the process for computing all of the MPEG-7 audio features is beyond the scope of this paper, and calculating process for many of them is strictly prescribed by the standard. That is because we will only describe features, which are programmed by non-strictly standardized algorithm, e.g. Audio Fundamental Frequency and features dependent on it. Nonetheless, authors want to stress up the fact, that the proposed toolbox computes all MPEG-7 audio features, and the results for some of them are depicted on following figures (Fig. 1 – Fig. 11).

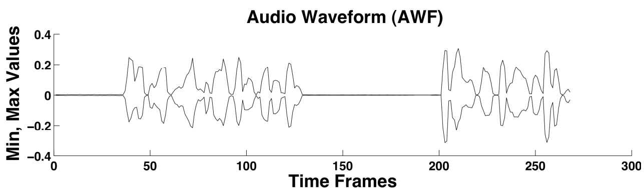


Fig. 1. Audio waveform for the speech signal.

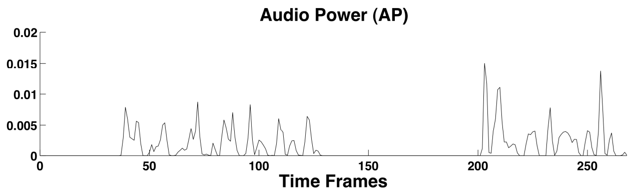


Fig. 2. Audio Power for the speech signal.

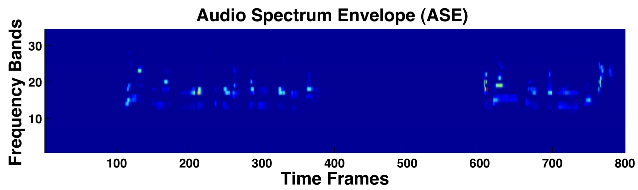


Fig. 3. Audio Spectrum Envelope representation for the speech signal.

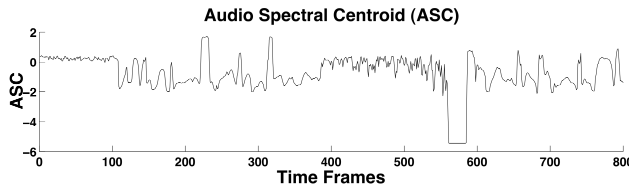


Fig. 4. Audio Spectral Centroid for the speech signal.

2.1. MPEG-7 Audio Fundamental Frequency (AFF) and AFF based MPEG-7 Features

The Audio Fundamental Frequency is one of the MPEG-7 features, but it is not defined by the standard how the fundamental frequency should be estimated. There exists several ways, how to estimate the pitch of the input signal [1].

In this work, we decided to use combination of two independent methods, because the proper estimation of the fundamental frequency is crucial for computation several other MPEG-7 audio features. First we use LPC prediction to calculate the LPC filter coefficients [2]. The LPC filter coefficients are used to predict the signal, and then the residual from the predicted signal and original signal are calculated. The autocorrelation from the residual and the original signal is computed in the next step. For the fundamental frequency estimation we simply combine the first lags from the autocorrelation of the residual and autocorrelation of the original signal. The estimated value is then stored, and will be used in the next steps.

As the second method for the AFF estimation we use the smoothed spectrum. Smoothed spectrum can be defined as the simplified power spectrum of the power spectrum. The simplest method how to smooth power coefficient is simply average several successive frames. In the smoothed spectrum we are estimating the peaks in power spectrum, which most probably represent harmonic peaks. From the founded harmonic peaks we estimate fundamental frequency, and the value is stored for the next computation.

In the final step, we simply linearly combine the AFF values from the first and second method, and the final value is stored as the AFF value for the given time frame. The result from this method - the AFF for the speech signal is depicted on the Fig. 7.

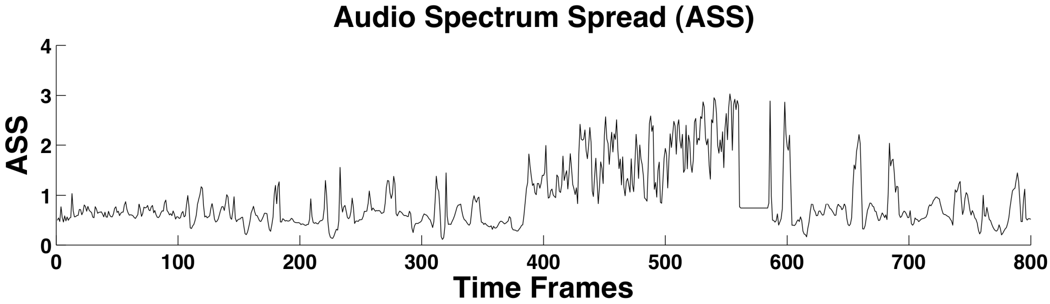


Fig. 5. Audio Spectrum Spread for the speech signal.

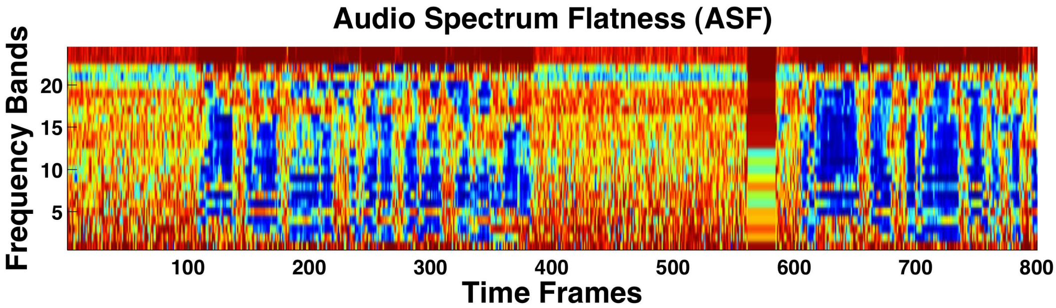


Fig. 6. Audio Spectrum Flatness for the speech signal.

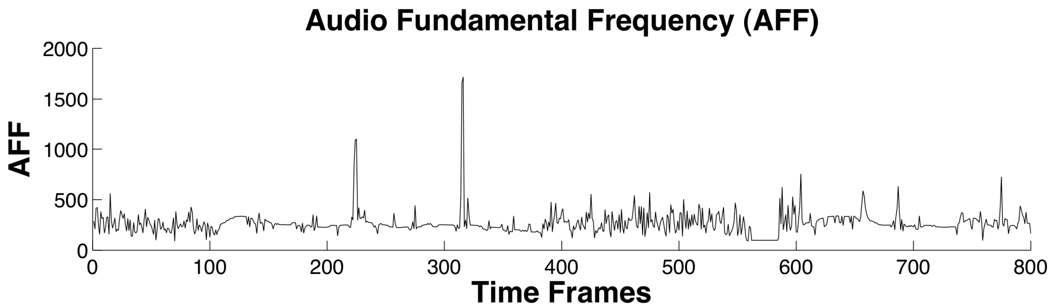


Fig. 7. Audio Fundamental Frequency for the speech signal.

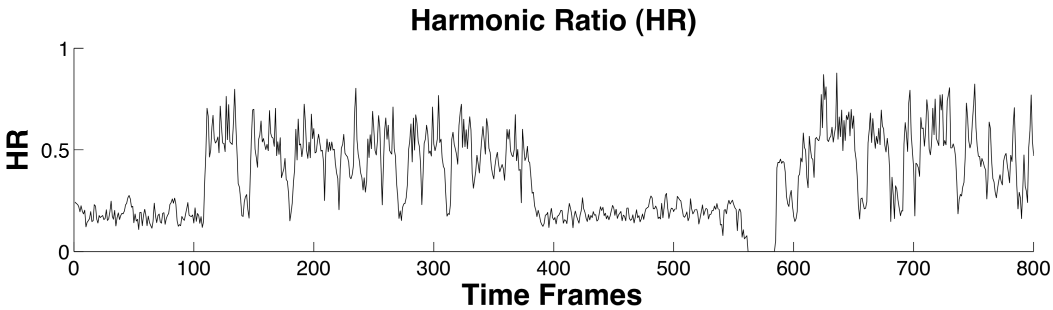


Fig. 8. Harmonic Ratio for the speech signal.

2.1.1 MPEG-7 Harmonic Spectral Centroid

For the estimation of the AFF we used estimation of the harmonic peaks of the signal. These peaks where stored, and now are used for calculation of the Harmonic Spectral Centroid (HSS).

The MPEG-7 HSS is defined as the average, over the duration of the signal, of the amplitude-weighted mean (on a linear scale) of the harmonic peaks of the spectrum [3]. The local expression $LHSC_l$ (i.e. for a given frame l) of the HSC is:

$$LHSC_l = \frac{\sum_{h=1}^{N_H} (f_{h,l} A_{h,l})}{\sum_{h=1}^{N_H} A_{h,l}}, \tag{1}$$

where $f_{h,l}$ and $A_{h,l}$ are respectively the frequency and the amplitude of the h^{th} harmonic peak estimated within the l^{th} frame of the signal, and N_H is the number of harmonic that is taken into account. Fig. 9 depicts MPEG-7 harmonic spectral centroid of the speech signal.

2.1.2 MPEG-7 Harmonic Spectral Deviation

The MPEG-7 Harmonic Spectral Deviation measures the deviation of the harmonic peaks from the envelopes of the local spectra. Within the l^{th} frame of the signal, where N_H harmonic peaks have been detected, the spectral envelope $SE_{h,l}$ is coarsely estimated by interpolating adjacent harmonic peak amplitudes $A_{h,l}$ as follows:

$$SE_{h,l} = \begin{cases} 1/2(A_{h,l} + A_{h+1,l}) & \text{if } h = 1 \\ 1/3(A_{h-1,l} + A_{h,l} + A_{h+1,l}) & \text{if } 2 \le h \le N_H - 1. \\ 1/2(A_{h-1,l} + A_{h,l}) & \text{if } h = N_H \end{cases} \tag{2}$$

$$LHSD_l = \frac{\sum_{h=1}^{N_H} |\log_{10}(A_{h,l}) - \log_{10}(SE_{h,l})|}{\sum_{h=1}^{N_H} \log_{10}(A_{h,l})} \tag{3}$$

On the Fig. 10 is depicted local deviation measure of the speech signal, which can be computed for the each frame:

2.1.3 MPEG-7 Harmonic Spectral Spread

The MPEG-7 Harmonic Spectral Spread is a measure of the average spectrum spread in relation to the HSC. At the frame level, it is defined as the power-weighted RMS deviation from the $LHSC_l$ defined in (3). The local spread value is normalized by the $LHSC_l$ as

$$LHSS_l = \frac{1}{LHSC_l} \sqrt{\frac{\sum_{h=1}^{N_H} [(F_{h,l} - LHSC_l)^2 A_{h,l}^2]}{\sum_{h=1}^{N_H} A_{h,l}^2}} \tag{4}$$

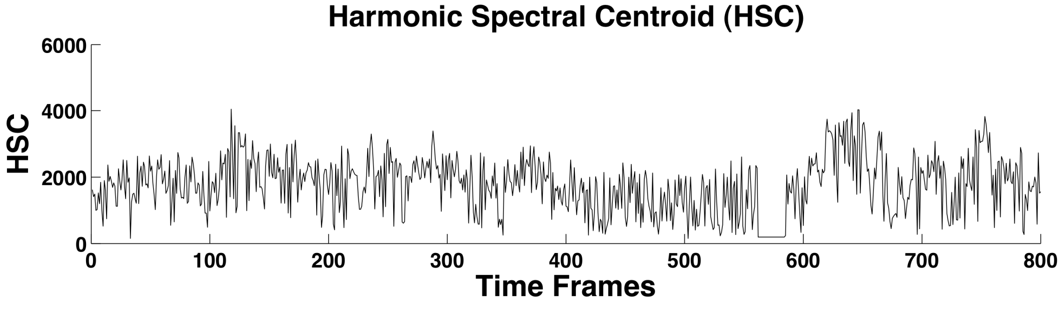


Fig. 9. Harmonic Spectral Centroid for the speech signal.

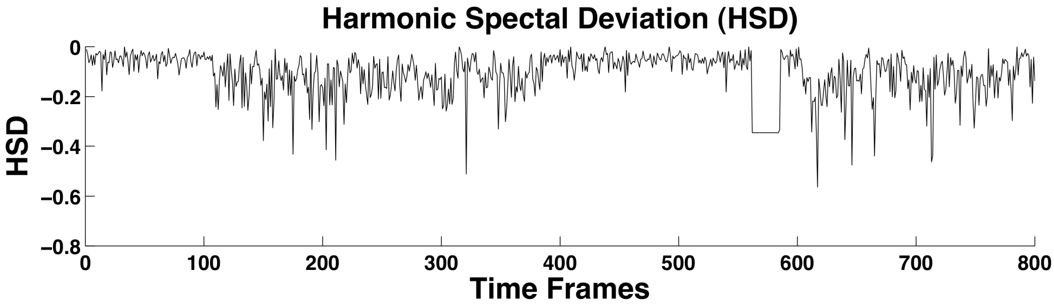


Fig. 10. Harmonic Spectral Deviation for the speech signal.

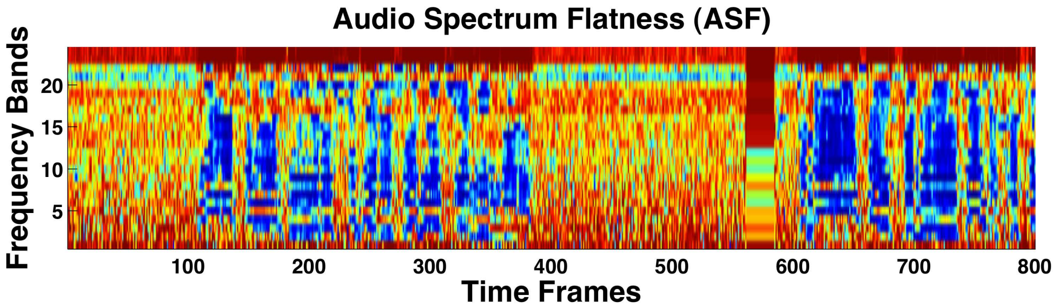


Fig.11. Audio Spectrum Flatness for the speech signal.

3 Conclusions

Matlab toolbox for computing MPEG-7 audio features is presented. The toolbox is easy to set, and uses advantages of the computing all of the MPEG-7 audio features in a single thread. It can be used for fast parameterizing of the speech corpuses, or audio databases. All the algorithms used for developing the script are MPEG-7 standard compliant, but instead of storing the results as XML files, the results are stored as binary Matlab files, which are more suitable for use in our research scenario.

References

[1] BENESTY, J., SONDHI, M.M., HUANG, Y. *Springer Handbook of Speech Processing.* Berlin, Springer, 2008.

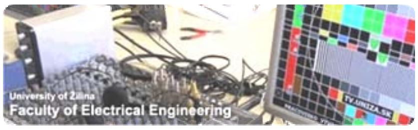
[2] MAKHOUL, J. Linear *prediction: A tutorial review*. Proceedings of the IEEE, vol.63, no.4, pp. 561- 580, April 1975.

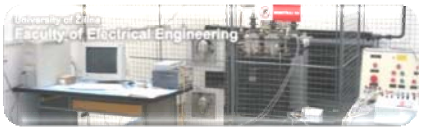
[3] KIM, H.G., MOREAU, N., SIKORA, T. *MPEG-7 Audio and Beyond. Audio Content Indexing and Retrieval.* London, Wiley, 2005.

University of Žilina
Faculty of Electrical Engineering
Univerzitná 1, 010 26 Žilina

About Faculty

The Faculty of Electrical Engineering was founded in 1953 as one of the basic faculties of the Railway College in Prague, and was re-established in 1992.The departments representing the basis of the Faculty exist for several decades. They continuously created with their activity the main trend of the Faculty both in pedagogic and scientific-research area. Original activities of these departments were oriented mainly for technical aspects of classical transport and later increased task of Communications in the form of technical operation of telecommunications. Present traditional activity is enriched with some changes characteristic for the development of science and technology in recent period. It concerns especially areas of information systems, modern telecommunication technologies, development of power semiconductive systems and modern control of electric networks. Study of interdisciplinary branches is also increasing, namely: mechatronics, telecommunication management, biomedicine engineering and specialization: oriented on information systems in electroenergetics.





Contact:

University of Žilina
Faculty of Electrical Engineering
Univerzitná 8215/1 010 26 Žilina,Slovakia
tel. : +42141 513 2051
tel.: +421 41 513 1111
fax. : +421 41 513 1515
e-mail : dean.office@fel.uniza.sk
web: www.fel.uniza.sk

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

The Concept of Elements Web Site as an Effective Tool for Internet Marketing

[*]Eva Gavláková, [*]Jiří Tengler
[*] University of Žilina, Faculty of Operation and Economics of Transport and Communications, Department of communications, Univerzitna 2, 01026 Žilina, Slovakia,
{eva.gavlakova, jiri.tengler} @fpedas.uniza.sk

Abstract. This paper focuses on the marketing elements for any web site. The internet also is opening up new marketing opportunities. Good web page design for an e-commerce business must sutisfy a number of criteria.

Keywords: marketing, web site, business, AIDA, web elements, design.

1. Introduction

The Internet was not originally created for personal computers, even though they have become its main delivery system. Businesses develop communication strategies and goals before they decide what and how to communicate to their audiences. Wireless devices, such as cellular phones and personal digital assistants (PDAs), also provide Access to the Internet but do not allow for rich complex content. Current cellular technology limits both the size of the viewing screen and the bandwidth available to carry data.

A key part of Internet marketing is a website. Developing the capability to create and maintain an effective online presence through a website. Effective in this case means that website and related communications must deliver relevance to its audience, whether this is trough new content for a portal, product and service information for a B2B site, or relevant products and offers for an e-commerce site. At the same time, “effective” means the website must deliver results for the company.

If you were in charge of the communication strategy for a business, you would need to consider the most cost-effective way to deliver a message to the audience. This becomes more complicated when you have multiple audiences and media from which to choose. Tie Internet offers a variety of new, cost-effective means of communicating.

A business communication goal should always be set based on the information needs of the market being served. Some examples of communication goals that businesses set include supporting the sales process, informing customers of products, reinforcing brand images, and recruiting employees. Often, a web site will be designed to reach all of these goals, allowing different audiences to find different information. Brochure sites, relationship sites, e-commerce sites, and web portals are each designed to reach specific communication goals.

Web site creator should be understand who the customer is, how they use the channel to shop, and understand how the marketplace works in that category. Very important in their works is online customer experience, which is combination of rational and emotional factors of using a company´s online services that influences customer´s perceptions of a brand online. You need continuous research, feedback and usability testing to continue to monitor and evolve the customer experience online.

The Christodoulides et al., provide an excellent framework which can be applied to assess and benchmark the quality of brand experience for different for different types of website. This analysis was performed across these five dimensions of brand equity assessed by asking the questions below:

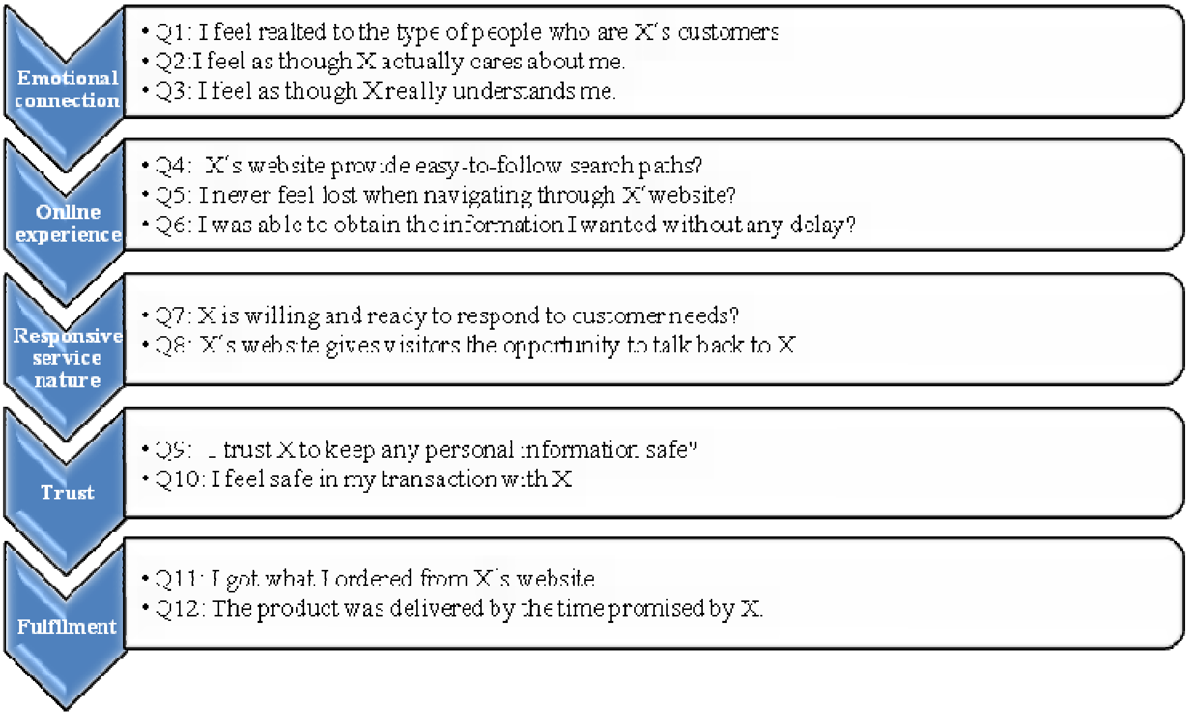


Fig.1 The five dimensions of brand equity assessed by asking the questions below.

In this figure we can see that many of the rational and emotional values are important to any website. It incorporates many of the factors, especially price and promotions which form web merchandising. Goals of web merchandising are to maximize the sales potential of an online store for each visitor, which means connecting the right products, with the right offer to the right visitor. The online store is part of a broader experience including online and offline advertising, in-store visits, customer service and delivery.[1][2]

When you start the web design process, it is useful to evaluate competitor´s sites and determine how your site can be more appealing. A design would be logical and consistent and, at the same time, must be enhance the image of the company, it must also allow the visitor to easily navigate the site. The home page is the first part of design process. There are a number of design considerations in developing good web pages:

- **Accessibility** – Access should be easy for Internet users with all tips of computer capabilities. An approach to site design intended to accommodate site usage using different browsers and settings. If your site has flashy components that require sophisticated plug-ins, you should have an alternative link for users whose systems are not advanced so that they may view the same information in a simpler format. There should also be links to sites where users can download the needed plug-ins.
- **Advertising** – The type and number of ads on a page should fit the market´s needs and should not overpower the main content for which the viewer is visiting.
- **Alignment** – Varying alignments will make the web site look messy. It is important to choose a style of alignment and stick with it.
- **Consistency and Repetition** - Repeat certain elements such as navigation buttons and company logo or basic text should be the same font, size and color on every page.
- **Content** – Type of information and the organization on the site should be based on the needs of the target market. Especially care that the content is clearly written and free of grammar, punctuation, and spelling errors.
- **Contrast** – Contrasting elements draw the viewer´s eye into the page, create interest, and show hierarchy. Contrast creates a focal point, emphasizing what is most important. Text should stand out clearly from any background colors or graphics.

- **Customization –** Delivering personalized content is best.
- **Feedback –** Contact information should be visible.
- **Links –** Links to other content should be appropriate, topical, and of interest to the target market.
- **Navigation –** Navigation would be logical and user friendly.
- **Ordering –** Purchasing should be simple, convenient, and secure.
- **Privacy –** There should be an affective privacy policy that is easy to find on the site.
- **Searches –** Search tools should be accurate and easy to use.
- **Speed –** Fast-loading graphics and text are important. Graphics originally created for high-quality print do not need the same resolution for good viewing on the Web. Instead, make smaller pages that load quickly and link to each other in a logical fashion.
- **Updates –** The site´s content and format should be updated frequently. Freshen the look of the site regularly but not drastically enough to confuse the regular visitor.[2] [3]

2. A structure of website

Successful business Web sites don't happen by accident. Companies with a sophisticated Web marketing staff deliberately place every item in a specific place on a page, think through each headline, and consider every graphic element and photograph for impact. They don't spend hundreds of thousands of dollars on the mere chance that a site will achieve its marketing and sales objectives. Those factors determine how a site looks on the screen and how visitors navigate through it, which is often called the *look and feel* of a site. [4]
What is important for the structure of web site?

The most important criterion for a successful business Web site is whether it accomplishes its objective. Your site doesn't have to be beautiful or cutting edge as long as it ultimately has a positive impact on your bottom line. The second most important criterion is how well the site works from the users' perspectives. The easier you make it for users to achieve what they want — whether buying a product, obtaining information, or connecting with others — the more likely your site is to succeed.

2.1. Using AIDA to guide visitors toward specific actions.

Direct marketing techniques are highly useful for coaxing users into taking actions you want them to take. The four standard steps of direct marketing (known as AIDA „Attention, Interest, Desire, Action") apply to the structure of Web sites:

- **Attention:** Get viewers' attention by using graphics, a headline that grabs, and a benefits-based lead. You have four seconds to convince them they'll find something of value on the site.
- **Interest:** Build interest with site design and navigation. Include intriguing options that pull people to additional pages on your Web site, giving you time and space to expose visitors to your products, services, and benefits.
- **Desire:** Create desire and a sense of urgency as visitors move themselves toward taking an action. If you think visitors are almost ready to buy, post a reminder to Buy Now for Free Shipping. If you think they're doing research, remind them to Bookmark This Page or Tell a Friend. It's tricky, but you can prod users to do what you want them to do. Use whatever content tools that will build desire in your audience, from marketing copy, photography, and special offers, to online activities or onsite entertainment.
- **Action:** Right from the beginning, make it obvious what you want visitors to do, whether it's to buy online, make a call, send an e-mail, or sign up for a newsletter. Then ensure visitors that it's extremely easy for them to take those actions. [3][4]

2.2. Applying marketing communications principles to your design

Marketing communications integrates marketing and sales principles with graphic design to achieve business objectives. It acknowledges that the presentation of information affects emotional response and thus influences buying decisions. Designers ask about your target audiences to be sure to select or create appropriate design elements. While essential for any type of sales collateral or packaging, marketing communications is particularly critical because of the short window for grabbing attention on the Web. Experienced Web designers intuitively adjust the font style, graphic style, colors, images, and white space to have a positive impact on your marketing process while reinforcing a brand.

In general exist 3 basic groups of element for success web site.

a) **Clean and Strong Coding**
HTML is the foundation of the web. And following closely is the CSS markup. Both the elements form the basic coding of the website and decide the functionality of the same. HTML and CSS markup form the basic structure of a website and their quality eventually determines the quality of the website. It is very easy to get carried away and incorporate complex functionalities in your website in order to show off your technological expertise. However, often it happens that the interface becomes so complex that users find it difficult to browse through and navigate easily.

b) **User-Friendly Interface**
Website interfaces should be kept as simple as can be. Complex features often tend to confuse the visitors and distract their attention. They only add up to the visual appearance and do absolutely nothing to enhance the functionality of the website. Fancy interfaces distract the purpose of the website. Therefore, it is best to build a simple and user-friendly website interface that serves to facilitate smooth performance of the website.

c) **Elements and Features on the Website**
Focus on creating a website that is clean, has a strong interface and can be easily understood. Instead of including cool features and jazzy site elements, build a website that would please your audiences and make them visit again and again. [4]

3. Conclusion

Today, web sites are the public face of companies. A business´s web site is often first place that an individual will look to find information about the company and its products, services, and employment opportunities. A single web site must be designed to meet all visitor´s needs and to grab the attention of those merely surfing.

Web page developers must be concerned with design and content and how they display in a variety of platforms. Good web page design for an e-commerce business must satisfy a number of criteria. As web moves toward broadband, new formats of content will be developed. A design should be logical and consistent and, at the same time, must enhance the image of the company. The design must also allow the visitor to easily navigate the site.

References

[1] CHAFFEY, D., *Internet Marketing, Strategy, Implementation and Practice.* 4th edition, Pearson education Limited, ISBN-13: 978-0-273-71740-9, 2009

[2] DIAMOND, S. *Web Marketing for Small Businesses.* Sourcebooks, INC. ISBN-13: 978-1-4022-1926-9, 2008

[3] KLEINDL, B. A., BURROW, J. *E-commerce Marketing.* South-Western, Part of Thomson Corporation. ISBN: 0-538-43808-8, 2005

[4] ZIMMERMAN, J. *Web marketing for Dummies*. Wiley publishing, Inc. ISBN: 978-0-470-04982-2, 2007

UNIVERSITY OF ŽILINA
THE FACULTY OF OPERATION AND ECONOMICS OF TRANSPORT AND COMMUNICATIONS

Department of communications

Conference DONT 2011 in November 2011

POŠTA, TELEKOMUNIKÁCIE A ELEKTRONICKÝ OBCHOD

Elektronický vedecký časopis Katedry spojov,
Fakulty prevádzky a ekonomiky dopravy a spojov,
Žilinskej univerzity v Žiline

Cieľom časopisu je publikovanie pôvodných vedeckých statí, odborných článkov a diskusných príspevkov zameraných na problematiku poštových a telekomunikačných podnikov, ako i na prudko sa rozvíjajúcu oblasť elektronického obchodovania v domácom, ale i európskom kontexte.

★★★★★★★★★

ISSN 1336-8281

Elektronický časopis vychádza polročne

Bližšie informácie na adrese:
http://ks.uniza.sk/casopis

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

RUM (Route Under MAC) in Real Conditions

[*]Peter Goga
[*]University of Transport, Faculty of Computer Engineering, Department of Technical Cybernetics, Univerzitna 2, 01026 Žilina, Slovakia, peter.goga@fri.uniza.sk

Abstract. This article deals about wireless network stack called Route Under Mac (RUM in later text), which is very usable for wireless sensor networks based on Zigbee wireless technology on chips from Atmel. It brings some good features to the conception of these networks, as IPv6, routing, easy network building. It has very good results in highly movable networks – such as VANET (Vehicle Ad-hoc Networks) and so on. This article will bring briefly information about this stack and will bring some tests, modifications and examples, which are developed to real use. In the end of article there is illustrated some real examples, which are done and working in real conditions.

Keywords: Zigbee, RUM, Routing, Wireless Sensor Network (WSN), 802.15.4, LR-WPAN, 6LoWPAN, IPv6

1. Introduction

Wireless sensor networks are the field of intensive research. We do not need to discuss, why is it so, because no need of cabling and possibility of moving sensors at small size and small energy consumption is very big point of this technology. Zigbee networks are developed to conform the requirements of idea mentioned before. Zigbee is part of IEEE 802.15.4 LR-WPAN specification. Dealing about conception of Zigbee networks is not goal of this paper, but in shortage, you have one coordinator (or more, depending on the software site), none or more routers (which are routing the data from end nodes to coordinator) and end nodes, which are containing the sensors and other electronic. If you get the hardware interface, you get the physical layer of communication, and software above is on your choice. There is many implementations of higher layers of network structure. Some of them are for free use (and many times it is from the same company which has developed the physical hardware), some are not free, but have some better features. One of implementation of higher layers is RUM – Route Under MAC (later only RUM) from Atmel company. This stack runs on microcontrollers from Atmel company. In my examples I use the modules from Meshnetics, these modules are described later. Within the stack, you can program any other function, which will meet your expectations.

2. RUM as Zigbee stack

Route Under MAC was developed for Zigbee low rate wireless personal networks. In real conditions this networks are at local, not wide area with randomly dispersed wireless nodes. Typical applications need to have small devices with low energy consumption. RUM meets these requirements and adds some other good features. For example, providing IPv6 address and functionality directly to end nodes (so the end node has fully reachable IPv6 address) enables high interoperability with existing infrastructure without any other investing and modifying. The main highlights of this technology are:

1. **Small object size:** the minimal compiled project with basic functionality of network, supporting sending information across the network, has only 6kB for AVR end node.
2. **Self-forming network:** fully automatically solved joining the network, based on the signal quality and hop count.

3. **Self-healing network:** network automatically detects transmit failures and automatically deals about it.
4. **Multi-hop routing:** end node can use routers in the communication path
5. **Source code:** you have all source code included, for free
6. **Very configurable and portable:** it is highly configurable and portable, based on applications

As mentioned before, the network consists from 3 types of devices: Coordinator, Router and End Device. The network is in tree-form. Every device can be in general the same – difference is only in software built in. In *figure 1* is example network, with all types of devices, and a gateway, which is serving the results from sensor network to real clients (as illustrated on the right of picture).

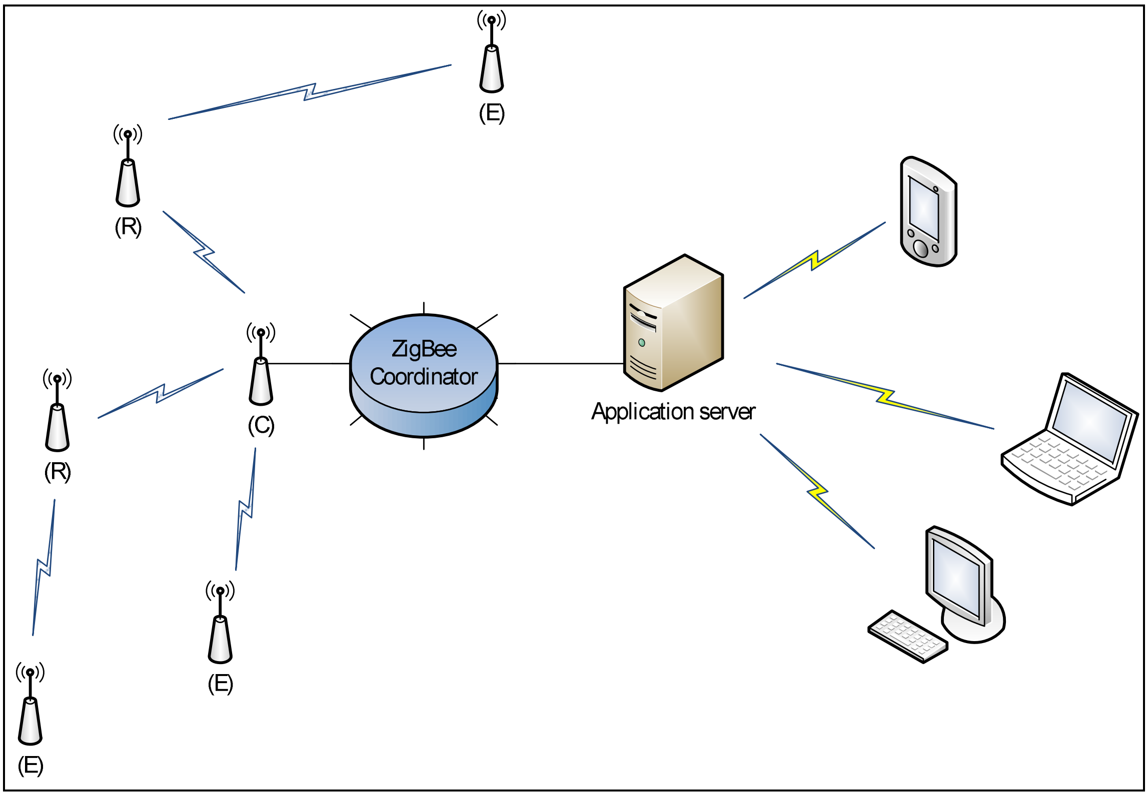


Fig. 1. Example network structure in Zigbee with RUM stack

RUM implements idea of 6loWPANs, which are low rate wireless personal networks using IPv6 as primary communicating protocol. In the real life when you need to transmit data, the base approach was to use some function and send it, and in the worst case you need to make sure if all information was good transmitted.

When you use 6loWPAN, you never mind about this, you simply create a socket to some other server in the IPv6 network (for example, somewhere on the internet, if you have IPv6-enabled internet connection), and simply via socket you send data. All things around will care the stack and you can be sure, that all is all right.

Maybe there will be question, why to use IPv6 and not IPv4. IPv6 is more simple and less resource-eating as IPv4, especially on the new low-latency and highly resistive networks. And it is the future, as we know, that IPv4 address space is fully occupied.

3. RUM routing solution

In RUM, there is goal to have simplest code and have low request to hardware features. Complete routing algorithm is based on simple idea. When the End Device needs to connect to the network, it scans channel for coordinator / router beacons. After this, the end device sends his information (2-byte short node address, 8-byte MAC address, 2-byte destination short address, command, type of packet and other) to the coordinator or router. The choice between router and coordinator is based on LQI (Link Quality Indication), which is number describing the quality of link (in hop count on one side and in signal strength on the other site). Packet goes to coordinator and the coordinator replies with the unique short address for the source node. When IPv6 compiled in, then the 2-byte short address is the last 2-byte of the full 16-byte IPv6 address. IPv6 coordinator deals about it and adds other 14 bytes, so everything goes well.

Coordinator has table of all nodes in this network, with information about all nodes. Each node in network has information about his parent. It means, that if the node is going to transmit any information, it simply sends the packet to the parent. Parent knows his other parent, and it forwards the packet to the next parent. If the parent is not reachable, it simply repeats last message for some time (it is fully configurable) and then makes re-associating to network. And this process repeats.

4. RUM in real conditions

Fig. 2. Frauenkirche Dresden

In real conditions we need to have to do some modifications. With colleagues from Romania we have implemented some better functionality into the routing stack. Especially, we targeted to have whole network sleeping as much time as possible. We were implementing this network into Frauenkirche in Dresden (fig. 2).

“Built in the 18th century, the church was destroyed in the firebombing of Dresden during World War II. It has been reconstructed as a landmark symbol of reconciliation between former warring enemies. The reconstruction of its exterior was completed in 2004, its interior in 2005 and, after 13 years of rebuilding, the church was reconsecrated on 30 October 2005 with festive services lasting through the Protestant observance of Reformation Day on 31 October” [1].

So this is a new building. There they have problem with climate and humidity, so they must do something to avoid these problems.

So our goal was to measure humidity and temperature in this church and do wireless communication without cables, because of not good looking cables. We needed to use batteries and therefore we must use as low energy consumption as possible.

We have done most modifications to real use. We deal about sleep-synchronized network, where each node knows when to get awake. In the first steps of our project we use only coordinator and end nodes in the network. We used model, where the end node sleeps all time and every minute wakes up, measures data, connects to network and sends this data to server.

Next step was to deal with routers. We solved this problem with the help of application server, which in answers for the data send the time, how long will the node sleep. With some other modifications (timeouts, message separation, IPv6 communication with server,…) the network is functioning and measuring real data.

Fig. 3. Zigbee Gateway for Dresden Frauenkirche

As the gateway we developed in cooperation with one local company in Dresden professional device, which they use as commercial product (fig.3). The final end nodes were in part of our stay not developed yet, but we have checked the functionality with our nodes (fig.4.).

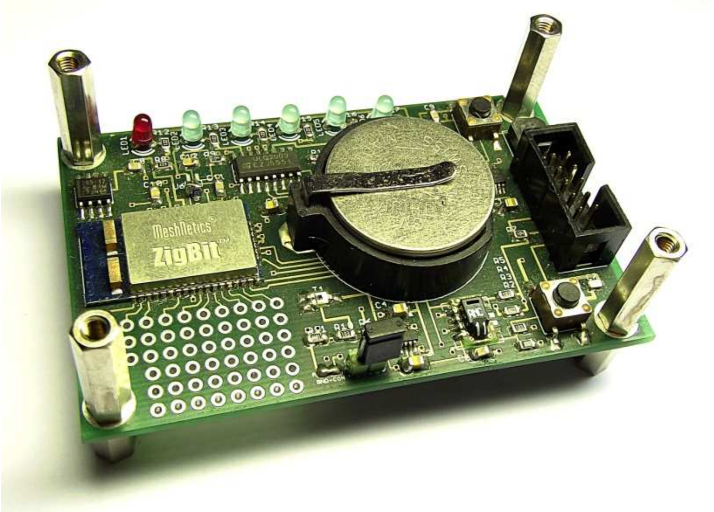


Fig. 4. Zigbee sensor node for Dresden Frauenkirche

5. Conclusion

We have developed and implemented full solution for measuring temperature and humidity in Frauenkirche Dresden. Next step is the Atmospheric Balloon http://universum.uniza.sk where we make picture and data transmission from balloon to the earth.

References

[1] Wikipedia page about Dresden Frauenkirche, ONLINE http://en.wikipedia.org/wiki/Dresden_Frauenkirche

[2] Route Under Mac datasheet, Atmel Corporation, 2009, ONLINE http://www.atmel.com/dyn/resources/prod_documents/doc8240.pdf

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Audio Events and Keywords Spotting Based on HMMs and Viterbi's Decoder

*Róbert Gubka, *Michal Kuba
*University of Žilina, Faculty of Electrical Engineering, Department of Telecommunication and Multimedia, Univerzitna 1, 01026 Žilina, Slovakia, robert.gubka@fel.uniza.sk

Abstract. In task of keyword spotting and acoustic events detection, the main accent is putted on accuracy and processing time of the system. In the paper, we present algorithm for audio events and keywords retrieval, based on statistical modeling using Hidden Markov Models and Viterbi's decoder. The paper offer basic introduction to HMM's and describe the process of model parameters and structure selection. Performance of proposed decoding system was evaluated on TIMIT database in task of keywords spotting in continuous speech.

Keywords: audio modeling, pattern retrieval, sequence decoding.

1. Introduction

Algorithms for keyword spotting and acoustic events detection can be applied in many different systems, where a large amount of audio documents must be treated for performing audio stream segmentation or classification.

Statistical modeling-based methods offer powerful tool for digital audio restoration, data mining and pattern searching [1]. Hidden Markov Models (HMM) proves to be a very effective method in tasks of speech recognition and keyword detection. However, such systems are highly language-dependent and conditioned by selecting proper speech fundamentals, which is followed by demanding training process.

Our effort was to create system which would not be pointed on speech, but can also search for generic audio patterns. In this paper, the proposed model structure, probability normalization method and decoding algorithm for spotting and decoding sequences of keywords in audio records are described.

2. Model selection

HMMs provide flexible method for modeling time-varying events, like acoustic signals. Parameters of the model are fitted in training process to represent statistical relationships within observations (audio features vectors) of searched keyword or acoustic pattern. Design of the model include the selection of structure, choosing the proper number of hidden states, and specification of state output probability function.

Markov process through the model is created by two related time sequences. The inner process consist of sequence of hidden states S. The outer process is created by sequence of observation $\mathbf{o}$, which are functions of state output probabilities, so that the probability of observing $\mathbf{o}$ in time t depend only on the state S, in which was the process in this time.

Generally the model consists of N hidden states, which are connected trough transition probabilities a_{ij} embedded in the transition matrix. Because of characteristic time progression of acoustic signals and for model simplification, left-to-right model structure was used. This means, that model at time t may occur only in the state, in which it was at previous time t-1, or in directly adjacent state. Number of hidden states can be determinate by unsupervised clustering algorithm K-Means K-Variables described in [2].

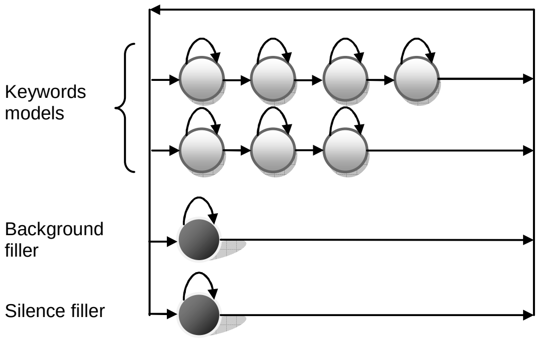


Fig. 1. Complex model network structure with audio pattern models, generic sound filler model and silence filler model.

Complex model structure consists from units - models of searched keywords or audio patterns and fillers (Fig.1). Fillers are represented by models of silence and several type of noises and model of general audio sounds, which parameters was obtained as secondary product of K-means clustering (Section 2.2). Similar structure was used in [3].

2.1. Probability function specification

State output probabilities are given by linear combination of 128 Gaussian PDFs. To reduce amount of required model parameters data and to speed up processing time, each state of the model shares the same mean values and covariance matrices of elementary PDFs components, obtained by clustering huge amount of data using K-means algorithm. HMM with such specification is called semi-continuous. The probability of observing feature vector **o** in state S_i can be computed as:

$$b_i(\mathbf{o}) = \mathrm{P}(\mathbf{o} \mid S_i) = \sum_k w_{ik} b_k(\mathbf{o}) \tag{1}$$

where w_{ik} is weight coefficient of Gaussian component k of state S_i and $b_k(\mathbf{o})$ is probability of observing vector **o** in Gaussian component k, identical for each state of model, computed by equation:

$$b_k(\mathbf{o}) = \frac{1}{(2\pi)^{D/2}\sqrt{\det \mathbf{U}_k}} \mathrm{e}^{-\frac{1}{2}(\mathbf{o}-\boldsymbol{\mu}_k)\mathbf{U}_k{}^{-1}(\mathbf{o}-\boldsymbol{\mu}_k)} \tag{2}$$

where $\boldsymbol{\mu}_k$ and $\mathbf{U}_k$ are mean value and covariance matrix of component k and D denotes for the dimension of observation **o** (number of elements in feature vector). Before weighting the probabilities $b_k(\mathbf{o})$, a normalization based on [4] was adopted, given by formula:

$$b'_k(\mathbf{o}) = \frac{b_k(\mathbf{o})}{\max\limits_k \{b_k(\mathbf{o})\}} \tag{3}$$

2.2. Parameters estimation

For model parameters training, the Baum-Welch re-estimating algorithm was used. This algorithm estimates the parameters of model according to maximum likelihood criterion, by finding the optimal set of model parameters Θ* that satisfy the condition:

$$\Theta^* = \underset{\Theta}{ArgMax}\{\mathrm{P}(\mathbf{O} \mid \Theta)\} \tag{4}$$

where **O** is the sequence of the observations. In our case, only transition probabilities a_{ij} and weight coefficients w_{ik} were estimated. The process of training is iterative; Fig.2 show the values of log. likelihood for different number of model states in estimation process.

Initial values of the weight coefficients for estimation come out from uniformly distributed data vectors of training examples into the states of model. For each state, a joint probability of all assigned vectors is computed in each mixture component respectively. These probabilities are normalized by dividing it with their summation, for determining initial values of weight coefficients.

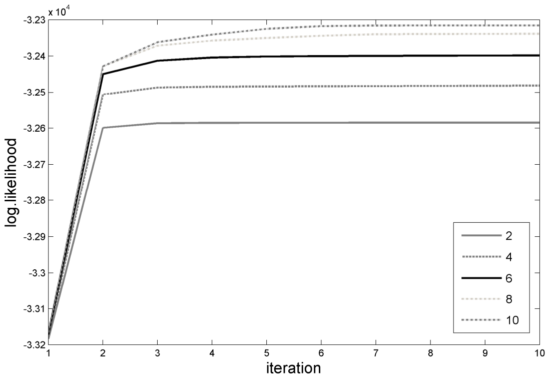


Fig. 2. Log. Likelihood in training process for different number of model states.

3. Pattern decoder

For decoding sequences of searched patterns we adopted system proposed in [3]. The decoding is based on the Viterbi's algorithm with propagation of accumulated score to adjacent states of the model.

Unlike the originally proposed work, we also put transition probabilities in use and make generalization in model structure, such that unit can start or end in any of its state, with respect to prior and final state probabilities of given unit.

4. Experimental results

Performance of the proposed system was evaluated using standard precision, recall and F-measure metrics defined as:

$$P = \frac{N_{CD}}{N_D}; R = \frac{N_{CD}}{N_T}; F = \frac{2 \times P \times R}{P + R} \quad (5)$$

where N_{CD} denotes for number of correctly detected keywords, N_D is number of all sequences detected as target keyword and N_T represent total number of target keywords in testing record.

Number of speakers	630
Training utterances	462
Testing utterances	168
Sampling frequency	16000 Hz
MFCC + Δ + ΔΔ coefficients	12 + 13 + 13
Parameterization window length	0.03s
Parameterization window overlap	0.01s

Tab. 1. TIMIT database specifications

Proposed system was tested on TIMIT database for detecting a set of ten different keywords. As audio features, MFCC coefficients with their time-derivatives was used. Closer specification of used database and parameterization parameters can be found in table 1. After feature extraction, the output probabilities were computed and forwarded to decoder for decoding the sequence of the model units. Experimental results listed in Tab. 2. are evaluated as the average value for 10 runs of algorithm on different testing set.

Precision	Recall	F-measure
0,81	**0,96**	**0,86**

Tab. 2. Average experimental results evaluated after 10 runs

In task of keyword spotting and acoustic event detection, also the processing time is putted in emphasis. Our system operates about 0.12 RT, including the parameterization process, with 10 keywords and 128 Gaussian components. Processing time grow with number of searched keywords.

5. Conclusion

In this paper, a system for keywords and audio events retrival, based on statistical modeling, has been proposed. We described complex model structure with searched units and fillers. Proposed approximation and normalization methods used for likelihood computation were evaluated. Since the system does not depend on type of audio signals modeled by units, it can be utilized for keyword spotting or audio events detection. In future work, we aim to improve the processing speed and modify the current system to propose a different method of adding and training unit models, based on Viterbi's alignment and "elementary sound" sub-models.

Acknowledgements

This publication is the result of the project implementation:

Creating a new diagnostic algorithm for selected cancer diseases, ITMS 26220220022 supported by the Research & Development Operational Programme funded by the ERDF.

Európska únia
Európsky fond regionálneho rozvoja
Operačný program VÝSKUM a VÝVOJ
Agentúra
Ministerstva školstva, vedy, výskumu a športu SR
pre štrukturálne fondy EÚ

References

[1] GODSILL, S.J., RAYNER, P.J.W. Digital audio restoration - A statistical model-based approach, New York: Springer-Verlag, 1998.

[2] REYES-GOMEZ, M.J., ELLIS, D.P.W. Selection, parameter estimation, and discriminative training of hidden Markov models for general audio modeling, *in Proc. IEEE ICME*, vol. 1, Baltimore, MD, July 2003.

[3] NOUZA, J., SILOVSKY, J. Fast keyword spotting in telephone speech, *Radioengeneering*, vol. 18, no. 4, Dec. 2009.

[4] JUNKAWITSCH, J., NEUBAUER, L., HOGE, H., RUSKE, G. A new keyword spotting algorithm with pre-calculated optimal thresholds, *Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings.*, , vol.4, no., pp.2067-2070 vol.4, 3-6 Oct. 1996.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Identification of Environment Parameters that Affects Localization Accuracy

*Michal Hodoň
*University of Žilina, Faculty of Management Science and Informatics, Department of Technical Cybernetics, Univerzitná 8215/1, 01026 Žilina, Slovakia, {Michal.Hodon}@fri.uniza.sk

Abstract. The key limitation currently facing GPS system is the latency of signal arrival caused by the effect of troposphere, where most of the world's weather takes place. This could be presented as an impact of GPS receiver's environment (surroundings) on the localization accuracy. For minimization of an impact of this error it is important to identify and measure all possible measurable parameters of local surroundings of GPS receiver (e.g. humidity, atmospheric pressure, temperature, light intensity, etc.). This could result to the information about correlation between environment parameters and localization accuracy and thereby it could be investigated influence of meteorological conditions on the quality and accuracy of positioning. For this purpose will be designed and implemented a special GNSS station with various sensors that will broadcast the data about actual conditions of environment according to the accuracy of GNSS signal.

Keywords: GNSS, GPS, atmospheric effect, environment monitoring, 32-bit ARM MCU, sensor

1. Introduction

Localization by GPS (or by another GNSS technology e.g. Galileo, GLONASS, etc.) is based on a precise timing of signals sent by GPS satellites high above the Earth. According to a transit time duration of each signal is computed the distance of each satellite to the measured position on the ground. GPS receiver compares the time when the signal was sent by the satellite with the time when the signal was received and from this time difference is calculated the distance between receiver and satellite - range (1):

$$\rho_k = (t_r - t_{ek}) * c\,, \tag{1}$$

where ρk is a range measurement between receiver and satellite k, t_r is the time when the ground receiver r received a signal, t_{ek} is the time when the satellite k sent a signal and c is a speed of light (c = 299 792 458 m/s).

Generally, they are taken into account data from several satellites, those known position is in an Earth-fixed frame (X_s, Y_s, Z_s) (Fig. 1) - the present position of receiver is calculated by trilateration (2):

$$\rho_r^s = \sqrt{(X_s - X_r)^2 + (Y_s - Y_r)^2 + (Z_s - Z_r)^2}\,, \tag{2}$$

where X_s, Y_s, Z_s is the known position of the satellites in an Earth-fixed frame, ρ_r^s is a range measurement between receiver and satellite s counted by (1) and X_r, Y_r, Z_r is the position of receiver converted to the geodetic system (for instance WGS84).

As the GPS (or another radio) signals propagate from the GPS satellite through the atmosphere to the appropriate receiver, on the ground, they are refracted in various ways. In general, they can be (as a common measurement system) affected by two kinds of failures - systematic and random. As concerns the influence of random failures on the localization accuracy, it is not as striking as an influence of systematic failures. They were nowadays identified all systematic error sources that can occur during localization process and it was examined that the biggest deflections are caused by the so-called atmospheric effect. This effect is ranked among physical influences of GPS surveying and is caused by the different densities of Ionosphere and Troposphere, which slow down the signal speed and therefore cause the signal delay.

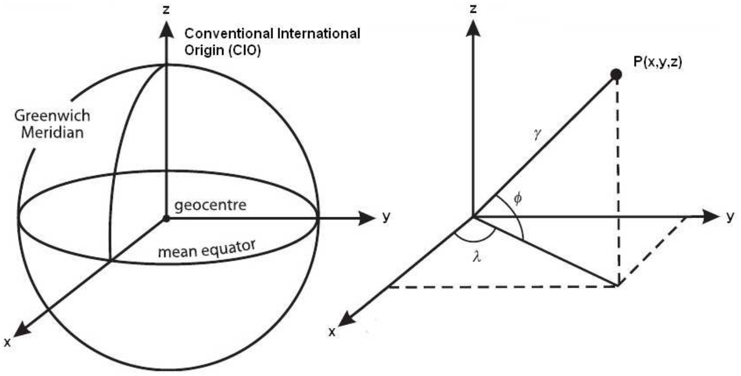


Fig. 1. Earth-Centred Earth-Fixed cartesian and spherical coordinates [1]

Various techniques for minimization of this effect are already known. Mostly are the delays counted from several models that as an input use average and seasonal variation data related to the receiver's position. An alternative is to use the real-time atmospheric data observations and by various communication systems (through satellites, GPRS,…) broadcast information about atmospheric conditions of monitored area to the receivers. But no one from these methods utilizes the precise monitoring of environment of smaller areas with the usage of local real-time data. Especially for monitoring of lower troposphere is better to use local real-time systems (e.g. LAAS - Local Augmentation Satellite) because only by them it is possible to obtain correct information about the actual state and potential inaccuracies there.

2. Atmospheric effect

Influence of atmospheric effect can vary, but in peaks it can cause localization inaccuracies for about 6m. By the use of mathematical models can be this failure minimized, however it will still depend on the elevation angles between the receiver and the satellite. On the other hand, when it occurs some unpredictable state of atmosphere, this method will be inconvenient, because especially the civil receivers are not capable of correcting unforeseen runtime changes (e.g. strong solar winds). Due to this, it is necessary to use the supplementary data, obtainable only by the real-time monitoring of atmosphere.

For the investigation of this effect, it was realized in the period of 13.-14.7.2010 27 hours remained measurement of localization accuracy. As a reference was used the ETRS89 (The European Terrestrial Reference System 1989) geodetic point 2631ZA-1006 located on the top of the hill, above the all housing developments, without any barriers that could block clear view on the sky (Fig. 2). ETRS89 geodetic data are the only data to be used for mapping and surveying purposes in Europe, so they are ideal for testing the accuracy of GPS positioning.

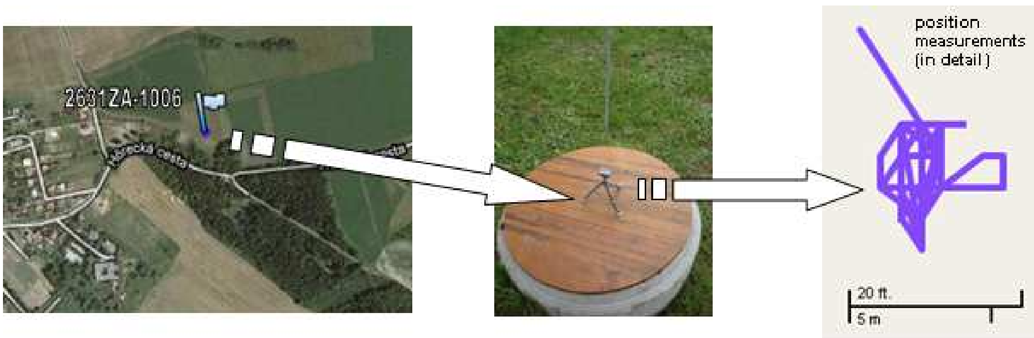


Fig. 2. Measurement system and measured positions in detailed view

In the Fig. 2 can be seen the inaccuracies that occurred during measurement. Detailed view is provided by the Fig. 3 and Fig. 4, where it is displayed measured latitude and longitude separately.

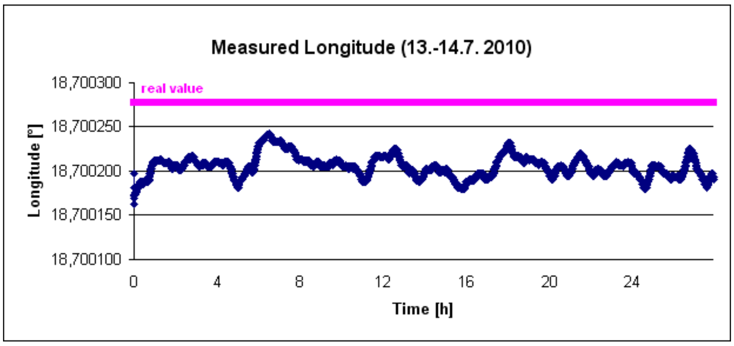


Fig. 3. Measured longitude (13.-14.7.2010)

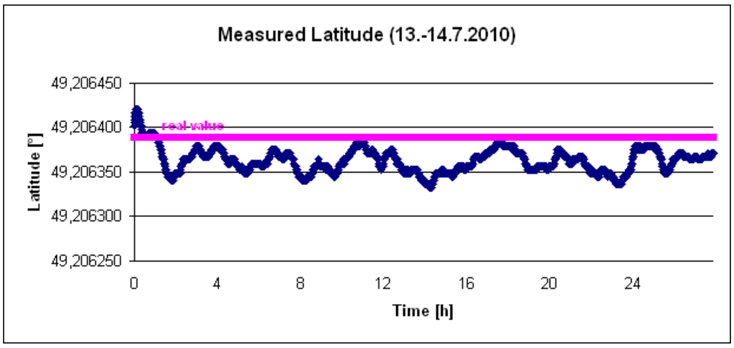


Fig.4. Measured latitude (13.-14.7.2010)

From the pictures above, it is evident that the accuracy of GPS static positioning is far from perfect. But when it would be considered positioning of dynamic objects (vehicles), it could be expected much worse results than they were. So it is necessary to implement additional filtrating methods that can make GPS localization more accurate and independent from environment conditions.

Since 1.10.2009 was launched an open service of the European SBAS (Satellite Based Augmentation System) EGNOS, the European Geostationary Navigation Overlay Service. With its implementation is possible to set up “maps” of atmospheric conditions over different regions in the European area and via the geostationary satellite transponders broadcast these information directly to the receivers with SBAS support. However, EGNOS is by implementation of so-called TEC maps good especially for ionospheric corrections of signal. It also counts with the tropospheric corrections of signal but because it covers larger area (it uses a network of about 40 ground station for all area) it cannot monitor conditions of local character. For this reason is used the combination of real-time data with tropospheric prediction models.

Generally the troposphere consists from two parts, known as a “wet” and “dry” (hydrostatic) troposphere, so the tropospheric delay can be separated into a hydrostatic and a wet component. While the hydrostatic component of the troposphere can be precisely determined and accurately modeled, the wet component, which is the region of the atmosphere below 8-10 km and contains significant levels of water vapor, cannot be sufficiently modeled due to the irregular distribution of water vapor in the atmosphere [2][3]. Real-time surface meteorological parameters are necessary to avoid failure of a tropospheric delay model under conditions when storm fronts pass, causing large gradients in temperature, pressure and humidity [4]. That is the reason for improving tropospheric delay estimation by employing various surface meteorological parameters in the appropriate tropospheric delay models [5].

3. GNSS station

Because the tropospheric propagation delay is mostly influenced by the distribution of water vapor in the lower troposphere, it could be generally said that it depends on temperature, pressure and humidity – environment parameters of GNSS receiver. So adding of these parameters, as additional unknowns, into the localization process could enhance the accuracy of positioning by itself. However, at first it is necessary to identify these parameters and measure their impact on the localization accuracy. This could be made by the special GNSS station, equipped with the EGNOS GPS receiver and with a series of different weather sensors. The station will be fixed on the geodetically precise-determined static reference point (with exact coordinates e.g. ETRS89) and on the base of measured position inaccuracies and environment conditions values could be then made the analysis of the possible impact of environment parameters on the localization accuracy. (Fig. 5)

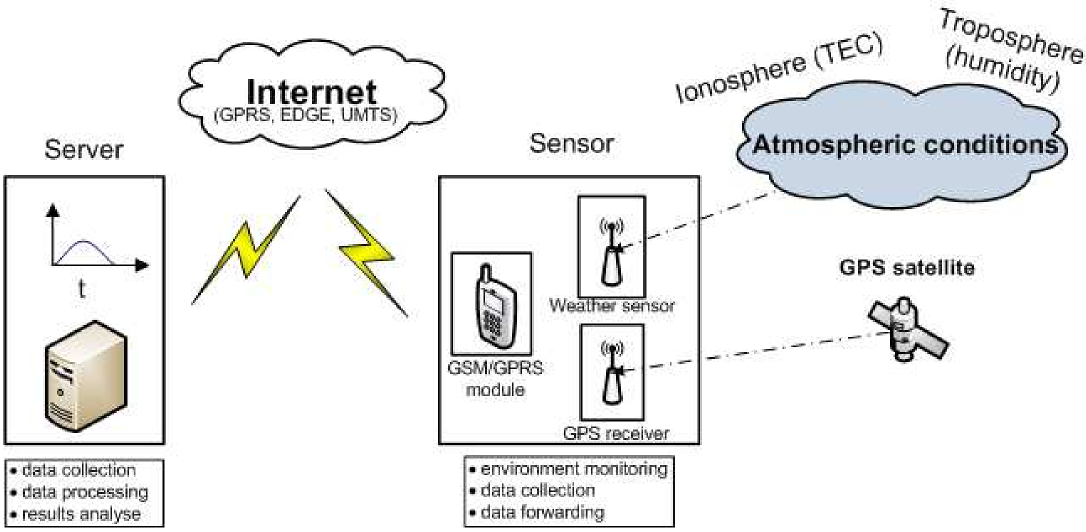


Fig. 5. Working principle of proposed GNSS station

In proposed GNSS station will be as a control unit implemented 32-bit ARM core microcontroller AT91SAM7S64, because of its sufficient amount of peripherals. The station will consist from several different parts that will require communication resources, so it will be necessary to use various communication standards. (Fig. 6)

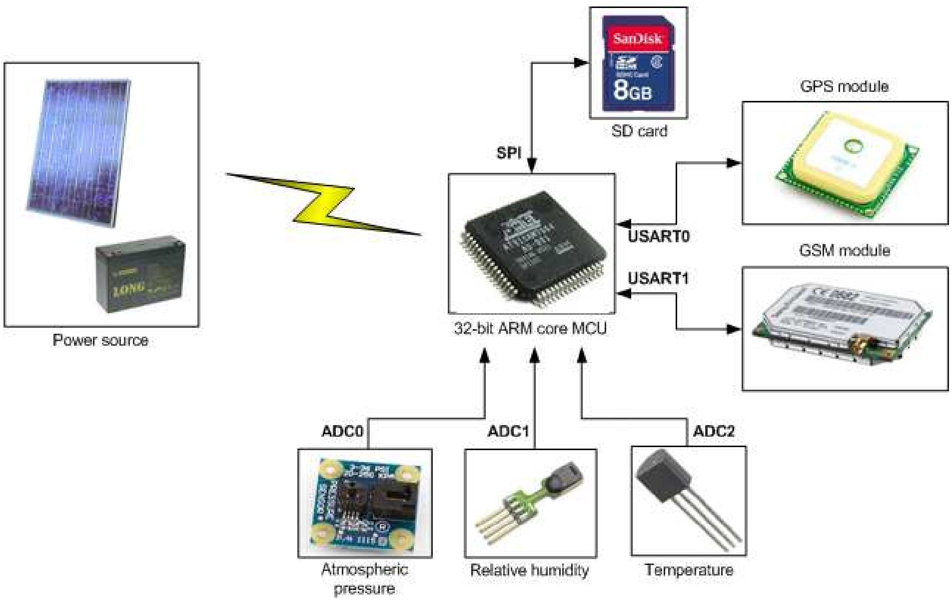


Fig. 6. Basic schematic diagram of proposed GNSS station with communication resources

For the weather monitoring will be implemented three kinds of sensors - monitoring of atmospheric pressure, humidity and temperature. Principle of the station working will be to store all measured data into the SD card and send them circa every hour together with position data via GSM network (through GPRS - General Packet Radio Service) to the server with public IP address for further processing.

Implementation of proposed GNSS station on the roof of the Faculty of management science and informatics of the University of Žilina (FRI-ŽU) is shown in the Fig. 7.

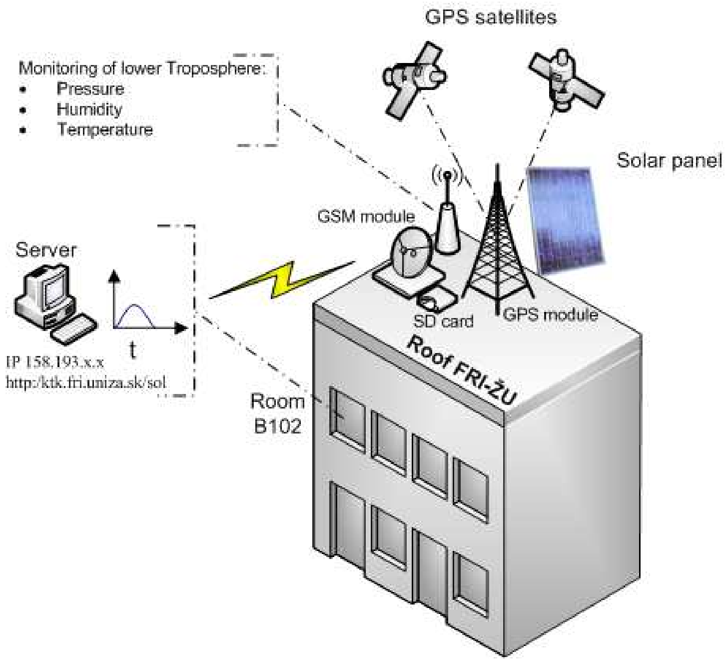


Fig. 7. Implementation of proposed GNSS station on the roof of FRI-ŽU

4. Conclusion

By developed GNSS station will be possible to identify and monitor impact of GNSS receiver's environment parameters on the localization accuracy. On the base of reached results it might be possible to form empirical (or mathematical) models that would allow minimization of GPS signal delay which is caused by the lower troposphere (its “wet” part). This will be made by including of investigated atmospheric parameters as additional unknowns into the analysis of GPS data process. The goal is to, by the use of a of simple, standard, common GNSS (C/A GPS) receiver, reach the localization accuracy suitable for most of quality demanding applications.

Acknowledgement

This work has been funded by the Tatra Banka Foundation grant program E-TALENT 2010 as the 10et001 project.

References

[1] XU G. *GPS theory*. Algorithms and Applications, Second Edition, Geo-ForschungsZentrum Potsdam (GFZ), ISBN 978-3-540-72714-9, 2007.

[2] GABOR M. *Remote sensing of water vapour from GPS receivers*. University of Texas at Austin, 1997, available on-line at http://www.csr.utexas.edu/texaspwv/midterm/gabor/gabor.html

[3] SCHÜLE T. *On ground-based GPS tropospheric delay estimation*. Dissertation, Universität der Bundeswehr München, Fakultät für Bauingenieur- und Vermessungswesen, 2001.

[4] MARKEZIC I., FILJAR R., JURICIC I. *Time distribution of the GPS signal tropospheric delay during passage of the warm front*. Proc. 2nd Congress Transport Traffic Logistics, Portoroz, Slovenia, 345-348, 2000.

[5] KOS T., BOTINCAN M., MARKEZIC I. *Evaluation of EGNOS tropospheric delay model in south-eastern Europe.* The Journal of Navigation, 62, 341-349, 2009.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Load-Balance and Data-Decomposition for Distributed Prime Number Algorithm

*Andrej Holúbek, **Filip Janovič
*University of Žilina, Faculty of Management Science & Informatics, Department of Technical Cybernetics, Univerzitna 2, 01026 Žilina, Slovakia, andrej.holubek@fri.uniza.sk
**European Polytechnical Institute, Institute of Applied Informatics, Osvobození 699
686 04 Kunovice, Czech Republic, janovic@edukomplex.cz

Abstract. This work is about parallelization of prime number algorithms on distributed systems. It is focused on data decomposition models. This article describes advantages and disadvantages of some of data decomposition models used in parallel prime number algorithms on distributed systems.

Keywords: parallel algorithm, prime number algorithm, distributed systems, decomposition models.

1. Introduction

Data used by parallel algorithms [1], in systems with distributed memory, has to be divided between processes, so that every process can work with proper part of data. Therefore data decomposition needs to be provided. There are many ways how to solve problem of data decomposition on distributed systems. In parallel prime number algorithms, we focused on two modes of data decomposition. First model is called interleaved data decomposition, second one is known as block data.

2. Interleaved data decomposition

This data decomposition model is based on interleaving data [2]. Each process will work on part of data interleaved of number of processes. Figure 1 shows interleaving of data for two processes. First process operate with items 2, 2+p, 2+2p, … and second process operate with items 3, 3+p,3+2p, where p is number of processes (p=2).

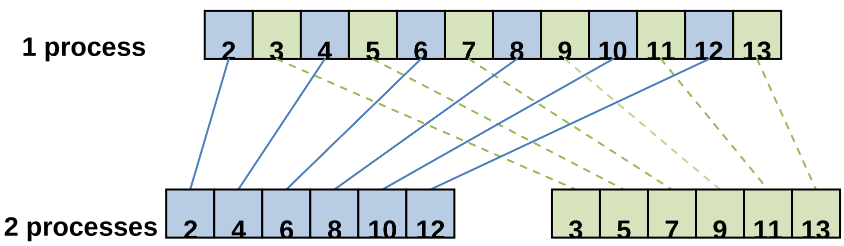


Fig. 1. Interleaved data decomposition.

Advantage of this method is, that for every index i is known, which process control it (i mod p). Disadvantage of this decomposition, in parallel prime number algorithms, is unbalanced distribution of items among the processes. From upper example first process will exclude n/2 items, second process will exclude only one item.

2.1. Experimental results

We used interleaved data decomposition for Sieve of Eratosthenes algorithm [3]. Experiments were done for 2, 4, 8, 16 and 32 processes for first 10^6 numbers. As shows Figure 2, workload of the processes is very unbalanced [4].

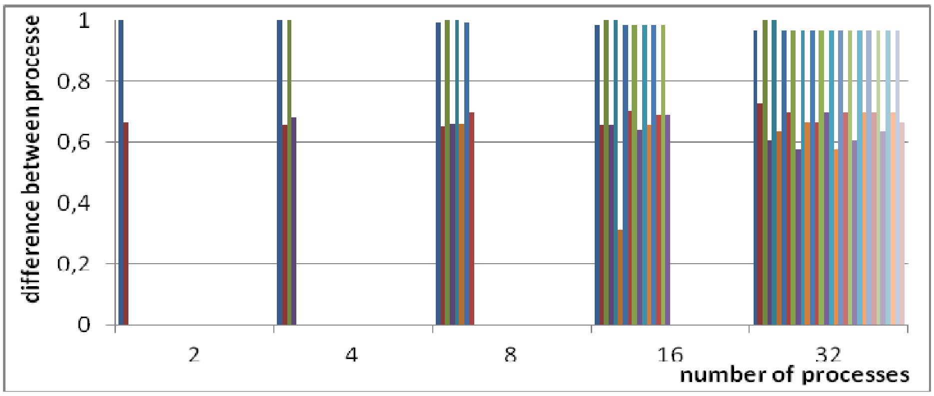


Fig. 2. Load balance of interleaved data decomposition for Sieve of Eratosthenes algorithm.

3. Block data decomposition

Second method of data decomposition is block data decomposition. This decomposition method is based on dividing data into p equal blocks. Afterwards the blocks are distributed among the processes, each process operate on one block of data. Block data decomposition for 3 processes is illustrated on Figure 3. Field of data is divided to p blocks of n/p items, where p is number of processes (p=3).

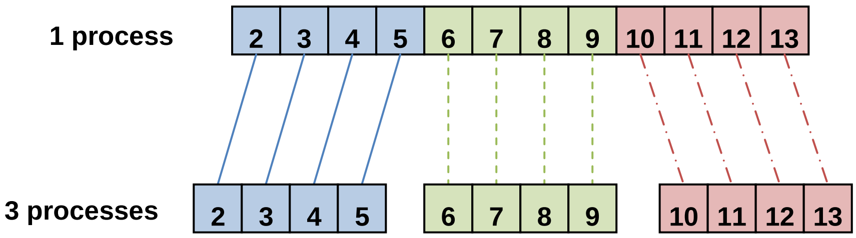


Fig.3. Block data decomposition.

In case that number of items is fold of processes (n mod p = 0), no other operation is needed to be done. In other case we can split the reminded items (r = n mod p) to the first or last r processes.

3.1. Experimental results

Also block data decomposition was used for Sieve of Eratosthenes [5]. This experiment was also made on 2, 4, 8, 16 and 32 processes for first 10^6 numbers. As seen on Figure 4, processes were much more balanced for this model of data decomposition [6].

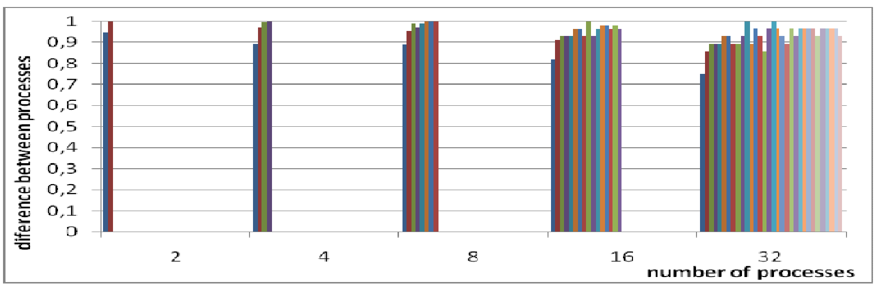


Fig. 4. Load balance of block data decomposition for Sieve of Eratosthenes algorithm.

4. Conclusion

Comparing interleaved data decomposition and block data decomposition for Sieve of Eratosthenes algorithm [7], better results in field of load balance of processes has block data decomposition method. Comparison of both methods is shown on Figure 5. Experiment was made on 32 processes for 10^6 numbers.

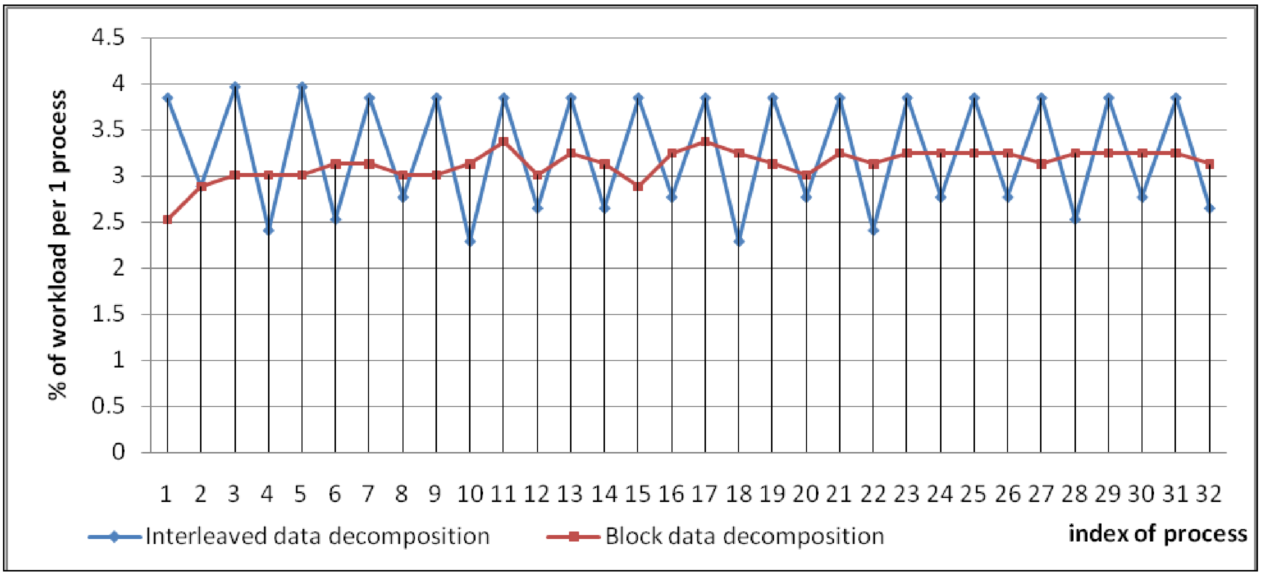


Fig. 5. Load balance of interleaved and block data decomposition for Sieve of Eratosthenes algorithm for 32 processes

As graph clearly illustrate block data decomposition is, for Sieve of Eratosthenes algorithm [8], much more balanced than interleaved data decomposition.

Acknowledgement

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I want to thank the Department of Technical Cybernetics, for giving me permission to commence this thesis in the first instance, to do the necessary research work and to use departmental data.

References

[1] BOHNSACK M., ERHARDT E. B.,*S 442 Introduction to Parallel Processing Project 1: MPI Sieve of Eratosthenes*, 2006.

[2] WILKINSON B., *Parallel Programming*, ISBN 9780131405639, Prentice Hall, 2004.

[3] O'NEILL, M.E., *The Genuine Sieve of Eratosthenes, Asoosiation* for Computing Machinery, 2006.

[4] SOONWOOK H., KYUSIK CH., DONGSEUNG K., *Load Balanced Parallel Prime Number Generator with Sieve of Eratosthenes on Cluster Computers*, Korea University, Soongsil University, 2007.

[5] BOKHARI S.H., *Multiprocessing the sieve of Eratosthenes*, IEEE Computer 20(4), pp.50-58, April 1984.

[6] HARDY G., WRIGHT E., *An introduction to the theory of numbers*, Clarendon Press, 1979.

[7] X. LUO, *A practical sieve algorithm for finding prime numbers*, Communications of the ACM, Vol. 32 No. 3, pp. 344-346, March

[8] BENGELLOUN S. A., *An incremental primal sieve*, Acta informatica, 23, 119–125.,1986.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Comparison of Automatic Speaker Identification Systems

*Martin Hric, *Michal Chmulík, *Roman Jarina
University of Žilina, Faculty of Electrotechnical Engineering, Department of Telecommunication and Multimedia, Univerzitná 2, 01026 Žilina, Slovakia, {martin.hric, michal.chmulik}@fel.uniza.sk

Abstract. This paper presents comparison of Automatic SpeakeR Identification (ASRI) systems using different classifiers: Gaussian Mixture Model (GMM), k-Nearest Neighbor algorithm (kNN) and Support Vector Machines (SVM). Every classifier represents different approach to the classification procedure. Features used in the experiment were Mel Frequency Cepstral Coefficients (MFCC). Classification precisions for each classifier were evaluated on frame and recording level. Experiments were conducted over dataset MobileDat-SK, which was recorded in mobile telecommunication network. Experiment shows promising results for SVM classifier.

Keywords: kNN, SVM, GMM, speaker identification.

1. Introduction

Nowadays speaker recognition is one of the basic task in various systems ASRI, automatic retrieval of audio documents, forensic purpose etc. System allows recognizing „who is talking„ from speech signal. Identification system consists from various parts working together. In this paper we deal with three different approaches for ASRI system and testing them on the same dataset.

The paper is organized as follows in section two is briefly discussed each of used classification method, section three presents result of classification and in subsection we depict database description, data preparation and parameters set to classifiers.

2. Classification techniques description

In this section we present three different methods of classification. Subsection gives a brief overview of GMM, k-NN and SVM classification.

2.1. GMM classification

In GMM classification Gaussian mixture model is used for statistical representation of speaker patterns. The distribution of feature vectors extracted from speech is modeled by a mixture of Gaussian density functions. For a D-dimensional feature vector x, the mixture density for speaker λ_r is defined as [1]:

$$p(x \mid \lambda_r) = \sum_{i=1}^{M} p_i^r b_i^r(x), \qquad (1)$$

Where M is number of components and p_i^r are mixture weights. Density is weighted linear combination of M component uni-modal Gaussian densities $b_i^r(x)$:

$$b_i^r(x) = \frac{1}{(2\pi)^{D/2} \left| \sum_i^r \right|^{1/2}} \exp\left\{ -\frac{1}{2} \left(x - \mu_i^r\right)' \left(\sum_i^r \right)^{-1} \left(x - \mu_i^r\right) \right\}. \qquad (2)$$

Each parameterized by a mean vector μ_i^r and covariance matrix Σ_i^r. Mixture weights must satisfy the constraint:

$$\sum_{i=1}^{M} p_i^r = 1. \quad (3)$$

Complete GMM is defined by mean vector, covariance matrix and mixture weights (4).

$$\lambda = \left\{ p_i^r, \mu_i^r, \sum \quad \right\} (4)$$

Every recognize speaker has its own model which is than using as its representation instead of utterances in identification procedure.

In computation of covariance matrix we use diagonal covariance matrix, which usually gives better results in recognition compared to full covariance matrix. The best results in parameter estimation are achieved by using the iterative Expectation Maximization (EM) algorithm [1], [2]. In this work we use 100 iteration steps for estimating of model.

The identification assignment is maximum likelihood classifier. Main task of system is to make a decision if input utterance belongs to one of the set of speakers, which are represented by its models $\lambda_1 \ldots \lambda_r$. This decision is based on computation of maximum posterior probability for input feature vector [1].

2.2. kNN classification

The kNN algorithm (k Nearest Neighbor) can be classed as a nonlinear nonparametric classification method [3]. This algorithm is based on very simple principle that similar data are close to each other in the searching or data space. In other words, the kNN finds for every object from test data set of k objects in the training data that are closest to the test object (nearest neighbors). The label assignment is usually based on the rule of majority voting, e.g. the most frequent class from the k nearest neighbors for given test object determines the class where this object should belong. A value of k dictates a number of closest objects from training data that are taking into account at the label decision. If the value is too small, then the result can be sensitive to noise points. If it is too large, then the neighborhood may include too many points from other classes. Besides a k value, the distance metric is important to the kNN algorithm. As can be clearly seen, the distance metric represents the measure of data similarity. The choice of particular distance metric usually depends on the given classification problem. Regardless simplicity of kNN, this method is well suitable for multi-modal classes, very flexible and belongs to top 10 data mining algorithms (IEEE Conference on data mining 2007 [3]).

2.3. SVM classification

SVM is learning procedure based on Vapnik's statistical learning theory [4] proposed in 1979. Classification task include a separating data into two sets first set consist of data for training process and second for testing procedure.

Training set instance – label pairs (x_i, y_i), $i=1, 2, \ldots, l$ where $x_i \in R^n$ and y={1, -1}^{l}, the SVM require the solution of the following optimization problem define [5]:

$$\min_{w,b,\xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \quad, (5)$$

subject to:

$$y_i \left(w^T \phi(x_i) + b \right) \geq 1 - \xi_i, \ \xi_i \geq 0. \quad (6)$$

Each instance in the training set contain features of observe data and class labels identifying class, in our task it is index of speaker. Term specify in (7):

$$K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j), \quad (7)$$

is denote the kernel function. Training vectors are mapped into higher dimensional feature space by the kernel function. There are four basic kernel functions linear, polynomial, radial basis function (RBF) and sigmoid. RBF kernel function which is used in our experiment is defined [4]:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \quad (8)$$

where $\gamma > 0$.

Aim of the SVM is to finds a linear separating hyperplane with the maximal margin in this higher dimensional space. C is the penalty parameter of the error term. Value of Penalty parameter must suffer condition $C > 0$. Not every function can be used as kernel, only those which suffer Mercer's Conditions [6].

For SVM classification system, every attribute of the data are scaled to range [1, -1]. The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric range [5].

SVM classifier requires setting up one or more parameters. In our experiment we used C-SVM formulation: included in implementation LIBSVM [7] with RBF kernel function; therefore we search for two model parameters C and γ. For parameter selection task we used Particle Swarm Optimization (PSO) technique [8].

3. Experiment

One of the crucial parts is classifier and features extractor. Classification task is to correctly identify speakers known to the system based on the previous learning procedure. This learning could be done by various techniques based on statistical modeling, distance measure or non-probabilistic linear binary classifier. Feature extractor is process where from speaker utterances are extracted feature vectors which better represents information of identity to system.

In evaluation process we use MobileDat-SK corpus [9, 10]. From corpus consisted of 1100 speakers we randomly select utterances from 20 speakers. From the speaker utterances, 22 MFC coefficients have been extracted as the speech features. Silent frames for each speaker utterance were dropped out using short time energy tresholding and Gaussian mixture modeling (GMM). The frames of 30 ms length and 10 ms overlap is using.

3.1. Test results

In our experiment we used 20 different speakers from database MobileDat-SK to task of identification. Each of tests for every classifier was run 30 times to achieve good statistical information about classification precision.

GMM classifier used 8 probability density functions (PDF), with diagonal covariance matrix. SVM classifier with RBF kernel function was used in test work. Model parameters selection were performed over parameters range $C = \{2^{-5}, 2^{-4.9}, \ldots, 2^{-19.9}, 2^{-20}\}$ and $\gamma = \{2^{-20}, 2^{-19.9}, \ldots, 2^{4.9}, 2^{5}\}$. Criterion function for model parameters selection was 5-fold cross validation accuracy. Classifier kNN with seven neighbors and Euclidian metric was used. Experiment results for classifiers is sown in Tab. 1

Method	GMM	kNN	SVM
Frame level accuracy [%]	16.58	43.21	49.90
Record level accuracy [%]	31.89	92.15	98.11

Tab. 1. Classification accuracy results

4. Conclusion

In this paper, we described three different classifiers used for speaker identification task. Classification accuracy for dataset MobilDat-SK were computed for frames of length 30 ms, and for whole recording of each speaker consisted from all frames. Best classification accuracy achieved SVM classifier. Worst classification achieved GMM classifier, reason could be limited amount (a few seconds) of training data.

Acknowledgements

This publication is the result of the project implementation:
Creating a new diagnostic algorithm for selected cancer diseases, ITMS 26220220022 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] REYNOLDS, D. *A. Speaker identification and verification using Gaussian mixture speaker models.* Speech communication. Vol 17, issues 1-2, 1995.

[2] BIMBOT, F. et al. *A Tutorial on Text-Independent Speaker Verification*, EURASIP Journal on Applied Signal Processing. Vol 4, pp. 430-451, 2004.

[3] XINDONG, W., VIPIN, K. *The Top 10 Algorithms in Data Mining*, Chapman & Hall/CRC, 2009.

[4] VAPNIK, V. Statistical Learning Theory, Wiley, New York, 1998.

[5] HSU, C. W., CHANG, C. C., LIN, C. J. *A practical guide to support vector classification*, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[6] JUNLI, C., LICHENG, J. *Classification mechanism of support vector machines*, 5th International Conference on Signal Processing Proceedings, 2000. WCCC-ICSP 2000, vol.3, 2000, pp.1556-1559.

[7] CHANG, C. C., LIN, C. J. *LIBSVM: A library for support vector machines*, 2001. Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[8] BLONDIN, J., SAAD, A. *Metaheuristic techniques for Support Vector Machines model selection*. In 2010 10th International Conference on Hybrid Intelligent Systems. 2010.

[9] RUSKO, M., TRNKA, M., DARJAA, S. *MobilDat-SK a mobile telephone extension to the SpeechDat-E SK telephone speech database in Slovak*. In Proceedings of the 11-th International Conference Speech and Computer (SPECOM'2006). St. Petersburg, 2006, p. 449-454.

[10]JUHÁR, J., ONDÁŠ, S., ČIŽMÁR, A., JARINA, R., RUSKO, M., ROZINAJ, G. Development of Slovak GALAXY / VoiceXML based spoken language dialogue system to retrieve information from the internet, In *Proceedings of the Interspeech 2006 - ICSLP*. Pittsburg (USA), Sept. 17-21, 2006, p. 485-488.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Parallel Complexity of Linear System Equations

* Húdik Martin, **Hanuliak Peter,
*University of Zilina, Faculty of Control and Informatics, Zilina, Slovakia, hudikm@gmail.com
**Polytechnic institute, Dubnica nad Vahom, Slovakia, phanuliak@googlemail.com

Abstract. The using of parallel principles is the most effective way how to increase the performance of applications (parallel algorithms). Therefore the paper describes the developing steps of parallel algorithms and then it summarised the basic concepts for parallel complexity of linear system equations.

Current trends in high performance computing (HPC) and grid computing (Grid) are to use networks of workstations (NOW) as a cheaper alternative to traditionally used massively parallel multiprocessors or supercomputers. In such parallel systems workstations are connected through widely used communication networks and cooperate to solve one large problem. Each workstation is threatened as a processing element in a conventional multiprocessor system. To make the whole system appear to the applications as a single parallel computer (virtual parallel system), run-time environments such as OpenMP, MPI (Message passing interface) and Java are often used to provide an extra layer of abstraction.

Keywords: parallel computer, parallel algorithm, performance modelling, system of linear equations, complexity

1. Introduction

Parallel organisations of processors, cores or independent computers and a use of various forms of parallel processes [4, 9, 11] at developing parallel algorithm are dominant nowadays. There has been an increasing interest in the use of networks of workstations (NOW) connected together by high speed networks for solving complex computation problems.

Network of workstations (NOW) [5, 7, 12] has become a widely accepted form of high performance computing (HPC). Each workstation in a NOW is treated similarly to a processing element in a multiprocessor system. However, workstations are far more powerful and flexible than processing elements in conventional multiprocessors (Supercomputers). To exploit the parallel processing capability of a NOW, an application algorithm must be paralleled. A way how to do it for an application problem is to build its strategy of decomposition. This step is the most important in developing parallel algorithm [6, 7] and to their performance modelling and optimization (Effective parallel algorithm).

2. Parallel algorithm

The role of programmer is to develop the effective parallel algorithm for the given parallel computer and for the given application task. In general development of the parallel algorithms include the following activities [4, 6]

- decomposition - the division of the application into a set of parallel processes and data,
- mapping - the way how processes and data are distributed among the computational nodes,

- inter process communication (IPC) - the way of cooperation and synchronization,
- tuning – performance optimisation of a developed parallel algorithm.

The most important step is to choose the best decomposition method for a given application problem [4].

3. The role of complexity

Quantitative evaluation and modelling of hardware and software components of any parallel systems are critical for the delivery of complexity and high performance of used parallel algorithms. To evaluate parallel algorithms there have been developed several following fundamental concepts

- analytical
  - asymptotic analysis [1, 3, 6]
  - Petri nets.
- simulation [8, 10]
- experimental measurement.

4. System of linear equations

Let us consider system of n linear equations with n unknown $x_1, x_2, x_3, \ldots, x_N$, in form

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{0,N}x_N &= b_{1,} \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,N}x_N &= b_{2,} \\ &\vdots \\ a_{N,1}x_1 + a_{N,2}x_2 + \cdots + a_{N,N}x_N &= b_{n.} \end{aligned} \qquad (1)$$

where a_{ij} , i=1, 2, ..., n, j=1, 2, ..., n+1 are real constants. If we use the matrix notation then we can rewrite the system of linear equations in matrix form as

$$\mathbf{A} \,.\, \mathbf{X} = \mathbf{B} \qquad (2)$$

, where **A** is a square matrix of coefficients, **B** is a vector of a right side and **X** is a solution vector. In this manner the Fig.1. illustrates the principal possible decomposition models of square matrix. There exist many various ways how to solve a system of linear equations. But there does not exist any universal optimal way of solving it. The existed methods can be divided into

- exact (finite) [2, 4, 6]
- iterative methods.

4.1 Exact methods

These methods come after deterministic number of steps to the exact solution. To these methods belong

- Cramer rule
- Gaussian elimination methods (GEM) and their alternatives.

4.1.1 Cramer rule

Application of this method to solving system of linear equations has its bottlenecks in the extensive calculation of subdeterminants. If the number of unknown n is high, the whole computation time for individual subdeterminants increases exponentially [6].

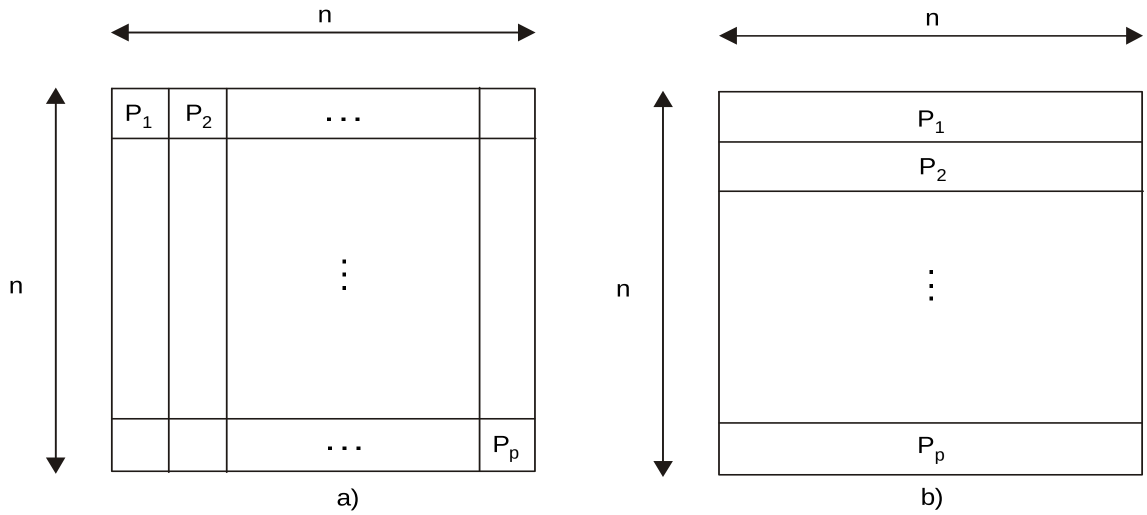


Fig. 1 Matrix decomposition a) blocks b) strips

4.1.2 Gaussian elimination method

Gaussian elimination methods GEM (LU factorisation) [2, 6, 11]) with their known alternatives (Pilot element, Gauss-Jordan elimination, Cholesky factorisation) [2, 11] belong to the most used exact methods of solving a system of linear equations. For this methods were developed sequential and parallel versions of application algorithms (BLAS, LINPACK, LAPACK, ScaLAPACK, and PBLACS).

4.2 Iteration methods

Iteration methods are another way, even more complicated methods, to find computer supported solution with a given accuracy. Own steps for the solution could be found in [2, 6].

5. Conclusions

Parallel complexity modelling of linear system equations have used classical parallel computers with shared memory (supercomputers, massively parallel systems) in which they were not considered any influences of parallel overheads supposing that they are lower in comparison to the latency of executed calculations. In this sense parallel complexity analysis were rationalised only to the analysis of own computation T_{comp} according the relation (3).

$$w(s) = \max\left[T_{comp}, h(s,p) < T_{comp}\right] = \max\left[T_{comp}\right] \tag{3}$$

On the contrary we have been analysing parallel complexity of linear system equations for the actual dominant parallel computers (NOW, SMP, Grid) in which it is necessary to analyse at least the most important parts of overheads (parallelisation, communication and synchronisation overheads). These overhead parts build overhead function h (s, p). In general nonlinear influence of h (s, p) could be at modelling parallel complexity of linear system equations dominant. Then for asymptotic isoefficiency analysis it is true

$$w(s) = \max\left[T_{comp}, h(s,p)\right]. \tag{4}$$

For illustration (Fig. 2) of the used overhead function h (s, p) we have compared evaluation of communication overheads for matrix decomposition according Fig. 1.

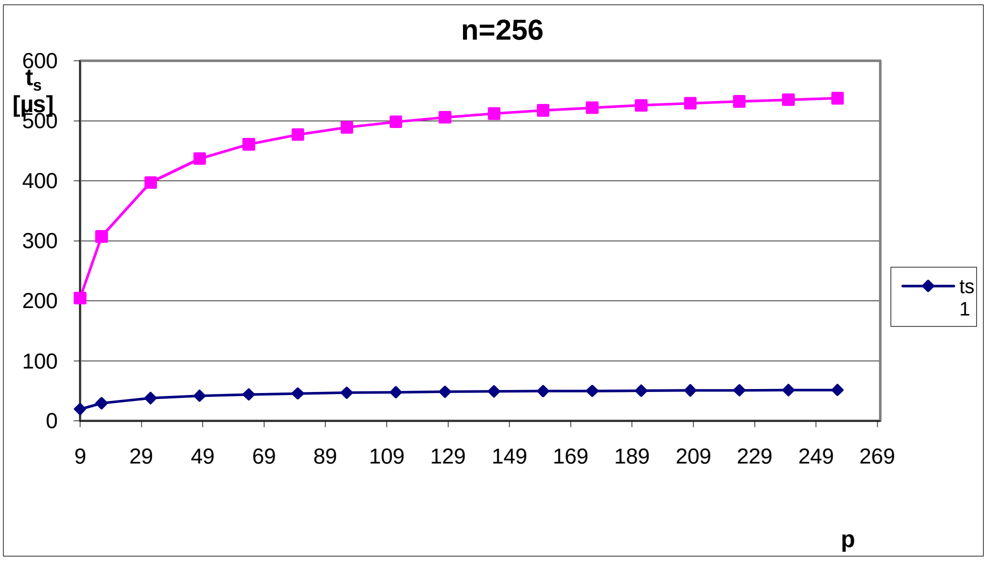


Fig. 2 Comparison of suggested decomposition methods.

References

[1]. ARORA, S., BARAK, B. *Computational complexity - A modern approach*, Cambridge University Press, 573 pp., 2009

[2]. CODENOTI, B., LEONCINI, M. *Parallel complexity of linear systems*, 228 pages, Imperial college press, United Kingdom, April 1991.

[3]. FORTIER, P., HOWARD, M. *Computer system performance evaluation and prediction*, 544 p., Digital Press, 2003.

[4]. HANULIAK, I. *Parallel computers and algorithms*, 327 pp., Publ.: ELFA Košice, 1999.

[5]. HANULIAK I., HANULIAK P. *To performance evaluation of DPA in NOW*, In Proc. GCCP 2005, pp. 30 – 39, SAV Institute of Informatics, Bratislava, 29.9 – 2.12. 2005.

[6]. HANULIAK I., HANULIAK P. *Performance evaluation of iterative parallel algorithms*, Kybernetes, Volume 39, No. 1, pp. 107 – 126, United Kingdom, 2010.

[7]. HANULIAK J. *To performance evaluation of distributed parallel algorithms*, Kybernetes, West Yorkshire, Volume 34, No. 9/10, pp 1633-1650, United Kingdom, 2005.

[8]. HILLSTON J. *A Compositional Approach to Performance Modelling*, University of Edinburg, 172 pages, Cambridge University Press, United Kingdom, 2005.

[9]. KIRK D. B., HWU W. W. *Programming massively parallel processors*, Morgan Kaufmann, 280 pages, 2010.

[10]. KOSTIN A., ILUSHECHKINA L. *Modelling and simulation of distributed systems*, 440 pages, Imperial College Press, Jun 2010.

[11]. KUMAR V., GRAMA A., GUPTA A., KARYPIS G. *Introduction to Parallel Computing*, Addison Wesley, 856 pp., 2001.

[12]. KUMAR A., MANJUNATH D., KURI J., *Communication Networking* , 750 pp., Morgan, 2004.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Informative Attribute Obtaining for Detection Period Boundaries of Human Speech

[*]Martin Hyben
[*]University of Žilina , Faculty of Technical Cybernetics, Univerzitna 2, 01026 Žilina, Slovakia,
martin.hyben@fri.uniza.sk

Abstract. This document deals with the possibilities of obtaining the informative attributes, which allow locating individual period boundaries of human speech. It is obvious that based on position of individual period boundaries, it is possible to segment the speech to the level of vowels. By recognizing the periods of speech it is possible to recognize the vowel itself. During the designing of the method, the emphasis was placed to retain a low computational complexity, so the results could be used on embedded systems and mobile devices. The presently known approaches will be described and compared with our proposed methods. The experiment will be presented for comparison of all described methods to find the best fit and results will be described in conclusion.

Keywords: Recognition of the speech, informative attributes, segmentation, units of the speech

1. Introduction

Recent systems of speech recognition are focused mainly on the recognition of the whole words. Mainly the systems of speech recognition in Slovak and Czech languages are also focused on this type of recognition. Compared to English, which is a dominant language and most of scientific researches pay attention to it, Slovak and Czech contain much more words and their variations. This fact results to a bigger size of reference set for Slovak and Czech languages, compared to English. This is the reason why we have decided to focus on segmentation of the speech on smaller speech units rather than on the whole words. It seems that ideal candidates are vowels because of their small amount. (In Slovak, there are approximately 50 normally used vowels). Recognition of the vowels is very complicated because of the presence of coarticulation. This is the reason why it would be necessary to design new methods which could segment the speech on the appropriate speech units just to identify them even through there is coarticulation.
The idea of the segmentation on the vowels comes from the example of human acoustic and auditory apparatus. Human naturally segments the speech on voiced and unvoiced vowels despites the presence of coarticulation which can be defined as conjunction of two neighboring vowels. For instance, voiced vowel “o“ in word “so“ is affected by high frequency of the previous unvoiced vowel “s“. It is caused by vocal apparatus which is adjusted to pronounce sound “s“. Pronouncing needs some time to be set up to a new position for pronouncing the sound “o“ and that is how the conjunction arise. Additional information for recognition is provided by fundamental frequency $f_{0.}$ This frequency is generated by pressed air from the lungs, passing through contracted larynx. The result is something like the highest intensity carrier frequency, on which the effects of articulation apparatus are modulated. When we recognize the speech back, it is necessary to search just for mentioned fundamental frequency, which each period carries the information about the presence of individual articulation apparatus, and so the information about the reference of the period to one of the voiced vowel. Unvoiced vowels, which are generated by open larynx, are characterized by specific features which also have to be considered.

2. Autocorrelation function

Nowadays, most commonly used method with the lowest computational complexity for determination of the periodicity of signal is clearly autocorrelation function. It is based on the principle of resonance where, in case of the time conjunction of two peaks from the signal, the addition of their amplitude and also a huge increase of final intensity occur.

2.1. Autocorrelation function

Autocorrelation function is mostly used for detecting the presence of periodicity. In case of audio processing, it is used for processing a single channeled audio signal. Thus the formula for computing an autocorrelation function for single channeled signal is defined as:

$$R(m) = \sum_{k=-\infty}^{\infty} s(k).s(k+m), \qquad (1.1)$$

where *s(k)* is a sample of a signal in the time *k* and *s(k + m)* is a sample *k* moved on a timeline by *m*. Autocorrelation function of periodical signal with the period *P* acquire its maximum values just for *m = 0, P, 2P ...* .(Psutka, 2006).

2.2. Mutual comparison of two channels

During the recording of the speech, it is possible to use a stereo microphone and to record the signal for two channels. In that case, we can use the fact that both channels contain required information which is distorted by different amount of unwanted effects. Thanks to this function, it is possible to highlight the information common for both channels. Then the formula for computation the autocorrelation function for both channels is defined as:

$$R_k = s_1(k) + s_2(k), \qquad (1.2)$$

where $s_1(k)$ is a sample of the left channel in time *k* and $s_2(k)$ is a sample of the right channel in time *k*. Mutual comparison function of periodical signal with the period *P* acquire its maximum values just for its periods.

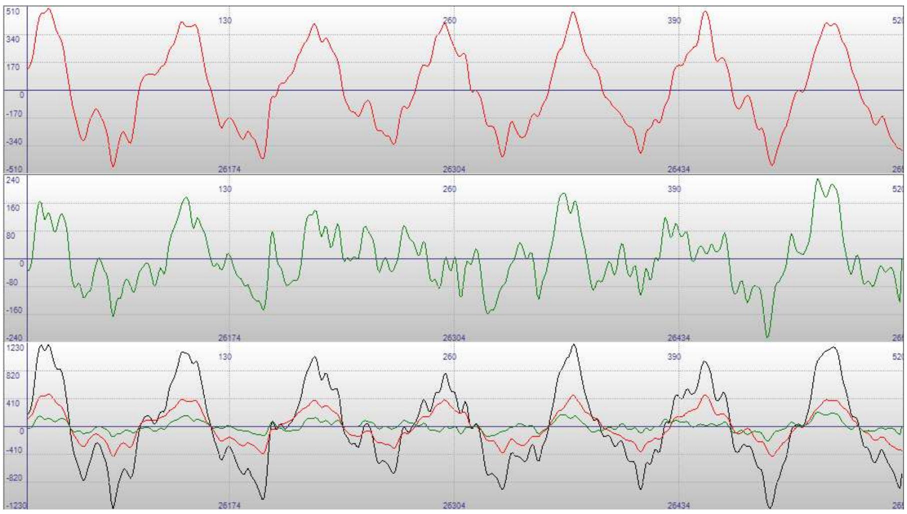


Fig. 1. Autocorrelation function of both channels. Red signal represents the left channel, green represents a right channel and black represents mutual comparison function.

3. Dominant peak function

The principle of the function is based on the domination of fundamental frequency f_0 compared to the other frequencies, which are present in voiced part of speech. Similarly unvoiced parts of the speech often posses a higher intensity in short time. The function is therefore focused on peak detection in the speech by additional selection, just by rule of domination. Detection of particular peaks is performed by using a following criterion:

$$\text{if } s(i)<s(i+1)>s(i+2), \text{ then } s(i+1) \text{ is concave peak.}$$

$$\text{if } s(i)>s(i+1)<s(i+2), \text{ then } s(i+1) \text{ is convex peak.}$$

where $s()$ is signal and $s(i)$ is it's *i-th* sample for $i = 0, 1, 2 \ldots n$.

The residual effects of articulation tract are removed by repeating the application of the function on its previous results. Consequently, the fundamental frequency presence is highlighted as shown on fig. 2.

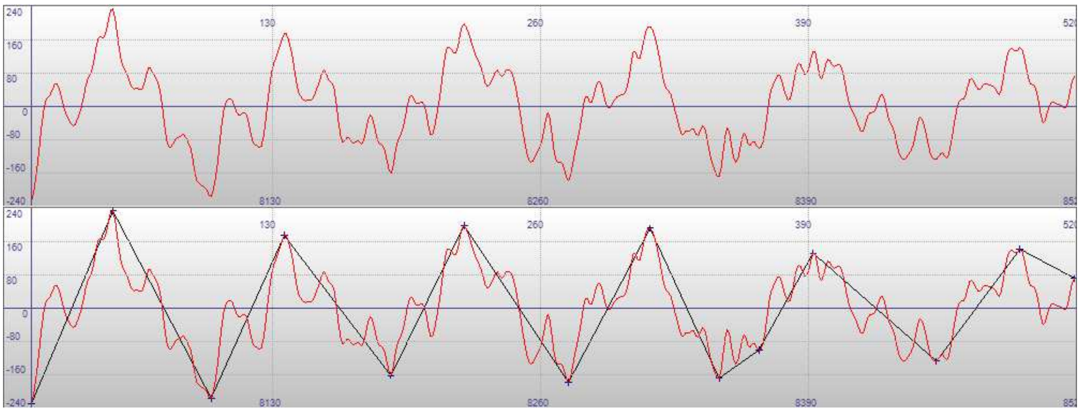


Fig. 2. Fundamental frequency detection. Red signal represents a sound signal and black one represents periods of fundamental frequency.

Dominant peak function principle can be applied also for detection of unvoiced part of the speech. Fig.3. shows highlighted dominant peak of unvoiced vowel „t".(Gold, 2000).

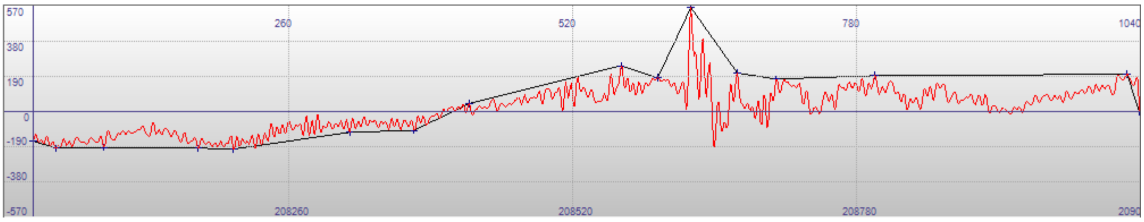


Fig. 3. Unvoiced vowels detection. Red signal represents a sound signal and black one represents a dominant peak of unvoiced vowel „t".

4. Choosing an appropriate method of segmentation

As mentioned before, fundamental frequency f_0 is a semi periodic frequency, which can be characterized by sinus signal. It is unique for each person. Since most of energy of flowing air is concentrated just on vocal chords, the harmonic component of the signal posses greatest intensity. Therefore it is dominant across entire voiced section of speech. The principle could be seen as modulation of carrier wave in order to transfer information. Individual articulation apparatus (e.g. tongue, teeth, nasal cavity) yet specifically affect fundamental carrier wave in the form of fundamental voiced frequency and distort it into resulting shape of human speech. The influence of the individual articulation apparatus is then perceived as representation of individual voiced vowels.

4.1. Segmentation of voiced section of the speech

The goal here is the segmentation of voiced section of the speech into individual periods of fundamental frequency and detection of the boundaries of unvoiced vowels.

The First alternative to achieve this goal is the use of *autocorrelation function* for detection periodicity and consequential detection of local extremes of result function. The second possibility is the *method of mutual comparison* of both channels. Similar to autocorrelation function the highlighting of fundamental frequency f_0 occurs. The last used approach is *dominant peak function*, designed and described in this document. The results of all methods are local extremes, which indicate period boundaries of fundamental voice frequency.

To compare the results of all three methods we chose to perform a simple experiment. We have generated a perfectly periodic sinus wave which represented fundamental frequency $f_{0.}$ Consequently, we gradually distorted this wave by random noise, while increasing its intensity. By using all three methods we have tried to determine the frequency f_0 period boundaries.

The results show that from the point of accuracy the autocorrelation function is clearly the most suitable. Even at a very high level of distortion, even for human eye, it was able to detect the periodicity. But in speech recognition problematic, the precision is often not as needed as low computational complexity. From point of computational complexity, remaining two methods markedly outperform the autocorrelation function, having almost similar computational complexity between them. However *dominant peak function* compared to *method of mutual comparison* achieved better accuracy. So if the goal is to find a balance between accuracy and computation complexity, a *dominant peak function* appears to be the most suitable.

4.2. Detection of boundaries of unvoiced speech sections

Unvoiced sections of the speech are random and characterized mostly by sharp increase of intensity and its sudden drop in short time, or higher frequency compared to common noise. Since the noise is random, the use of autocorrelation function is not relevant in this case. On the other hand *dominant peak function* can be used even for the case of unvoiced speech sections (see. fig.3). Compared to *method of mutual comparison* it achieved similarly as for voiced speech sections, better accuracy along with similar computation complexity.

5. Conclusion

In this document we designed a method for segmentation of speech into individual frequency periods. We have performed a comparison of our *dominant peak function* with *autocorrelation function* and *method of mutual comparison*. The results shown, that from ratio of accuracy and computational complexity, the *dominant peak function* best meets our requirements.

References

[1] PSUTKA, J.,MÜLLER, L., MATOUSEK, J. RADOVA, V. *Mluvíme s počítačem česky*. Praha : ACADEMIA, ISBN 80-200-1309-1, 2006.

[2] GOLD, B., MORGAN, N. Speech and Audio Signal Processing Wiley 2000.

[3] SHAMMA, S. *On the role of space and time in auditory processing*. Theoretical Comput. Sci., 2001.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

MFCC vs MPEG-7 Features in Speaker Recognition Task

*Michal Chmulík, *Roman Jarina
*University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitná 2, 01026 Žilina, Slovakia, michal.chmulik@fel.uniza.sk

Abstract. In this article, short comparison of MFCC and MPEG-7 features is presented. We have chosen speaker recognition task as a classification problem to reveal how suitable are the audio features of MPEG-7 standard for this task, compared to well established MFCC features. We have applied evolution optimization techniques – Particle Swarm Optimization and Genetic Algorithm – for the purpose of "optimal" features subset selection for both groups of features. As a classifier, we have employed kNN algorithm. Test results show that MFCC features outperform quite significantly MPEG-7 audio features at the speaker recognition task.

Keywords: audio features, MFCC, MPEG-7, speaker recognition

1. Introduction

Speech, as a basic communication tool with primary informational nature, allows people to express their thoughts and feelings. Due to this fact, a lot of researchers are interesting in speech processing area from where the speaker recognition is one of the typical classification problems. Speaker recognition, when we attempt to recognize a person from a spoken phrase is very useful for instance for radio and TV broadcast indexing. As another examples related to the speech we can exemplify speech recognition, speaker verification or speaker gender identification.

2. Audio Features

For any classification task, we need to extract from input audio signal some particular properties or features that can describe sufficiently and precisely given audio signal. These features should have discriminative capability and should be decorrelated as far as possible. Basically, we can divide audio features into 3 groups [1]:

- temporal
- spectral
- statistical

To the group of temporal features obtained directly from input audio signal belong *zero crossing rate*, *short time energy* or *temporal centroid*. Spectral features group obtained from signal transformed from temporal to spectral area covers *audio spectrum envelope, DFT, MFCC coefficients* or *spectral centroid.* Statistical features describe statistical properties of input audio signal and besides *correlation* or *covariance* functions *skewness* or *kurtosis* can be mentioned as an example. Selection of audio features used for classification depends greatly on the particular problem and classification efficiency is proportional to the richness of the audio signal description expressed by given audio feature or group of features.

2.1. MFCC Features

One of the most commonly used audio feature applied in many classification problems from the area of audio signal processing is the group of *Mel-frequency cepstrum coefficients* (MFCC). It is conventional features vector for the representation of human voice and partially musical signals because it is a compact representation of the spectral envelope of audio signals which respects the nonlinear human perception of the pitch [2, 3]. Therefore, this feature is very often used in speech and speaker recognition task and it also is able to analyze and represent the musical signals. Process of audio signal parameterization by *MFCC* features is depicted on Fig. 1.

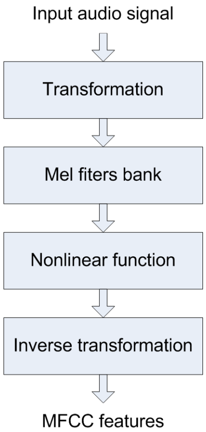


Fig.1. MFCC features extraction

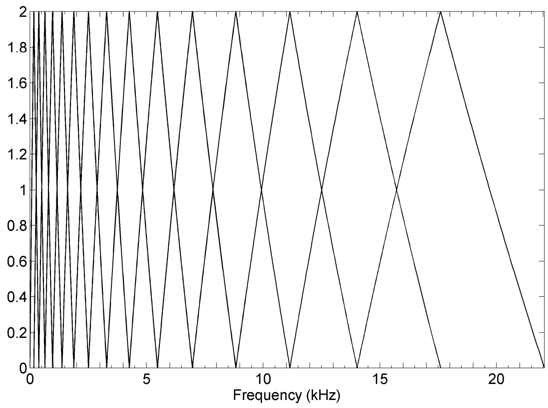


Fig. 2. Mel filters bank

Filters bank with mel frequency scale is included into signal chain. This bank respects human ears perception and is depicted on Fig. 2. *MFC coefficients* are counted according to the following equation:

$$c_i = \sum_{j=1}^{N_f} \left\{ \log(E_j) \cos\left[i\left(j - \frac{1}{2} \right) \frac{\pi}{N_f} \right] \right\} \quad 1 < i < N_C \tag{1}$$

where c_i is the *MFCC* order, E_j is the spectral energy of the signal in *j*-th mel filter, N_f is total filter number and N_C is total coefficients number extracted from one frame. *MFCC* features are usually supplemented with dynamic features – *derivative* (Δ) and *acceleration* (ΔΔ) *MFCC* - describing temporal changes of spectrum.

2.2. MPEG-7 Features

The MPEG-7 standard provides a rich set of standardized tools to describe multimedia content and includes tools for audio as well as images and video data [4]. Audio part of this standard consists from several elements. Basic layer is formed by 17 low – level descriptors that are simple, low complexity audio features that can be used to characterize any type of sound. We can divide it into 6 groups:

- basic: *audio waveform, audio power,*
- basic spectral: *audio spectrum envelope, audio spectrum centroid, audio spectrum spread, audio spectrum flatness,*
- basic signal: *audio harmonicity, audio fundamental frequency,*
- temporal timbral: *log attack time, temporal centroid,*
- spectral timbral: *harmonic spectral centroid, harmonic spectral deviation, harmonic spectral spread, harmonic spectral variation, spectral centroid,*
- spectral basis representation: *audio spectrum basis, audio spectrum projection.*

More detailed description of particular features is out of the scope of this article. Besides these low - level descriptors (features), standard *MPEG-7* also offers higher - level description schemes that are combination of various low - level features. There are 5 sets of description schemes that cover different application areas like for instance melody description, general sound recognition or musical instrument timbre.

3. Experiments

As a classification problem, the speaker recognition has chosen. The database, excerpted from MobilDat Sk [5], consisted from the set of 1100 speakers where every speaker had 3 samples. Every sample (utterance) was uncompressed PCM wave file with sample frequency of 8 kHz with 16 bits depth and had length of 8 seconds. The database was parameterized by *MFCC* features - 23 *MFCC* + 23 Δ *MFCC* + 23 ΔΔ *MFCC*. The same database was also parameterized by *MPEG-7* audio features. The window length at the parameterization process was 30 ms with a 10 ms overlapping. Frames that did not contain speech were removed. In the experiments were used 2 samples as a reference or training data and 1 sample as a test data. As a classifier, we have chosen the k Nearest Neighbor (kNN) algorithm. Although this method is very simple, it remains its success in the data mining and is comparable with other more complicated methods and belongs to the top 10 data mining algorithms [6]. We have chosen 20 classes for the tests so the *kNN* classifier should distinguish 1 between 20 speakers. All speakers were randomly selected from the database for every test. As a confidence measure for the test evaluation, we used *precision*, *recall* and *f measure*. *Precision* can be described as follows:

$$P = \frac{N_{CD}}{N_{TD}} \tag{2}$$

where N_{CD} denotes the number of correctly detected speakers and N_{TD} denotes the total number of detected speakers. The precision value expresses how many detected speakers truly belong to the given class. *Recall*, defined by the equation:

$$R = \frac{N_{CD}}{N_{TC}} \tag{3}$$

where N_{TC} means total number of correctly speakers to detect, says how many speakers were detected. Finally, *f measure*, described as follows:

$$F = \frac{2PR}{P+R} \tag{4}$$

is the harmonic mean of *precision* and *recall* and gives the overall classification rating. The *precision*, *recall* and *f measure* values are from the range of (0, 1). We have employed evolution optimization techniques – namely relatively new approach called Particle Swarm optimization as well as typical representative Genetic Algorithm – to optimize feature space. As a criterion function, we used *f measure* counted from the whole utterance for every speaker, as well as mean precision value counted from the frames of given utterance for every speaker. We also took into consideration the number of features. For *MFCC*, more features combinations gave almost equal result and we used 19 + 5 + 10 *MFCC* respectively Δ *MFCC* and ΔΔ *MFCC* in our tests. For *MPEG-7* features, the best subset was the combination of *audio spectrum centroid, audio spectrum spread, audio harmonicity* and *audio fundamental frequency.* It should be noted that at the optimization process of *MPEG-7* features we discarded basic and temporal timbral features because of their irrelevancy in speaker recognition task. We also discarded *harmonic spectral centroid, harmonic spectral variations* and *harmonic spectral spread* from the group of spectral timbral features. All tests were performed in the Matlab program and every test consisted from 100 runs to secure the statistical credibility of obtained classified data. Fig. 3 depicts mean values of classification accuracy for both audio features - MFCC reached value 84.41 % while MPEG-7 53.8 %.

4. Conclusion

As it can be clearly seen on Fig. 3, the combination of *MFCC* features reached significantly better classification accuracy value than the combination of MPEG-7 features and the difference is more than 30 %. From this point of view, *MFCC* features remain the best parameters to describe speech when we take into consideration that the MPEG-7 features is combination of more parameters chosen by MPEG group of ISO from large set of available audio features.

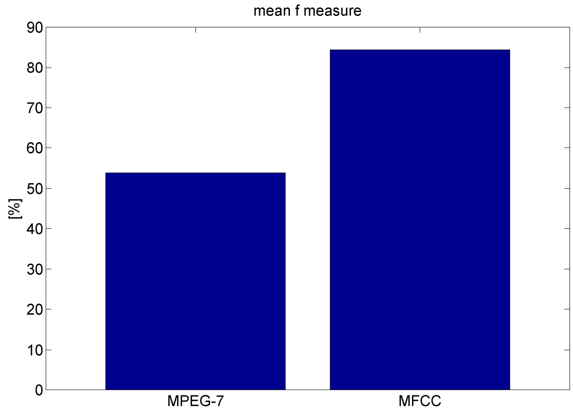


Fig. 3. Classification accuracy

In the future work, we will employ MPEG-7 features for the purpose of general sound recognition. For this application, MPEG-7 features should be more suitable in comparison with speech recognition task and we expect noticeable increased classification accuracy.

Acknowledgements

This publication is the result of the project implementation: Centre of excellence for systems and services of intelligent transport, ITMS 26220120028 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] MINGSAIN R. BAI, MENG-CHUN CHEN.: Intelligent Preprocessing and Classification of Audio Signals, Journal of Audio Engineering Society, Vol. 55, 2007.

[2] KADLEC, F.: Zpracování akustických signálů, Praha ČVUT, 2002.

[3] EVEREST F. A.: Master Handbook of Acoustics, McGraw-Hill, 2001.

[4] HYOUNG-GOOK KIM: MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval, Wiley, 2005.

[5] RUSKO M, TRNKA M., DARJAA S.: MobilDat-SK – a Mobile Telephone Extension to the SpeechDat-E SK Telephone Speech Database in Slovak, Proc. of SPECOM, St. Petersburg, 2006.

[6] W. XINDONG, K. VIPIN: The Top 10 Algorithms in Data Mining, Chapman & Hall/CRC, 2009.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

The Information System of the Data Boxes and its Influence on Financial Results of Czech Post

[*]Barbora Chromcová, [**]Libor Švadlenka
*University of Pardubice, Jan Perner Transport Faculty, Department of Transport Management, Marketing and Logistics, Studenská 95, 532 10 Pardubice, Czech Republic, E-mail: barbora.chromcova@student.upce.cz
**University of Pardubice, Jan Perner Transport Faculty, Department of Transport Management, Marketing and Logistics, Studentská 95, 532 10 Pardubice, Czech Republic, E-mail: libor.svadlenka@upce.cz

Abstract: The paper deals with the computerization of the delivery of official documents, specifically the information system of the data boxes and its influence on the situation of holder the postal license. In this time the holder of postal license is Czech Post, s.e. It is generally assumed that the volume of letter items will continue to fall thanks to e-mail and electronic communications. The information system of the data boxes is a revolution in the delivery. The official document in paper form is replaced by an electronic document. The aim of this system is to facilitate communication between authorities and citizens and firms - legal entities and entrepreneurs. Through operation of the system of the data boxes the large volume of guaranteed communication will remain in the hands of the Czech Post, s.e. This is an advantage for the Czech Post, s.e., at least within the validity of postal license.

Keywords: The information system of the data boxes, the data message, holder of the postal license

1. Introduction

The Act No. 300/2008 on e-Government was adopted on 1st July 2006. This law represents a big breakthrough in communication among public authorities, legal entities and citizens. Egon is a symbol of e-Government and the computerization of public administration. It is a living organism whose existence and life functions are ensured by:

- Fingers: Czech Point as a system of easily accessible points of contact,
- Circulatory system: communication infrastructure of public administration ensuring secure data transmission,
- Heart: e-Government Act - Act No. 300/2008 Coll. on Electronic Operations and Authorized Conversion,
- Basic: basic registers of public administration - secure and up-to-date database of data about citizens and government or non-government subjects.

2. The information system of the data boxes

The information system of the data boxes (ISDS) is a technical revolution in the guaranteed electronic delivery of official correspondence between public authorities (ministries, central state administration subjects, local and county governments, health insurance companies, courts, court executors, etc.) and public (individuals, entrepreneurs, legal entities - firms).

The establishment and operation of the ISDS is regulated by Act No. 300/2008 Coll. on electronic Communication and authorized conversion of documents. This Act came into force on 1th July 2009. The Act defines the ISDS as public administration information system which contains information about the data boxes and its users. The manager of the system is Ministry of Interior of the Czech Republic and its provider is the holder of postal license. Costs associated with the operation of ISDS are covered by the state. A data box can be established voluntarily and free of charge by other legal entities, individuals or entrepreneurs.

The obligation to establish a data box applies to public authorities and all legal entities established by law and legal entities registered in Czech Commercial Register.

The Act was amended on 1st January 2010. Until 30th June 2010 individuals, business individuals and legal entities could supply only invoices or similar requests for payment to a data box. And from 1st July 2010 it is also possible to deliver not only invoices but also other documents, i.e. there are no restrictions on content.

The issue of the ISDS is modified with the following notices:

- Notice No. 192/2009 Coll., which implements certain provisions about the archives and records service and amends some laws,
- Notice No. 191/2009 Coll., about details of records service,
- Notice No. 193/2009 Coll., about determination of details of the implementation of authorized conversion of documents,
- Notice No. 194/2009 Coll., about determination of details of use and operation of the ISDS, including amendments.

For the basic milestones of the start of operation of the ISDS are considered:

before May 1st 2009	development and implementation of ISDS
May 1st 2009	trial and piloting operation with selected offices
June 1st 2009	public testing enviroment for archive andrecords service and other applications
July 1st 2009 - September 28th 2009	start project; taking data from a data manager, establishment of data boxes from the Act and also at the request of individuals
November 1st 2009	reguired activation data boxes establish from the Act (public authorities, legal persons registred in the Commercial Register)
January 1st 2010	use of ISDS for private communication - the possibility of sending electronic invoices between private parties - Postal data report
July 1st 2010	the possibility of using the ISDS for private communication without content restrictions

Tab. 1. Milestones of the information system of the data boxes.

The data box evokes imagination that an electronic delivery is within the meaning of “emailing”. This notion is false. The data box is a data repository. Basically, it is not data messages sending but a message insertion to a specific data box. The data message thus does not leave the information system and its provider is able to guarantee delivery of messages. The main purpose of establishing a data box was the introduction of the institution of electronic transactions. This is defined as the binding force of a data message, where legal action represents the same weight as if the document was delivered in a paper form personally. Furthermore, it is a guaranteed delivery including delivery of legal fiction and finally the transience as the messages are automatically deleted from the data box after 90 days of delivery. The user is responsible for archiving, which can only be converted to the documentary form at an authorized place or ordering of a paid commercial service called the data safe. Submission via data box has the same effect as a signed written submission.

The data message consisting of an envelope containing the appropriate electronic mark (e-stamp) and timestamp (qualified stamp) can be only sent by an authorized person.

The aim of introducing electronic delivery via the data box is to reduce bureaucracy for citizens and the use of electronic delivery rather than the classical one.

Use of data boxes:

- G2G - for authorities compulsory,
- G2B - compulsory communication,
- B2G - compulsory communication, but for not-using the data box there is no punishment,
- C2G, G2C - for citizens not compulsory, everyone can choose a way of preferred communication,
- C2B, B2C, C2C, B2B - optional, it is not supposed to be used as a classical email communication.

The project worth 432 million CZK has its drawbacks. The director of the project claims that this system will save the authorities at least one third of the cost of postage and ideally - with internal computerization of the authorities - up to 64 per cent of the cost (sending of one data message is 18 CZK). However, most authorities still deal with the records service by using paper forms and thick binders. In practice this means that most offices have to print a delivered report and continue working with the document in a paper form. Therefore, the cost of office supplies increases. The cost increase also applies to the user - individuals, who must pay for the document conversion, it is 30 CZK per page.

Another disadvantage represents the existing legislation which defines the documents that must be sent in a paper form (letters) only. This includes for example construction documents or a court decision of 50 pages. Using of data boxes is not possible for obvious reasons.

There are other disadvantages, for example: safety concern, the installation program and the need to have an Intel processor.

3. The information system of the data boxes in numbers

On 27th February 2011 a total of 401 012 data boxes were made available, of which 18 332 individuals, 10 955 entrepreneurs, 7 676 public authorities and 371 185 legal entities.

The number of data messages sent from 1st November 2009 to 27th February 2011 is shown below.

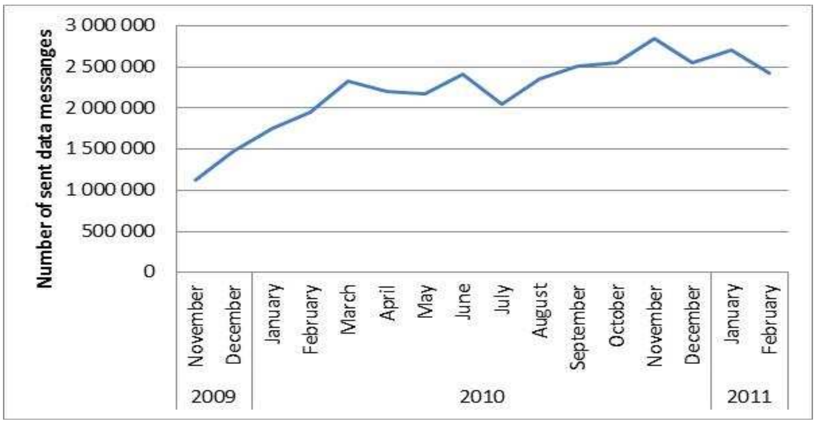


Fig. 1 Data messages sent from 1st November 2009 to 27th February 2011[Source: Czech post internal material]

From the total number of sent data messages was 96 per cent delivered to the logging system and 4 per cent was delivered through the legal fiction.

According to the nature of the message we can distinguish public data messages and postal data messages. The following number of data messages was sent among particular users of the system.

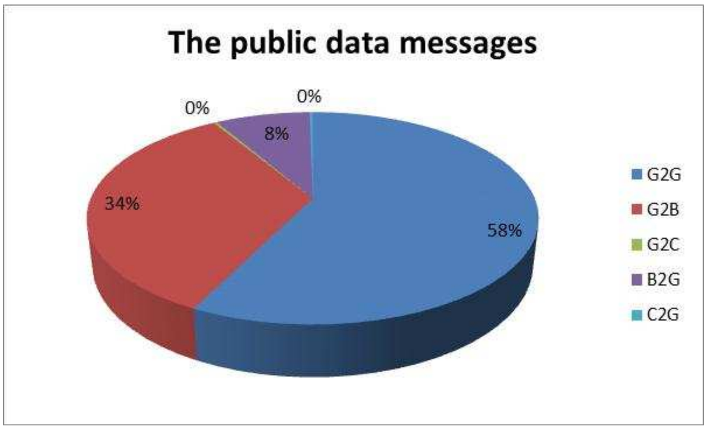


Fig. 2 Number of sent public data messages [Source: Czech post internal material]

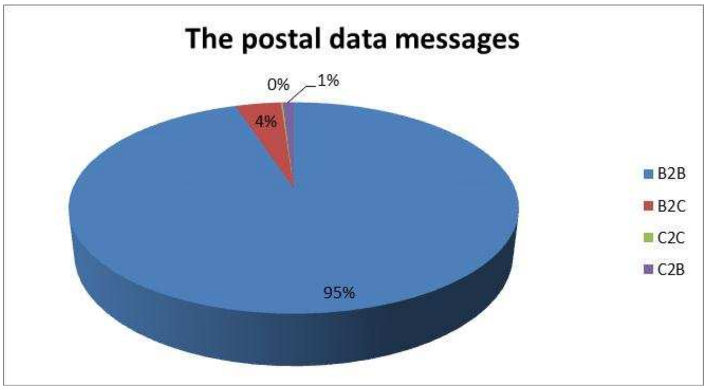


Fig. 3 Number of sent postal data messages [Source: Czech post internal material]

4. Conclusion

The provider of the information system of the data boxes is the holder of the postal licence – the Czech Post s. e. Transition from written to electronic communication involves reduction in yields. The Czech Post took into account the loss of revenues from letter items of 1,5 milliard CZK. However, the Czech Post has received from the state 432 million CZK. This amount covered the acquisition of the system, including flat payment and postal charges from the beginning to the end of April. The Ministry of Interior paid the Czech Post, s.e. a monthly flat fee of 15 million CZK excluding VAT for the operation of the system from 1st May 2010. And also for each the data box payable 18 CZK. Another income are revenue of communication between individuals and also revenue of additional services offered by the Czech Post, s.e. The loss of revenue is largely replaced by revenue from the operation of the system. The important advantage for the provider is winning a new market segment, i.e. electronic communication. That means that the transition to electronic form will reduce the revenues of the universal service provider, but the potential market (B2B and B2C segments) will be reduced for future alternative postal providers, too. The data boxes will be mandatory for lawyers and tax advisors from 1st July 2012. From 1st January 2013 can be holder of postal license a different company than the Czech post, s.e. However, it is expected that the provider of the information system of data boxes will be still Czech Post s.e.

References

[1] Konec obálek s modrým pruhem. *Ekonom*. 2009, 25, s. 34-35. ISSN 1210-0714.

[2] Ministerstvo vnitra [online]. c2010 [cit. 2011-02-27]. *Datové schránky*. Dostupné z WWW: <http://www.mvcr.cz/datove-schranky.aspx>.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Distributed Database Systems in the Dynamic Networks Environment

[*]Ján Janech
[*]University of Žilina, Faculty of Management Science and Informatics, Department of Software Technologies, Univerzitná 1, 010 26 Žilina, Slovakia, jan.janech@kst.uniza.sk

Abstract. This article describes a new architecture of a distributed database system. The architecture is designed especially for the dynamic networks environment. It explains problems with the architecture of a traditional distributed database system, and it offers a solution.

Keywords: dynamic network, ad-hoc network, distributed database system, system architecture, broadcast communication, query processing.

1. Introduction

This article aims at describing a new architecture of distributed database systems. This architecture specializes in operating in dynamically changing networks (such as mobile/vehicular Ad-Hoc networks (MANET/VANET), mesh networks or dynamically changing local networks). A system with such architecture is able to work even after some sites have been separated from the rest of the system, or after the network gets divided into several parts.

A database of parking places integrated into vehicles can serve as an example of such a system. Each vehicle acts like one site in a distributed database system. After the vehicle is completed and leaves the car factory, its inbuilt database is empty. After the system becomes activated, the vehicle starts to query the distributed database for a list of parking places periodically. When the vehicle receives some data, it saves them into its own inbuilt database. The sites providing these data could be either other vehicles or parking places.

Another example is a fragmented database stored in a group of PDA computers. If the PDAs are interconnected through Wi-Fi (managed or Ad-Hoc), bigger part of the database can be available for queries.

What is considered a very interesting possibility is a distributed database of local maps. The user of the system can acquire the map of their surroundings through the Ad-Hoc Wi-Fi network.

The goal of the article is not to evaluate existing solution. Instead it aims to identify problems of the classic architecture and to propose solutions for them.

2. The Architecture of DDBS

To demonstrate the problems of implementing the classic DDBS architecture in a dynamic networks environment, some of the DDBS architecture basics are presented below. They are not supposed to provide an exhausting description of the architecture, but to focus on the parts causing problems with the queries in dynamic networks.

Figure 1 shows the DDBS architecture from the viewpoint of data organization. It is a simple multi-layer model with four layers. Each one of the layers represents a certain view of the data themselves.

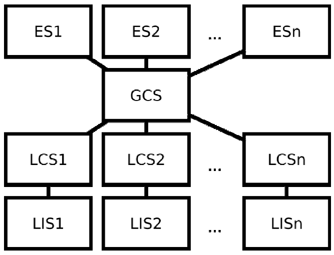


Fig. 1. DDBS reference architecture [4]

1. LIS (Local Internal Scheme) is a physical representation of the data stored in one site. It is an analogy of the internal scheme from centralized databases.
2. LCS (Local Conceptual Scheme) describes the logical organization of data in one site. It is used to handle the data fragmentation and replication.
3. GCS (Global Conceptual Scheme) represents the logical organization of data in the whole distributed database system. This layer is an abstraction of the fact that the database system is distributed.
4. ES (External Scheme) represents the user view into the distributed database. Each external scheme defines which parts of the database the user is interested in.

The fact that the user uses only the global conceptual scheme, and they do so by means of the views defined in the external scheme, assures that the user can manipulate with the data regardless of their position in the distributed database system. Therefore it is necessary to have a mapping from every local conceptual scheme to the global conceptual scheme. This mapping named GD/D (Global Directory/Dictionary) is defined as a part of the distributed database system. There has to be some way how to approach the GD/D as a whole, whether the GD/D is stored in one place or distributed throughout the system in any way [1].

This is not possible in a dynamic network. Either in an Ad-Hoc network, or in a dynamically changing managed network, there is no way to ensure communication between all sites in the system. Under these circumstances the GD/D cannot be used to locate data in a distributed system.

3. Solution

The biggest problem for deployment of distributed databases in the dynamic network environment is the necessity of the knowledge of all the sites in the system caused by GD/D. So the only way to make sure it is possible to use the distributed database system in the dynamic networks environment is to remove the GD/D from the system and replace it with a different principle. As it has been said already, the GD/D describes the mapping between the local and global conceptual schemes. Without the mapping the system does not know where the data are located and how to query them.

Using the GD/D in the dynamic networks environment is impossible because it requires knowledge of the whole system (global directory). In a dynamic network every site knows its immediate surroundings only. So querying of a distributed database is fairly limited in such environment. The only sites which can be addressed to with queries, except the one which requests information, are those in the immediate surroundings in the network. So the system naturally creates virtual clusters of sites that can communicate with each other. The clusters might overlap, so each of the sites of the cluster can communicate with another set of sites.

This implies a possibility to move the directory from the global level to the cluster level. This principle may be called Cluster Directory/Dictionary (CD/D). In this way the directory contains only the mapping of the global conceptual schemes of the sites accessible to the local conceptual schemes exclusively. However, there might be problems in this system due to clusters overlapping. Therefore the local conceptual scheme of each site has to be mapped onto the global conceptual scheme in several CD/D. So the question remains how to store the CD/D in a cluster. The cluster has no central site as it is strictly a peer-to-peer system, so the only possibility from the ones described in the previous chapter is a local directory. Each site has a mapping from the global conceptual scheme onto its own local conceptual scheme. The CD/D then consists of all of the local mappings in the cluster.

Executing local queries in such a system is very simple because all of the accessible data and all of the mappings are available, so there is no need to communicate with the other sites in the cluster. But since the queries are normally done globally all over the whole accessible part of the database [5], the communication between the sites in the cluster is necessary.

4. Broadcast Query Principles

In the new distributed database system architecture queries are sent via the network as broadcast or multicast messages. This way a query site does not know who is to process the query, or if anyone is to process it at all. This restricts the query process in two ways:

- Asynchronous communication is required. There is no way how to determine the number of data sites that will respond to the query. Therefore, the querying site has to process responses after they come, instead of waiting for the known number of responses, like it would have to do in the traditional architecture.
- The query syntax needs to be well structured, so it can be easily divided into smaller subqueries. The data site which has received the query can only contain a part of the data required for the response (e.g. when the data are horizontally fragmented). So it has to send one part of the query back to the cluster.

The UML sequence diagram in Figure 2 shows an example of the communication between querying sites and three data sites. A query in a form of an expression formatted in a certain way is sent by means of a broadcast message into the whole cluster. The data site 1 has some fragments of the queried data, thus it returns them directly to the query site. The data site 2 does not have any data, so it stops the query processing immediately. The data site 3 also has some data; therefore it returns them to the query site. It is up to the query site to decide when to stop receiving answers. It is important to be aware of the fact that the connection between any of the sites and the rest of the cluster can be interrupted at any time of processing the query.

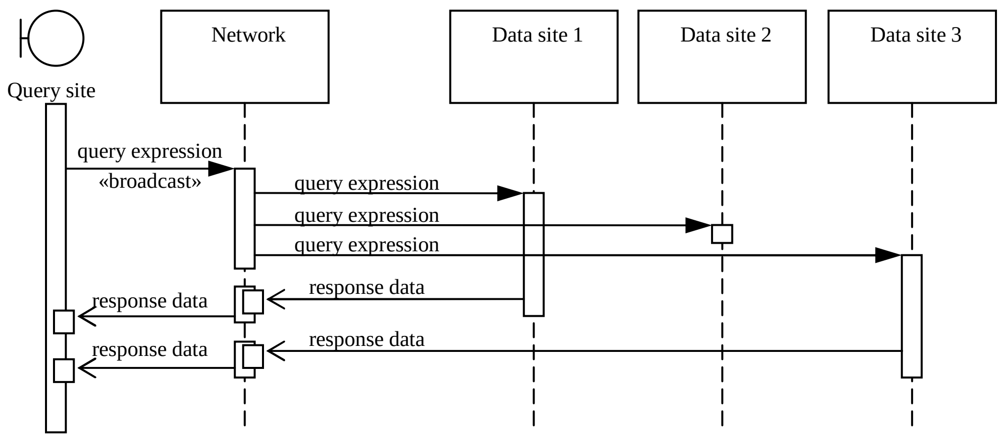


Fig. 2. Processing a query in the cluster

Another question is formatting the query. A classic SQL query is not structured very well, and dividing such a query into smaller parts is not a simple task. However, the relational algebra allows doing so. The only problem is the bad format of the query, which resembles natural language rather than a relational algebra expression. Another possibility is using a relational algebra expression directly, in a form of lambda calculus [2][3]. The lambda function syntax allows to extract any part of the expression, and to evaluate it individually. In this way the user can easily create queries, send them over the network and divide them if necessary.

5. Limitations of The New Architecture

The architecture introduced in this article does not intend to be a universal database system. It is a specialized solution for the implementation of distributed database systems in the dynamic networks environment. Therefore it has some problems which may prevent using the architecture in some situations.

It is impossible to do a lot of optimizations due to the fact that the GD/D does not exist. The sites do not even have access to the whole CD/D. Broadcast messages can also overload the network in case many queries are processed at the same time.

In most situations the querying sites do not have access to the whole database.

For the query site there is no way of how to find out whether it has already received all the data it has requested or if it has to wait longer.

It is impossible to provide referential integrity because the distributed database can be partially or even fully inaccessible.

This architecture allows just the read-only access to the distributed database. Destructive operations can be executed only locally on the data sites.

These limitations are results of the environment rather than the defects of the architecture. Thus the architecture can only be used when these problems are not limiting.

6. Conclusion

We can find many options how to use database systems in dynamic network environment. Ability to share public data stored in internal database of car (e.g. information about parking places), to provide data specific to geographic location of the user (sharing maps) etc.

The present architectures are not suitable for such an environment. The proposed architecture removes their limitations; however it brings few of its own as a result of principles of dynamic networks.

In a short time a prototype of the system will be created so that its behavior can be tested experimentally. It would be possible then to compare this new architecture with the existing ones.

Acknowledgement

This research was supported by the Centre of excellence for intelligent transport systems and services, ITMS 26220120028, University of Žilina.

Operačný program VÝSKUM a VÝVOJ

ERDF - European Regional Development Fund

The project is being cofinanced by European Community

EURÓPSKA ÚNIA ŠTRUKTURÁLNE FONDY

References

[1] DATE, C. J. *An introduction to database systems*. Addison–Wesley, 6th edition, 1995.

[2] HILLEBRAND, G. G., KANELLAKIS, P. C., MARISON, H. G. Database query languages embedded in the typed lambda calculus. *Proc. of 8th IEEE Symphosium on Logic in Computer Science*, Montreal, Canada, June 1993, pp 332-343.

[3] MERUNKA, V. *Objektové modelování*. Praha : Alfa Nakladatelství, s.r.o., 2008.

[4] ÖSZU, M. T., VALDURIEZ, P. *Principles of distributed database systems*. Pearson Education, 2nd edition, 1999.

[5] SOKOLOVSKÝ, P., POKORNÝ, J., PETERKA, J. *Distribuované databázové systémy*. Academia, Praha, 6th edition, 1992.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Simulation Tools for Logical Systems Education – ConveyorFlash Composer

[*]Adam Jaroš
[*]University of Žilina, Faculty of Management Science and Informatics, Department of Technical Cybernetics, Univerzitna 2, 01026 Žilina, Slovakia, Adam.Jaros@fri.uniza.sk

Abstract. Simulation tools are favorite in now days. In computer world are used most widely. One on most important factor in spreading those tools is access to Internet supported with fast connection. There are quite a few environments for wide of use. They will be shortly presented on the beginning. A most important of them are Adobe Flash & Adobe AIR, SUN Java & JavaFX and Microsoft Silverlight. The newest is HTML5 Canvas as a new incoming standard defined by W3C Consortium. All those tools can be used for implementation of classic programs used on different platforms. Simulation tools can be then used on-line and off-line as well. Historically it was Java which brought this approach as first in wider scope. At present time it is Adobe Flash platform which stand for a new standard of creative tools for on-line publishing and content delivering.

In this paper I present the result of over two years of my research. Paper deals with design and implementation of a new tool for support of Logical systems education. Presented work is a part of a bigger concept let us say concept of tools for design of Moore and Mealy automata.

Keywords: ConveyorFlash Composer, on-line tools, Flash platform, finite state automata, logical systems, education, Moore, Mealy.

1. Purpose of Simulation Tools in Education Process

Here we are not speaking about all aspects. But to present some witch I consider as most important from my eleven years of experience as a teacher and at the same time from responses from my students. To second and last point I give highest priority.

Here is a list of aspects I take into consideration when I develop tools:

- visual demonstration for quicker and better understanding, less abstract, more practical,
- individual experience, area for exploration and investigation,
- on-line tools - immediate accessibility (no installation, only web browser), help as HTML & video, latest version, sharing of tasks and results, price, security,
- good design get us ability use tools off-line and on-line, synchronization at active internet connection,
- main goal is increase interest and concern of students.

2. Principles of Simulation in Logical Systems

As a teacher of subject Logical systems it seems to me that content of lectures and practices can be divided to three parts. I always try to make a clear coherency between theoretical and practical experiences for students.

First basic level of education contains logical functions (AND, OR, NOT, XOR), truth table, Karnaugh map, optimal-minimalization, drawing of electrical scheme.

Second intermediate level of education contains analysis of scheme, dynamic of combinational circuit - hazards, design of automata.

Third advanced level of education contains optimal design of synchronous finite state automata, design of asynchronous automata.

2.1. Using of Results In the Subject Logical Systems Education

My developed tools fall under last category. Modern computers work in principle as a complex state machine. Intention of development is to help students not only of technical and informatics specializations but also of a management science to better understand the principle of operation of modern computers and aspects of their design.

2.2. Existing Situation

There are plenty of complex professional tools available to simulate all three levels. All steps of design of electronics devices can be simulated. Almost all software tools are for off-line use only. On-line tools are very rare, to name some of them LogiFlash, Hades and LogicLAB.

Besides software tools for simulation and practical experiences there are plenty of electronic kit which contains real active and passive electronic circuits. We use them on practices to verify our designs.

3. New Trends in On-line Publishing - A State of Art

In a last few years we observe boom of on-line publishing tools. Each one wants to become most widespread and de facto standard. I believe that fair competition helps us in this long race run considerably.

Let us look shortly on most advanced platforms for implementation of simulation tools:

- Microsoft® platform based on Silverlight®, use development tools based on .NET Framework and languages as C# and Visual Basic for .NET, applications can work both on-line and off-line, multiplatform is supported too,
- Adobe Flash® platform based on Adobe AIR® and Flash Player®, in 2008 Flash platform become open source (Flex SDK), use EMACS programming language ActionScript 3.0, can run on-line and off-line,
- SUN platform based on Java and JavaFX®, sophisticated development tools, native language is Java, applications can work both on-line and off-line, software and hardware multiplatform is well supported,
- Canvas prepared by W3C Consortium as a part of incoming HTML5 web standard, many new features are already implemented in major web browsers.

3.1. Existing Situation And the Future

Up today most widespread are on-line tools based on Flash then Java and Silverlight. In coming years I believe will come forward a HTML5 and JavaScript with canvas support as a native part of all modern web browsers. There are quite a few multiplatform web browsers today and some of them aspire to reach all platforms – PC, tablet, mobile phone, multimedia devices and television. This is a new challenge for the rest of them.

4. Applying a New Technologies in Creative Process

Here I share my own view based on three years of development of on-line tools. Essential requirements for design of simulation tools can be defined as:

- availability of development tools, tutorials and literature for target platform,
- simplicity and short time to learn new techniques and programming language,
- utilize already known is advantage, for example C++,
- to have verified all complex algorithms on well-known platforms, which stands up behind tools like presented one.

4.1. What To Choose

The most advanced and reliable platforms for now days are Adobe Flash® and Adobe AIR® from Adobe. For development we can use free or commercial tools, both of high quality with excellent documentation and books [1], [2]. For implementation of on-line tools I chose Adobe Flash platform.

5. Concept of Proof - AutomatFlash Composer

In 2008 was initiated development of AutomatFlash Viewer to visualize finite Moore and Mealy automata. Later was added simple interactivity. This was a first idea of on-line tools for subject Logical systems. My researches about availability prove that there is a place for such tools.

Short history of on-line simulation tools for Logical systems:

- 1st version of project, shown on Fig. 1: consists of two Flash applications - Conveyor strand and AutomatFlash Viewer, represents static task-example where types and number of products to detect and visual representation of automata are fixed, started in November 2008, results was presented in [3],
- 2nd version of project: consists of two independent Flash applications - ConveyorFlash Composer and AutomatFlash Composer, started in December 2010, fully editable.

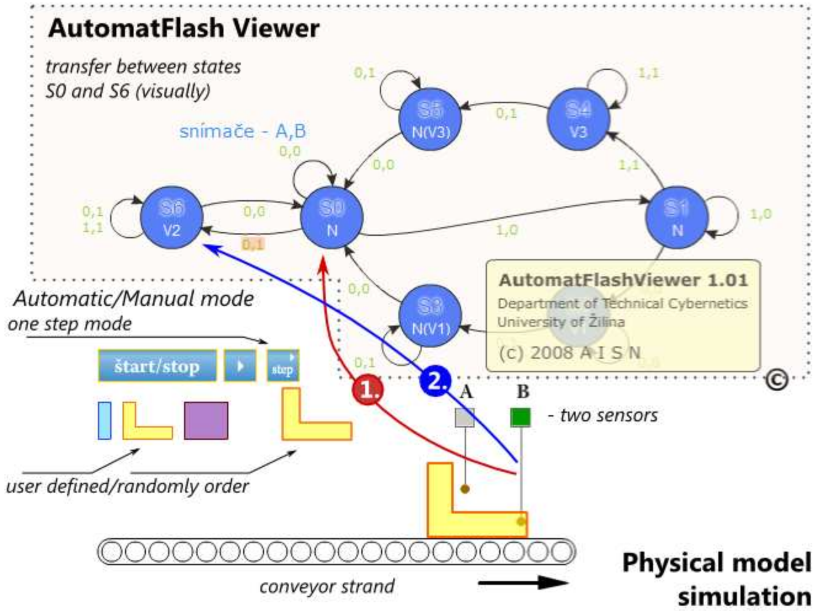


Fig. 1. Figure shows a static task with basic interactivity - Conveyor strand and AutomatFlash Viewer.

6. Practical Showcase - ConveyorFlash Composer

To create such a complex tool is necessary have to good design. It is important to define effective data structures and file formats. For web there are two possibilities to avoid database dependency. They are XML and JSON file formats. As we observe trends of newest devices like iPhone 2 and tablet they have very effective controls. I tried to implement well designed and at the same time simple controls too. Some aspects of design are shown on Fig. 2, Fig. 3 and Fig. 4.

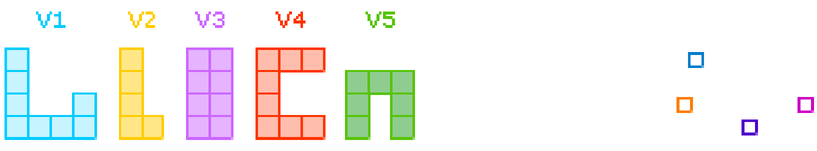


Fig. 2. Figure show a visual representation of XML data file from Figure 3 – shapes (left) and optical sensors (right).

```
<?xml version="1.0" encoding="utf-8" ?>
- <conveyor>
  - <shapes>
      <shape def="3,2,2,3,3,0,2,0,1,0,0,0,3,3,3,1" color="0x00CCFF" bgcolor="0xC7F4FF" shadow="0x00CCFF" enabled="1" name="V1" />
      <shape def="3,1,3,0,2,0,1,0,0,0" color="0xFFCC00" bgcolor="0xFFE787" shadow="0xFFCC00" enabled="1" name="V2" />
      <shape def="3,1,2,1,1,1,0,1,2,0,1,0,0,0,3,0" color="0xCC66FF" bgcolor="0xE6B3FF" shadow="0xCC66FF" enabled="1" name="V3" />
      <shape def="0,0,1,0,2,0,3,0,3,1,3,2,0,1,0,2" color="0xFF3300" bgcolor="0xFFBEAD" shadow="0xFF3300" enabled="1" name="V4" />
      <shape def="3,2,2,2,1,2,1,1,1,0,2,0,3,0" color="0x56C406" bgcolor="0x8FCC8F" shadow="0x8DC466" enabled="1" name="V5" />
    </shapes>
  - <sensors>
      <sensor x="0" y="2" color="0xFF7C00" active_color="0x63F700" enabled="1" name="A" />
      <sensor x="1" y="0" color="0x007CCC" active_color="0x63F700" enabled="1" name="B" />
      <sensor x="6" y="3" color="0x4F00CD" active_color="0x63F700" enabled="1" name="C" />
      <sensor x="11" y="2" color="0xCC00C7" active_color="0x63F700" enabled="1" name="D" />
    </sensors>
    <special_cases />
  - <about>
      <created>2011-01-18 16-01-53</created>
      <author contact="">Adam Jaroš, PhD.</author>
      <composer homepage="http://frtk.fri.uniza.sk/~aj/AutomatFlash-Composer.php">ConveyorFlash Composer</composer>
    </about>
</conveyor>
```


Fig. 3. Example of data structure of ConveyorFlash Composer XML file.

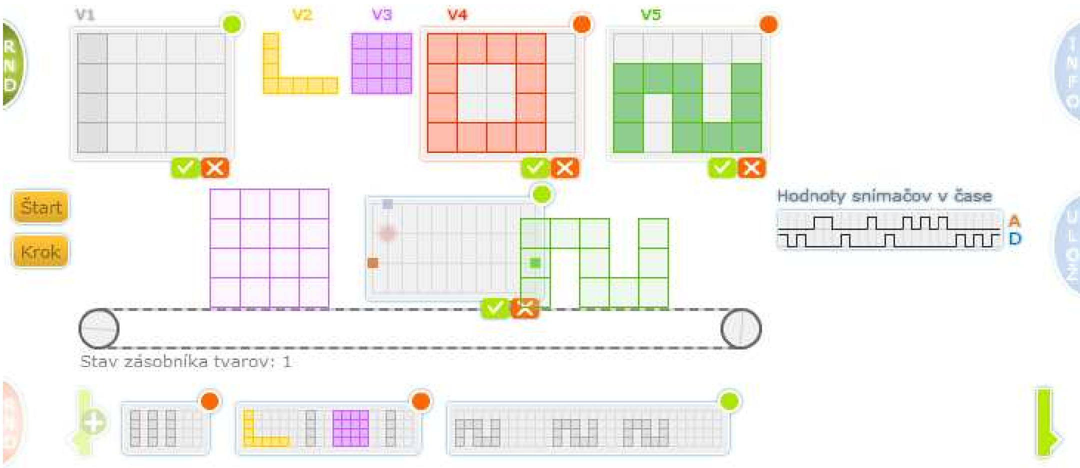


Fig. 4. Figure show a ConveyorFlash Composer at operation.

7. Conclusion

An essential research results since 2009 to 2011 has been presented here, mainly concept of proof for AutomatFlash Composer and practical showcase of ConveyorFlash Composer.

I believe that project at finish state will represent a unique tool for teachers and students in education of basic of logical systems.

7.1. What's Next

Main goal is to finish second part of AutomatFlash Composer for finite state automata design and create documentation. Develop more classes of tasks (already in preparation), for example: elevator, car-parking and green-house also available for students as diploma theses.

Acknowledgement

This work was supported by grant VEGA 1/1099/11: Modeling and simulation of Dynamical Interaction in Environment of Driver/Car/Traffic Situation.

References

[1] SHUPE, R., ROSSER, Z. *Learning ActionScript 3.0*. 2nd ed. O'Reilly, ISBN 978-1-449-39017-4, 2010.

[2] BRAUNSTIEN R. *ActionScript 3.0 Bible*. 2nd ed. Wiley, ISBN 978-0-470-52523-4, 2010.

[3] JAROŠ, A. *Online education tools a state of art*. TRANSCOM 2009, June 22-24, 2009, University of Žilina. Slovak Republic.

[4] ADOBE INC. On-line documentation: http://help.adobe.com/cs_CZ/flash/cs/using/index.html (date 10. 3. 2011)

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

The Principle of Canonical Correlation Analysis and Application for 2D Face Recognition

*Dominik Jelšovka
*University of Zilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitna 1, 01026 Zilina, Slovakia, jelsovka@fel.uniza.sk

Abstract. This paper provides an example of face recognition using CCA method. We applied one dimensional canonical correlation analysis (CCA) to image processing and biometrics. CCA is a powerful multivariate analysis method. For two sets of variables, CCA is to construct the CCA subspace to mutually maximize the correlation between these two sets variables. The goal is to extract the important information from the face data, to represent it as a set of new canonical variables. Canonical correlation analysis deals with the association between composites of sets of multiple dependent and independent variables. Our algorithm has been tested on database 10 subjects (faces). Test recognition rate for our database was 100%.

Keywords: canonical correlation analysis, face recognition, canonical correlation coefficient.

1. Introduction

In the recent years canonical correlation analysis arouse the growing interest of experts in biometrical technologies of people recognition, as a method which helps to relate sets of observations describing different aspects of appearance. Canonical correlation analysis is first proposed by Hotelling [1] in 1936. CCA represents a high-dimensional relationship between two sets of variables with a few pairs of canonical variables. It was intended to describe relations between two sets of one dimensional data sequences. The CCA method has been widely used in several fields such as signal processing [2], medical studies, and pattern recognition [6,7]. CCA can simultaneously deal with two sets of data, compared to principle component analysis (PCA) and linear discriminant analysis (LDA). In the case of a single feature set, CCA directly takes another set of variables as class labels. For multi-modal features, CCA can be used for feature level fusion. Proper feature fusion can increase recognition rate. Fast development of computer technologies increased memory capacity and processing speed, wide use of software packages for digital image processing (e.g.“MATLAB”, “Lab View”, “Statistics”), application of mathematical modeling has promoted application of this method in processing of the multidimensional data. Such data may include face images and images of human gaits, hand gestures, etc. CCA will find pairs of directions that yield maximum covariance resp. maximum correlation between the two random variables x, y.

2. Canonical correlation analysis (CCA)

CCA is a powerful multivariate analysis method [3]. It has various applications in pose estimation [8] and face matching [9]. For two sets of variables, CCA is to construct the CCA subspace to mutually maximize the correlation between these two sets variables. Canonical correlation analysis deals with the association between composites of sets of multiple dependent and independent variables. In doing so, it develops a number of independent canonical functions that maximize the correlation between the linear composites, also known as canonical variates, which are sets of dependent and independent variables.

2.1. What is CCA

Canonical correlation analysis can be defined as the problem of finding two sets of basis vectors, one for x and the other for y, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Given two zero-mean random variables $x \in R^p$ and $y \in R^q$, CCA finds pairs of directions w_x and w_y that maximize the correlation between the projections $x = w_x^T x$ and $y = w_y^T y$ (in the context of CCA, the projections x and y are also referred to as canonical variates). More formally, CCA maximizes the function:

$$\left\{ \rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[w_x^T xy^T w_y]}{\sqrt{E[w_x^T xx^T w_x]E[w_y^T yy^T w_y]}} \right., \quad (1)$$

$$\left\{ \rho = \frac{w_x^T C_{xy} w_x}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \right. \quad (2)$$

,

whereby E denotes the empirical expectation, $C_{xx} \in R^{p\times p}$ and $C_{yy} \in R^{q\times q}$ are the within-set covariance matrices of x and y, respectively, while $C_{xy} \in R^{p\times q}$ denotes their between-set covariance matrix [8].

2.2. Calculating canonical correlations

Consider two random variables x and y with zero mean. The total covariance matrix

$$\left\{ C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = E\left[\begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}^T \right] \right. \quad (3)$$

is a block matrix where C_{xx} and C_{yy} are the within-sets covariance matrices of x and y respectively and $C_{xy} = C_{yx}^T$ is the between-sets covariance matrix. The canonical correlations between x and y can be found by solving the eigenvalue equations

$$\begin{aligned} C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \hat{w}_x &= \rho^2 \hat{w}_x \\ C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} \hat{w}_y &= \rho^2 \hat{w}_y \end{aligned} \quad (4)$$

where the eigenvalues ρ^2 are the squared canonical correlations and the eigenvectors $\hat{w}_x$ and $\hat{w}_y$ are the normalized canonical correlation basis vectors.

Instead of the two eigenvalue equations (4), we can formulate the problem in one single eigenvalue equation:

$$\left\{ B^{-1} A \hat{w} = \rho \hat{w} \right. \quad (5)$$

where

$$\left\{ A = \begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}, B = \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}, \hat{w} = \begin{pmatrix} \mu_x & \hat{w}_x \\ \mu_y & \hat{w}_y \end{pmatrix} \right. \quad (6)$$

It can be shown that the solution $W= (\omega_x^T ,\omega_y^T)^T$ amounts to the extremely points of the Rayleigh quotient:

$$\left\{ r = \frac{W^T A W}{W^T B W} \right. \quad (7)$$

As a subspace learning method, CCA is inclined to overfit to the training data, especially when the sample size is small [3].

3. Experimental results

In this section, some experimental results with canonical correlation analysis method will be presented. The first experiment shows CCA subspace. We apply canonical correspondence to find cross-correlation between the same faces. The correlation between two faces (signals) is a standard approach to feature detection [6,7] as well as a component of more sophisticated techniques. The result cross-correlation is extraction face features such as nose, eyes, mouth and the result is shown in the Fig.1. Then CCA subspace is formed just these face features. This extraction is similar as in eigenfaces method, which for face recognition using the face features.

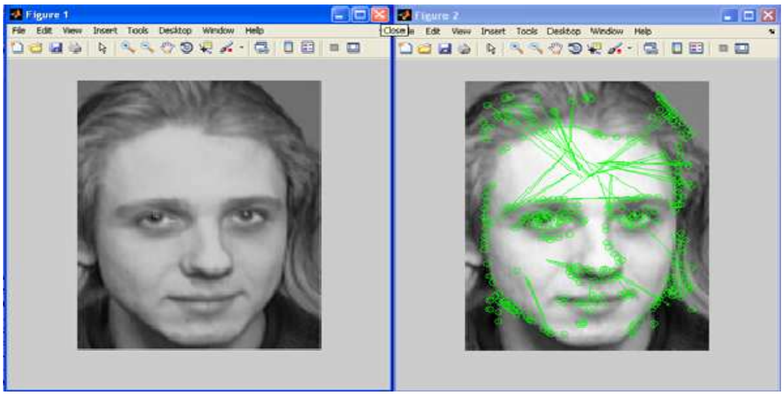


Fig. 1. Extraction face features using cross-correlation.

The next experiment shows simple face recognition using canonical correlation method. The proposed method has been tested on two input face images. We applied CCA method on 2D face images from our training database. We applied canonical correlation analysis for face recognition on our database in Matlab. Our experiment can describe by symbolic notation

$$[W_x, W_y, r] = cca(X, Y) \tag{8}$$

where X,Y are vectors representing 2D face images. The output of CCA is a pair of direction ω_x and ω_y to maximize the correlation between the two face images and the canonical correlation coefficient r. We determined the canonical correlation coefficient using equation (1), (2) and the correlation coefficient r determines correlation match between face images. The experimental result is shown in the next Fig.2. The figure 2 shows face recognition rate, where r is the correlation coefficient acquires value from interval (-1,1). Maximum correlation between two face images is while correlation coefficient has value 1 or -1. Then the faces are identically.

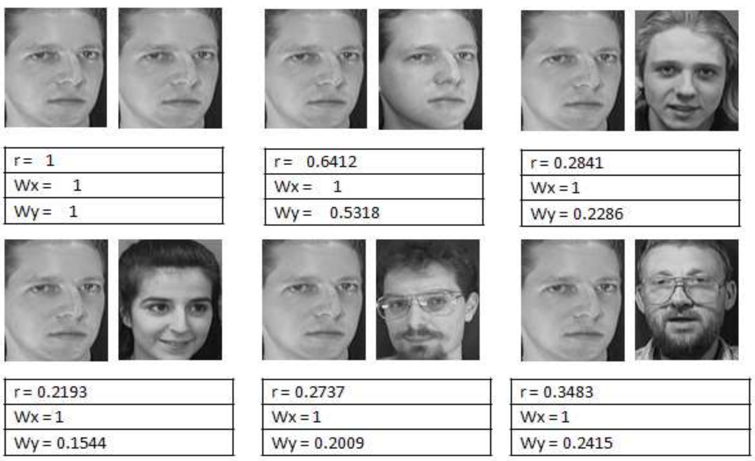


Fig. 2. Face recognition results using CCA method.

4. Conclusion

Although little known in the field of pattern recognition and signal processing, CCA is a very powerful and versatile statistical tool that is especially well suited for relating two sets of measurements. CCA can also be regarded as a linear feature extractor. In section 3, we applied CCA for face recognition of 2D images and used cross-correlation for face feature extraction. The experiments show that the CCA method have achieved relatively good recognition rate on database with 10 subjects.

Acknowledgement

This paper has been supported by the Slovak Scientist project VEGA grant agency, Project No. 1/0570/10 "Algorithms research for automatic analysis of multimedia data" and Slovak Scientist project VEGA grant agency no.1/0655/10, "Algorithms for capturing transmition of 3D images for 3D IP television".

References

[1] HOTELLING, H. *Relations between two sets of variates*. Biometrika 28, 321, 1936.

[2] BORGA, M. *Canonical correlation a tutorial*. January 12, 2001, http://www.imt.liu.se/~magnus/cca/tutorial/tutorial.pdf.

[3] MELZER, T., REITER, M, BISHOF, H. *Appearance model based on kernel canonical correlation analysis*. Pattern Recognition and Image Processing Group, Vienna University of Technology, 183, Austria, 2008.

[4] HARDON, R, SZEDMAK, R., SHAWE, R. J. *Canonical correlation analysis: An overview with application to learning methods*. 2639-2664, 2004.

[5] HARRIS, C. G., STEPHENS, M. J. *A combined corner and edge detector*. Forth Alvey Vision Conference, University of Manchester, England, 15: 147-151, 1992.

[6] LEI, Z., BAI, R. H., LI, S. Z. *Face shape recovery from a single image using CCA mapping between tensor space.* In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008.

[7] GONZALES, R.C., WOODS, E.R. *Digital Image Processing* Reading, Massachusetts: Addison-Wesley, 2002.

[8] YANG, W. LEI, Z. SANG, J. *2D-3D face matching using CCA method.* National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, Beijing, 2008.

[9] YI, D. LIU, R. CHU, R. LEI, Z. *Face matching between near infrared and visible light images.* 523-530, 2007.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Depth Map Recovery for Stereo Vision Using Segmentation Method

*Patrik Kamencay, *Martin Breznan, *Peter Lukac, *Miroslav Fric, *Martina Zachariasova
*University of Zilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitna 1, 01026 Zilina, Slovakia, {Patrik.Kamencay, Martin.Breznan, Peter.Lukac, Miroslav.Fric, Martina.Zachariasova}@fel.uniza.sk

Abstract. Stereo vision refers to the ability to infer information on the 3D structure of scene from two or more images taken from different viewpoints. This paper describes procedure for depth map creation based on rectified stereo images and segmentation algorithm belief propagation (BP). The depth recovery is important part of image analysis, which involves extracting information about shapes, textures or distances and it is optimized by using the belief propagation techniques based on Markov Random Fields (MRFs). Very necessary step to depth map creation is camera calibration. Calibration of the stereo camera system consist from a two parameters: intrinsic parameters, which characterize the transformation mapping an image point from camera to pixel coordinates in each camera and extrinsic parameters, which describe the relative position and orientation of the two cameras.

Keywords: stereo vision, depth map, disparity map, stereoscopic cameras, calibration, belief propagation.

1. Introduction

Stereo vision system is a set of two or more cameras, which extract depth information of a 3D scene as viewed from different vantage points and stereo vision is an imaging technique that can provide full field of view 3D measurements in an unstructured and dynamic environment. It is an area that deals with the reconstruction of depth information (third dimension) of two or more two dimensional images [1]. The inputs to stereo vision system are two 2D images. In a stereo vision system, cameras are horizontally aligned. Fig. 1 shows an example of stereo vision system with two cameras. By using modern stereo vision systems and algorithms, is possible to estimate the depth of most visible structures.

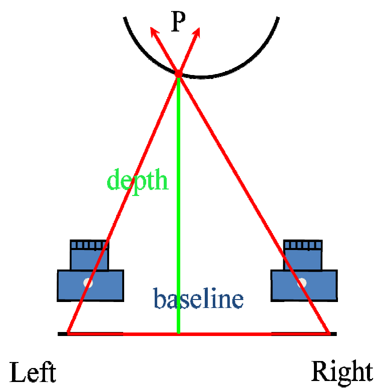


Fig. 1. Stereo vision system. With known correspondence between image pair, we can found the depth information from their disparity.

The fundamental idea behind stereo computer vision is the difference in position of a unique point in two different images. When a distant object is viewed by two cameras positioned in the same orientation but separated by a distance known as the baseline, that object will appear in a similar position in both images. As the object moves closer to the camera(s), the relative position of object will change, and the positions in each image will move away from each other. In this way, is possible to calculate the distance of an object, by calculating its relative positioning in the two images. This distance between the same object(s) in two images is known as disparity [2].

The imaging points of 3D point P in image coordinate system of the left and right cameras are $p_l(x_l,y_l)$ and $p_r(x_r,y_r)$ respectively. Because the left and the right camera image planes locate in the same plane, the y coordinates in these two images are the same $(y_l=y_r)$, and the disparity equals to the difference between the horizontal coordinate (x_l-x_r). According to the geometry of stereo vision imaging, you can easily find the world coordinates of point P:

$$\left\{ X = \frac{b(x_l + x_r)}{x_l - x_r}, Y = \frac{b(y_l + y_r)}{x_l - x_r}, Z = \frac{bf}{\underbrace{x_l - x_r}_{d}}, \right. \tag{1}$$

where Z is distance along the camera Z axis, f is the focal length, b is the baseline and d is disparity. Triangulate on two images of the same scene point to recover depth.

2. Stereo camera system

Stereo camera system will be based on extracting dense depth information from the input image pair. These images are captured using a pair of digital cameras – left image is an color image captured by the left camera and right image is an color image captured by the right camera. The left and right images are rectified and color segmented. Then we apply a stereo matching algorithm on the rectified left and right images to obtain an disparity map. This reconstruction algorithm is base on belief propagation techniques.

2.1. Stereo matching algorithm

In this section, we describe a 3D reconstruction algorithm that input images first segments and then find the same features in the left and right images.

- We first mathematically remove radial and tangential lens distortion – this process is called camera calibration, then we segments an image into regions using belief propagation.
- Finding all correspondence (matching point) in the left and right camera view, a process known as correspondence.
- The output of stereo matching algorithm is a disparity map, where the disparities are the differences in x coordinates on the image planes of the same feature viewed in the left and right cameras.

2.2. Belief propagation method

Segmentation is application dependent because it needs image interpretation. It is a process that allows to interpret spatially close parts of the image as objects. Regions are important building steps towards segmentation and object is everything what is of interest in the image, the rest of the image is background [3]. The belief propagation (BP) segmentation algorithm Fig. 2 passes messages throughout a graphical model via a series of messages sent between neighboring nodes. Basic structure BP algorithm is created from vertices and edges [4, 5].

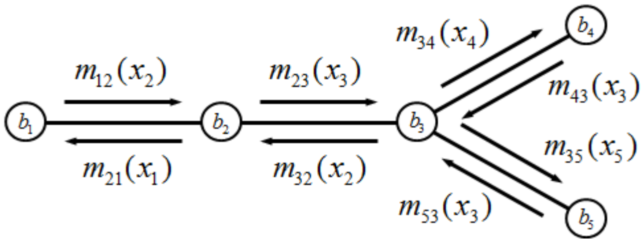


Fig. 2. Basic structure of Belief Propagation algorithm.

- Initialize all messages with a uniform distribution:

$$m_{ji}(x_i) = \sum_{x_j \in X_j} \phi_j(x_j, y_j)\psi_{ji}(x_j, x_i) \prod_{k \in N(j) \setminus i} m_{kj}(x_j), \tag{2}$$

- Update all messages iteratively.
- Compute passes messages – beliefs:

$$b_i(x_i) = \phi_i(x_i, y_i) \prod_{j \in N(i)} m_{ji}(x_i), \tag{3}$$

where $\phi_j(x_j, y_j)$ and $\psi_{ji}(x_j, x_i)$ is compatibility between states and observed values or neighboring vertices *i* and *j*.

3. Experimental results

In this section, some of the obtained experimental results rectifying, matching points and generating disparity map will be presented. The proposed architecture Fig. 3 has been tested on two input images. In our experiment, we combine image information with the pixel disparities to get a best result for finally disparity map. We segment the reference image (in our case, the left image) using a segmentation method belief propagation. Then, for each segment we look at the associated pixel disparities.

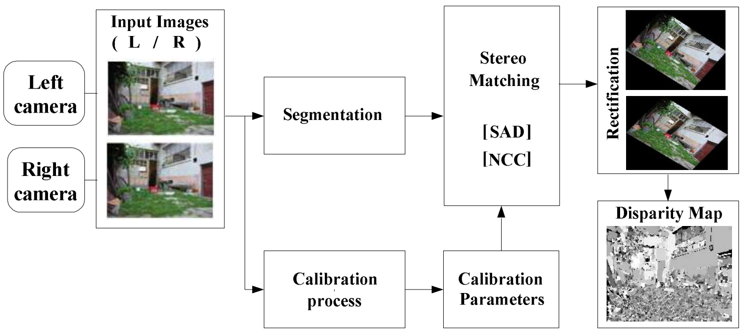


Fig. 3. Architecture for depth map recovery.

The detection feature point is based on the Harris principles [6] and then these points must be matched. Match points can be obtained using correlation approach along the epipolar line. The matching methods include Normalized Cross Correlation [9] and Sum of Squared Differences [9]. Epipolar geometry determines a pair relative orientation [7, 8]. Next step is image rectification. It is a transformation which makes pairs of conjugate epipolar lines Fig. 4(a) become collinear and parallel to the horizontal axis (baseline). Finally, dense matching is performed, where a disparity map is obtained. A disparity map codifies the distance between the object and the camera – closer points will have maximal disparity and farther points will get zero disparity.

For both objects, 200 image features were extracted using the Harris corner detector [6]. The result obtained can be observed in Fig. 5(a). After, both stereo pairs were rectified using the algorithm presented in [8]. As observed in Fig. 4(b) rectified image. Then, dense matching was performed using Stereo matching algorithm.

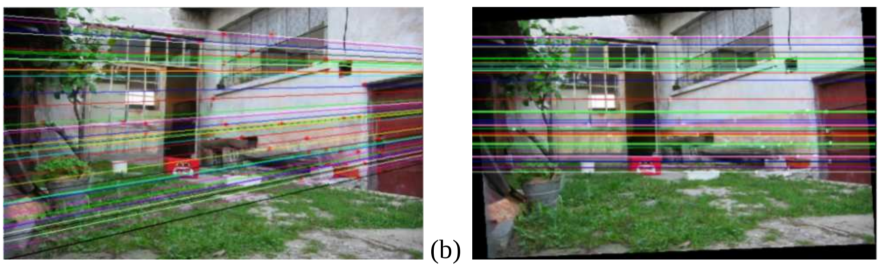
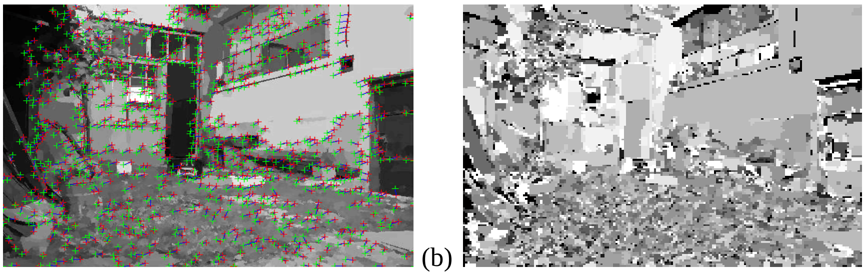
(a)  (b)

Fig. 4. Rectification results for stereo images. (a) Epipolar geometry (b) image rectification.

The disparity map is shown in Fig. 5(b), where brighter areas represent objects that are closer to the camera, darker areas further away. If we have found corresponding points between image pair, we can recover the depth information from their disparity.



(a) (b)

Fig. 5. Result of the feature points matching for both objects considered: Green crosses represent the matched feature points of the first image and the red crosses represent the correspondent matched feature points of the second image (a). Disparity map (b).

4. Conclusion

In this paper, the method for recovering depth map using segmentation method belief propagation from two input images acquired by a stereo camera system was presented. From the experiment result we can see a disparity map, which is presented displacements between two images and finally used to estimate the depth value. A Matlab environment has been used for final realization of all experiments. The applications of recovery methods are very useful in different sphere, for example detection and identification objects. Future task we could speed up computation time and improve precision of BP algorithm.

Acknowledgement

This paper has been supported by the Slovak Scientist project VEGA grant agency, Project No. 1/0655/10 “Algorithms for capturing, transmition and reconstruction of 3-D image for 3-D IP television”.

References

[1] SEXTON, I., SURMAN, P. *Stereoscopic and autostereoscopic display systems*. IEEE Signal Processing Magazine, vol. 16, no. 3, pp. 85-99, May 2000.

[2] SUMAN, P.; SEXTON, I., HOPF, K., WING KAI LEE, BUCKLEY, E., JONES, G., BATES, R. *"European Research into Head Tracked Autostereoscopic Displays,"* 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008, vol., no., pp.161-164, 28-30 May 2008.

[3] BENCO, M., HUDEC, R. *The advanced image segmentation techniques for broadly useful retrieval in large image database*., NSSS IX, Tatranske Zruby, Slovak Republic, pp. 40-44, May 2006, ISBN 978-80-8040-344-7.

[4] GUAN, S., KLETTE, R. *Belief propagation on edge images for stereo analysis of image sequences*. In Proc. Robot Vision, LNCS 4931, pages 291-302, 2008.

[5] GUAN, S., KLETTE, R., WOO, Y. W. *Belief propagation for stereo analysis of night vision sequences*. In Proc. PSIVT, LNCS 5414, pages 932-943, 2009.

[6] HARRIS, C. G., STEPHENS, M. J. *A combined corner and edge detector*. Forth Alvey Vision Conference, University of Manchester, England, 15: 147-151, 1992.

[7] CASTILLO, C. D., JACOBS, D. W. *Using Stereo Matching with General Epipolar Geometry for 2D Face Recognition across Pose*. Pattern Analysis and Matching Intelligence, IEEE Transactions on, vol. 31, no. 12, pp. 2298/2304, Dec. 2009.

[8] ISGRO, F., TRUCCO, E. *Projective rectification without epipolar geometry*. IEEE Conference on Computer Vision an Pattern Recognition, Fort Collins, Colorado, CO, USA, 1: 94-99, 2001.

[9] POLLEFEYS, M., GOOL, L. V., VERGAUWEN, M. *Visual Modeling with a Hand-Held Camera*. International Journal of Computer Vision, 59(3): 207-232, 2005.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

MDA Approach: Science and Research as a Business Process and Possible Transformation of Its Sub-part to IS Design Model

*Martin Kardoš, *Michal Mokryš
*University of Žilina, Faculty of Management Science and Informatics, Department of Information Networks, Univerzitna 8215/1, 01026 Žilina, Slovakia, {Martin.Kardos, Michal.Mokrys}@fri.uniza.sk

Abstract. Universities as the other organizations support their activities by IS/ICT. There are many business processes (BP) in university environment that should be transformed to their automated form. A research of strategic objectives of Slovak universities has showed that there are two main processes: 1. Education, 2. Science and Research. Many universities are trying to support them by IS/ICT. In this contribution we would like to introduce a possible approach to transformation of business processes to information system (IS) design models in context of science and research and its sub-processes at university. This approach is demonstrated on a sub-process of organizing of scientific conferences and presenting scientific results. MDA is a modern approach to the development of IS which models it at different levels of abstraction. This paper deals with a possibility of the traceability from business processes modeled in BPMN to IS design models expressed in UML in context of model driven architecture (MDA).

Keywords: MDA, CIM, PIM, BPMN, UML, transformation.

1. Introduction

A university is a specific type of organization providing the education and scientific / research service. The analysis of the strategic objectives of Slovak universities showed, that the goal of innovation and changes at universities are those activities which have to be carried out in following processes: Education and Science and Research. [1]

In context of science and research process of the university (fig. 1) we can find several sub-processes: Publications, Organizing of Scientific Conferences and Presenting Scientific Results, PhD Study and Career, Research Projects. All of them can be marked as business processes because they describe a sequence or flow of activities that have to be fulfilled with the objective of carrying out work. For effective functioning of the science and research process, it is necessary to have an IS/ICT[1] support for all of them. In this paper, we focus on a sub-process of organizing of scientific conference and presenting scientific results with several supporting processes. One of them is a process of registering participants and processing scientific papers of the conference. This process is relatively easy and many universities have IS/ICT support of it, but it is a good example to presenting our approach to transformation of business processes (CIM level of MDA) to IS design model (PIM level of MDA).

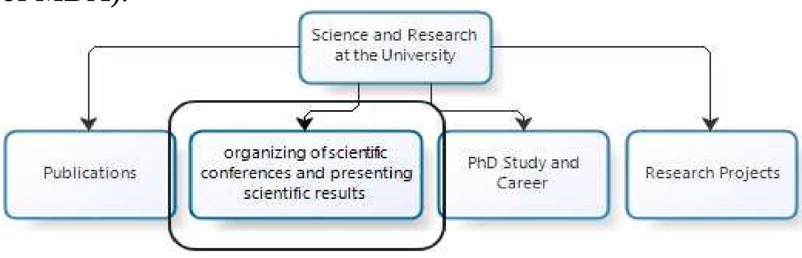


Fig. 1. Science and Research process

[1] Information System / Information Communication Technology

Generally, there are many approaches to development of IS[2] supporting this process. One of them is Model driven architecture (MDA) [2]. It is a modern software development framework based on creation of models and transformations between them. MDA is defined by a standardization body in software engineering – Object Management Group (OMG). MDA levels of abstraction are called: Computation Independent Model (CIM), Platform Independent Model (PIM), Platform Specific Model (PSM) and Implementation model (IM) - Code. All of them are described in [2]. MDA describes certain levels of abstraction and their relations, but it does not specify which exact models and notations should be used for their representation and how to transform between them. In our mention the CIM layer is represented as a business process model described in BPMN [3] notation. The PIM layer is represented by several UML [4] models. We mention an UML use case diagram and a general class diagram (analytical class diagram). In this contribution, we present only an analytical class diagram.

As we mentioned above, the process of registering participants and processing scientific papers of the conference is relatively easy but we would like to demonstrate how we can model this process in BPMN and a possible approach of the traceability from business processes to IS design models. In other words, we would like to present our approach to defining of rules for semi-automatic CIM - PIM transformation in MDA.

The rest of this paper is organized as follows. Basic knowledge about modeling of the process of registering participants and processing scientific papers of the conference is presented in Section 2. In section 3, we present our approach to transformation of business process model (CIM) to IS design model (PIM). Finally, Section 4 concludes this paper with a short summary. The process of registering participants and processing scientific papers of the conference is described in the next chapter.

2. Creation of model of Business Processes presenting MDA CIM Level

In our approach, we present Business Process Modeling Notation (BPMN) [3] to business process modeling. The BPMN and its business process diagram (BPD) were created by OMG in 2004 and are used for graphical modeling of business processes. The primary goal of BPMN is to provide a notation that is readily understandable by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business people who will manage and monitor those processes. Thus, BPMN creates a standardized bridge for the gap between the business process design and process implementation [3].

There are three participants in the process of registering participants and processing scientific papers of the conference that participate in three sub-processes - Administrator, Reviewer and Participant. Execution of all three sub-processes is not necessary simultaneously. Each of participants performs his activities / tasks as it is presented on figure 2. The process is started by Administrator which creates conference in IS system and defines its required attributes. His important role is in assigning of contributions (papers) to Reviewers. Participant may register for the predefined conference and send his contribution (paper) for a review. Sub-process of Reviewer starts when reviewer receives a message about assignment of the contribution (paper) from Administrator. The main role of Reviewers is to prepare the review of the contribution (paper) and to send a message about its acceptance or rejection. The Participant's process is ending by receiving of such a message. In our case this created business process model in BPMN presents CIM level of MDA.

[2] Information System

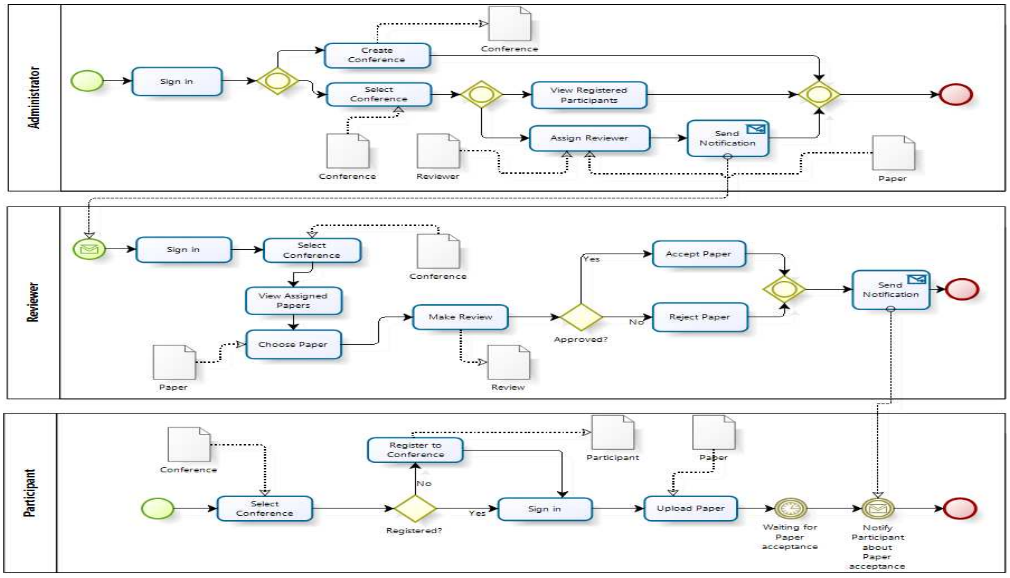


Fig. 2. Business process diagram of the process of registering participants and processing scientific papers of the conference.

3. Our semi-automatic Approach to Transformation of Business Process Model (CIM) to IS Design Model (PIM)

As we mentioned in Section 1, the CIM level is representing by the model of business processes modeled in BPMN notation. The main problem of BPMN is that it is only a graphical representation of business process (BP). When we transform BP modeled in BPMN to UML model by automatic or semi-automatic approach we need a standardized semantic formalism for BP collaborating with BPMN notation. XML Process Definition Language (XPDL) [5] is the serialization format for BPMN. It is defined by the Workflow Management Coalition (WfMC) to interchange business process definitions between different modeling tools [6]. XPDL provides a file format that supports every aspect of the BPMN process definition notation including graphical descriptions of the diagram and also it is extensible so that it allows each different tool to store implementation specific information within the XPDL [7]. Mentioned features, an extensibility of XPDL and a support of this format in many BPMN tools led us to use it as a semantic formalism of BP.

Our CIM to PIM transformation rules are based on mapping of chosen BPMN elements to elements of certain UML models (in our case the analytical class diagram) via XML Metadata Interchange (XMI) format for UML and XPDL for BPMN. Metadata of UML meta-model can be expressed in Meta-Object Facility (MOF). It can be used as an interchange format between models. UML models are therefore expressed in their own standardized written format XMI and it can be presented as MOF. The XMI format for UML meta-model has two main parts: “UML model” and “UML diagram”. The UML model carries semantic information about the model itself and its elements. The UML diagram stores data for a graphical representation of UML model. This XMI format can be imported to the UML modeling tools. As we mentioned above, a business process diagram (BPD) in BPMN notation is also only a graphical representation. Semantic information of every BPMN element of BP is in a standardized written format - XPDL. The both - XMI and XPDL are based on XML language. This aspect is necessary for creation our transformation rules. They are written in XSLT (Extensible Stylesheet Language Transformations). An analytical UML class diagram model presenting PIM level of MDA (figure 3) is a result of our semi-automatic transformation. For implementation of extensibility of XPDL and applying of mentioned transformation rules we work on our mapping software. However, we note that the resulted generated analytical UML class diagram model is only initial PIM level of MDA which will be further transformed to the design UML class diagram model manually.

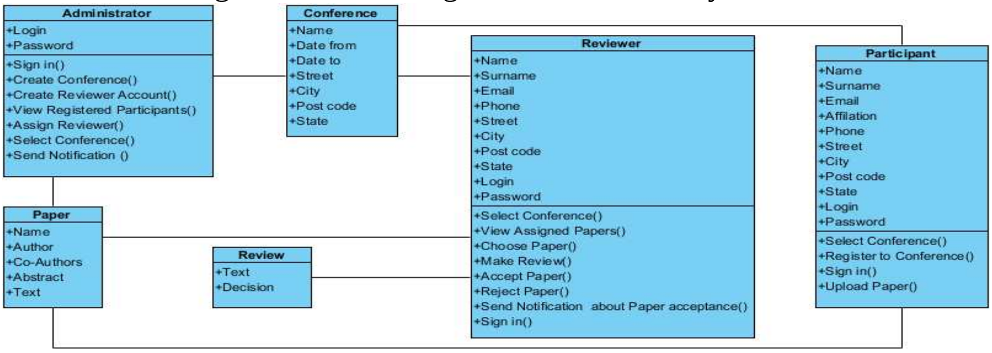


Fig. 3. Generated analytical UML class diagram model presenting initial PIM level of MDA.

4. Conclusion

Possible approach to transformation of business process model to IS design model in context of MDA is described in this contribution. In other words, we talked about possible approach to CIM-PIM transformation in MDA. We presented universities as a specific type of organization with two main processes: Education, Science and Research. The IS/ICT support of these processes is necessary for effective functioning of every university. We focused on IS/ICT support of the Science and Research process. We described one of the supporting processes: the process of registering participants and processing scientific papers of the conference. We presented business process model of this process in BPMN notation. This model presents CIM level of MDA. For semi-automatic approach to transformation of business process model (CIM) to IS design model (PIM), we described transformation rules based on XMI exchange format for UML meta-model and on XPDL for BPMN notation. The result of our transformation is an analytical UML class diagram model presenting initial PIM level of MDA. Usage of this way of transformation is discussed and we assume that it should be a good foundation for better and more detailed design of IS and software applications. We hope that this contribution opens a scientific discussion of this problem.

References

[1] KARDOS, M., HRBANOVA, K., DROZDOVA, M. *Information system's architectures in the University IS engineering*. In: WMSCI 2010: the 14th world multi-conference on systemics, cybernetics and informatics : June 29th-July 2nd, 2010 - Orlando, Florida, USA : proceedings. Vol. III. - [Orlando]: International Institute of Informatics and Systemics, 2010. - ISBN 978-1-936338-00-9. - S. 265-267, 2010.

[2] OMG INC. *Model Driven Architecture (MDA) FAQ...* 2005, http://www.omg.org/mda/faq_mda.htm, downloaded: September, 2011.

[3] WHITE A. S. *BPMN Fundamentals*. 2005, http://www.omg.org/docs/pm/05-12-06.ppt, downloaded: February, 03rd 2011.

[4] OMG INC. *Introduction to OMG's Unified Modeling Language™ (UML®).* http://www.omg.org/gettingstarted/what_is_uml.htm, 2005.

[5] WORKFLOW MANAGEMENT COALITION: Process *Definition Interface -- XML Process Definition Language (XPDL 2.1 Complete Specification),* 2010, http://www.wfmc.org/index.php?option=com_docman&task=doc_download&Itemid=72&gid=132, downloaded: February, 03rd 2011.

[6] VAN DER AALS, W. M. P. *Patterns and XPDL: A Critical Evaluation of the XML Process Definition Language*. QUT Technical report, FIT-TR-2003-06, Queensland University of Technology, Brisbane, 2003, Eindhoven University of Technology, http://is.tm.tue.nl/research/patterns/download/ce-xpdl.pdf, downloaded: January, 2011.

[7] http://www.xpdl.org/

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Acoustic Analysis of Male and Female Voices

*Jacek Kołodziejski
*Kielce University of Technology, Department of Mechanics, Al. 1000-lecia P.P. 7, 25-314 Kielce
j.kolodziejski@tu.kielce.pl

Abstract. The aim of the paper was to compare male and female voice signals during phonation. The analysis was made by using chosen parameters: mean fundamental frequency, jitter, shimmer. Special attention was paid to harmonic-to-noise ratio (HNR). The parameters were extracted from speech of samples representing sustained vowel /a/. Samples were collected from 33 singers of Kielce University of Technology Choir.

Keywords: Acoustic parameter, normal speech, phonation

1. Introduction

A healthy speaker can produce a sustained vowel /a/ with HNR of around 20 dB. The mean fundamental frequency parameter (mF0) equals to 120 Hz for male voices with minimum value 80 Hz and maximum 200 Hz [1]. Mean values for male acoustic parameters were 0.64% for jitter, 3.16% for shimmer, and 17.5 dB for HNR [2]. Female voices are higher than the men voices reaching 390 Hz for maximum value, and 110 Hz for minimum. The mF0 equals to 200 Hz for female voices. In the short term analysis for healthy non trained female voices, mean values for acoustic parameters were 0.5% for jitter, and 2.6% for shimmer [3]. The phonation time parameter (PHT) should exceed 10 seconds [1-5].

The aim of the paper was to compare male and female voice signals during phonation. Results will be used for preparation speech pattern.

2. Methods

Samples were collected from 16 female and 17 male singers of Kielce University of Technology Choir according to methodology described [2,4]. Recordings included phonation of sustained vowel /a/. The subjects were deemed to have a normal functioning larynx with professional trained voices. Sustained vowel /a/ was uttered three times by the speaker.

Recordings were made in regular conditions by using digital recorder. The sampling rate was 44 kHz and the signal resolution was 16-bit. Speakers were sitting and the mount-to-microphone distance ranged of around 0,25m.

Acoustical analysis included measures of mean fundamental frequency, Jitter local, Shimmer local, harmonics-to-noise ratio. The jitter' acoustic parameter was calculated in order to show the short-term changes occurring in the time of domain and also in the field of speech signal frequency. The Shimmer parameter allows to see the short-term changes occurring in the amplitude of the speech signal. The Harmonics-to-Noise Ratio parameter (HNR) represents the degree of acoustic periodicity. The parameters were extracted and edited using Audacity, Matlab and Praat softwares [6-8]. The source code of parameters' extraction in Praat software is available in study [4]. Measurements of mean fundamental frequency, Jitter local, Shimmer local and HNR are based on pitch extraction and first three can only be calculated for voices with a clear fundamental frequency [4].

3. Results

The maximum phonation time obtained by female equals to 24 seconds, for male equals to 30 seconds. Minimum phonation time obtained for female speaker, with 10 seconds. As for the male, minimum was 9 seconds. Mean value of phonation time was 16.25 second for female, and 17.33 seconds for male. All of female and male speakers in case of phonation time parameter meet the standard [1-5].

In tab. 1. mean values, standard deviation, maximum value, minimum value and levels of trust (0.05) of acoustic parameters extracted from signal samples representing sustained vowel /a/ of female and male voices are presented.

Female	Mean	SD	Level of Trust (0.05)	Min	Max
mF0[Hz]	244	51.177	14.631	188.52	388.67
Jitter[%]	0.39	0.001	0.0004	0.09	0.80
Shim[%]	4.94	0.022	0.006	1.66	11.50
HNR	21.55	3.926	1.122	15.09	36.557
Male	**Mean**	**SD**	**Level of Trust (0.05)**	**Min**	**Max**
mF0[Hz]	123.86	15.183	4.049	98.84	161.78
Jitter[%]	0.75	0.011	0.003	0.19	6.47
Shim[%]	5.46	0.029	0.007	2.24	18.29
HNR	17.73	5.189	1.383	0.29	27.04

Tab. 1. Mean values of parameters calculated for female and male speech signal based on sustained vowel /a/, with level of trust equals 0.05, maximum values, minimum values and standard deviation (SD)

We observed the significant correlation for values of mean fundamental frequency between our recordings and standards for healthy female speaker in paper [3] and for male speaker in paper [1]. The rest of acoustic parameters presented in tab. 1. shows no significant differences to normal speech for female and male acoustic parameters standards.

In Fig. 1. and 2. mean values of HNR (in decibels) parameter during phonation sustained vowel /a/ for female and male speaker are shown.

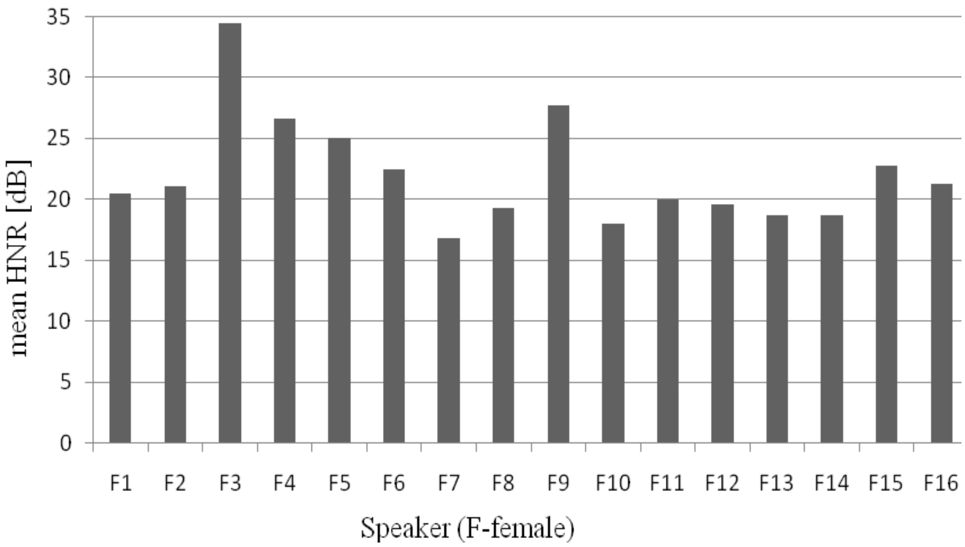


Fig. 1. Mean values of HNR (dB) for female choral voices

Female HNR analysis meets the defined standard of around 20 dB, getting only in one case (F3) result of 34.42 dB.

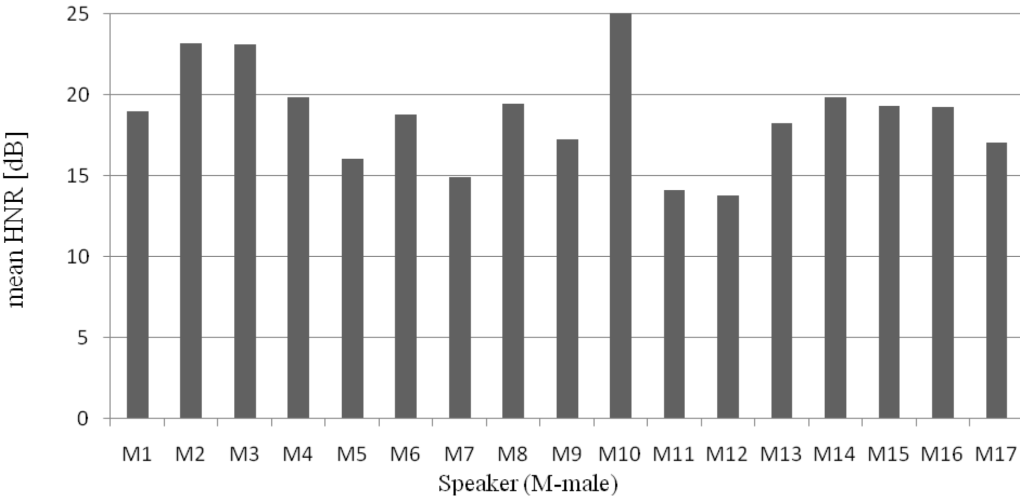


Fig. 2. Mean values of HNR (dB) for male choral voices

Male HNR analysis also meets the defined standard of around 20 dB, getting the highest result of 25.15 dB for tenth speaker.

4. Conclusions

Authors compared male and female choral voices in case of acoustic parameters for sustained vowel /a/. Short-term spectral analysis shows the small difference in the middle section of frequency level, where in female - soprano voice increase was greater than in male – bass voice. It should be also mentioned the differences on peaks in low frequency level, where especially in female - soprano voices the peaks are more complex and present in larger quantities. More research on this subject author provides in his further work.

Measurements of time phonation parameter for male and female were similar, with average of 10 to 20 seconds. These results are in agreement with literature data [1-5].

Mean values of acoustic parameters extracted differences between male and female choral voices, shows almost twice higher results for mean fundamental frequency (mF0): female - 244 Hz, male – 123.86 Hz. The standard deviation and level of trust for 0.05 was also much higher for female as for male. Minimum and maximum values were smaller dispersed for male (difference were only 62.94 Hz) than in female (difference was 200.15 Hz). These results are similar with data from [1-4].

Jitter parameter shows inverse relationship, where mean value of this parameter is smaller for female than for male. Difference between the minimum and maximum values is significantly greater in male than in female. Similar differences are observed in Shimmer parameter.

HNR parameter demonstrated slightly higher mean value and significantly higher minimum value in female choral voices results. In individual speaker HNR analysis during phonation, male choral voices classify in defined standards, around 20 dB. Female HNR analysis also meet the defined standard of around 20 dB, getting only in one case result of 34.42 dB.

Preparation of speech pattern needs further investigations.

References

[1] OKŁA, S., *Chirurgiczna rehabilitacja głosu i mowy osób po całkowitej laryngektomii*, Wydawnictwo Lekarskie PZWL, 2007

[2] KOŁODZIEJSKI, J., MIĘSIKOWSKA, M., *Acoustic analysis of normal speech*, proceedings of InterTech 2010, 19-21 May, Poznań, 2010

[3] NIEDUBEK-BOGUSZ, E., FISZER, M., KOTYŁO, P., JUST, M., ŚLIWIŃSKA-KOWALSKA, M., *Voice acoustic analysis in healthy women*, Otorynolaryngologia, 3 (1), 33-39, Wrocław, 2004

[4] MIĘSIKOWSKA, M., *Analiza sygnału mowy u chorych po całkowitym usunięciu krtani*, thesis, Kielce, 2009

[5] JASSEM, W., Po*dstawy fonetyki akustycznej*, PWN, Warszawa, 1973

[6] AUDACITY – darmowy editor i dekorder audio, *http://audacity.sourceforge.net/*

[7] MathWorks – MATLAB and SIMULINK for Technical Computing, *http://www.mathworks.com/*

[8] PRAAT – doing Phonetics by Computer, *http://www.fon.hum.uva.nl/praat/*

Jacek Kołodziejski is a holder of scholarship from European Union Fund with EFS programme realized in Kielce University of Technology “Education for success”.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Heuristics Methods in Time Series Forecasting

*Alexandra Kotillová
*University of Žilina, Faculty of Management Science and Informatics, Department of macro and microeconomics, Univerzitná 8215/1, 01026 Žilina, Slovakia, kotillova@gmail.com

Abstract. In this paper I provide a short overview of the Radial Basis Functions (RBF) neural networks, their properties and the motivation behind their use. RBF networks have been employed for functional approximation in time series modelling in pattern classification.

Keywords: RBF neural networks, K-mean clustering algorithm, back-propagation, heuristics methods

1. Introduction

Radial Basis Function Networks have been successfully applied to a large diversity of applications including interpolation, chaotic time-series modelling, system identification, control engineering, electronic device parameter modelling, forecasting, channel equalization, medical diagnosis, pattern recognition, speech recognition, image restoration, shape-from-shading, 3-D object modelling, motion estimation and moving object segmentation, data fusion, etc [1]. They perform excellent approximations for curve fitting problems and can be trained easily and quickly.

2. RBF Neural Networks and learning algorithm

RBF neural network is feed forward neural network and it is interesting because it's involving supervised and unsupervised learning. K-mean cluster is an unsupervised learning method and it is applied from input layer to hidden layer. It is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The centres are then located as described above and the forecasts are calculated. K-step forecasts are calculated iteratively. The back-propagation learning algorithm is a supervised learning method and it is applied from hidden layer to output layer. This algorithm is divided into two phases: propagation and weight update. The first phase is about the spreading of the input signals to the outputs of the network and computing the output errors. The second phase is about the back error propagation to individual neurons in the lower layers with a subsequent modification of the weights [4]. More information about algorithm and algorithm itself you can find in [2].

3. RBF Activation Function

RBF network is a feed forward neural network that consists of one hidden layer. The activation function of hidden neurons is based on cloud activation function given by the form

$$\psi_2(u_j) = o^j = \exp\left[-\frac{(x-w_j)^2}{2\sigma_j^2}\right] \qquad (1)$$

where o^j are the outputs from the hidden layer, x is an input vector, w_j are weights of the centers and σ_j are spreads of the centers. Fig.1 shows the three-dimensional graph of hidden neuron and the function is radially symmetric around the center w_j.

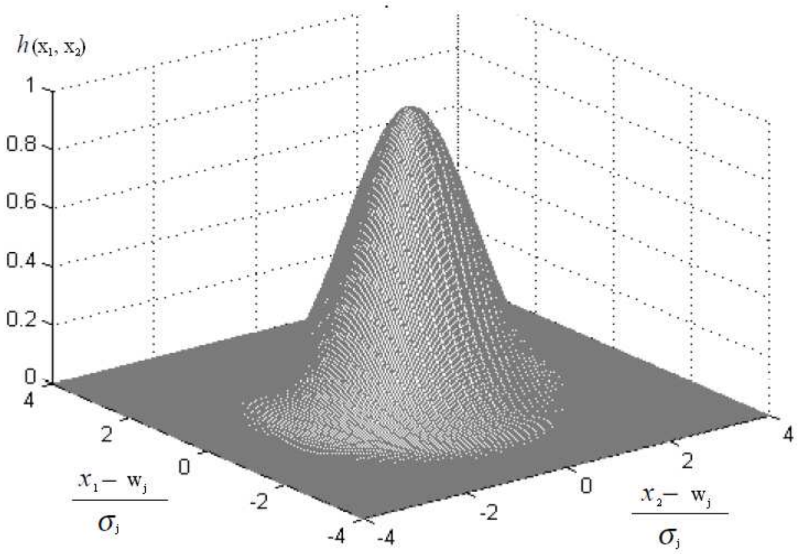


Fig. 1. Gaussian activation function for the neuron with two inputs

Learning of RBF network with Gaussian activation function is generally divided into two phases. The first one is an unsupervised learning phase in which appropriate locations for the centers of the rbf functions in the hidden layer and the standard deviations (spreads) are estimated. We can use K-means clustering algorithm or method based on competitive learning for finding the centers of activation function. Competitive learning is a class of unsupervised learning algorithms based on the idea of adjusting a weight matrix in such a way that the weights represent cluster centers [3]. This method is based on Kohonen's adaptive rule.

The activation function of output neuron also called predicted value is given by

$$\hat{y}_t = \sum_{j=1}^{s} v_t^j o_t^j \qquad (2)$$

where v_j are the weights between the output neuron and neurons of hidden layer at time t. During the second phase is used a supervised learning in which mentioned v_j weights are calculated using back-propagation algorithm [5].

4. Future plans

The following described neural network is used for purposes of forecasting time series. Exact methods are always trying to find the optimal solution. In general, the problem of these methods is that they are able to address only certain types of tasks and their execution time grows with the complexity of the problem and it is not guaranteed end in the short time. Example for RBF neural network is: gradient method, the method of back-propagation, Newton-Rapshon ...

Disadvantage of gradient method is slipping into a local minimum. Described in two-phase learning method is the effort of finding a global minimum supported execution of several epochs of learning. Sometimes mean square error reaches a local minimum then begins to rise and after a certain number of epochs of learning again falls to the new minimum level. In finding a global minimum it is necessary to make more learning epochs or use heuristics.

Heuristic methods with high probability will not find optimal solution but they will find a tolerable solution in acceptable time. The best known includes: the climbing algorithm, tabu search, simulated annealing, evolutionary strategies and genetic algorithms.

Future plans involve implementation of heuristics. Then they will be compared and confronted with optimal solutions obtained via ex post forecast. Better approach will be chosen and new algorithm or improvements in existing algorithms will be done.

Conclusion

Many experiments show that RBF networks are superior over other neural networks approaches because of the following reasons: RBF networks are capable of approximating nonlinear mappings effectively. The training time of the RBF networks is quite low compared to that of other neural network approaches. This follows from the fact that the input layer and the output layer of an RBF network are trained separately and sequentially.

On the other hand critics of this model claim that estimating behaviour of economy is possible to predict only with taking into account economic theory.

Acknowledgement

This work has been supported by Slovak grant foundation under the grant VEGA No. 1/0667/10.

References

[1] BORS, A.G. *Introduction of the Radial Basis Function Networks algorithms*. Online Symposium for Electronics Engineers. Feb. 2001. Pages 1-7. http://www.osee.net/.

[2] MARČEK, D., MARČEK, M. *Neurónové siete a ich aplikácie*. Edis - University of Zilina. 2006. ISBN: 9788080704971.

[3] SAHIN, F. *A Radial Basis Function Approach to a Color Image Classification Problem in a Real Time Industrial Application*, Master's thesis, 1997, Virginia polytechnic institute, Blacksburg.

[4] RUMELHART, R.E., McClelland, J.L. *Parallel distributed processing explorations in the microstructure of cognition*, Cambridge: MIT Press, 1980.

[5] ZHANG, G.P. *An investigation of neural networks for linear time-series forecasting*. In Computer and Operations Research. Vol. 28. No. 12. Elsevier. October 2001. pp. 1183-1202.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Modeling of the Chromatic Dispersion Through the Use of Split-step Fourier Method in Matlab

*Ján Litvík, *Milan Dado, *Jozef Dubovan
[1]University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitná 8215/1, 010 26 Žilina, Slovakia, {Jan.Litvik, Milan.Dado, Jozef.Dubovan}@fel.uniza.sk

Abstract. A method of the chromatic dispersion modeling in single mode optical fibers is presented. Simulations are designed for the optical fibers corresponding to recommendation ITU G.652.A, G.653.A and G.656.A. For each fiber is created its characteristic dispersion curve around the optical carrier frequency. It is created a mathematical description for chromatic dispersion. The proposed model is in some respects partly idealized for its simplicity. This approach presented here can be extended to other phenomena such as the Self Phase Modulation, Cross Phase Modulation and Four Wave Mixing. The interactions of these phenomena indicate the direction of further development of complex optical systems and devices for ultra-high-speed optical networks using advanced modulation formats.

Keywords: split-step Fourier method, chromatic dispersion, nonlinearities

1. Introduction

There are various methods for modeling effects in optical fiber. This method presented here is based on nonlinear Schrödinger equation. The accuracy of this method is dependent on the correct setting steps that determine the value of the operation of elemental phenomena in optical fiber. They are simulated three types of optical fibers: G.652.A – standard single-fiber, G.653.A – dispersion-shifted single-mode optical fiber and G.656.A – non-zero dispersion for wideband optical transport.

2. Theory

Spreading of the optical pulses in the time domain in single mode fiber is inter alia caused by chromatic dispersion. Consequently, bandwidth is limited ultimately by spectral width of the optical source. In contrast with the situation in multimode fibers, mechanism of chromatic dispersion in single mode fiber tends to be with each other complex. Transmission time or a specific group delay τ_g for pulse distributed per unit length of single mode fiber can be given as

$$\tau_g = \frac{1}{c}\frac{d\beta}{dk}, \tag{1}$$

where c is the speed of light in vacuum, β is the mode propagation constant in the fiber core with a refractive index n_1, k is the propagation constant for same light in vacuum.

The total value of the first dispersion parameter or chromatic dispersion in single mode optical fiber, D_T, is given by the derivation of a specific group delay with the regard of the wavelength in the vacuum λ

$$D_T = \frac{d\tau_g}{d\lambda}. \tag{2}$$

The total dispersion of the fiber, which depends on the fiber material, refraction index profile and dimensions, can be minimized by the material and waveguide dispersion and simultaneously we can minimize dispersion profile [1].

3. Analytical model

We begin from the nonlinear Schrödinger equation, which is given by

$$i\frac{\partial A}{\partial z}+\frac{i\alpha}{2}A+\frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2}+i\gamma|A|^2A=0, \tag{3}$$

where A is the slowly varying envelope of the electric field,β_2 is group-velocity dispersion parameter, γ is nonlinear parameter, α is attenuation constant.

To understand the philosophy of the split-step Fourier method, it is appropriate to write (3), formally in the form

$$\frac{\partial A}{\partial z}=\left(\hat{D}+\hat{N}\right)A, \tag{4}$$

where $\hat{D}$ is differential operator, which is responsible for dispersion and loss in the linear medium $\hat{N}$ is nonlinear operator, which governs the impact of nonlinearities on the optical pulse. Since we are dealing with chromatic dispersion modeling will be addressed only differential operates within the linear media. This operator is given

$$\hat{D}=-i\frac{\beta_2}{2}\frac{\partial^2}{\partial T^2}-\frac{\alpha}{2}. \tag{5}$$

Generally, the dispersion and nonlinearity operate together along the fiber. Split-step Fourier method obtained an approximate solution, provided the spread of the optical field at a small distance h, dispersion and nonlinear effects can be assumed to operate independently. More precisely, the spread of z to $z+h$ is performed in two steps. In the first step (4), the nonlinearity operate alone, and $\hat{D}=0$. In a second step, dispersion operate alone, and $\hat{N}=0$ [2].

Mathematically expressed as

$$A(z+h,T)\approx e^{h\hat{D}}e^{h\hat{N}}A(z,T). \tag{6}$$

3.1. Ideal linear case

If $\gamma=0$, then (4) can be solved by Fourier transformation. In this case, if $\hat{A}(z,\omega)=\int_{-\infty}^{\infty}A(z,t)e^{j\omega t}dt$ indicates transformation, then first derivation is

$$\frac{\partial\hat{A}(z,\omega)}{\partial z}=i\beta_2\omega^2\hat{A}(z,\omega)/2, \tag{7}$$

whose solution is

$$\hat{A}(z,\omega)=\hat{A}(0,\omega)e^{j\beta_2\omega^2 z/2}. \tag{8}$$

Then $\left|\hat{A}(z,\omega)\right|=\left|\hat{A}(0,\omega)\right|$for all z. This means that the Fourier transform has a constant module, and especially in the size of frequency components does not change over spread [3].

4. Simulations

Input parameters of the simulation program are fiber length, transmitted wavelength, and bit rate. In this case, has been for simplicity only the monochromatic source considered. This implies that the signal spectrum is given only by the pulse spectrum. At the input was used the pulse with Gaussian shape. The output pulse was degraded by chromatic dispersion without considering attenuation. Simulated was the chromatic dispersion of standard optical fibers G.652.A, G.653.A and G.656.A. As results of simulations are displayed the input and output pulses for different types of optical fibers after passing through. Fig. 1 shows the simulation when the fiber length L = 50km, wavelength λ = 1550 nm and bit rate *BR* = 20Gb/s. Fig. 2 shows the simulation when the fiber length *L* = 70 km, the wavelength λ= 1550nm and bit rate *BR* = 33 Gb/s. Fig. 3 shows the simulation when the fiber length *L*=50km, wavelength λ = 1550 nm and bit rate *BR* = 60Gb/s. Fig. 4 shows the simulation when the fiber length *L*=50km, wavelength λ = 1460 nm and bit rate *BR* = 60Gb/s. The simulations show that the dispersion is dependent on fiber length, bit rate and wavelength. Pulse shape distortion is negligible with regard the type of pulse and transmission speed. Since the spectrum of Gaussian pulse is enough narrow, pulse shape distortion in our model, taking into account the ideal monochromatic source have no effect.

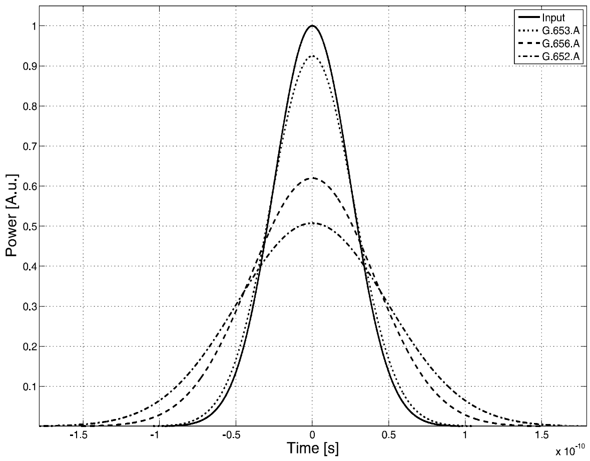


Fig. 1. Simulation of the output pulse degradation due to dispersion: λ=1550 nm, *L*=50km, *BR*=40 Gb/s

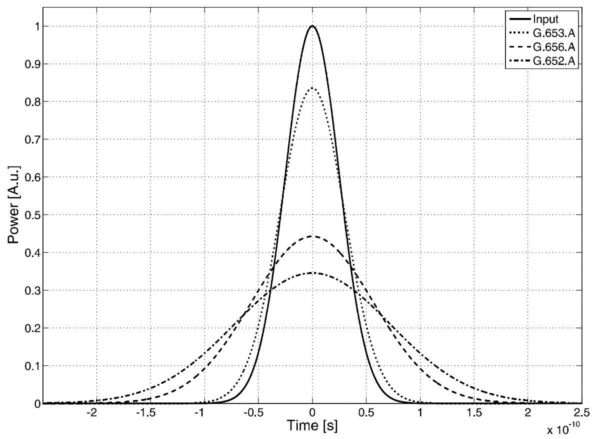


Fig. 2. Simulation of the output pulse degradation due to dispersion: λ=1550 nm, *L*=80km, *BR*=40Gb/s

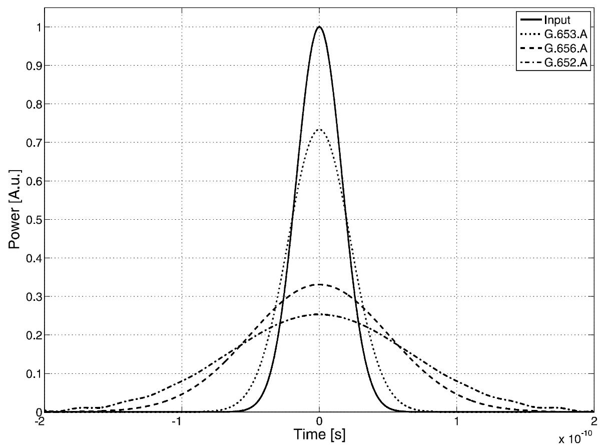


Fig. 3. Simulation of the output pulse degradation due to dispersion: λ=1550 nm, *L*=50km, *BR*=60Gb/s

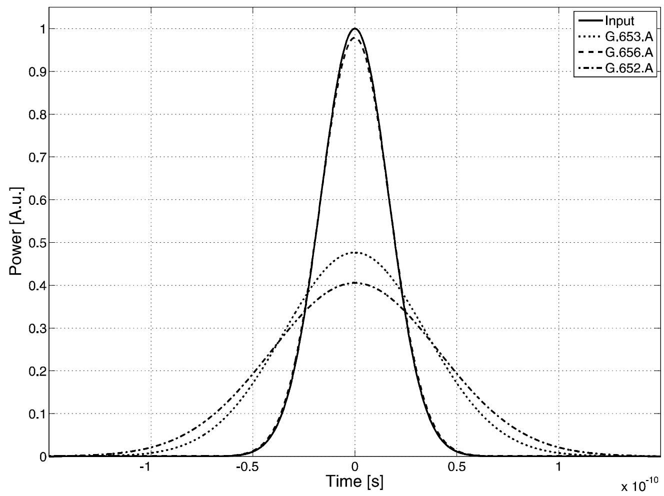


Fig. 4. Simulation of the output pulse degradation due to dispersion: λ=1460 nm, *L*=50km, *BR*=60Gb/s

5. Conclusion

In this article the numerical computer model for linear phenomena calculation of the influence of chromatic dispersion was proposed. The model is solved by Slit-Step Fourier Method. This method is based on spread over very short distances. This method is designed to use not only for chromatic dispersion calculation but also for the nonlinear effects and their interaction. As can be seen from the results of simulations the chromatic dispersion is dependent on the wavelength, the length of optical fiber and transmission rates. Input pulse was defined as a Gaussian pulse for simulation. Slit-Step Fourier Method was designed to calculate the interaction of chromatic dispersion and nonlinearities. Hence our model will be extended in the future for calculation both linear and nonlinear phenomena and the calculation of the influence on different modulation formats.

Acknowledgement

This work has been based supported by Slovak grant agency VEGA 1/0411/10.

References

[1] SENIOR J. M.*Optical Fiber Communications: Principles and practice, 2nd ed.* Harlow : Pearson Education Limited, ISBN 0-13-635426-2, 1992.

[2] AGRAVAL, G. P. *Nonlinear Fiber Optics,4th ed.* San Diego: Academic Press, 2007.

[3] SHAW, J. K. *Mathematical Principles of Optical Fiber Communications*. Blacksburg: Regional conference series in applied mathematics, 2004.

[4] ITU-T Recommendation G.652.*Characteristics of a single-mode optical fibre and cable*. 06/2005.

[5] ITU-T Recommendation G.653. *Characteristics of a dispersion-shifted single-mode optical fibre and cable.* 12/2006.

[6] ITU-T Recommendation G.656. *Characteristics of a fibre and cable with non-zero dispersion for wideband optical transport*. 12/2006.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Safety Mechanisms and Solutions within Profile OpenSAFETY

*Juraj Ľupták
*University of Žilina, Faculty of Electrical Engineering, Department of Control and Information Systems, Univerzitna 2, 01026 Žilina, Slovakia, juraj.luptak@fel.uniza.sk

Abstract. This paper points to the safety open solution, which will be in the future compatible with several types of safety products. Moreover it deals with the safety mechanism and solutions within the safety profile openSAFETY compared to the safety profile PROFIsafe. Practical part is oriented to transmission of subframe 1 and subframe 2 across to noise channel via Matlab.

Keywords: safety mechanism, openSAFETY, Ethernet, CRC.

1. Introduction

The industrial network is in many cases a part of a system, which takes place in the management of Safety-Related Critical Processes for example the management of certain manufacturing processes in mechanical engineering, chemical industry, nuclear power, traffic management. The system must be designed to guarantee the required safety integrity level SIL [1]. Recently many different families of the industrial Ethernet networks occur. In the case of industrial machines safety, production lines safety, ranged from control up to the automated production, enormous costs are given to the safety on the purpose of using several types of communication networks. From this arise the effort to find a safety communication profile which will be possible to use in any case. Capable nominee for this position can be the safety profile openSAFETY. It has achieved the safety integrity level (SIL) 3. OpenSAFETY is open source software and it is possible to customize it.

2. The safety communication profile openSAFETY

This method of protection is used for the Sercos III, EtherNet / IP, Modbus-TCP, Powerlink, compatible for all Ethernet types. OpenSAFETY is open safety protocol also in the technical respects:

- given the protocol's bus-independence,
- openSAFETY can be used with all fieldbuses,
- industrial Ethernet solutions.

OpenSAFETY uses a frame with a uniform format, no matter whether for payload data transfer, or time synchronization purposes and for a configuration. Frame length depends on the amount of data to be transferred. It uses the function called “black channel” which means to achieve the desired SIL by adding a safety layer into the seventh layer of the communication model (Ethernet Powerlink, Modbus-TCP and Profinet). This principle rises from the basic economical request; not to change anything within the hardware or software of the communication systems, which are already well or very well protected against any breakdown.

Safety nodes are solved on the network automatically, recognition of frame types and lengths do not have to be configured.

OpenSAFETY supports several communication models:

- Producer/consumer (see Fig. 1.a)
- Expanded producer/consumer (see Fig. 1.b)

- Client/server (see Fig. 1.c)

2.1. Basic OpenSAFETY frame

OpenSAFETY frame (see fig. 2.) consists of two subframes with identical content. Two identical subframes duplicate into one frame and is able to transport data up to 254 bytes of payload data. Each subframe is provided with an individual checksum. The identical content of these frames is compared by the receiver [3].

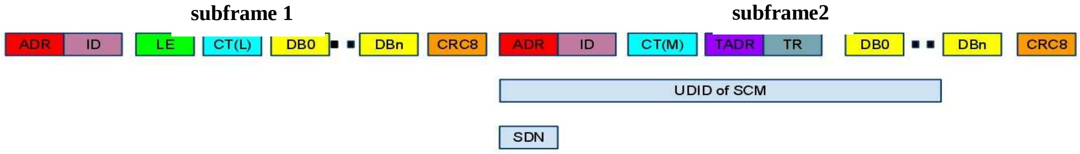


Fig.2. OpenSAFETY Frame

Subframe consists of:

- ID (Frame ADR (Address field) – this field is a 10 bites data, it defines the SADR of a node. ADR in 2nd subframe transmit SDN (Domain Number) information as well using a logical operation XOR. ADR field contains SADR of a node, from which it was sent or to which it was assigned. Non-critical time data are transferred without subsequent analysis of a specific part of the OpenSAFETY frame.
- Identification of the type of frame and message. The size is 8 bites.
- LE (Length field) – specifies the number of transferred data in a frame and specifies the used CRC type.
- CT (Consecutive Time field) – consists of 2 parts:
  - LSB – subframe 1
  - MSB –subframe 2
- Both of these parts have a size of 8 bites. That means it is possible to adjust CT in the range from 0 to 65535.It is possible to use SPDO (Process Data Object) and SSDO (Service Data Object) services for this.
- DB (Data byte) –transferred data have length from 0 to 254 bytes. Size 0 is usually used for time synchronization.
- CRC (Cyclic Redundancy Check) – the choice of CRC code depends on the length of LE.
- TADR (Time Request Address) – expanded address field (SADR). For SPDO it contains addresses of nodes that respond to time synchronization application. For SSDO and SNMT they contain the address of a source or of a receiver.
- TR (Time Request Distinctive Number) – unique number in the range from 0 to 63. Its function is to distinguish a certain message from the other ones (Time Request).
- UDID of SCM (Unique Device Identification of Safety Configuration Manager) –contains 6 bytes that are encoded by a logical operation XOR for SPDO and SSDO services. SNMT is not encoded. They are put into the 1st six bytes of a 2nd subframe. This coding is useful to detect the difference in SD (openSAFETY Domain) division.

2.2. Cyclic Redundancy Check (CRC)

The most reliable means to identify changes to the original content is the CRC procedure, which uses a key to generate a checksum for each data set, and attaches that as well as the key as a bit sequence to the data set. This checksum is a distinctive encoding of the data set itself. Using the bit sequence and the key, the receiver calculates the original data set, and checks the result against the data set that was received in the clear. If any deviations from the original data content are detected, the message will be ignored [3].

That is why these types of generator polynomial g(x) for CRC are suggested with OpenSAFETY:

Field Length for the value 0 – 8 is CRC – 8:

$$g(x) = x^8 + x^5 + x^3 + x^2 + x + 0 = 12F\ hex \tag{1}$$

Field Length for the value 9–240 is CRC – 16:

$$g(x) = x^{16} + x^{14} + x^{12} + x^{11} + x^8 + x^4 + x^2 + 1 = 15935\ hex \tag{2}$$

Field length for the value 241-255 is reserved.
Related with it are the rules for creating CRC subframes of the openSAFETY where:

- MSB (Most Significant Bit) of the first byte is moved to the first place
- CRC is created by the whole field of subframes except CRC
  1. subframe consists of: ADR, ID, LE, CT(L), DB0 – DBn
  2. subframe consists of: ADR, ID, CT(H), TADR, TR, DB0–DBn
- Completion of the null bites into data selected for transfer (if they are smaller than LE).

3. Practical part

I have created a programme in a software called Matlab for transmission of openSAFETY frame (subframe 1 and 2) across the BSC (binary symmetric channel) noise canal with the $p_b = 0.01$ error probability. I have designed the programme according to suggestions in [4] and presented CRC Coding Example. I have used suggested polynom (1) mentioned above. I have created an m.file in which I have used following functions (see fig. 3): deconv, bitor, bsc. The testing was running for: subframe1: 0010001111001000000010000011010000010001
subframe2: 00100010110010000000100100101011000110000000010001

```
[q2 r2]=deconv(uimseg,poly)
r2=abs(r2)
for i=1:length(r2)
    aa=r2(i);
    if ( mod(aa,2)== 0 )
        r2(i)=0;
    else
        r2(i)=1;
    end
end
crc2=r2(length(uimsg)+1:end)
frame2 = bitor(uimseg,r2)

frame=[frame1 frame2];

frame = bsc(frame,0.001)
```

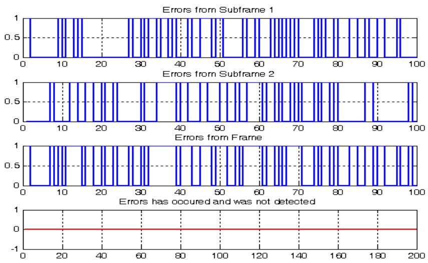


Fig.3. Left side-example shows source code via Matlab and right side-graph first measurement via Tab.1.

Under each measurement 100 messages were sent. Results from every individual measurement are presented in the table. The number of detected errors was evaluated.

S.N. detected errors	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
subframe1	35	44	37	37	38	35	32	33	31	32	35	34	37	38	39	28	35	35
subframe2	51	46	44	51	49	44	35	36	42	45	36	47	39	46	39	33	44	49
frame(sub 1,2)	74	71	65	73	67	66	56	54	60	61	63	65	60	66	58	55	64	67
undetected errors	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tab.1. Detected errors transmission of openSAFETY frame (subframe 1, subframe 2) and undetected errors across BSC.

Fig. 3 on the right shows selected signal flows from the tab.1 necessary for verification of the detecting ability of a given code.

4. Conclusion

OpenSAFETY chooses the combination of safety mechanisms, mostly CRC and data redundancy, which are used in other safety profiles (CIPsafety, PROFIsafe) as well. I would choose PROFIsafe to compare with OpenSAFETY, because both of them are implemented in our laboratories (Siemens – PROFIsafe, B&R – OpenSAFETY). Here we operate with real measuring and subsequently with real comparison of the quality of the safety mechanisms.

PROFIsafe is a method of protection. It is compatible with the PROFIBUS or the PROFINET safety networks. Standard transmission system includes all hardware and basic protocols of the reference ISO/OSI model [2]. Safety-relevant and safety-irrelevant applications share one transmission system (PROFIBUS, PROFINET IO) at the same time. The Safety-relevant functions consist of safety mechanisms which reveal errors of a safety-irrelevant transmission system or keep the error rate below the required level [6]. To eliminate the communication errors of a transmission system of industrial networks like repetition, deletion, insertion, delay, reordering, disruption, message masking and errors caused by the memory elements of a network, the following security mechanisms are recommended for the PROFIsafe profile [5]:

- Identification of the sender and recipient.
- Consecutive number (virtual).
- Data integrity check (CRC8 and CRC16).
- Timeout with Receipt).

OpenSAFETY enables cross checking in individual subframes. Through this it supports safety data transmission from sender to receiver. The probability of breakdown of both subframes is very low and with rising size even lower. OpenSAFETY driver profile for Powerlink devices is certificated by TÜV organization in the level of integrity safe SIL 3 [1].

To verify the correctness of data and to enable the receiving device to identify any stochastic error, the following principles will be used [4]:

- CRC-calculation (CRC8 and CRC16).
- Comparison of data-information of both subframes.
- Checking of time stamp.
- Checking of received address (comparison with internal address or address in look-up-table).

From these arise the following advantages of the safety profile openSAFETY:

- One safety standard for various technologies used in entire corporation
- Only one network development.
- Low initial costs thanks to the open source.

Disadvantages:

- It is a new profile and unlike PROFIsafe (with major distribution) it is still rarely.
- So far, it is implemented only in four types of Ethernet.

References

[1] IEC 61508 Series on the Safety of Industrial Automated Systems: Safety-related electrical, electronic and programmable electronic (E/E/PE) systems. 2005.

[2] ZEZULKA, F. - HYNČICA, O.: Průmyslový Ethernet V, Automa, 2007, 13th, n. 12, p. 58–61.

[3] OpenSAFETY, EPSG [online]. 2010, http://www.open-safety.org/

[4] EPSG W D P 304 V1.1.3 OpenSAFETY-protocol source [online 07.12. 2010], http://www.ixxat.de/zugangsdaten_powerlink_safety_de.html

[5] FRANEKOVÁ M.- KÁLLAY F.- PENIAK P.- VESTENICKÝ P.: Komunikačná bezpečnosť priemyselných sietí. Edis, ŽU Žilina, ISBN 978-80-8070-715-6, 2007.

[6] DIN EN 954-1: Safety of maschinery - Safety-related parts of control system. Part 1: General principles of design.1996

[7] Siemens AG 2009: Comunications – PROFIsafe profile [online 12.01. 2011]. http://www.automation.siemens.com/cd/safety/html_76/produkte/kommunikation/PROFIsafe.htm

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Digital Signage in Urban Mass Transportation in Žilina

[*]Radovan Madleňák, Lucia Madleňáková
[*]University of Transport, Faculty of Operation and Economics of Transport and Communications, Department of Communication, Univerzitná 1, 01026 Žilina, Slovakia, {radovan.madlenak, lucia.madlenakova}@fpedas.uniza.sk

Abstract. This article is dedicated to potential of digital signage and its influence on marketing communication in urban mass transportation in Žilina.

Keywords: digital signage, GEM, urban mass transportation system, marketing

1. Introduction

There is no guaranteed evidence of the first advertisement. Some specialists consider wall paintings in the caves as some sort of advertisement. Advertisement on papyrus was used in Egypt to spread the commercial messages. Similar method was used in Pompeii and ancient Arabia especially to spread political messages. Advertisement experienced rapid development after the year 1447. It was the year when Johannes Gutenberg invented printing. Therefore sellers could spread their messages faster, easier and more efficient. The very first television commercial was shown on WNBT station in New York. "Bulova" company advertised its watches before the baseball game on July 1st 1941 with the slogan "America runs on Bulova time". Sponsoring widely used in radio broadcasting was employed in television as well. In the late 1990s the Internet boom overloaded the advertising market. Target marketing in the condition of Internet achieved new dimension. One-on-one marketing enable communication with potential customer more effectively and geographic distances are no longer obstruction for purchase.

1.1. Commercial signage

Commercial signage is any kind of message, information, graphic, picture or sign. It usually promotes particular brand or store but we can also find some which encourage people to live healthier e.g. "Got milk?" This is a very popular American campaign for milk consumption. Over two hundreds famous stars including actors, singers, musicians or athletes have participated on the campaign since 1993. However signage is mostly used for business. Advertisers which adopted new form of marketing (e.g. guerrilla marketing) skipped the usage of traditional commercial signage including billboards or neon signs and shifted to more innovative and more progressive methods. Skywriting is one of the most attractive commercial signage but it is very expensive therefore is used mostly for mass sport events.

After adoption of ICT for advertising, advertisers were looking for new possibilities how to deliver interactive and multimedia content to customers. They needed something progressive at fair price. The lifestyle changed a lot during last ten years and people spend more time out of their homes. It is more difficult to catch people "on the go" and deliver them the advertising content because the motto of nowadays is: "Time is money, money is time!" Advertisers soon realized they need to find niche place in this fast era. Therefore any free time and free place that is suitable for advertising was utilized. As the target marketing and interactive content is one of the most effective methods, marketers wanted to bring this concept outside. The best way how to achieve this goal showed up with expansion of big screens. Digital screens started to appear on private and public places as well. As these screens bring multimedia advertising content they are called multimedia signage or according to the technology **digital signage.**

1.2. Digital signage

The digital signage can bring different content to the customers and therefore it can be used in many places for various purposes. The content that is shown on the screens is easy to modify. People sometimes do not realize they are in contact with digital signage. As the final product that comes into view is mostly ordinary flat screen. There are many possibilities where to implement digital signage. The list is not completed and places depend on the company which is installing this type of advertisement. However there are certain places where we come across digital signage more often.

One of the common applications of digital signage is for public information. Municipalities install interactive touch screens on the places with high visit rate. Inhabitants and tourists can easily find shops, historic monuments, sport center or even look for free job positions. This type of digital signage considerably complements info center and increases amount of points of contact with citizens. Digital signage is implementing as internal information channel in companies. Employees are informed about corporate news, goals, missions, visions and internal rules via digital screens placed on frequented areas in company buildings. As companies want to increase their brand recognition they place advertising screens in their stores and buildings. We can come across this type of digital signage in banks for example. Companies are promoting their own products and direct customer's attention to special offers.

Flasma is another innovative method of digital signage implementation. The idea arises from the walls overloaded with commercial messages. Customers therefore walk with their eyes stick to the floor. And floor has not been utilized for commercial purposes enough yet. This innovative approach can be successfully used in shopping malls and pedestrian zones.

2. GEM systems

GEM is a shortcut for Global Entertainment Media. It is a software solution which enables its users to display different type of digital content on the screens. It is most suitable for public locations e.g. urban mass transportation, train or bus station, airport or even shopping malls. GEM combines *three technologies* in one solution. First there is positioning, usually GPS. If you want to deliver up to date content according to the place, you need to know what your actual position is. Second there is mobile telecommunication. As long as the concept is interactive, there has to be a possibility for back coupling from the users. And finally there is digital signage.

One of the most suitable places where to apply GEM system is urban mass transportation. Anyhow all kinds of public transport are appropriate for digital signage. There are clearly several benefits for passengers. Traveling and especially commuting could be tiresome. In public transport there is no time and no place to start any kind of work. Passengers are therefore looking around for some entertainment. Advertising posters become boring very soon. Changing them every other day is very expensive and therefore impossible. Advertisers are searching for new ways of delivering the content to the customers more effectively. They discovered opportunity in using digital screens on frequented places, e.g. buses, trams or trains. They can capture broad masses of different segments of customers. It is much easier to deliver the right content to the right segment because it is predictable who is traveling on which bus line on which time.

GEM system is based on two different triggers – time trigger and GPS trigger. GPS triggers initiate advertising content with regard to the location. As the bus enters the geographic cell of opera house for example, its computer gets information about it. As the result the commercial for new opera play appears on the digital screen. Passenger can even book the tickets via SMS. Very efficient method how to attract customers is offering them bonuses. Shops in big malls can use

GEM Interactive to offer discounts in the form of number code as a counter value for SMS. In the geographic cell of shopping mall the commercial for clothing store appears on the screen. It invites passengers to visit the store. If they send SMS with the key word on the given number they receive SMS with number code which can be used for discount in the store.

Second type of trigger is based on time. In some areas where no shopping malls, no theatres and no other amusement facilities are is no need for commercials. For that reason news or weather information can be shown. Time triggers initiate the content according to expected time in timetable. However this type of trigger is not that reliable because of traffic jams for example.

3. Survey of GEM implementation in Žilina's urban mass transportation

We will consider on implementation of digital media network in urban mass transportation in Žilina, as GEM System is designed for public location and urban mass transportation. As digital signage is not wide spread in Slovakia and nor in Žilina we need to collect information about the awareness among residents and their interest in this advertising tool. The tool for collecting requested information will be used **questionnaire**.

For the purposes of the research we will consider on two customer's segments. Segments will be divided on the basis of demographic – specifically on the age. Millenials is so called demographic group which members were born between 1979 and 1994. This group is very promising for marketers. More than 90% members of this group use Internet and therefore we will conduct online survey. As Millenials is big group we will divide it in two subgroups. First subgroup (segment A) will include 15 to 21 years old people. Second group (segment B) will include 22 to 30 years old people. The questionnaire is divided to three parts. First part of the questionnaire is oriented on the public transportation. We need to know whether Millenials use urban mass transportation in Žilina and if so how often they travel. Second part is focused at digital signage and how do Millenials perceive it in comparison with printed posters and third part of questionnaire is dedicated to awareness of digital signage.

3.1. Travelling by urban mass transportation in Žilina

As investment into digital media network requires lot of money and effort we wanted to be sure that urban mass transportation is suitable for that. Advertisers are looking for places with big amount of passing people who spend some time there. 85% of respondents use urban mass transportation in Žilina on a regular basis. 3,5% of respondents use it sporadically and only 11,5% of respondents do not use urban mass transportation at all.

The most utilized bus line is number 14 which serves the area of ZOC MAX and TESCO hypermarket. Second one is line number 4 which serves the same area. On the third place is bus line number 6 which operates in the area of OC Dubeň. The most utilized lines serve areas with existing shopping malls. Implementing the digital signage in these lines at the beginning of the project would have impact on many passengers and potential customers of shopping malls.

3.2. Awareness of digital signage

According to results of questionnaire, the average age gap between segment A and segment B was three years and eight months. This age gap represents considerable difference in awareness of digital signage between the segments. 59% of all respondents did not come across digital signage or were not sure about that. However in the segment A it was 64%. On the other hand 54% respondents of segment B did not have experience with digital signage. This might be caused by greater opportunities of the older segment to travel abroad for example because of exchange studies or internships. This segment contributes of employed members more than segment A.

The people have already noticed multimedia screens in shops. Many of them, especially in segment B, know digital signage from foreign countries. Many respondents commute to Žilina and have noticed the screen in train station which display short informative presentations about railways. Printed posters still have strong position on the advertising market. However the power of moving pictures is stronger. Both segments are attracted with digital signage more than with printed posters.

Another question was dedicated to interest in printed posters in buses. This helped us to ensure that advertisement in urban mass transportation has impact on viewers and still captures their attention. 25% of respondents notice printed posters regularly and 71% notice sometimes. Only remaining 8% do not pay attention to printed posters in buses at all. When advertisers would be able to bring more attractive commercial content to the buses they could gather even more viewers and therefore customers. Digital signage offers this in one solution and urban mass transportation provides viewers which have to stay at the same place for a longer time.

3.3. Opportunities for passengers

It is necessary to make sure whether passengers even want to have digital media network in buses and whether they would utilize the services which are offered by digital signage. More than three quarters of all respondents would appreciate digital signage in public transport in Žilina. The difference between segment A and segment B is 11%. GEM System offers two opportunities for passengers how to interact with it. First is entertainment SMS. Interest in this service is almost the same in both segments. 53% members of segment A and 52% members of segment B would use this service. However the difference is in the certainty. More members of segment A are sure they would use entertainment SMS if available than members of segment B. Second opportunity for passengers offered by GEM System is discount SMS. This service is offered in cooperation with advertisers. 63% respondents from segment A are ready to use the opportunity to obtain discount as a counter value for sending SMS according to instruction on digital signage. 50% respondents from segment B are ready to do the same. Offering discount and small presents would lure more customers to the shopping malls.

4. Conclusion

This article was only introduction to the whole GEM concept. Before the investment into digital media network in urban mass transportation further investigation is needed. We considered only on passengers in this article. Whether there even is interest into digital signage. Now as we know that people are ready to accept this type of advertisement and would like to use its opportunities it is desirable to focus on advertisers, shopping malls and provider of public transport in Žilina. Financial calculations are needed as well. Although we pointed out that ROI of digital signage is much bigger that ROI of printed posters it is necessary to support the project with other calculations according to the conditions in Slovakia and Žilina.

References

[1] KOTLER, P., ARMSTRONG, G.: Principles of Marketing. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2007. ISBN 0132390027, p. 408.

[2] STRAUSS, J., FROST, R.: *E-marketing*, fifth edition. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2008. ISBN 0136154409.

[3] TUNGATE, M.: *Adland: a global history of advertising*. Kogan Page Publishers, Great Britain, 2007. ISBN 0749448377.

[4] VACULÍK. J.: *Marketing v prostredí internetu*, In: Zborník príspevkov KIT 2003: 26.-28. novembra 2003, Liptovský Mikuláš: Vojenská akadémia, 2003. - ISBN 80-968711-4-5.

[5] ČOREJOVÁ, T., MAJERČÁKOVÁ, M.: *Rozvoj informačno-komunikačnej infraštruktúry SR vo vzťahu k budovaniu znalostnej ekonomiky*. In: AIESA medzinárodná vedecká konferencia: 2008, Bratislava, ISBN 978-80-8078-233-7.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Automatic Image Annotation to Reduce Semantic Gap

*Ellen Molitorisová
*University of Žilina, Faculty of Management Science and Informatics, Department of transportation networks, Univerzitna 1, 01026 Žilina, Slovakia, {Ellen.Molitorisova}@fri.uniza.sk

Abstract. The number of available data brings challenge of its effective and efficient storage and retrieval. Reaching this aim in area of visual data is complicated by semantic gap. To reduce this gap process of automatic image annotation is utilized where set of keywords representing the semantic content are added to images. In this paper is depicted and discussed process which consists of set of these steps: segmentation, low-level feature extraction, annotation step and post-processing. Additional clustering is proposed to enrich text labels related to images and so improve the effectiveness of image retrieval.

Keywords: content based image retrieval, automatic image annotation, semantic gap.

1. Introduction

The way of efficient access to information, in time of its overloading, has become a serious topic for research. The effective way to manage and store images is important in many cases: management of the personal digital photos, in cultural sphere, in medicine, in Large-scale Image Collections at NASA [1, 2]. One way how to handle the problem of the image retrieval is to acquire images with related keywords, so the image retrieval process is transformed into the text retrieval. Indexing images with relevant text descriptors is aimed by image annotation techniques.

2. Content Based Image Retrieval

Currently proposed frameworks for the image retrieval, which are embedded into the content-based image retrieval (CBIR) systems, are taking advantages from the analysis and recognition of the visual content of the images. CBIR systems work with visual content represented by low-level features such as colour, texture, shape, spatial layout, etc. Input queries in the image retrieval could be image, sketch or keywords. Naturally, the most convenient way is to use keywords, because to interpret images, humans tend to use high-level features (concepts) such as text descriptors.

The consequences were described by Smeulder [3] as a semantic gap: "The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation." The situation is illustrated in Fig.1. Datta in [4 7] note that the semantic gap is in its fundamental dynamic in a real world context. Information extracted from the visual content is static, but interpretation of the image could change through people and time. So, to ensure the effective image retrieval and management it is necessary to join and use both of the image aspects – low-level visual features addressing more detailed perceptual aspects and high-level semantic features underlying more general conceptual aspects of visual data.

According [1, 5] the core techniques applied in the image retrieval include:

- visual signature
- similarity measures
- classification and clustering and
- relevance feedback-based search paradigms
- using object ontology to define high-level concepts

- generating semantic template to support high-level image retrieval
- fusing the evidences from HTML text and the visual content of the images for WWW image retrieval.

These techniques are applied in the proposed frameworks of the annotation systems in various combinations.

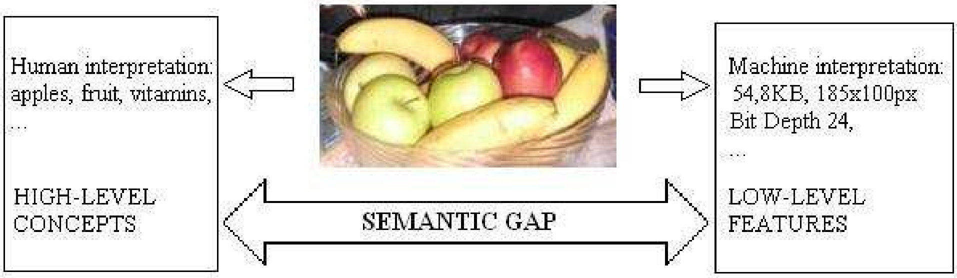


Fig. 1. Semantic gap

3. Automatic Image Annotation

Automatic image annotation utilizes a visual content with aim to bridge the semantic gap. The whole process could be divided according to [6] into several steps (see Fig. 2). The segmentation component partitions images into the local contents via either a block or region based method. The feature extraction component extracts low-level features from the segmented images, so each block or region is represented by feature vectors. Next, the annotation component assigns the (low-level) feature vectors to the high-level concepts. Finally, the post-processing component (dependent on the application) uses the output of the annotation component to decide on a recommended action for the final decision.

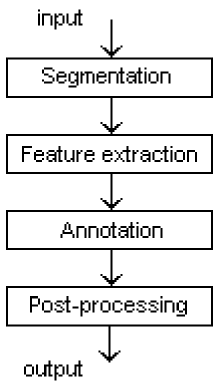


Fig. 2. Block diagram of the image annotation systems

3.1. Segmentation Step

There are two strategies applied in the segmentation: either divide image to a set of the fixed sized blocks or tiles or partitioning into the variable shaped regions of the interest [6]. Variety of the techniques have been used such as basic k-means clustering, normalized cuts criterion, expectation-maximization (EM) algorithm, blob world segmentation, JSEG algorithm [1, 5]. Segmentation methods could involve some drawbacks such as computational cost or reliability of a good segmentation. Some researchers try to go around these liabilities using global features or applying additional methods [1].

3.2. Low-level Features Extraction

The basic types of the visual features are colour, texture and shape descriptors. Colour features are in annotation systems the most widely used. Common colour descriptors include colour histogram, colour moments, colour coherence vector [5, 7] and colour descriptors defined in standard MPEG-7. Texture features are intended to capture the granularity and repetitive patterns of the surfaces. In addition to the colours, it has been extracted to classify and recognize objects and scenes. Effective extraction of the shape features depends on the segmentation methods. Compared to the colour and texture, it is difficult to apply shape characteristics due to the inaccuracy of the segmentation [5, 6]. Except for colour and texture features standard MPEG-7 defines also three shape descriptors.

3.3. Annotation Step

Annotation approaches are by various authors classified in different ways. According to the structure in [8 9] they are classified into: (1) free text descriptors – no pre-defined structure for the annotation (2) keywords - chosen from controlled vocabularies (3) classifications based on ontology – keywords belongs to ontology (large classification systems that classify different aspects of life into the hierarchical categories). Datta et al. in [1] recognize two schools: Joint Word-Picture Modelling Approach and Supervised Categorization Approach. In the first one, image is represented by properties of each of their segments or blobs, thus the pictures under such models are treated as bags of words and blobs. After learning joint word-blob probabilities the annotation problem is reduced to the likelihood problem relating blobs to words and solved by applying statistical tools. Following models belong in this group: machine translation model, Latent dirichlet allocation (LDA) model, relevance models (Cross Media Relevance Model, Continues-space Relevance Model, Multiple Bernoulli Relevance Model, Dual Cross-Media Relevance Model).

An alternative approach treats image annotation as a supervised categorization problem, where annotation systems classify visual features into some pre-defined classes. Supervised learning models as artificial neural networks, decision trees; k-Nearest Neighbour and Support Vectors Machines (SVM) are used [6].

3.4. Post Processing

The last component following annotation step is post-processing. As an optional and a very specific process, it depends on the concrete annotation framework and covers the actions taken after labelling images.

4. Selected Approach and Future Work

Annotation can facilitate the image search through the use of the text. Several core techniques are applied in this process. Segmentation is an optional stage. It could help to exclude non-important parts of the image or to define objects presented in the image. On the other hand, inaccurate image segmentation could involve difficulties in next processing steps. Extraction of the low-level features is a necessity. Decision about characteristics of the best-representing image content with acceptable computational cost is another point which significantly influences the annotation result. The choice of the image signature to be used should depend on the desirability of the system.

Annotation step is crucial in the whole process. There were presented several approaches applying different methods. In the recent years, classification with SVM became one of notable techniques. These binary classifiers separate data by hyper-plane and in combination solve multiclass problems. SVM could be successfully utilized in the image retrieval or more precisely in the automatic image annotation process. Many researchers have already involved SVM into their annotation framework [9, 10, 11, 12]. SVM appears as promising and practical method to solve above mentioned tasks.

To improve future retrieval as the cause of annotation next stage is proposed. After the annotation step images are clustered into collections according to text labels which have already been added. More general concepts describing these collections are linked to the images. Selection of suitable source of concepts depends on a concrete matter. The universal lexical database of English WordNet where semantic relations are included is one promising adept for this purpose. In situations with indistinct text queries which are rather used at the start of searching is the need for this kind of concepts. Finally, including of clustering step is expected to improve search result.

There are still open issues in the whole process which should be the area of the research in my future work.

5. Conclusion

Need of the effective and efficient tools for multimedia management is in many fields significant. The image management and retrieval are challenging research areas because of the existing semantic gap between low-level features and high-level concepts. The automatic image annotation tries to bridge this gap by labelling images with the text. Due to many implemented constraints in this process, there is still a long way to go. However, with collaboration of computer vision, image processing techniques and annotation, the semantic gap will be gradually bridged.

References

[1] DATTA R., JOSHI D., LI J., WANG J.Z: Image Retrieval: Ideas, Influences, and Trends of the New Age, ACM Computing Surveys (CSUR), Vol. 40, No.2 ACM, 2008.

[2] HALASCHEK-WIENER CH., SIMOU N., TZOUVARAS V. (2007) Image Annotation on the Semantic Web. [Online]. Available: http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation/

[3] SMEULDERS A. W. M., WORRING M., SANTINI S., GUPTA A., JAIN R.: Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, IEEE Computer Society, 2000.

[4] DATTA R.: Semantics and Aesthetics Inference for Image Search: Statistical learning approaches, PhD Thesis, The Pennsylvania State University, 2009.

[5] LIU Y., ZHANG D., LU G., MA W.-A..: A survey of content-based image retrieval with high-level semantics, Pattern Recognition, 40, Elsevier Science Inc., 2007.

[6] TSAI CH. F., HUNG CH.: Automatically Annotating Images with Keywords: A Review of Image Annotation Systems, Recent Patents on Computer Science, Vol.1, No.1, Bentham Science Publishers Ltd., 2008.

[7] ZHAO R., GROSKY W. I.: Bridging the semantic gap in image retrieval, Content-based retrieval and image database techniques, Distributed Multimedia Databases: Techniques and Applications, T. K. Shih (Ed.), Idea Group Publishing, Hershey, Pennsylvania, 2002.

[8] HANBURY A.: A Survey of Methods for Image Annotation, Journal of Visual Languages and Computing, Vol.19, No.5., Academic Press, Inc., 2008.

[9] LINDSTAEDT S., MÖRZINGER R., SORSCHAG R., PAMMER V., THALLINGER G.: Automatic image annotation using visual content and folksonomies, Multimedia Tools and Applications, Vol. 42, No.1, Kluwer Academic Publishers, 2009.

[10] GOH K.-S., CHANG E. Y., Li B.: Using One-Class and Two-Class SVMs for Multiclass Image Annotation, IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.10, IEEE Educational Activities Department, 2005.

[11] GUAN H., ANTANI S., LONG L. R., THOMA G. R.: Bridging the semantic gap using ranking SVM for image retrieval, Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro, IEEE Press, 2009.

[12] GUO J., LIAO X.: Cross-Media Image Retrieval via Latent Semantic Indexing and Mixed Bagging, CSIE '09: Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering, Vol. 04, IEEE Computer Society, 2009.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Iris Detection for Automatic Acquiring of Eye Template

*Tibor Moravčík
*University of Žilina, Faculty of Electrical Engineering, Department of Control and Information Systems, Univerzitná 1, 010 26 Žilina, Slovak Republic, tibor.moravcik@fel.uniza.sk

Abstract. This paper deals with a problem of image template acquiring which included part of eye from center to iris boundary. In my application a template is used as the first step in system for eye center detection in video sequence based on template matching. Method of iris boundary detection is the most important to obtain template. There are many methods to solve this problem. This paper presents two most popular approaches (Daugman's integro - differential and circular Hough transform). The first part of the paper includes work related to this topic and chosen methods are theoretically described with pre-processing techniques. Both methods have been implemented in programming environment MATLAB. An accuracy of these methods on the set of eye region images has been tested in the second part of the paper. Results from the testing, pros and cons of mentioned methods and practical experiences are compared in the conclusion.

Keywords: digital image processing, iris detection, circular Hough transform, integro – differential operator

1. Introduction

The number of real-world applications needs the ability to non - invasive localization of the human eye. Information about center and boundary of pupil and iris are the most significant for monitoring of eye movement, for iris extraction used for biometrical identification purpose, for blink detection etc. Template of eye is used for iris recognition as one of the most reliable biometric measures and also it can be used for quick eye detection through the template matching in frames of a video sequence for further analysis of eye behavior in the time. Many factors affect accuracy of the detection. The most important is a quality of input image what depends on lighting, noise, quality of camera and resolution. The resolution of input image is an important factor, which limits the choice of adequate methods. Processing time is next important parameter of these methods. It depends on application type to prefer one of them.

1.1. Related work

Many different works have been done in this field by using image processing tools for implement such systems. One of the most relevant approaches has been presented by Daugman [1]. He introduced the integro - differential operator to find both inner and outer borders of the iris. This operator remains still actual, and was proposed with some minor differences in 2004 by Nishino and Nayar [2]. Camus & Wildes [3] with a similar way identify the iris borders through maximization of equations over an N3 space. Wildes [4] uses for iris segmentation binary edge map followed by circular Hough transform. Liam et al. [5] have proposed a simple threshold method with function maximization to obtain iris inner and outer borders. In other approach is finding of approximated pupil center as minimum value of the summation of intensity along each row and each column. Then Canny edge detection and Hough transformation are applying for detection of exact pupil center. Also morphologic operators, Laplacian or Gaussian operator for edge detection with median filter can be applied to obtain iris borders.

Based on mentioned previous work and with respect of our use case, we chose to use integro – differential operator and circular Hough transform as two methods to compare each other. Both are able to detect iris and pupil border, in our application is this reduced only to searching for an iris edge. Methods with their pros and cons use circularity and contrast different between iris and sclera as main features for detection.

For increase of accuracy and speed of proposed system is used procedure presented in [6], which consists from followed pre-processing techniques:

- **Histogram equalization -** This operation improves the contrast between eye's regions, thus contributing to the correct algorithm segmentation.
- **Specularity suppression -** Attenuates the influence of specularities in the image, which can disrupt a pupil-iris or iris-sclera boundary.
- **Binarization -** Applying a threshold on an image before the operator's execution enables the maximization of the contrast between the regions belonging to the iris and the remaining ones. This process is highly dependent on the threshold value.

1.2. Facts about human eye

The iris is the annular part between the pupil and the sclera. Iris and pupil can be taken approximately as non-concentric circles. Two additional eye features are the upper and lower eyelid. In normal state is iris not completely visible, because its top and bottom part are covered by eyelids and eyelashes. Pupil size varies depending on the light conditions and it is darker than the rest of the eye, even in brown or dark eyes. The iris is always darker than the sclera no matter what color it is. In this way the edge of the iris can be detected as the set of points that are disposed on a circle. In Fig. 1 are shown all main features in the human eye.

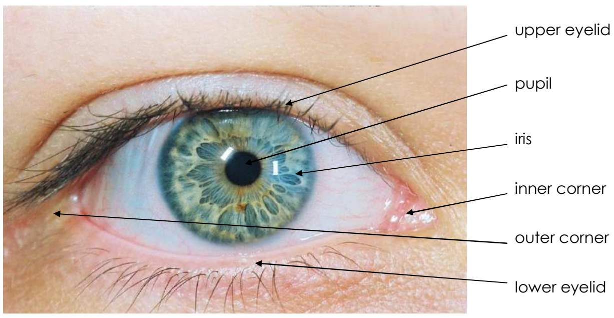


Fig. 1. Eye features

2. Technical approach

Eye image regions of both left and right eyes intended for iris localization are obtained from face image with previous face detection [7] process and proportionally selected from this area. These eye regions converted to gray scale then entering to pre-processing part. Histogram equalization maps the intensity values in gray-scale image to new output values such that 1% of data is saturated at low and high intensities of input image. This increases the contrast of the output image J.

2.1. Suppression of specularity

Specularity is a very bright area in the image, which is formed from reflections of the cornea of the eye, or of the lenses of glasses or contacts if present. These bright areas can significantly affect detection of iris boundary, because degrade of iris shape. The way of specularity filling is based on finding of very bright parts in the image via thresholding with threshold value 250 from range [0, 255]. Inside of these parts is filled with interpolated image intensities from surrounding pixels (Fig. 2). Interpolation is realized through solving of system of linear equations with matrix representation

$$x = A^{-1} b, \quad (1)$$

$$A = \begin{bmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \vdots & \vdots \quad \ddots & \vdots \\ a_{m1} & a_{m2} \cdots & a_{mn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

where A is matrix m x n, b is vector of sums of four surrounding intensities, x is vector of calculated intensities for inside pixels.

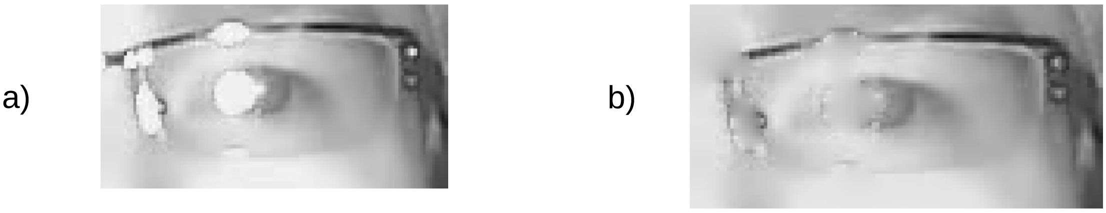


Fig. 2. Specularities filling in image a) with visible specularities, b) with suppressed specularities

2.2. Selection of candidates

Candidates are selected as points that correspond to the local minima of image intensity. For current purposes, a local minimum is defined as an image intensity value below a certain global threshold and also one that is the smallest value within a 5x5 pixel region. For cases discussed in this paper, the global threshold was set to 128 from a range of [0, 255]. Fig 3. shows result from selecting.

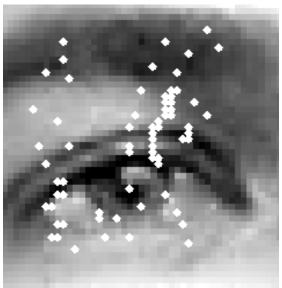


Fig. 3. Candidates for center of eye

3. Integro - differential approach

For localization of outer iris circle Daugman [1] introduced an integro - differential operator which implements circular edge detection with pseudo-polar coordinate transform. This operator allows to obtain a first approximation of the both iris and pupil boundary. This method searches in an N^3 space for the circumference center and radius that have the highest derivative value in compare to neighbor radius.

It is not possible to know the exact diameter of the iris since people can have different iris dimensions and also the system has to manage variable distances between people and the camera. For this reason a range $[r_{min}, r_{max}]$ is set to tackle different iris radius. But larger range of radius causes more computational time and memory consumption. These values are estimated from ratio of iris size to eye height h: $r_{min} = 0.1h$, $r_{max} = 0.3h$

Operator is for each from selected candidates with coordinates (x_0, y_0) and with increasing radius in range $r \in (r_{min}, r_{max})$ defined as

$$max(r, x_0, y_0) \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right|. \qquad (2)$$

where $I(x, y)$ are intensities of the gray-scale image, $G_\sigma(r)$ is a smoothing function, in this case is used Gaussian function with parameter ó, and the contour integral is along circles circular arc ds given by center (x_0, y_0) and radius r.

In other words operator searches maximum in the blurred partial derivative what correspond with boundary of circle in one candidate. Then maximum from all this values for all candidates represents center of iris. In Fig. 4 are compared two cases: correct (Fig.4a) and incorrect (Fig.4b) detection of the eye center via integro - differential operator. First image from left on Fig.4 shows input eye region with center point and two circles with radius r_{min} and r_{max}. Middle image includes all pixels inside radius range without top and bottom part affected with eyelids. Average values from each circle arc created through integral are represented with one row. Lower values are brighter (sclera) and higher with darker (iris) pixels. Boundary between iris and sclera is visible on differential row with higher (the brightest) value. This transition is not visible in Fig. 4b, because there is no circle in the radius range.

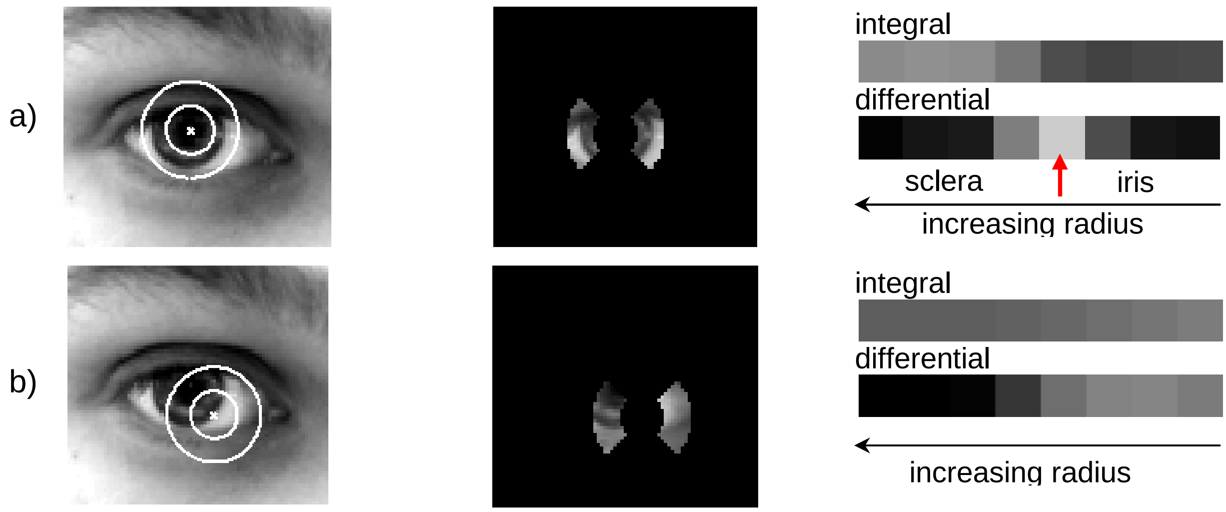


Fig. 4. Results from apply of integro - differential operator: a) correct center of eye, b) wrong center of eye

4. Circular Hough transform approach

This technique is finding imperfect instances of objects, in our case circles, within a certain class of shapes by a voting procedure. This voting procedure is carried out in a parameter space, from which object candidates are obtained as local maxima in an accumulator space [8]. The main advantage of the Hough transform technique is that it is tolerant to gaps in feature boundary descriptions and is relatively unaffected by image noise.

A thresholding on the gradient magnitude is performed before the voting process of the circular Hough transform for removing of the 'uniform intensity' image background. So pixels with gradient magnitudes smaller than gradient threshold are not considered in the computation. Also for this method a range of radius is needed to tackle different iris radius.

The result from Hough transform is an accumulation array in which higher values represents areas with the occurrence of circles in image with specific radius. To build the accumulation array is need to compute the gradient and the gradient magnitude of the selected eye region. It is the first derivative of 2D image. For pixels whose gradient magnitudes are larger than the given threshold are created the linear indices, as well as the subscripts. The accumulation array of the image consists of the gradient magnitude of the image and its linear indices. Fig. 5 shows all steps of circular Hough approach.

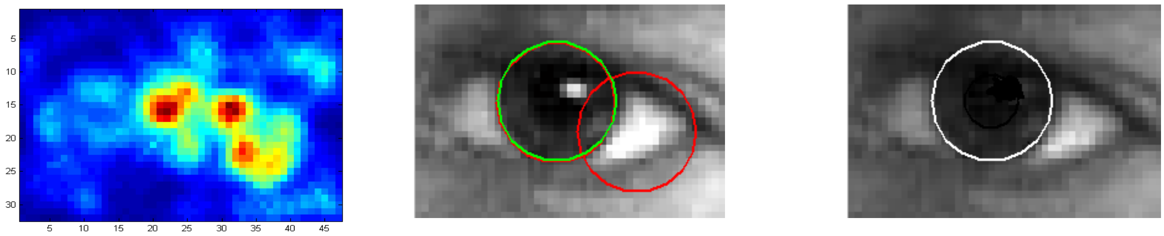


Fig. 5. Example of a) accumulator array, b) candidates for iris boundary, c) results from iris detection

5. Tests and results

Accuracy of both methods integro - differential (next in the figure only indiff) and circular Hough transform for iris boundary detection have been tested and compared in MATLAB. Tests are based on comparison between detected and real center and radius of iris. Real position of center and length of radius have been defined by user with placing of circle on iris boundary into the each image from set. There is obviously included influence of human factor in the phase of marking, so it is impossible to get fully objective results. For this reason set of eye regions have been created. The set includes about 400 left and right eye regions selected from face images, which have been captured with infra-illuminated camera from 3 persons. Size of eye images (about 60x60 pixels) varies depends on face size. Comparison of center points (real and detected) is done by calculating of the distance between these two points (Fig. 6a). Radii are compared as difference between real and detected radius (Fig. 6b). If values of comparison are equal to zero, circles are the same. Detection accuracy decreases with increasing of these values.

a)

b)

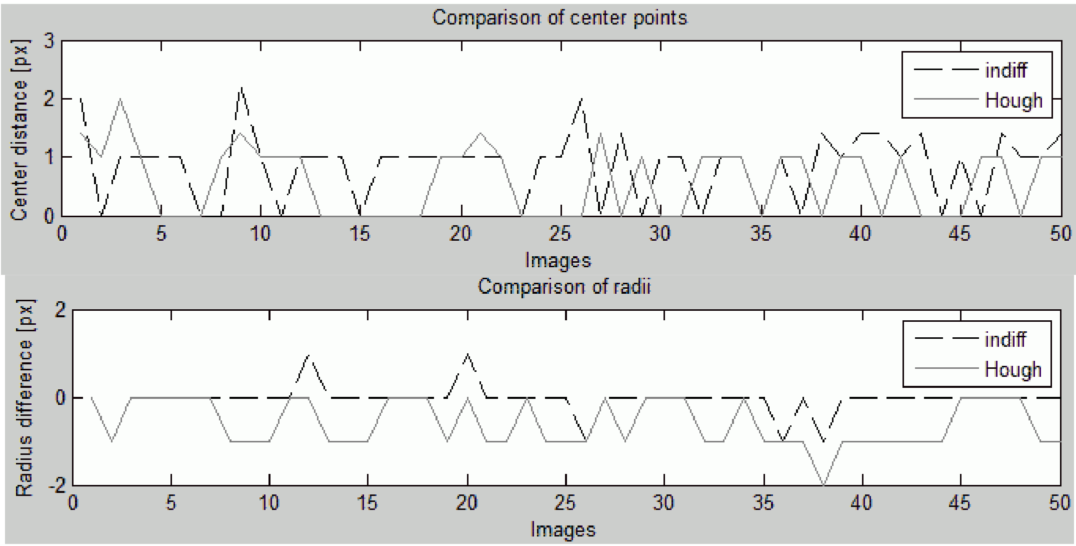


Fig. 6. Example of accuracy testing for both methods center and radius comparison

Tab. 1 shows all results obtained from tests. For different conditions the data are reported separately for left and right eye because there were some differences in image lighting. Average percentage values are presented for all cases. If detected center is outside of real circle there is 0% and from center to radius 100-0%. When radii are equal, then accuracy is 100%, and radius higher or lower than half of reference radius accuracy is 0%. Average processing time is also included to compare performance of both methods. This time obviously depends on PC configuration (in our case was Intel C2D 2,2GHz) and MATLAB (2009b) possibilities.

Test conditions		Center comp. [%]		Radius comp. [%]		Proc. time [ms]	
		Indiff	Hough	Indiff	Hough	Indiff	Hough
centered iris	right eye	91,4	94,1	97,7	88,5	278	38
	left eye	92	86,4	93,1	96,4	292	30
different radius	right eye	91,6	80,1	93,9	84,9	352	47
	left eye	94,7	83,9	94,5	81,1	344	43
eye movements	right eye	91,5	85	93,9	89,1	272	36
	left eye	89,3	83,7	94,5	90,9	250	34
with eyeglasses	right eye	71,9	40,85	81	79,6	274	43
	left eye	59,6	51,9	80,9	77,6	271	40
Average values		85,3	75,7	91,2	86	291,6	38,9

Tab. 1. Results from comparison between user marked and detected center and radius in multiple images

6. Conclusion

From results of Tab. 1 is obvious that both tested methods are very good usable for detection of iris boundary. In the first test condition with centered iris in the eye was almost equal accuracy for the methods. Second test shows that Hough approach is more sensitive to change of radius size in images so more specific radius range is needed for better results. In third condition where were images with non-centered iris both methods have a bit lower accuracy because iris affects its shape in the corner of the eye what complicates detection process. The last test was aimed to testing ability of the methods to detect iris in eyes covered with eyeglasses significantly affected with specularities from camera infra-illumination. In both methods were specularities suppressed (chap. 2.1). Better detection ability in this case has integro-differential approach; overall this method has better results in all conditions. Hough approach has advantage in shorter processing time what is about 10 times less than in case of integro-differential approach use. Higher processing time can be reduced with reduction of selected candidates through changing of threshold value. Based on these facts integro-differential approach is preferred for template acquiring application because accuracy is more important than processing time.

Acknowledgement

This work has been supported by the Grant Agency VEGA of the Slovak Republic, grant No. 1/0023/08 “Theoretical apparatus for risk analysis and risk evaluation of transport telematic systems".

References

[1] DAUGMAN, J. G.: *High confidence visual recognition of persons by a test of statistical independence*, IEEE Trans. Pattern Anal. Mach. Intell., 15, (11), pp. 1148–1161, 1993.
[2] NISHINO, K., and NAYAR, S.K.: *Eyes for relighting*, ACM Trans. Graph, 23, (3), pp. 704–711, ISSN 0730-0301, 2004.
[3] CAMUS, T.A., and WILDES, R.: *Reliable and fast eye finding in closeup images*, IEEE 16th Int. Conf. on Pattern Recognition, Quebec, Canada, pp. 389–394, 2004.
[4] WILDES, R.P.: *Iris recognition: an emerging biometric technology*, Proc. IEEE, 85, (9), pp. 1348–1363, 1997.
[5] LIAM, L., CHEKIMA, A., FAN, L., DARGHAM, J.: *Iris recognition using self-organizing neural network*, IEEE, 2002 Student Conf. On Research and Developing Systems, Malaysia, pp. 169–172, 2002.
[6] ŠONKA M., HLAVÁČ V., BOYLE R.*: Image Processing, Analysis and Machine Vision.* Thomson, Toronto, ISBN 978-0-495-24428-7. 2008.
[7] ORAVEC, M.: *Face Recognition*, In-Teh, printed in India, ISBN 978-953-307-060-5, 2010.
[8] BALLARD, D. H.: *Generalizing the Hough transform to detect arbitrary shapes*, Rochester, USA, ISBN 0-934613-33-8, 1980.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Quality of Synthesized Speech: A Brief Overview of Methods

*Miroslava Mrvová
*University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications an Multimedia,
Univerzitná 8215/1, 01026 Žilina, Slovakia, {miroslava.mrvova}@fel.uniza.sk

Abstract. This contribution deals with the issue of synthesized speech. It introduces principles and approaches of creating this type of speech. In addition, the paper also focuses on the basic methods and techniques used to assess the quality of synthesized speech. Finally, this article also offers short overview of relevant experimental studies discussing issues of this kind of speech and its quality assessment.

Keywords: synthesized speech, synthesizer, text-to-speech systems, quality assessment.

1. Introduction

Nowadays, synthesized speech achieves massive increase of interest in the case of development and utilization. The reason might be the fact, that speech is the most natural human form of communication and therefore there are efforts to imitate human voices. Systems used for speech synthesis offer wide range of utilization, for example in a place where other way of communication can not be used or in the human computer interaction systems involving higher number of modalities. Therefore the synthesized speech is implemented in many applications of daily life, where this kind of speech replaces real human speaker. The synthesized speech is mainly deployed, for example in systems providing reports containing frequently changing and routine information (weather forecast, timetable), in systems offering different dialogue situations (games) or reading various scripts (SMS-reader, e-mail reader). Their aim is to make this part of communication available for people with some disabilities, such as visually handicapped people or people with dyslexia. This can be the reason, why we should deal with synthesized speech, transmitted through telecommunication channel and its quality assessment.

This paper is organized as follows: Section 2 presents synthesized speech in general. In particular, the paper describes the principles of synthesized speech formation and brief comparison between quality assessment of synthesized speech and naturally-produced voices. Section 3 shows short overview of relevant experimental studies discussing the issues of quality assessment of synthesized speech. Finally, Section 4 concludes this paper.

2. Synthesized Speech

2.1. Principles of creating synthesized speech

What does it mean synthesized speech? Synthesized speech represents artificially made speech, i.e. given text utterance spoken by computer. It is created by unifying pieces of speech, recorded by speaker and stored in speech database. These systems are also termed as speech synthesizers. They are based on transformation technology called text-to-speech systems (TTS). In order to realize this transformation, TTS consists of many algorithms and modules. Figure 1 shows schematic representation of text-to-speech system.

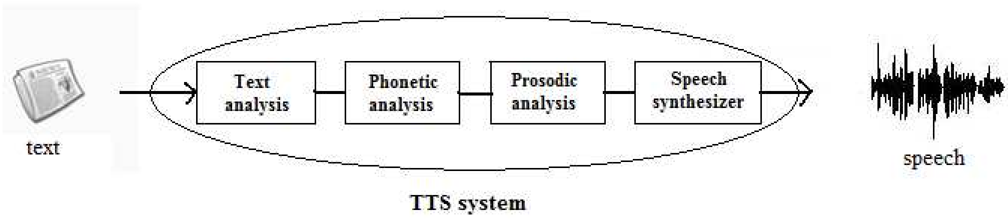


Fig. 1. Schematic representation of text-to-speech system (TTS).

In principle, functions of the TTS system can be divided into the following parts:

- Text analysis (normalization) – characterizes analysis of the text, which is separated into sentences. Numbers, abbreviations, symbols are replaced by their own word transcription,
- Phonetic analysis – transforms the text to voice (phonemes),
- Prosodic analysis – implements prosodic language characteristics, such as melody, speaking rate, volume, emphasis, pauses, accent, etc.
- Synthesis of the speech – generates speech signal from given sequence of prosodic modified phonemes.

Nowadays, there are three different approaches available to create this type of speech. The currently most widely used of them is *concatenate synthesis*, which is based on combining short speech strings to form a longer one. Output of this synthesis is the most naturally sounding synthesized speech. There are three main types of concatenate synthesis: unit-selection synthesis (uses large database of recorded pieces of speech, such as diphones, words, etc.), diphone synthesis (uses database all diphones finding in individual language) and domain-specific synthesis (database consists of recorded words and phrases). Other approach is *formant synthesis*, which is based on the fundamental frequencies of amplitude spectrum of voice. Systems deploying this synthesis generate artificial, robotic sounding speech, which can not be confused with natural speech. Finally, *articulation synthesis* represents new approach, which deals with straight human vocal track imitation, i. e. overall speech generation process. This approach has been poorly investigated, at the cost of its complexity. Definitively, synthesized speech should be indistinguishable from the human actual speech. It should characterize the most reliable copy not only in case of quality as well as in speaking style. There are efforts to ensure that synthesized speech will be the most natural, not fatiguing, not monotonous, and does not make efforts with respect to listening or comprehension [12].

2.2. Quality assessment of the synthesized speech

In principle, the methods designed for assessing the quality of naturally-produced speech can be also applied for assessing the quality of synthesized speech. Naturally, there are some limitations and modifications linked with their application on different kind of speech (e.g. accuracy, etc.). Those methods can be divided into two groups: *subjective methods* and *objective methods*. Quality of degradation of speech signal in case of *subjective method* is evaluated statistically. Method is based on opinions of group of test subjects (at least 24 people), who listen to given samples and fill out the questionnaire presented in Recommendation *ITU-T P.85*. This Recommendation is defined for subjective performance assessment of quality of speech voice output devices. Subjects are asked to express their opinion using 5–point opinion scale. ITU-T Rec. P.85 recommends the following rating scales: overall impression, acceptance, listening effort, comprehension problems, articulation, pronunciation, speaking rate and voice pleasantness. Assessment is based on rating MOS (Mean Opinion Score). MOS represents the average values representing opinions of testing subjects expressed on the 5-point quality scale varied from bad (1) to excellent quality (5). The Speaking Rate uses 5-point scale varied from too slow (1) to too fast (5) and the Acceptance uses only 2-point scale (yes - no). Each sample is played twice to each subject. In first phase subjects answer questions on the information finding in samples (e.g. train number, price the item). In second phase subjects are asked to assess the speech quality using one or more rating scales. For assessing the quality, two types of questionnaires, namely type I (intelligibility) and Q (quality) are used. Even though that, this method is criticized for its shortcomings; it is still frequently used [4].

In general, the quality of synthesized speech is evaluated in terms of intelligibility (how well the listener understands given samples) and naturalness (overall speech quality assessment). SUS (Semantically Unpredictable Sentences) belongs to the group of famous intelligibility tests. The semantically nonsense sentences with correct syntax are presented to subjects and their task is to correct the presented sentences. Each utterance is played only once. The most widespread naturalness test is MOS see details above (ITU-T Rec. P.85). Other example of them is Paired Comparison test (PC), where each sample in two variants is presented to subjects. Listener task is to choose one, which he prefers.

Finally, *objective methods* are based on the model of human perception. The final score is computed by mathematical model, which predicts the quality of the speech signal perceived by a user. At this moment, there are not standardized models for objective quality assessment of synthesized speech. However, there is ongoing research dealing with this issue, e.g. work presented in [6]. On the other hand, there are also ongoing efforts to verify whether the existing models designed to assess the quality of naturally-produced speech, like PESQ (Perceptual Evaluation of Speech Quality), P.563, ANIQUE+ (Auditory Non-Intrusive Quality Estimation Plus), are capable to predict the quality of synthesized speech to a certain degree [7-10, 13].

3. Overview of the studies dealing with quality of synthesized speech issues

As mentioned in the previous section, nowadays in the synthesized speech research area is tendency to verify ability of the models designed for evaluating the quality of naturally-produced speech and apply them to evaluate quality of synthesized speech. In order to realize this, many experiments were performed. Basically, the studies can be divided into three groups. First group of articles investigates subjective quality assessment, especially the accuracy and reliability of approach defined in ITU-T Rec. P.85. Secondly, the next group contains articles comparing the behaviour of objective models (like intrusive PESQ, nonintrusive P.563) with subjective assessments when the synthesized speech is deployed. Finally, there are also available studies dealing with the impact of various speech quality impairments (like noisy-type degradations, low bit rate codecs, etc.).

Authors in [1-3] investigated subjective assessment of quality of synthesized speech. In [1], the approach presented in ITU-T Rec. P.85 was compared with other available methods (test of intelligibility (SUS) and test of naturalness (MOS)) for evaluation of text-to-speech systems. Their aim was to investigate, whether this approach provides the better performance than SUS and MOS test. In particular, the reliability of this standard for evaluation of text-to-speech systems was investigated in [2]. Authors monitored how the ranking of TTS is changing across different text genres and listening sessions. Outputs were compared with pair-comparison test, using above mentioned aspects. In [3], the authors have compared naturally-produced speech and synthesized speech with respect to type of the speaker (male, female) and finally have observed gender stereotyping effects.

The investigation of behavior of objective models can be found in [5-9]. In [9], model PESQ was applied to assessment of quality of synthesized speech. Authors concluded that PESQ model can be used for evaluation of synthetized speech, without subjective tests. On the other hand, PESQ can not be used for small size of diphone samples. The behavior of model P.563 in case of assessment of synthesized speech is described in [5-8, 10]. Based on the results presented in [8], P.563 is better for predicting impact of transmission channel on quality of naturally-produced voice, but it has lower accuracy in prediction of the overall voice quality. Further, P.563 achieves low correlation with subjective quality ratings for synthesized speech (especially in case of female synthesized voices [5]). In [11], Sebastian Moeller focused on the following issue: whether the effects of overall amount of degradations caused by transmission channel is similar in case of synthesized speech in contrast to naturally-produced speech. The investigation was focused on e.g. noisy-type degradations, which affected the quality both synthesized and naturally-produced speech in the same amount; and on low bit rate codecs, which had a bit different impact on the quality. Additional results can be seen in [13]. In [10], the authors also compared the results from various auditory tests with the predictions provided by three single-ended models (P.563, Psytechnics, ANIQUE+) using naturally-produced and synthesized voices. The samples used in this study were transmitted through different telephone channels (same impairments as used in study published in [11]). They found out, that these models provide distinct correlation with results of auditory tests in case of particular experiments.

4. Conclusion

The paper gives a brief overview of assessment of quality of synthesized speech. Two basic principles of the quality assessment have been introduced namely subjective and objective approach. In addition, a overview of the current state-of-the-art of synthesized speech research has been described, summarizing the experimental studies investigating performance, accuracy and reliability of existing approaches and models (mainly designed for evaluating the quality of naturally-produced speech) to evaluate the quality of synthesized speech.

References

[1] SITYAEV, D., KNILL, K., BURROWS, T. *Comparison of the ITU-T P.85 Standard to Other Methods for the Evaluation of Text-to-Speech systems*. INTERSPEECH 2006-ICSLP, Pittsburgh, Pennsylvania, 17-21 September, 2006.

[2] VAZQUEZ ALVAREZ, Y., HUCKVALE, M. *The Reliability of the ITU-T P.85 Standard for the Evaluation of Text-to-Speech Systems.* In Proc. of ICSLP, 2002.

[3] MULLENNIX, J. W., STERN, S. E., WILSON, S. J., DYSON, C. *Social Perception of Male and Female Computer Synthesized Speech.* Computers in Human Behavior, Vol.19, p. 407-424, 2003.

[4] ITU-T Recommendation P.85 *A Method for Subjective Performance Assessment of the Quality of Speech Voice Output Devices*, International Telecommunications Union Publication, 1994.

[5] MOELLER, S., FALK, T., H. *Quality Prediction for Synthesized Speech: Comparison of Approaches.* NAG/DAGA, Rotterdam, pp 1168-1171, 2009.

[6] FALK, T. H., MOELLER, S. *Towards Signal-Based Instrumental Quality Diagnosis for Text-to-Speech systems*. IEEE Signal Processing Letters, Vol. 15, pp 781-784, 2008.

[7] ITU-T Contribution COM 12 – C 180 – E. *Single-Ended Quality Estimation of Synthesized Speech: Analysis of the Rec. P.563 Internal Signal Processing.* Federal Republic of Germany (Authors: S. Moeller, T.H. Falk), ITU-T SG12 Meeting, 22-29 May, Geneva, 2008.

[8] ITU-T Contribution COM 12 – D 174 – E. *Estimating the Quality of Transmitted Synthesized Speech with the Single-Ended Quality Prediction Model According to ITU-T Rec. P.563.* Federal Republic of Germany (Authors: S. Moeller), ITU-T SG12 Meeting, 5-13 June, Geneva, 2006.

[9] CERNAK, M., RUSKO, M. *An Evaluation of Synthesized Speech Using the PESQ Measure.* Proc. Forum Acusticum, Budapest, pp 2725-2728, 2005.

[10] MOELLER, S., KIM, D.-S., MALFAIT, L. *Estimating the Quality of Synthesized and Natural Speech Transmitted Through Telephone Networks Using Single-Ended Prediction Models.* Acta Acustica united with Acustica, Vol. 94, pp 21-31, 2008.

[11] MOELLER, S. *Telephone Transmission Impact on Synthesized Speech: Quality Assessment and Prediction.* Acta Acustica united with Acustica, Vol. 90, pp 121-136, 2004.

[12] PSUTKA, J., MUELLER, L., MATOUŠEK, J., RADOVÁ V. *Mluvíme s počítačem česky.* Academia, Praha, ISBN 80-200-1309-1, 752 p, 2006.

[13] MOELLER, S. *Quality of telephone-based spoken dialogue systems*, Springer, New York (USA), Chapter 5, ISBN 0-387-23190-0, pp. 201-236, 2005.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Time-space Scheduling Problems in Grids

*Michal Murín
*University of Žilina, Faculty of Management and Informatics, Department of Mathematical Methods, Univerzitná 1, 010 26 Žilina, Slovakia, {michal.murin}@fri.uniza.sk

Abstract. Scheduling problems play important role in all the fields of industry. In the computer science a lot of research nowadays is oriented into the field of Grid systems as a promising way of the utilization of the available computer resources. Part of this research is also oriented in the scheduling problems in Grid. In this article will be presented difference between classical and time-space scheduling problems. At the end of the article, Grid specific time-space scheduling problems will be introduced and considarations that should be thought of, when modeling such problems, will be presented.

Keywords: scheduling, time-space, Grid, Data Grid, scheduling classification, scheduling problems

1. Introduction

Scheduling problems arise from the fact that the time is in the present one of the most valuable resources. These problems and their solutions play an important role in supplying, pruduction, distribution, transport, information processing and communication.

By words ''scheduling problems'' scheduling theory means time scheduling problems in which only time matters. But a lot of scheduling problems in the present are actualy time-space scheduling problems even though this term is not used when referencing them. These problems are specific in the way that subjects of scheduling are distributed in some kind of network infrastructure and for scheduling to take effect they have to be transfered through this network. From this network arise additional constraints that has to be taken into account (network transfer time, network capacity constraints, etc.). Time-space scheduling problems can be found in the production (scheduling of assembly lines or centres with commodities moving between them), transport (train, bus, metro schedules), computer networks and systems (scheduling in distributed and Grid systems), etc. These problems are usualy very complicated and they have to be decomposed into smaller problems and solved independently. Problems become even more complicated when models with multiple criteria are introduced.

2. Scheduling problems

According to the Pinedo [7]: ''Scheduling is a decision-making process that is used on a regular basis in many manufacturing and services industries. It deals with the allocation of resources to tasks over given time periods and its goal is to optimize one or more objectives.''

In classical scheduling theory, scheduling problem is usually formulated as follows:

Let us have three sets: set $\mathcal{J} = \{J_1, J_2, \ldots, J_n\}$ of n jobs (set $\mathcal{T}$ of tasks), set $\mathcal{M} = \{M_1, M_2, \ldots, M_m\}$ of m machines (set $\mathcal{P}$ of processors) and set $\mathcal{R} = \{R_1, R_2, \ldots, R_s\}$ of s types of additional resources. Scheduling, generally speaking, means to assign machines from $\mathcal{P}$ and (possibly) resources from $\mathcal{R}$ to jobs from $\mathcal{J}$ in order to complete all jobs under the imposed constraints. There are two general constraints in classical scheduling theory. Each job is to be processed by at most one machine at a time (plus possibly specified amounts of additional resources) and each machine is capable of processing at most one job at a time. [4, 2].

Such an assignement is called a schedule S. A feasible schedule satisfies all the imposed constraints. There exists usually more than one feasible schedule for a scheduling problem. Quality of the schedule is measured by one or more objective (criteria) functions $f(S)$. Optimal schedule is the one with the best value of objective function. When there are multiple objectives such schedule is pareto optimal.

In the 1979 Graham et al. [4] intorduced classification of scheduling problems. Classification was extended to cover more classes of scheduling problems through the years and is used in the present. It contains three greek alphabets: $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$.

$\boldsymbol{\alpha}$ - defines characteristics of machines used. Main classes are:

ø - single processor
$\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R}$ - identical, uniform, unrelated parallel machines
$\boldsymbol{G}, \boldsymbol{J}, \boldsymbol{F}, \boldsymbol{O}$ - dedicated (specialized) machines (general shop, job shop, flow shop, open shop)

$\boldsymbol{\beta}$ - defines characteristics of jobs (preemption, resources, precedence, release date, processing time, deadline. sequence dependent setup times, batch processing ...)

$\boldsymbol{\gamma}$ - determines objective function used ($C_{max}, L_{max}, \sum C_i, \sum L_i, \sum w_i C_i, \sum w_i L_i$)

For more information regarding scheduling classification see [7, 2, 3, 4]

3. Time-space scheduling problems

There is no explicit definition of time-space scheduling problems in literature. In a general meaning, nearly all scheduling problems can be formulated as time-space scheduling problems. The reason is that jobs and machines are distributed in some area and jobs have to move between machines in this area to be processed by them (or vice versa).

In the narrower meaning, time-space scheduling problems are those, in which such movement creates another constraints like movement time that significantly influences objective function, transport capacity constraint etc. According to this meaning time-space scheduling problem can be formulated as a scheduling problem with jobs or machines moving on a network and additional constraints arising from the network limits.

Such a network may be a road, railroad or water network, but also computer network, phone network etc. The network has its constraints (throughput of the edges or nodes, orientation of the edges, etc.), that have to be satisfied. Transportation on the network produces an additional time constraint, that may vary according to the situation on the network. These additional constraints have to be considered otherwise unfeasible schedules couuld be produced by optimization process.

Example: (School and university schedules)

One of the classical scheduling problems is the problem of creating a schedule of school classes. Elementary and secondary schools usually arrange their lessons in one or two buildings that are near to each other. The time needed to move between the classes is usually less than 10 minutes. Therefore this time can be included into the schedule as a fixed 10 minutes block. In contrast, universities may own much more buildings that can be quite distant. Therefore the transfer time between the buildings has to be taken into consideration when creating a schedule. We can model this problem by a complete graph with nodes representing buildings and weight of edges representing number of unit time blocks needed for transportation between buildings.

Time-space scheduling problems occur in transportation (vehicle routing problems, railroad timetables, regional bus timetables, ...) [6, 9], in production (production with transportation, production centres) [8, 5], in computer systems (routing data streams, distributed and grid systems with data transfers) [12, 11, 10].

The Graham classification of the scheduling problems introduced in section 2. is not appropriate for the classification of the time-space scheduling problems. Knust[5] in his PhD thesis extended the classification so it can incorporate also scheduling problems with transportation, but this extension covers only part of time-space scheduling problems.

4. Time-space scheduling problems in Grid systems

The Grid can be viewed as a sytem that unites widely distributed computer resources connected by network infrastructure (LAN, WAN, Internet) so the users of the system will see it as seamless, integrated computational and collaborative environment. The users interact with the Grid resource broker to solve problems, which in turn performs resource discovery, scheduling, and the processing of application jobs on the distributed Grid resources[1].

Time-space scheduling problems occur in the data Grid. The Data Grid provides the service of processing large datasets. Because of the large data beeing transfered, the transfer time and the bandwith constraints of the network need to be considered when scheduling in the Data Grid. Data Grids are commonly used by communities of researchers in domains such as high-energy physics, astronomy and biology.

The Data Grid can be viewed as a set of datasources (data nodes) that provide datasets required by the jobs submited to the Grid, a set of computional nodes that will be executing submited jobs and a computer network that interconnects data nodes and computional nodes. One node can act as a data and computional node at a same time. A set of jobs has been submited to the Grid in a specified time interval and has to be scheduled in the Grid. Each job requires a subset of the datasets to be transfered to the computional node at which it will be executed.

The Grid scheduler has to assign each job to a computional node that will execute it. For every dataset that the job requires, the scheduler has to determine the data node that will be used to transfer the dataset to the computional node. If a job consists of a set of operations. Every operation of the job has to be scheduled in the manner mentioned. Also precedence requirements of the operations have to be met. The scheduler has to create the best assignemnt according to the selected criteria.

When modeling the scheduler of a Data Grid system certain considerations that affect the complexity of the model have to be taken into account:

– Data transfers will highly influence the overall schedule therefore the data transfer times have to be incorporated into the model.
– Do all the required datasets have to be available at the computional node before the execution of the job begins or can some datasets be transfered as a streams during the execution?
– Will a data node be considered occupied during the tranfer of a dataset to a computional node or can more datasets be transfered simultaneously from the data node?
– Will also decrease in transmision rate from the data node be modeled, while there will be multiple transmissions from the data node?
– Will a computional node be considered occupied during the dataset transfer to the node or will simultaneous execution of scheduled job and transfer of datasets for the next job be modeled? If the second case, will the decrease in computional speed of computional node be modeled?
– Will the decrease of the network throughput be considered when transfering multiple datasets at the same time through the network?
– Will jobs be scheduled at their arrival to the Grid system (online scheduling) or will all jobs that will arrive during a spacified time interval be scheduled at the end of the interval (batch scheduling)?

5. Conclusion

By words ‘‘scheduling problems’’ scheduling theory means time scheduling problems in which only time matters. But a lot of scheduling problems in the present are actualy time-space scheduling problems even though this term is not used when referencing them. This paper formulates the term time-space scheduling. It demonstrates the differences between the classical and time-space scheduling problems. At the end of article time-space scheduling problems that occur in Grid system are presented and the problems that have to be considered when modeling such a problems are formulated.

Acknowledgement

The research of the author is supported by Slovak Scientific Grant Agency under grant VEGA 1/0374/11

References

[1] BAKER, M. - BUYYA, R. - LAFORENZA, D.: Grids and grid technologies for wide-area distributed computing. In: *Journal of Advanced Computational Intelligence and Intelligent Informatics of Software-Practice & Experience*, vol. 32, no. 15, 2002, p. 1437 - 1466.

[2] BLAZEWICZ, J. - ECKER, K. H. - PESCH, E. et al.: *Handbook on Scheduling: From Theory to Applications*. Springer, NY, USA, 2007, ISBN 978-3-540-28046-0.

[3] BRUCKER, P.: *Scheduling Algorithms*. Springer, NY, USA, 2007, ISBN 978-3-540-69515-8.

[4] GRAHAM, R. - LAWLER, E. - LENSTRA, J. et al.: Optimization and approximation in deterministic sequencing and scheduling: A survey. In: *Annals of Discrete Mathematics*, vol. 5, 1979, p. 287 - 326.

[5] KNUST, S.: *Shop-Scheduling Problems with Transportation*. Ph.D. thesis, Universität Osnabrück, 1999.

[6] PEŠKO, T.: *Optimalizácia NP-ťažkých dopravných rozvrhov*. Habilitation thesis, Žilinská Univerzita, Žilina, 2002.

[7] PINEDO, M.: *Scheduling: Theory, Algorithms and Systems*. Springer, 3rd ed., 2008, ISBN 978-0-387-78934-7.

[8] SAUER, J. - APPELRATH, H. J.: Integrating transportation in a multi-site scheduling environment. In: *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, 2000.

[9] TORMOS, P. - LOVA, A. - BARBER, F. et al.: *Metaheuristics for Scheduling In Industrial and Manufacturing Applications*, chap. A Genetic Algorithm for Railway Scheduling Problems. Computational Intelligence, Springer, 2008, p. 255 - 276.

[10] VENUGOPAL, S. - BUYYA, R.: An SCP-based heuristic approach for scheduling distributed data-intensive applications on global grids. In: *Journal of Parallel and Distributed Computing*, vol. 68, no. 4, 2008, p. 471 - 487, ISSN 0743-7315, doi:10.1016/j.jpdc.2007.07.004.
URL `http://www.sciencedirect.com/science/article/B6WKJ-4P7FSCK-1/2/21afaf9fe3b6bd451174b5a53ae6622b`

[11] WU, Q. - GAO, J. - ZHU, M. et al.: Self-adaptive configuration of visualization pipeline over wide-area networks. In: *IEEE Trans. Comput.*, vol. 57, no. 1, 2008, p. 55 - 68, ISSN 0018-9340, doi:http://dx.doi.org/10.1109/TC.2007.70777.

[12] YIN, Y. - WU, W.: A real-time measurement algorithm for available bandwidth. In: *International Journal of Communications, Network and Systems Sciences (IJCNS)*, vol. 2, no. 8, 2009, p. 746 - 753.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Inginiering Approach to Modelling of Qos Mechanism - Weighted Round Robin Markov's Chains

*Dušan Nemček
*University of Žilina, Faculty of Management Science and Informatics, Department of InfoComm Networks, Univerzitná 8215/1, 01026 Žilina, Slovakia, dusan.nemcek@fri.uniza.sk

Abstract. The paper deals with the creation of an analytical model for QoS mechanism – Weighted Round Robin in a system with two infinite queues. Into the particular queues arrive independent streams of packets. The described model is designed to calculate an estimate of a packet delay (a response time) which is one of the key QoS parameters. It is shown here, how the MM1∞ queues can be used for this purpose.

Keywords: Weighted Round Robin, WRR, Quality of Service, QoS, Markov's chains, analytical model, service intensity

1. Weighted Round Robin

Weighted round robin (*WRR*) is the foundation for a class of queue scheduling disciplines that are designed to address the limitations of the *Fair queuing* (*FQ*) and *Priority queuing (PQ)* models. WRR addresses the limitations of the *FQ* model by supporting flows with significantly different bandwidth requirements. With *WRR* queuing, each queue can be assigned a different percentage of the output port's bandwidth. *WRR* addresses the limitations of the strict *PQ* model by ensuring that lower-priority queues are not denied access to buffer space and output port bandwidth. With *WRR* queuing, at least one packet is removed from each queue during each service round.

In *WRR* queuing, packets are first classified into various service classes (for example, real-time, interactive, and file transfer) and then assigned to a queue that is specifically dedicated to that service class. Each of the queues is serviced in a round-robin order. Similar to strict *PQ* and *FQ*, empty queues are skipped. The packet is taken from serviced queue into transmitter and then it is sent bit by bit through output port. A rate of sent packets from transmitter is given by bandwidth of output port. The priorities called "weights" are assigned particular queues. The weight determines a number of bytes which can be transmitted from queue during the service round. If a cumulate sum of transmitted packets exceeds the weight, any other packets are not taken from queue within the service round. Therefore assigned weights determine distribution of total bandwidth among particular services classes.

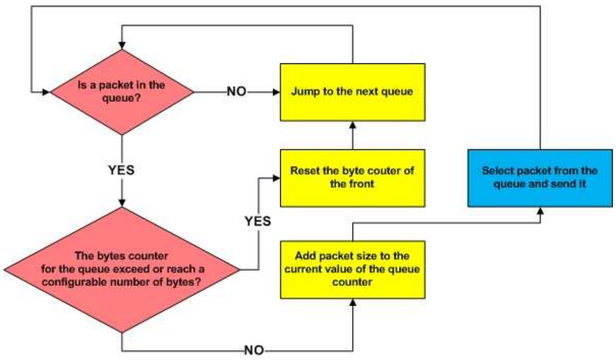


Fig. 1. Flow graph of WRR which send more packet from queue during its service

2. Analytical model of WRR

Mechanism *WRR* can be modelled using different ways, similarly as other *QoS* mechanisms. This section show, how Markov's chains can be used for this purpose. For the sake of simplicity, model with infinite queues is assumed here. The goal is to find an estimate of the mean response time for packets of particular queues. It is important performance measure. The methodology of calculation is based on obtaining average service intensities for packet of particular queues. The final formula is not quite exact, but it gets sufficient result relatively.

2.1. Basic terms

Necessary conditions for using *Markov's chains* are *Poisson* distribution of packet arrivals and *exponential* distribution for service time. Thus if *A(x)* denotes the number of packets that arrive during any time interval into the *k*-th queues of length *x*, the Poisson distribution of packet arrivals is given by

$$Pr[\,A(x) = n\,] = \frac{(\lambda_k x)^n}{n!} e^{-\lambda_k x}\,. \tag{1}$$

Parameter λ_k is intensity of packet arrivals into the *k*-th queue and is measured in packet per second. Compared to original, packets are not transmitted bit by bit but as a whole. Time of packet in transmitter (interval between selection packets to transmitter and his transmission on output port) is given by packet size and it is called *service time*. It has the exponential distribution for Markov's system Thus service time for packets of *k*-th queue I_k with mean $1/\mu_k$ is given by

$$Pr[\,I_k \le x\,] = 1 - e^{-\mu_k x}\,. \tag{2}$$

where μ_k is *service intensity* for packets of *k*-th queue. The mean *service time* for packet of *k*-th queue is an inverse of service intensity and here is denoted as $\tau_{k.}$

$$\tau_k = 1/\mu_k. \tag{3}$$

Waiting time W_k is the interval from the arrival time of packet for *k*-th queue to the time when its transmission is started. In other words, W_k is the total time of packet in the queue. *Response time* T_k is defined as the time interval from the arrival time of an arbitrary packet for *k*-th queue to the time, when the packet leaves the system after the transition is finished. The response time consists of the waiting time and the service time.

Packet taking probability p_k determines the probability of taking packet from *k*-th front immediately after completion of a packet service. It determines the allocation of different amounts of bandwidth to different service class. E.g., if we have configured 3, 2, 1 packets respectively for service of particular classes in one service round, packet taking probabilities would be $p_1=1/2$, $p_2 = 1/3$, $p_3 = 1/6$. We have $p_2 = (1 - p_1)$ in considered two-front system.

2.2. Transition graph

A state of considered system is given by three indexes *(k,i,j)*, where $k = \{1,2\}$, $i=\{1..L_1\}$ and $j=\{1..L_2\}$. *k* determines queue, that packet is actually transmitted, *i* determines number of packets of first queue and *j* the number of packet of second queue that are actually in the system. A value of state is determined by probability that system is in this state. In further text, only the steady-state probabilities are considered. If system can pass from one to another state, a transition exists between these states. Value of transition is determined by the mean intensity of arrival ($\lambda_{k.}$) or departure (μ_k). All possible states of system and transition between them can be represented by transition graph

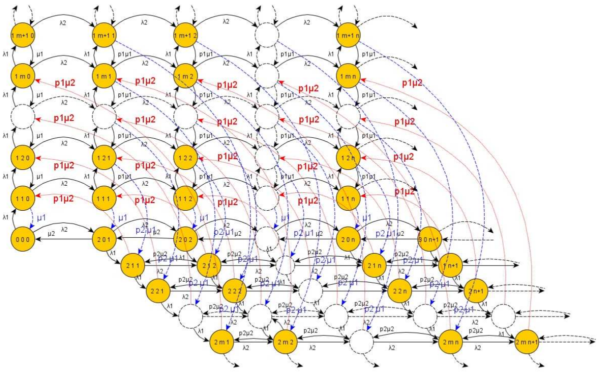


Fig. 2. The transition graph of WRR with two queues

2.3. Solution

To find the probability generating function for states in Fig.2 applying method in [4], it is not possible. Any types of cuts don't lead to solutions. Therefore it is necessary to simplify the model. The simplify lies in the fact that we don't distinguish type of transmitted packet. Hence is needed to define a mean *service intensity of all packets* in model μ. It can be shown that for μ applies the following:

$$\mu = \frac{\mu_1 \mu_2}{p_2 \mu_1 + p_1 \mu_2} \tag{4}$$

The neglecting of recognition of a transmitted packet results in a simplification of transition graph. Let

$$\mu_i^C = p_i \mu \tag{5}$$

is *service intensity* for packet of *k*-th queue, if all fronts are nonempty. Since empty queues in *WRR* are not serviced and they are skipped, μ_k is *service intensity* for packet of *k*-th queue, if only *k*-th queue is nonempty. The graph in Fig.3 corresponds to these facts.

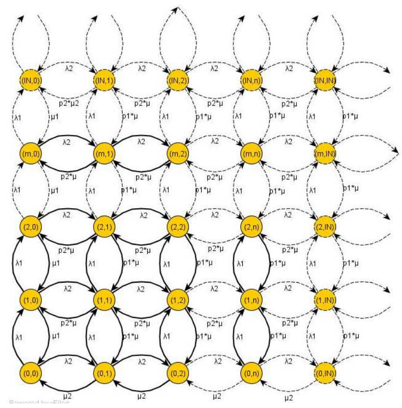


Fig. 3. The simplified transition graph of WRR with two queues

However such simplification does not lead to the solution either. Thus it will be necessary to simplify the model more. The simplification will be based on a division transition graph in Fig.3 into two transition graphs which correspond to two *M/M/1∞* queues. Further efforts lead to finding appropriate service intensities for these queues. Based on the statements above, it can write for them:

$$\mu_1^A = (1 - \pi_{20})\mu_1^C + \mu_1 \pi_{20} \tag{6}$$

$$\mu_2^A = (1 - \pi_{10})\mu_2^C + \mu_2 \pi_{10} \tag{7}$$

where π_{10} and π_{20} are probabilities, that first and second queue are empty, respectively.

The probability of empty system $MM1\infty$ is given by:

$$\pi_0 = (1-\rho) = 1-\lambda/\mu \qquad (8)$$

Therefore we get for μ_1^A and μ_2^A:

$$\mu_1^A = \frac{\mu_1(\lambda_2 p_2\mu_1^2 + \lambda_1 p_2\mu_1\mu_2 + \lambda_1 p_1\mu_2^2 - p_2\mu_1^2\mu_2 - p_1\mu_1\mu_2^2)}{\mu_2(\lambda_1-\mu_1)(p_2\mu_1 + p_1\mu_2)} \qquad (9)$$

$$\mu_2^A = \frac{\mu_2(\lambda_2 p_2\mu_1^2 + \lambda_2 p_1\mu_1\mu_2 + \lambda_1 p_1\mu_2^2 - p_2\mu_1^2\mu_2 - p_1\mu_1\mu_2^2)}{\mu_1(\lambda_2-\mu_2)(p_2\mu_1 + p_1\mu_2)} \qquad (10)$$

At obtained intensities it can be looked as at the service intensities in $MM1\infty$ queues.
Response time (*T*) for system MM1∞ is given by:

$$T = \frac{1}{\mu-\lambda} \qquad (11)$$

After substituting $\mu_1^A(\mu_2^A)$ into μ and $\lambda_1(\lambda_2)$ into λ, we obtain the estimate of *response time* for particular queues finally.

$$T_1 = \frac{\mu_2(\lambda_1-\mu_1)(p_2\mu_1 + p_1\mu_2)}{\mu_1(\lambda_2 p_2\mu_1^2 + \lambda_1 p_2\mu_1\mu_2 + \lambda_1 p_1\mu_2^2 - p_2\mu_1^2\mu_2 - p_1\mu_1\mu_2^2) - \lambda_1\mu_2(\lambda_1-\mu_1)(p_2\mu_1 + p_1\mu_2)} \qquad (12)$$

$$T_2 = \frac{\mu_1(\lambda_2-\mu_2)(p_2\mu_1 + p_1\mu_2)}{\mu_2(\lambda_2 p_2\mu_1^2 + \lambda_2 p_1\mu_1\mu_2 + \lambda_1 p_1\mu_2^2 - p_2\mu_1^2\mu_2 - p_1\mu_1\mu_2^2) - \lambda_2\mu_1(\lambda_2-\mu_2)(p_2\mu_1 + p_1\mu_2)} \qquad (13)$$

3. Conclusion

This paper show, how can be modeled the queuing mechanism WRR using $MM1\infty$ queues. The final result of the paper is formula for response time of packets of particular queues. For the sake of simplicity are assumed only two infinite queues here. The obtained result will be compared with result of simulation in a further research. Since this model accepts skipping of empty queues, it should get better results than model in [8].

References

[1] TAKAGI, H. *Analysis and application of Polling Model*: *Origins and Directions*. In: Proceeding Performance Evaluation. London: Springer-Verlag , ISBN: 3-540-67193-5, 2000.

[2] TAKAGI, H. *Queueing analysis - Volume1: Vacation and Priority System.* Amsterdam: North-Holland, ISBN: 0-444-88910-8, 1991

[3] TAKAGI, H. *Stochastic Analysis of Computer and Communication Systems*. Amsterdam: North-Holland, ISBN: 0444884793, 1990.

[4] NEMČEK, D. *Analytical model of QoS mechanism – Priority Queueing using Markov's chains.* In: Sborník příspěvku z medzinárodní vědecké konference MMK 2010. Hradec Králové: Magnanimitas, ISBN: 978-80-86703-41-1, 2010.

[5] CISCO Systems. *Implementing Cisco Quality of Service. Volume 1*. San Jose: CISCO Press, 2008.

[6] CISCO Systems. *Implementing Cisco Quality of Service. Volume 2*. San Jose: CISCO Press, 2008.

[7] SEMERIA, Ch. *Supporting differentiated service classes: Queue scheduling principles.* Sunnyvale: White Paper, 2002.

[8] NEMČEK, D. *Analytical model of QoS mechanism – Weighted Round Robin using Markov's chains.* In: Sborník příspěvku z medzinárodní vědecké konference QUAERE 2011. The Czech Republic: Magnanimitas, 2011.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Modern Information and Communication Technologies in Business

[*]Ľuboš Oláh, [*]Lukáš Šveda, [*]Veronika Fogašová
[*]University of Economics in Bratislava, Faculty of Commerce, Department of Marketing, Dolnozemská cesta 1, 85235 Bratislava, Slovakia, {olah.lubos, lukas.sveda, veronika.fogasova}@gmail.com

Abstract. This presented paper deals with trade issue, retail in particular. Retail is in a narrow interaction with the information and communication technologies and is interconnected through means of communication, and is more mobile than ever before. The aim is to identify and characterize the most important and relatively "modern" information and communication technologies for the consumer. These technologies fundamentally change not only the functioning conditions of the market environment, but also the consumer's behavior and his/her everyday life. Digital price tags, Digital Signage technology or usage of NFC payment technology are one of the new technologies.

Keywords: consumer, ICT, digital price tags, digital signage, contactless credit cards, mobile payments

Introduction

Not only marketing workers, managers of various brands, and manufacturers of various products, but also the entire political and economical life pay attention to the 21th century's consumer and his/her behavior.

From a macroeconomic view, the consumers' demand on goods and services is the driving force of the economy which accelerates investments and influences the entire market environment. Therefore, the effort of producers and traders should mainly focus on supporting customers' needs and purchasing goods and services, recognizing their preferences, behavior, and optimizing purchasing conditions for them.

The digitization process and development of new technologies, in particular the internet as the global phenomenon and interactive media, bring new options and opportunities in this direction changing the functioning conditions of the market environment in a great extent.

From individual activities of the domestic trade, retail represents an environment which is perceived by the customer in the most sensitive way. Regarding this issue's extent of implementation of innovative technologies in retail, this paper deals with selected forms of new technologies introduced in retail.

1. Retail Store as a Consumer's Point-of-Purchase

Retail functions as a mediator between goods and services, and end consumer. *Retail sells goods to end consumers and represents the last interconnection that ensures that the good leaves the circulation sphere and enters the end consumption sphere. The greater part of retail transactions is performed in retail stores which represent the basic business operation unit.*[1] A retail store is equally a consumer's point-of-purchase. From this point of view, the retail store can be evaluated as an attractive sphere of influence on the end consumer. It is a place where in-store communication and presentation of specific products or brands is performed as well it is a medium to gain and retain customers for retail chains. The concept of shopper marketing is built on the same grounds which presents the store as a communication medium with the customer.

[1] Own translation; see VIESTOVÁ, K. 2001 *Teória obchodu*. Bratislava, Sprint vfra, pp. 46-47. ISBN 80-88848-88-1.

The enlargement of market supply may be perceived as a significant global factor influencing the purchasing behavior of consumers – i.e. strong competition fight, market fragmentation, as well the financial crisis. The consumer is aware of his/her importance in this greatly competitive environment which leads to loyalty decrease and mainly rational and selective approach to choose a particular product or brand. Today's consumer is demanding and busy, therefore, s/he expects that the point of sale will make their decision to purchase a product easier for them. Simple availability of goods is the main consumers' requirement when purchasing. The current purchasing behavior of the consumers is characterized by the following variables[2]:

- 68% of the total number of purchases is unplanned;
- 70% of the decisions on choosing a specific brand are made right at the point of sale;
- Only 5% of the customers are loyal to a respective brand.

Various innovative technologies are being gradually implemented in order to improve the quality of purchasing and promote retail sale. These technologies can change the existing functioning, organization, and perception of purchasing in a revolutionary way. Thus we may talk about modern information and communication technologies which *provide efficient data and information transfer, process, and storage in connection with automatic identification*[3].

Regarding the extent of this issue, the following parts are devoted to selected forms of modern innovative technologies in retail.

2. Introduction of Selected Information and Communication Technologies in Retail

The current consumer is confronted with numerous technologies in retail. Some of them, such as automatic identification systems through contactless scanning of goods (RFID) or self-service cash registers, are employed in practice. However, others are just being introduced – digital price tags, Digital Signage technology, or NFC payment technologies.

2.1. Digital price tags as a medium in retail

The digital price tag, or electronic shelf tag, is another innovation besides the self-service cash registers in the field of retail sale. The digital price tag itself reflects a natural development, which means that paper is gradually replaced with digital media in almost every field.

The consumers can find all the information on this digital tag as they are used to on the paper tags. That means that the deposit amount on bottles, for example, will not be omitted. It can be assumed that gradually the prices of individual products will be distinguished in colors, e.g. discounts of particular products will be red.

Straightaway there are some advantages of these digital price tags:

- Automatic system changes of prices of most of the products in the store,
- Considerable increase in work efficiency of correctness checking and price updating,
- Option to introduce discounts at the end of the day by pressing one key button,
- Display of stock condition of individual products, daily/weekly sale, facing.

However, the consumer might come across some disadvantages as well:

- Risk of using simple display which can significantly decrease the legibility of information,
- Significant initial investment into terminal, system of aerials, transmitters, and last but not least electronic tags.

---

[2] DELOITTE. 2007 *Shopper Marketing Study: Capturing a Shopper´s Mind, Heart and Wallet*. Deloitte Development LLC.

[3] SIXTA, J., MAČÁT, V. 2010 *Logistika - teorie a praxe.* Brno, Computer Press, 261 pp. ISBN 80-251-0573-3

2.2. Digital Signage technology in retail

The Digital Signage system presents active and dynamic POP (Point of Purchase) and POI (Point of Information) solutions not only for retails. Information or advertisements are in electronic form – e.g. mainly electronic poster, store sale guide, or flat screen projection.

The goal of this information and communication technology is to deliver targeted messages to specific locations at specific times. It enables to react immediately on the actual clients' needs and synchronize direct sale with the actual running campaigns.

A content management server together with a central message and administration of the entire digital signage network are the key functions of the digital signage. The played content is defined in playlists. These playlists are created in a general way for the entire network, its individual parts, or media players for particular individual digital signage (e.g. There are situations where there are many media players at one place and, of course, many displays. In such cases it is necessary to structurally modify the content and correctly harmonize the point of sale with the digital signage content).

The asset of this technology consists in:

- Increasing the value and efficiency of communication with the consumers,
- Increasing the marketability of the presented products by tens of percents,
- Creating new sources of revenues (to be used also for advertisements),
- Allowing implementing "tactical marketing".

2.3. Contactless credit cards

The PayPass technology is becoming one of the latest most important trends in the field of payment operations in retail. This new technology enables to pay for the goods and services comfortably and fast by a simple tap of payment card on a point-of-sale terminal reader. Although this technology appeared in the world in 2006, it was introduced in Slovakia only two years later and it has been more intensively promoted on the Slovak market on the turn of 2010 and 2011thanks to an extensive advertising campaign by the Slovak bank sphere.

Many banks such as OTP, VÚB, and UniCredit Bank have come together with the Slovak capital to introduce such contactless cards and launched the Bratislava city card using the contactless PayPass technology. However, only the residents with a permanent residence in Bratislava can receive this card. They can use this card for contactless payment for goods and services, withdrawal from ATMs, public transport, or tickets to cultural or sport events.

MasterCard and Visa are the main supporters of this technology. Their main task is to increase the interest and use by retailers through terminal installation. Currently approximately 500 stores use the contactless payment system in Slovakia.

Advantages of launching for the consumers:

- PIN number is not needed as well there is no need to sign the bill,
- The payment is executed within 5 seconds,
- Data are protected by a chip technology with the highest level of cryptography.

Advantages of launching for the retailers:

- Speedup of customers' ordinary shopping,
- Higher safety and lower costs on cash handling,
- A competitive asset tool for the retailer,
- A store's image building tool,
- Dealing with a larger amount of customers at the same time,
- Creation of a new customer profile – contactless card owners.

2.4. Mobile payments

Contactless mobile payments are based on the use of mobile/ cell phones to pay for goods and services. It is an automated system of micropayments based on SMS. The whole transaction will be included in the invoice or will be withdrawn from the credit. After the SMS is sent with a key word generated by the system, an SMS confirmation on the settled payment for goods or services will be delivered in a few seconds.

A monthly limit is set for € 150, including VAT, per each customer, i.e. one telephone number. Micropayments for goods and services can be charged within predefined price levels from € 0.15 up to € 20.00, including VAT. This service is available for customers of all mobile operators in Slovakia.

Conclusion

Information and communication technologies present an inseparable part of today's world and their continuous development fundamentally changes their form. Their development in retail as a mediator of goods and services to the end consumers brings new opportunities to all participating market entities. From retailers' point of view, these new technologies are considered as a driving force of the competitiveness thanks to process dynamism, greater cost efficiency and transparency of control processes, speed of operative reaction period to market demands, image building in a relation to the consumer.

A complete replacement of in-store sale with online sale is a highly discussed topic and presents potential risk. Against all advantages of online sale, a standard department store will not disappear since nowadays stores represent not only the fulfillment of basic needs of purchasing goods and services, but also certain relax.

The implementation of the abovementioned innovative technologies is influenced by the economic crisis and initial investment demands. Also the extent of their acceptance from the consumer's side is questionable. However, their development and functioning for retail is indisputable.

References

[1] DELOITTE. 2007 *Shopper Marketing Study: Capturing a Shopper´s Mind, Heart and Wallet*. Deloitte

[2] PALFIOVÁ, A. 2011 *Počet akceptačných miest stále rastie.* In: Obchod, vol. 16, No. 1. 2011. 17 pp. ISSN 1335-2008

[3] PALFIOVÁ, A. 2010 *Zaostrite na spôsoby prezentácie tovarov.* In: Obchod, vol. 15, No. 11. 2010. 12 pp. ISSN 1335-2008

[4] SIXTA, J., MAČÁT, V. 2010 *Logistika - teorie a praxe.* Brno, Computer Press, 261 pp. ISBN 80-251-0573-3

[5] VIESTOVÁ, K. 2001 *Teória obchodu*. Bratislava, Sprint vfra, pp. 46-47. ISBN 80-88848-88-1.

[6] DMARKETING. (n.d.) *Tesco má digitální cenovky.* Retrieved March 7, 2011, from http://www.dmarketing.cz/2010/08/tesco-ma-digitalni-cenovky/

[7] ETREND. May 14th, 2006. *Technológie automatizujú maloobchod.* Retrieved March 3, 2011, from http://podnikanie.etrend.sk/podnikanie-riadenie/technologie-automatizuju-maloobchod.html

[8] PLATBAMOBILOM. (n.d.) *O službe.* Retrived March 3, 2011, from http://sms.platbamobilom.sk/?page=o-sluzbe

[9] SZINTEZIS PLUS. (n.d.) *Elektornická cenovka.* Retrieved March 7, 2011, from http://www.szintezis.hu/?p=elektronicka.cenovka

[10] TATRA BANKA. (n.d.) *Bezkontaktné platby: Benefity pre obchodníkov.* Retrieved March 4, 2011, from http://www.bezkontaktnaplatba.sk/

This paper is an output of the VEGA project No. 1/0333/11 *Modern approaches and methods in marketing management* as well of the VEGA project No. 1/0652/10 *Dynamic aspects of the Slovak population's purchasing behavior in unsteady economic conditions.*

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Security Protocols in ZigBee Technology

*Tomáš Ondrašina
*University of Zilina, Faculty of Electrical Engineering, Department of Control and Information Systems, Univerzitna 1, 01026 Žilina, Slovakia, {Tomas.Ondrasina}@fel.uniza.sk

Abstract. There are many applications of wireless sensor networks that collect and disseminate sensitive and important information. In order for many implementations of these applications to operate successfully, it is necessary to maintain the privacy and security of the transmitted data. This paper considers security protocols used in ZigBee Networks. It describes progressive security architectures and authentication mechanism.

Keywords: ZigBee, security, wireless, advanced encryption standard, authentication.

1. Introduction

ZigBee is an industrial consortium, which was designed to build a standard data link communication layer for use in ultra low power wireless communications [1]. The members of this organization came together because they felt that “existing standard technologies were not applicable to ultra-low power application scenarios” [2]. The ZigBee network layer (NWK) is designed to operate just above the PHY and MAC layers, specified in the IEEE 802.15.4 standard.

The main responsibilities of the ZigBee NWK layer include the mechanisms used to join and leave a network, apply security to frames and to route frames to their intended destinations. The ZigBee specification also details extra security services, including the processes of key exchange and authentication, in addition to those provided under the IEEE 802.15.4, upon which it is built. These will also be examined further in the next section.

The 802.15.4 standard is the IEEE specification for low-rate wireless personal area networks (LR-WPANs). Unlike wireless local area networks (WLANs), connections effected via WPANs involve little or no infrastructure. This is set to become the standard communications protocol for use in wireless sensor networking. Features allow small, power efficient, inexpensive solutions to be implemented for a wide range of devices. The main objectives of an LR-WPAN are ease of installation, reliable data transfer, short-range operation, and extremely low cost and reasonable battery life, whilst maintaining a simple and flexible protocol [3].

2. IEEE 802.15.4 Security

Under this standard, a link layer security protocol provides four basic security services. These include access control, message integrity, message confidentiality, and replay protection. An application sets its security requirements by setting the appropriate parameters into the radio stack. If the application does not set any parameters then, by default, there is no security enabled. Access control and message integrity imply that this protocol should prevent unauthorized parties from participating in the network. Legitimate nodes should be able to detect messages from unauthorized nodes and reject them. Message integrity protection implies that if an adversary modifies a message from an authorized sender while the message was in transit, the receiver should be able to detect the tampering. To ensure message authentication and integrity, a message authentication code (MAC) is appended to each message sent. This MAC is viewed as a cryptographically secure checksum of the message [4].

Computing the MAC requires senders and receivers to share a secret cryptographic key, and this key is part of the input to the computation. The sender computes the MAC over the packet and includes it with the packet (using the secret key). A receiver sharing the same key re-computes the MAC and compares it with the MAC in the packet. If the two are the same then the receiver accepts the packet, or rejects it otherwise. Message authentication codes must be difficult to forge without a secret key and, resultantly, if an adversary to the network changes a valid message or introduces a phoney message, then it would be unable to compute the corresponding MAC, and authorised receivers will reject any of their attempts to damage the network [5].

The standard defines 8 different security suites. See Tab. 1 below. The security suites can be more broadly classified by their properties. The first of these is the Null suite and provides no security. The next is encryption only (AES-CTR), followed by authentication only (AES-CBC-MAC), and finally encryption and authentication (AES-CCM) [4].

Name	Description
Null	No Security
AES-CTR	Encryption only, CTR Mode
AES-CBC-MAC-128 AES-CBC-MAC-64 AES-CBC-MAC-32	128 bit MAC 64 bit MAC 32 bit MAC
AES-CCM-128 AES-CCM-64 AES-CCM-32	Encryption & 128 bit MAC Encryption & 64 bit MAC Encryption & 32 bit MAC

Tab. 1. Security suites defined by IEEE 802.15.4

3. ZigBee Security

ZigBee uses all of the basic security elements of the IEEE 802.15.4 standard. In addition, the ZigBee security specification employs a simpler and unified mode of operation of CCM, defines key types (Master, Link, Network) and describes key setup and maintenance (Commercial, Residential) [6].

Additionally, ZigBee provides freshness through the use of freshness checks. These checks prevent replay attacks, as ZigBee devices maintain incoming and outgoing freshness counters. Whenever a new key is created, the counters are reset. It is postulated that devices that communicate once per second will not overflow their freshness counters for 136 years [5]. Message integrity and encryption are also provided under the ZigBee security specification, the operations of which are documented in [1] and [5]. Under the ZigBee specification, authentication is defined to provide assurance about the originator of a message. This prevents an attacker from mimicking the operation of another device in any attempt to compromise the network.

Authentication is a mechanism whereby the identity of a node in a network can be identified as a valid member of the network and as such data authenticity can be achieved. This is where the data is appended with a message authentication code (MAC) and can only be viewed by valid nodes capable of decrypting the MAC. Authentication is possible at both the network level and the device level. At the network level, authentication is achieved using a common network key, thus preventing outside attacks whilst adding very little in memory cost. Device level authentication is achieved by using unique link keys between pairs of devices [7].

Privacy of data can be vital to the success of many of the applications that such networks are currently being used for. Data encryption and node authentication are the main defences against attack.

3.1. ZigBee Security Architecture

The concept of a “Trust Center” is introduced in the specification. Generally the ZigBee coordinator performs this duty. This device allows other devices to join the network and also distributes the keys. There are three roles played:

1. Trust Manager, whereby authentication of devices requesting to join the network is done,
2. Network Manager, maintaining and distributing network keys,
3. Configuration Manager, enabling end-to-end security between devices [8].

It operates in both Residential Mode and Commercial Mode. The Trust Center running Residential Mode is used for low security residential applications. Commercial Mode is designed for high-security commercial and safety–related applications.

In Residential Mode, the Trust Center will allow devices to join the network, but does not establish keys with the network devices. It therefore cannot periodically update keys and allows for the memory cost to be minimal, as it cannot scale with size of the network. In commercial mode, it establishes and maintains keys and freshness counters with every device in the network, allowing centralized control and update of keys. This results in a memory cost that could scale with the size of the network [5].

There are three types of keys employed, the Master Key, the Link Key and the Network Key. Master keys are installed first, either in the factory or out of band. They are sent from the Trust Center and are the basis for long-term security between two devices. The Link key is a basis of security between two devices and the Network keys are the basic of security across the entire network. Link and Network keys, which are either installed in the factory or out of band, employ symmetrical key-key exchange (SKKE) handshake between devices [9]. The key is transported from the Trust Center for both types of keys. This operation occurs in commercial mode, as residential mode does not allow for authentication.

3.2. ZigBee Security Mechanisms

The main elements of security mechanisms provided in ZigBee are summarized in Fig. 1 [10].

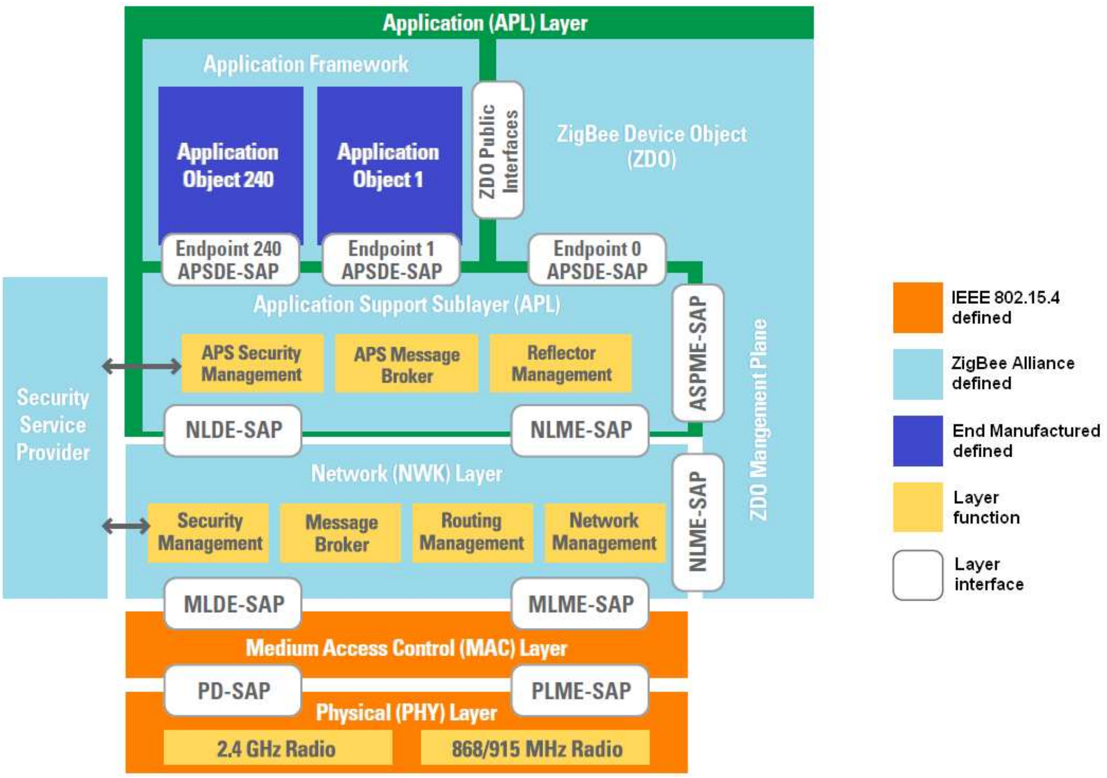


Fig. 1. ZigBee security architecture

Freshness. Freshness check prevents replay attacks (an attacker from replaying messages). ZigBee devices maintain incoming and outgoing freshness counters. Counter is reset when a new key is created. Devices that communicate once per second will not overflow their freshness counters for 136 years.

Message Integrity. Message integrity prevents an attacker from modifying the message in transit. We can use 0, 32, 64 or 128 bits for integrity check. Integrity options allow tradeoff between message protection and message overhead [11].

Authentication. Authentication provides assurance about the originator of the message. It prevents an attacker from modifying a hacked device to impersonate another device. Authentication is possible at network level or device level. Network-level authentication is achieved by using a common network key. This prevents outsider attacks while adding very little in memory cost. Device level authentication is achieved by using unique link keys between pairs of devices: This prevents insider and outsider attacks but has higher memory cost.

Encryption. Prevents an eavesdropper from listening to messages. ZigBee uses 128-bit AES encryption. Encryption protection is possible at network level or device level. Network-level encryption is achieved by using a common network key. This prevents outsider attacks while adding very little in memory cost. Device-level encryption is achieved by using unique link keys between pairs of devices. This prevents insider and outsider attacks but has higher memory cost. Encryption can be turned off without impacting freshness, integrity or authentication.

4. Conclusion

ZigBee is designed as a global hardware and software standard for wireless networking devices. Its main features are: highly reliable, low cost, low power, low data rates and highly secure. The main focus of this work was ZigBee Security, which is based on a 128-bit AES algorithm. ZigBee's security services include methods for key establishment and transport, device management, and frame protection.

References

[1] ZIGBEE ALLIANCE: *ZigBee Specification v1.0*, San Ramon, CA, USA, 2005.

[2] HILL, J. L. *System Architecture for Wireless Sensor Networks*. University of California, Berkeley, USA, 2003.

[3] IEEE Standard for Information Technology, Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Networks, Specific Requirements, *Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, 2007.

[4] SASTRY, N., WAGNER, D. *Security Considerations for IEEE 802.15.4 Networks.* Proceedings of the 2004 ACM Workshop on Wireless Security, Philadelphia, PA, USA, 2004 & New York, USA: ACM Press, 2004.

[5] ZIGBEE ALLIANCE: *ZigBee Security Specification Overview*, San Ramon, CA, USA, 2006.

[6] DARGIE, W., POELLABAUER, C. *Fundamentals of Wireless Sensor Networks*. Wiley Series on Wireless Communications and Mobile Computing, UK, 2010.

[7] LIU, D., NING, P., ZHU, S., JAJODIA, S. *Practical Broadcast Authentication in Sensor Networks.* The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'05), San Diego, California, USA: IEEE Computer Society Press, 2005.

[8] AKYLDIZ, I. F., VURAN, M. C. *Wireless Sensor Networks.* Georgia Institute of Technology & University of Nebraska–Lincoln, USA, 2010.

[9] BOYLE, D., NEWE, T. *Security Protocols for use with Wireless Sensor Networks.* The Third International Conference on Wireless and Mobile Communications (ICWMC'07), Guadeloupe, French Caribbean, 2007.

[10] LABIOD, H., AFIFI, H., DE SANTIS, C. *Wi-Fi, Bluetooth, ZigBee and WiMax*. France: Springer, 2007.

[11] DAINTREE NETWORKS: *Getting Started with ZigBee and IEEE 802.15.4*. Mountain View, CA, USA, 2008.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Internal Infrastructural Requirements for Fully Efficient and Automated Structured Electronic Invoicing in Small and Medium Size Companies

*Igor Pihir, *Neven Vrček
*University of Zagreb, Faculty of Organization and Informatics Pavlinska 2, 42000 Varaždin, Croatia,
igor.pihir@foi.hr, neven.vrcek@foi.hr

Abstract. Electronic invoicing (e-invoicing) between companies in Europe is in raise and broader development. Companies that implement this new way of interoperability, information service providers that offers new kind of services and countries that advance their legislation to support electronic interchange of structured electronic invoices, dream about automation and cost-efficiency of companies' processes and cross-border invoices. What infrastructure is needed in middle sized company to achieve full automation and rid of paper invoice for real and forever? Is this really possible and what is the purpose of invoice in enterprise practice? As a result of literature review and case studies of business processes in few small and medium size companies this article presents internal infrastructural requirements and obstacles in a way of completely replacing paper-based invoice with structured electronic invoice.

Keywords: electronic invoice, process automation, infrastructural requirements.

1. Introduction

Unlike paper-based invoices, e-invoices provide all data in digital format. Such e-invoicing offers substantial benefits over paper invoicing. It allows for shorter payment delays, fewer errors, and reduced printing and postage costs and, most importantly, fully integrated processing. One distinctive feature of the e-invoice is therefore its potential for automation, especially if the invoice is sent in a structured format: e-invoices can be generated and transferred automatically and directly from the issuer's or service provider's financial supply chain systems to those of the recipient. Most of the economic benefits therefore do not arise from savings in printing and postage costs but rather from the full process automation and integration from order to payment between trading parties. [1]

This nice lines from Brussels describes in general what is overall purpose of e-invoicing and its' final goal. Most companies in Europe today have the possibility to print invoice on the paper through their information system which can easily be improved by printing invoice to the file respectively to the not-structured PDF format which is in practice digital picture of invoice. Such digital invoices could be sent over to the customer in compliance to legislation and security requirements and companies could benefit from postage costs and printing. Legislation of this kind is not jet established in all EU countries and there are still some unsolved problems. Real benefits are expected in structured e-invoice which is also digital but readable through the information systems with known structure and required fields and syntax. In 2010 EU countries e-invoicing by unstructured e-invoice (receiving and sending) is in average at 8% to 41% [1] of penetration. Real structured e-invoicing is barely at 5% - 7% [2][3]. If we exclude some major problems for a moment such as: legal requirements for e-invoicing, electronic signatures and time stamps problems, problems of one unique accepted structure of the e-invoice (EDI, ebXML, structured PDF, …), information service providers legislation and purpose in e-invoicing for medium and small companies, return of investment and cost saving analysis of e-invoicing, digital storage for long term tax audit purposes, cross border invoice recognition problems. The one question would be still open, and that is, does the small and medium size company has adequate information infrastructure to support such cause and would that still make us paper free? What is missing for the company to do so?

2. Invoice application in business processes and infrastructural requirements

By the definition of invoice – invoice is an application for payment, issued by a taxable person or his agent, in respect of the taxable supply (including Zero Rate) of goods or services. The document evidences the amount of VAT that has been charged by the taxable person and is used to support the recipient's entitlement to VAT recovery, subject to rules on deductibility. [4]

What is obvious is that invoice is a document witch has more than one purpose. If we try to research its application we could ask our self in how many copies would the paper-based invoice be issued? In normal circumstances that could be from 2-6 copies. Let's see some scenarios for that:

a) Scenario one is simple business transaction for example at gas station where you could buy a gasoline for companies' car.

- One copy of original invoice from legal entity is for customer (his accounting office and tax deduction). You can take it immediately on paper. What is if you want that they send it to you as e-invoice in electronic form?
- Second copy is for issuer and his (accounting office and tax payment). E-invoice scenario?

In this simple scenario with payment at scene (cash or card) everything is simple and if customer receives e-invoices, Gas Company could send him invoice by electronic means. Does the customer needs some kind of proof of payment at scene? Do we still need paper proof of payment? This was a simple scenario but what is happening if we use e-invoice in more complex scenario? For example cross border shipment and invoicing.

b) Scenario two is more complex case of cross border shipment of produced goods to another country. Situation is more complex and invoice is issued in more copies, for instance in 6 copies.

- First two copies of original invoice are shipped together with shipment goods. Purpose of these copies is in transportation to prove the source of goods, who is selling what and to whom. Of course there can be other documents to support this purpose like dispatch order or delivery list. But invoice is also doing the thing. One copy is for the shipping company and one for customer.
- Third copy is needed for customs purposes and there evidence of trade and customs payments.
- Forth copy is attached to the working / production order to conclude the production process tracking or to be attached to the customer order and to conclude delivery tracking.
- Fifth copy is send trough post to the final customer who will process the invoice and finally pay the goods. Invoice is used for conclusion of his business deal with seller. There is also tax obligations and long term storage for tax audits.
- Sixth copy of original invoice stay in company and it is processed by accounting office. It is used as a legal evidence of business deal and its is stored and processed for evidence of obligation of customer to pay us, tax record and tax obligation and long term tax audits purposes.

For more detail consideration we could conclude that the first two invoice copies, in scenario two, maybe could be replaced by other documents such as dispatch order or delivery list. In that case we do not need e-invoice but of course we still have paper documents for transportation documentation. If this would be still e-invoice, transporters would need a digital form of document. It is know in large delivery companies to carry out palm computers or smart phone sized logistics computers with digital documents and touch screens to sign the delivery with no need of any paper-based document. But it's not cheap and the question is, does this kind of technology is common in all transport companies. Second copy will not be needed with the transported goods, because the e-invoice would be available in customer warehouse in time and almost instantly as it is issued electronically. If invoice is used for the goods delivery confirmation it is needed in warehouse storage so that goods can be checked in the process of admission to the warehouse.

Problems and needs for third copy of invoice in electronic form are in customs information system. Customs must use e-invoicing to accept our goods to customs and we also must send the e-invoice to them prior the transport arrived to customs check. All paperwork at customs office could benefit from structured electronic invoicing in same manner as business parties. This also requires that customs department, in the state and abroad, receives this kind of documents in electronic form.

Forth copy is not a problem of high magnitude because in highly integrated information systems, (integrated in every common process inside the company that can or should be supported by information systems, IT support from order to delivery) the conclusion of production order and matching it with invoice could be solved internally in information system and it is common in ERP systems that integrate production and accounting processes. ERP system stands for common word in use as integrated information system in general, and not only the original meaning as Enterprise Resource Planning System.

Fifth copy of original invoice issued as e-invoice is the copy that is commonly known as "e-invoice" and it would be issued to the customer that can and wants to receive the e-invoices. Benefits are great, if he also can process the electronic invoice automatically into his own ERP system and match the invoice data with his order and receipt note in his ERP system. Payment could be easier if his ERP system allows also automatically produced payment order which can be processed through internet banking system or similar solutions. Customer as a taxable person must also provide insight into its invoices in its long term storage of e-invoices in original electronic form for tax audits. In some EU member states storage obligatory period is 11 years. [4].

Sixth copy of original invoice, now in electronic form (processed in ERP system prior to issuing) is also stored in long term archive as in customers' case, and it has purpose for tax audits. As an issuer of e-invoice all data of such invoice would be stored in companies ERP system and they can be used for managing the billing processes and customer obligation fulfilment. For full automation of invoicing process automatic receiving and matching of payments notes is also great thing. For this purpose bank payment received notes for our company must also be structured and manageable through our ERP system. Its possible and common practice in good ERP solutions but not for cross border payments. As a conclusion companies need a very good ERP system to support new way of doing business and this is requirement, independent from requirements for e-invoicing software or e-invoice service providers' solutions.

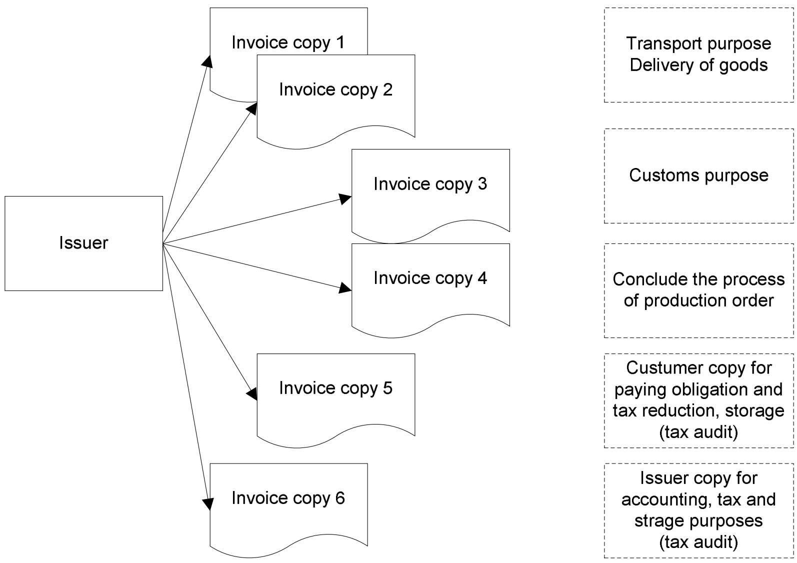


Fig. 1. Invoice purpose in complex business transactions

3. Conclusion

Business problem of crossing from full paper-based invoicing to full paper-less or e-invoicing is not ease and companies are struggling to cut cost and raised efficiency in any possible way and one of them is to "go live" with e-invoice. Although there are still many unsolved problems one of them is the fact that electronic interchange of invoices and its problems is one thing and the overall change of some companies business processes to support new way of doing business is something else. Through this simple scenarios described above there is lots of detail to care off in real business processes. Many research studies like [2] [5] are describing the e-invoice process at high level and they describe it as easy and simple. They mainly focus on transmission of invoice and not at business use of document called invoice and printed in many copies for different purposes.

As a conclusion we can say that there must be very serious information infrastructure implemented and used for fully automation of business processes by electronic means. Many small and medium size companies does not have this kind of solutions and their efficiency and cost cutting abilities in case of using e-invoicing are questionable and return of investment is also doubtful. As a result of this thinking we could conclude that enterprises, which would like to automate and improve their processes by use of structured e-invoicing, would need modern ERP system that supports internal processes electronically with paper-free document processing and that this system must be integrated with e-invoicing software in such a way that transmission of invoice data is fully automated and potential benefits are maximized. At the end of every invoice cycle ERP system would need also integration with internet banking software to fulfil payments and payment notes processing.

Further steps could be made in form of detailed process analysis and development of cost-benefit methodology for small and medium size companies to reach the reasonable level of automation with a real savings in case they will implement e-invoice.

References

[1] EUROPEAN COMMISION, *Communication from the Commission to the European parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Reaping the benefits of electronic invoicing for Europe*. Brussels, 2010.

[2] NIENHUIS, J.J. BRYANT C. *E-invoicing 2010: European market guide*. Euro Banking Association (EBA) and Innopay, 2010.

[3] MAI, H. MEYER T. *E-invoicing: Final step of an efficient invoicing process.* Deutsche Bank Research, Frankfurt, Germany. 2010.

[4] *CEN Workshop Agreement CWA 15580: Storage of Electronic Invoices.* European Committee for Standardization, Brussels 2006.

[5] *INTERIM CEN WORKSHOPAGREEMENT: CEN/Fiscalis e-Invoicing Good Practice Guidelines.* CEN/Fiscalis, Brussels, 2008.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

A Performance Study of the Multi-interface Wireless Mesh Networks

[*]Štefan Pollák, [*]Vladimír Wieser, [*]Tomáš Čakan
[*]University of Zilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitna 8215/1, 01026 Žilina, Slovakia, {pollak, wieser, cakan}@fel.uniza.sk

Abstract. Wireless mesh networks (WMNs) are emerging as one of promising technologies for broadband wireless networks. These networks offer many advantages in terms of scalability, connectivity and reliability. WMN provides multi-hop wireless connectivity between mesh nodes and also allows access to the internet via gateway nodes. Mesh nodes are typically equipped with the single radio interface. In this configuration exists several limitations, such as lower throughput. In this paper, a performance study of the multi-interface WMN is presented. The simulation model of static WMN was created in NS-2 simulator, and the performance of the multi-interface WMN based on chosen quality of service parameters (QoS) was studied.

Keywords: WMN, QoS, multi-interface, NS-2

1. Introduction

WMNs have emerged as a highly flexible, reliable and cost efficient solution for covering large areas wirelessly. These networks are characterized by dynamic self-organization, self-healing and self-configuration to enable quick deployment and high scalability. The WMN consists of mesh routers and mesh clients. The mesh routers are usually stationary and form multi-hop wireless backbone network (static WMN). Only few nodes in WMN backbone have direct access to wired network and serve as Internet gateways for the rest of the network. These nodes are called mesh portal points. Each mesh router operates not only as a host but also as a router. This function allows to forward packets to nodes that may not be within direct wireless transmission range of their destination. It also guarantees the existence of multiple paths, which increase the network reliability [1].

WMNs, based on standard IEEE 802.11, are typically configured to operate on a single channel using a single radio interface, which decreases the network capacity due to interference from adjacent nodes in the network [2]. One of the most promising approaches for improving capacity of WMN is the using of multiple interfaces. In our work we have created the simulation WMN model in NS-2 simulator environment in which the performance of multi-interface WMN is studied.

In [5] the study of multi-interface WMNs performance was described but only with one QoS metric – the throughput. Several QoS parameters we have included in our simulations to determine the performance of the multi-interface WMN for services required the real time transmission (e.g. video conference).

The rest of the paper is organized as follows. The section 2 describes the simulation model and in the following chapter the simulation results are presented. Finally, the section 4 concludes the paper.

2. Simulation Model

A simulation WMN model was developed in NS-2 network simulator with additional function to support multi-channel and multi-interface solution [3]. Each mesh node used 1 to 8 interfaces and the same number of channels. Our simulation model consisted of 25 static wireless mesh nodes placed in an area of 1000 x 1000 m (Fig. 1). We have used the WMN with 25 nodes, because of the typical number of mesh nodes in WMN between 25 to 30 [ ]. Transmission range for each node was set to 200 meters. For traffic generation, 5, 10, 15 and 20 CBR (Constant Bit Rate) flows and the packet size 512 bytes were used. Flow was created from chosen nodes and all traffic was routed to one mesh gateway. The radio default parameters in NS-2 [4] were used except that we set the channel data rate to 11 Mbit/s. Simulation parameters are summarized in Table 1.

Parameter	Value
Test area	1000 x 1000 m
MAC protocol	IEEE 802.11
Propagation model	Two ray ground
Routing protocol	AODV
Antenna type	Omni-directional
Traffic type	CBR
Packet size	512 bytes
Simulation time	100 seconds

Tab. 1. Simulation Parameters.

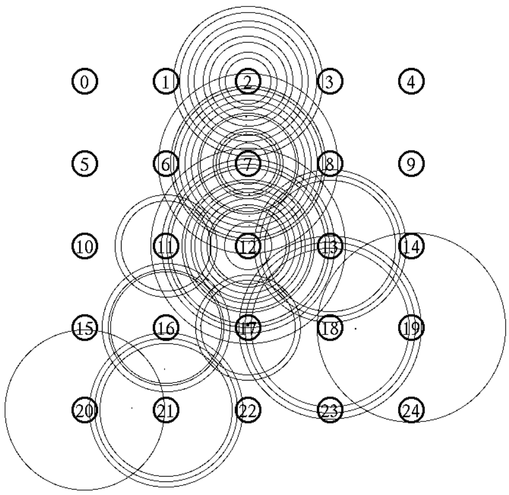


Fig.1. Simulation model of WMN created in NS-2 simulator.

3. Simulation Results

In this section, the results of our experiments are presented. The purpose of simulations was to determine the performance of the multi-interface WMN expressed by QoS parameters.

For simulation evaluations we have chosen following QoS parameters:

- *Average End-to-end Delay* – The average time taken for a packet to reach the destination. It includes all possible delays in the source node and in each intermediate host, caused by queuing at the interface queue, transmission at the MAC layer, routing discovery, etc. Only successfully delivered packets are counted.
- *Average Throughput* – The sum of data packets delivered to all nodes in the network in a given time unit (second).

Packet Loss – Occurs when one or more packets being transmitted across the network fail to arrive at the destination.

3.1. Average End-to-end delay

Figure 2 shows the average values of end-to-end delay for the different number of data flows. From results it is seen that the best performance achieved multi-interface WMN with five or six interfaces, when the number of flows changed.

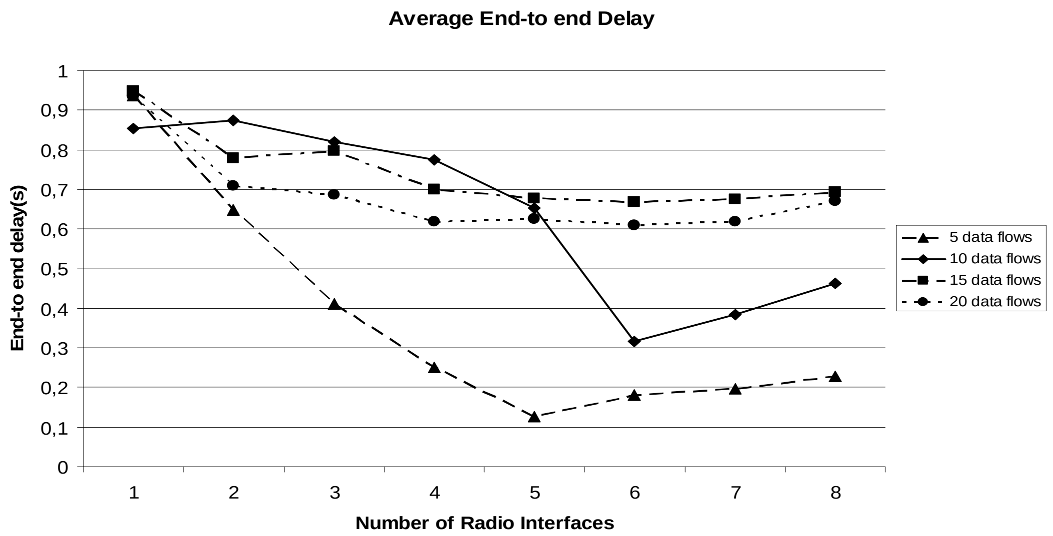


Fig.2. End-to-end delay for various data flows versus number of radio interfaces.

3.2. Average Throughput

Figure 3 shows the simulation results of average values of network throughput for the 5, 10, 15 and 20 data flows. From results it is obvious that the highest value of average throughput was reached in the multi-interface WMN with six radio interfaces. In the WMN with more than six interfaces the network performance is decreasing.

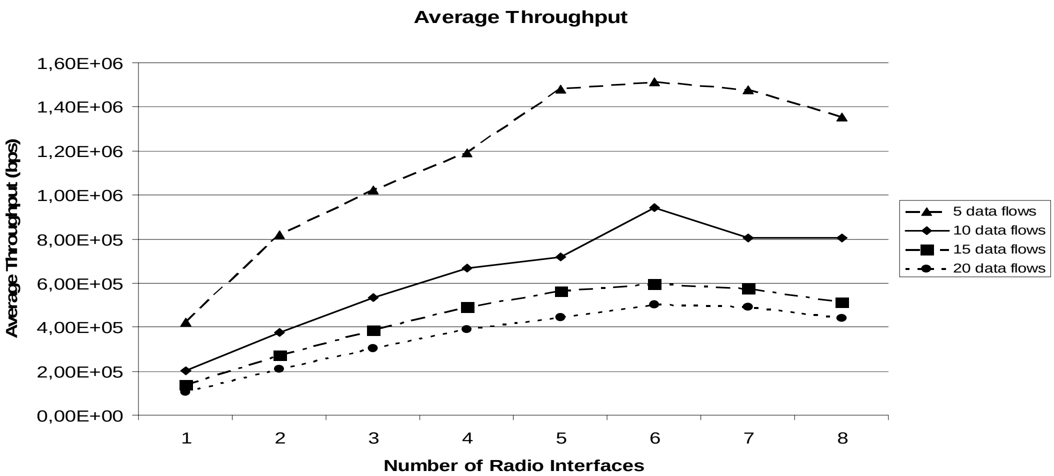


Fig.3. Average Throughput for various data flows versus number of radio interfaces.

3.3. Packet Loss

As we can see from Figure 4, the best value of packet loss was reached in multi-interface WMN with six radio interfaces. The highest value of packet loss was reached in WMN, where nodes have used for transmission one radio interface.

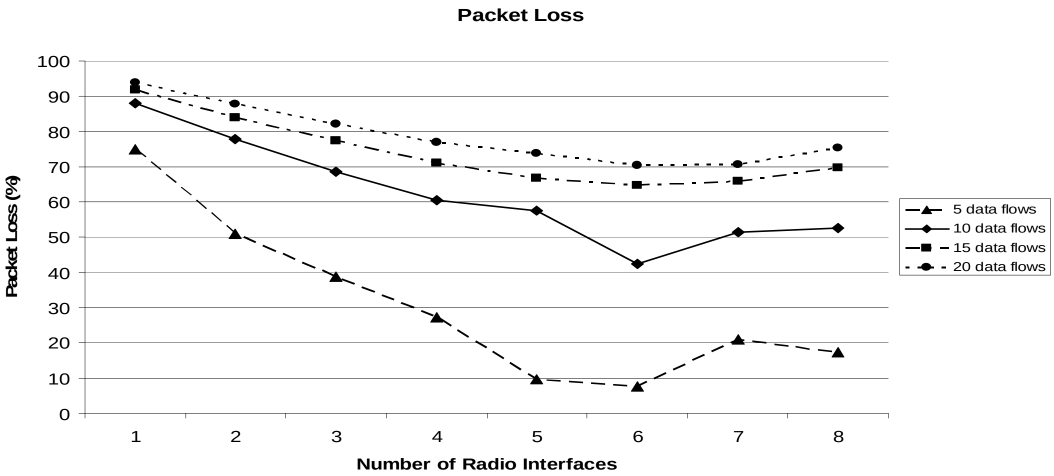


Fig.4. Packet Loss for various data flows versus number of radio interfaces.

4. Conclusion

In this paper a performance study of the multi-radio WMNs for various data flows was presented. The study was based on increasing number of radio interfaces (1 to 8) for all mesh nodes. The static WMN scenario was created in NS-2 simulator where the common channel assignment method was used. For simulation evaluations the average end-to-end delay, average throughput and packet loss metrics were chosen.

The results show that by increasing the number of interfaces it is possible to increase network performance, but increasing number of data flows mitigate performance of WMN. This is due to an increased number of transmissions between nodes, which caused increasing of interference in the network. Simulation results show that the best performance was achieved in multi-interface WMN with 5 or 6 radio interfaces. From simulation results we can see that the network performance depends on the number of radio interfaces and number of data flows.

Acknowledgement

This paper was supported by the Scientific Grant Agency VEGA in the project No. 1/0336/10

References

[1] HOSSAIN, E., LEUNG, K. *Wireless Mesh Networks – Architectures and Protocols.* New York: Springer, 2007. 333 pages. ISBN 978-0-387-68839-8.

[2] P. GUPTA, P. KUMAR. *Capacity of Wireless Networks*. In IEEE Transactions on Information Theory, volume 46, pp. 388–404, March 2000.

[3] RAMON AGUERO CALVO, JESUS PEREZ CAMPO. Adding Multiple Interface Support in NS-2, Jan. 2007.

[4] The Network Simulator ns-2, http://www.isi.edu/nsnam/ns/, 2008.

[5] CHI MOON OH, HWA JONG KIM, GOO YEON LEE, CHOONG KYO JEONG. A Study on the Optimal Number of Interfaces in Wireless Mesh Network. In *International Journal of Future Generation Communication and Networking, IJFGCN Vol. 1, No. 1*, Dec. 2008, pp. 59 – 66.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Simulation Toolkits for Multiagent Systems

[*]Jana Púchyová
[*]University of Žilina, Faculty of Management Science and Informatics, Department of Technical Cybernetics, Univerzitna 1, 01026 Žilina, Slovakia, {Jana.Puchyova}@fri.uniza.sk

Abstract. The multiagent systems have been very highly spread area of research in recent years. This approach to problem solving can be used in many different groups of tasks and in some cases these systems can find solution where the others do not work. Sometimes there can be problem to define the right characteristics of the whole system. And before implementing multiagent system there is very important to verify the correctness of its behavior. For this reason special toolkit software for simulating multiagent systems was developed. So this paper deals mostly with questions as how some of available toolkits work, and when and how can we use them.

Keywords: Multiagent systems, agent, simulation.

1. Introduction

Multiagent systems (MASs) are used in many areas in the last years where other systems are not able to find a problem solution. At first we need to know, that agent according to [1] is autonomous actor which can be able to interact with environment and other agents. MAS is then a group of communicating agents in environment trying to solve defined problem. Here exists many types of agent and MAS, this paper is focused on the simulation of software agents regarding to their future robotic form. Some basic divisions you can find in [2].

2. Simulation and modelling

Because of unsteady terminology, the definition of simulation will be based on terminology by [3]. Simulation and modelling are used to study those types of objects which in real world exist already and we want to improve them or they detect the characteristics of objects we want to create in the future.

Simulation is considered as a research method that replaces dynamic, time-dependent system by its simulating model. Thanks to this change we try to find information about original system.

In order to use simulation, we need to create a suitable model, which creates the analogy between the modelled and modelling system. In this process we neglect irrelevant characteristics of the system and we work only with those of them that are necessary for behaviour which has to be verified.

For work with MASs there have been developed special simulation toolkits. In some cases, the limitations of these special simulators can be the reason why to use rather one of object-oriented languages and use any of the libraries for the simulation. Here will be described the special simulators designed for groups of agents. All three of these simulators are freely available, more about them will be described below. Of course each of the simulators has limited capabilities and for its use it is necessary to know the structure for own model you will create. All of these simulators are object-oriented which is in terms of properties of agents the best way. In one class are encapsulated not only the characteristics of the agent but also its possible behaviour in situations arising from the environment.

3. Tools for multiagent systems simulation

For all simulators, which will be described, the environment consists of a grid composed of square cells, and thus each cell has eight neighbouring cells. Each of them is using its own model structure, which is written below.

3.1. SSR-Swarm

Simulator was developed at Harvard University by Ian Rose, especially for colonies of living organisms simulation [4]. The group of libraries is written in Java. Results of simulation can be presented in text but also graphic mode. The main idea for creating the simulator was to separate behaviour and characteristics of the agent from the environment in which it is located. The structure of the model used in this simulator is as follows:

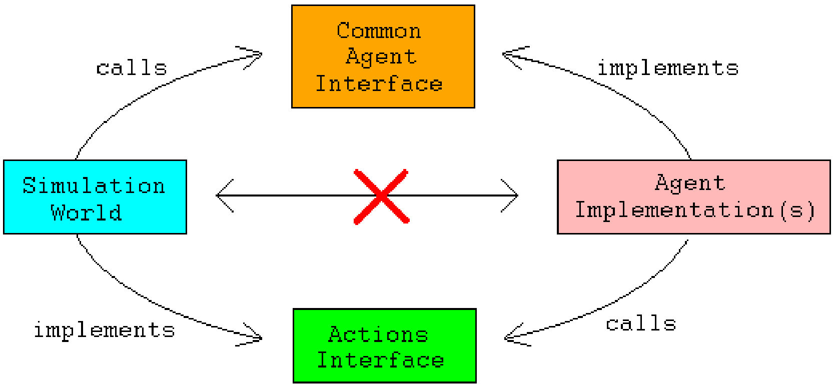


Fig. 1 Structure of model used in SSR-Swarm, [4].

The class "Simulation World" contains the simulation environment and all the rules agents must obey. "Agent Implementation" is class completely separate from the environment and they cannot communicate directly but through the interfaces "Common Agent Interface and "Actions Interface". The relationship among them you can see from the figure.

3.2. Swarm 2.2

Described simulator has better manual and more about you can found in [5]. Swarm simulator is written in Objective-C but has been modified so it is possible to write your code in Java also. It was created as a hierarchical simulator, model may vary slightly according to each task, but the main structure is the same. Also in this case it is possible to run the simulation in text or graphic mode.

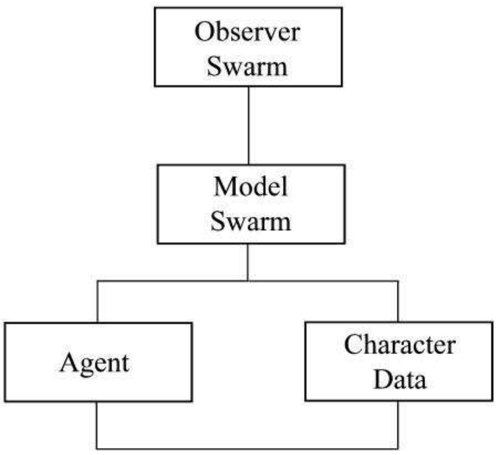


Fig. 2 Structure of model for the simulator Swarm 2.2.

- “Observer Swarm” creates the desired simulation mode and other components. It is considered the highest class of the model.
- “Model Swarm” contains the whole simulation mechanism, organizes hierarchically lower classes, collects information it needs and provides them also to the higher layers.
- “Agent” is the class for definition of the characteristics and behaviour of each agent. Model may contain several agents of this class, in the case of multiagent systems it is a necessity. For a heterogeneous system, each different type of agent has a new class.
- “Character Data” is the class containing information about the observed characteristic of the environment, which is examined in the system. Information from this class uses not only upper “Model Swarm”, but also agents for their decisions.

During simulation it is possible to set the values of selected probes to adjust the simulation conditions.

3.3. NetLogo

NetLogo is the modelling environment for characteristics of simulation project study of various kinds of problems from nature and society [6]. NetLogo runs on Java virtual machine and is also a multi-agent modelling language. Model can be displayed in 2D and also in 3D and you can rotate your graphic output according to all the axes.

The World used in NetLogo is composed of four types of agents:

- turtles- mobile agents, while running the simulation they can die or hatch,
- patches- forming agent environment, in graphic mode they are cells in a grid, turtles move over them,
- links- agent that connects turtles for graphs and networks making,
- observer- agent looking after everything happening in the world.

The main NetLogo advantage is update after every code change. There is no need to close each application individually and compile separately.

Before starting the actual simulation you can use the command line, which is used for experiments with the goal to find out better improvements of the simulated system, rarely to save the changes.

According to the possibilities of the simulation toolkits and a better graphic output was NetLogo used in several different studies [7], [8].

3.4. Why to use the simulation

Simulation is used to examine the behaviour of MAS under some conditions. Everything depends on the characteristics we want to study. Here are mentioned two examples for comparison, how can characteristics of the system change the state of whole system. In first example you can see how a big number of agents can overheat the environment in a special task. In the second example is visible, that in some problems can be the use of more agents better way because they can help to shorten time of terrain exploration.

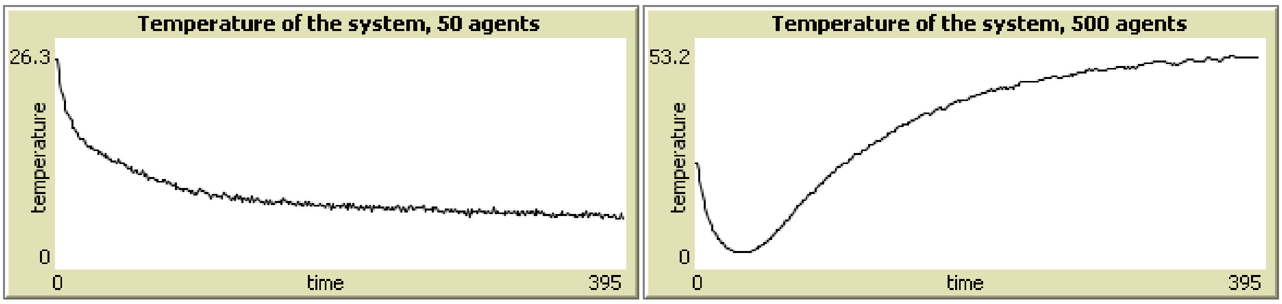


Fig. 3 Dependence of temperature of the system on the number of agents in special task.

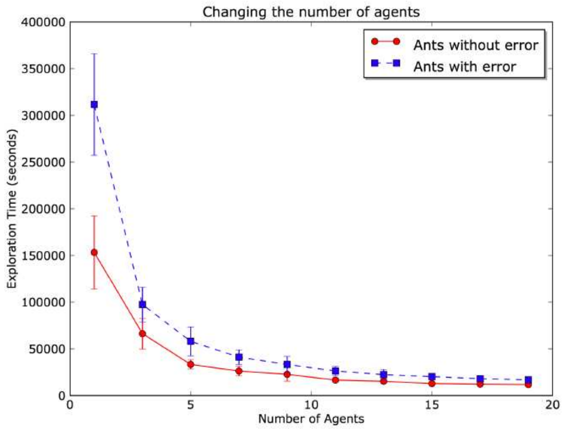


Fig. 4 Dependence of exploration time on the number of agents in special task, [9].

4. Conclusion

Simulation can be used to verify the behaviour of multiagent systems. As it is visible from the written, the most extensive simulator of these is NetLogo. Sometimes it is better to use not as complex simulator and programme your own needs. Another way how to choose the simulator is the choice according to the structure of a model. But in this case, for later studies we will use the NetLogo at first. For our studies we need to create the universal model of multiagent system for the most covering terrain problems, find how failure of agents can influence the solution and verify its characteristics.

References

[1] SCHMOTZER, M. *Samoorganizujúce Sa Skupiny Rozumných Agentov.* Dissertation Thesis. Košice. 2005.

[2] PÚCHYOVÁ, J. Behaviour of Multiagent systems. *Winter school MICT*. 2011.

[3] KAVIČKA, A., KLIMA, V., ADAMKO, N. *Agentovo Orientovaná Simulácia Dopravných Uzlov.* Žilina : EDIS, 2005.

[4] ROSE, I. *SSR-Swarm Simulation Environment*. [Online]. 2007. http://www.eecs.harvard.edu/~ianrose/cs266/.

[5] JOHNSON, P., LANCASTER, A.[Online] 2000. http://www.swarm.org/swarmdocs-2.1.1/userbook/userbook.html.

[6] WILENSKY, U. NetLogo. *Center for Connected Learning and Computer-Based Modeling.* [Online] 1999. http://ccl.northwestern.edu/netlogo/.

[7] JANOTA, A., SPALEK, J., HRBČEK, J. NetLogo – prostredie na tvorbu multiagentových systémov a jeho využitie na simuláciu riadenia železničného priecestia. *AT&P Journal, PLUS7.* 2005.

[8] TREVERTON, R., MENEZES, R. Evaluating failure in terrain coverage by autonomous agents. *IEEE Workshop on Robotic Intelligence in Informationally Structured Space.* Nashville, TN. p. 79 - 86. 2009.

[9] FERRANTI, E., TRIGONI, N. The Impact of Localization Errors on the Performance of the Ants Exploration Algorithm. *Workshop on Agent Technology for Sensor Networks (ATSN), a workshop of the International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).* 2009.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Implementation Mes into Intelligent Manufacturing System

*Ján Rofár **Jana Strápková
* University of Žilina, Faculty of Mechanical Engineering, Department of Industrial Engineering, Univerzitná 1, 010 26 Žilina, email: jan.rofar@fstroj.uniza.sk
**University of Žilina, Faculty of Mechanical Engineering, Department of Industrial Engineering, Univerzitná 1, 010 26 Žilina, email: jana.strapkova@fstroj.uniza.sk

Abstract: This article is focused on possibilities of collaboration MES with Intelligent Manufacturing System. Complicated information flow which execution has became difficult demand to find ways to ensure transmission of data at all levels of the company by information technology. One of the ways is implementation systems of manufacturing control (MES). These systems represent the middle layer between the subordinate systems to be at the lowest level of the enterprise and superior information systems. Implementation of this type of system enables efficiency all processes. Correct functioning of systems and a requirement that the information should be in the correct range, quality and the right time and right place are certainty. Global changes have forced manufacturing companies to respond to their challenges. Many projects in the areas of research and development which are based on creating new information technologies and intelligent manufacturing facilities have started. Seeing that they are based on an intelligent communication of all their components currently are revaluated possible links and cooperation with MES.

Keywords: Manufacturing Execution System, Automation, Intelligent Manufacturing System

1. Introduction

Constantly changing condition in business environment leads companies to change ways to increase their competitiveness and effective utilization of emerging changes. Be successful and to resist changes is requiring rapid response to emerging situations. As a result is progressing discussions about the direction of development of production and have been defined major areas which involved in their development. The basis for success in global markets is the adaptation of production and customers relations. Customer requirement and theirs behavior are unpredictable, because many companies were forced to find solutions to adapt their source to markets and customers without losing productivity. IMS (Intelligent Manufacturing System) is identifying as a concept "new production", which is acting in service sector with new business systems architecture and with highly integrated enterprise functions. Transformation of the conventional type factory into a modern service center has led to problems with management. Modern value creation is no longer a matter of the product itself, but the whole process. A decisive potential of companies became not the production capacity, but their operational ability. Ensure high transparency of the processes led to the requirement to mapping the value of real-time without unnecessary processes that involve high costs. Consequently the market began to appear modern manufacturing information systems that met these requirements.

2. System of manufacturing control MES

Development of information systems passed significant changes. While in the past information system have focused on improving the use of computer support, today the accent on obtaining the real value stream mapping. Changes in business environment and increasing complexity of production began to require a holistic view of manufacturing, services and options. Companies started to realize the need to link information between the different levels and thereby increase efficiency their use. This needs and requirements have resulted in creation of system of manufacturing control MES. Manufacturing Execution Systems (MES) are represented by an independent organization MESA International. [4] This organization began them to standardize and with them established three application levels within the corporation (enterprise, operational, procedural). Although these systems are known for several years still exists new knowledge in this areas, which contribute to the efficiency of information transmission across the enterprise. Continuous development of information systems has increased effort to integrate many information solutions. To ensure right functioning of systems is the need support of information at the right time, right place, in the correct place and quality. This ensures easy, fast processing and flexible exchange between workers. [6] The most important organization in this field is MESA International which first defined these systems, their basic model and described their basic functions. Its aim is to assist in implementing these systems. It groups together companies and organizations which deal with these issues. Regularly carries out surveys and edits many publications about direction and trends of MES. Another organizations deal with MES is ISA (International, Systems and Automation Society). Its edited standards which are important for systems of manufacturing control MES. It has defined five layer which corresponding basic layer in enterprise. (Fig. 1)

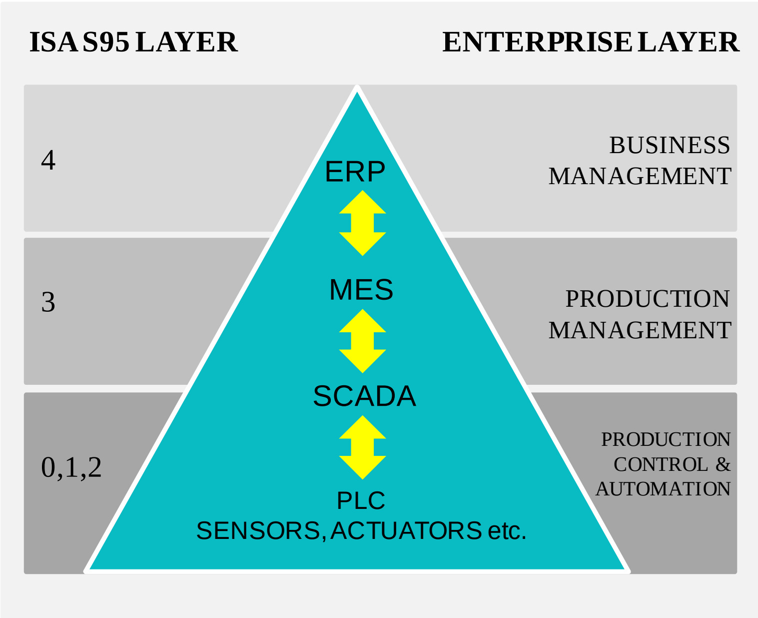


Fig. 1. Layer by ISA 95

ISA-95 is used by many companies in creating new solutions. It divided into models of UML (Unified Modeling Language) which are basis for the development of standardized interfaced between ERP (Enterprise Resource Planning) and MES. [1] MES solutions are not simple applications, but they consist of an integrated set of manufacturing activities and support applications that were developed using the right closed technologies. The main entity of systems is a simple automated information loop between shop and enterprise information systems. MES provides up to minute information available at the shop floor layer to other systems what enables rapid response to the conditions and requirements. It constrains database of real-time data for monitoring processes in the production and also information about continuous improvement in various activities. Data obtained in real-time are providing a continuous overview of the manufacturing operations and with them we will get complete look at the shop floor. The transfer of information inside enterprise layer is realized in different time horizons. MES with ERP systems together providing the information which are generating precise and realistic plans, shorter production cycles, less work in process and lower inventories. MES also provides shop floor control and a necessary overview to effective solving requirements. MESA International defined eleven basic functions from which in new c-MES model are eight. [5] Application of MES brings to companies more benefits and with them can improve their market position. Every year MESA organizes surveys of benefits after implementation MES and it has demonstrated large number of contributions to companies.

3. System of manufacturing control MES

Manufacturing organizations are faced with constantly changing environment and dynamically growing complexity of operations. Professor H. Yoshikawa of the University in Tokyo initiated creation of an Intelligent Manufacturing System (IMS). This system is capable of providing the flexibility which is increasing with performance. They can facilitate the very difficult manufacturing systems as well as variety degrees of functional products. [2] By them we can execute changes and adapting as soon as possible to market changes and customer requirements. Short periods of production cycles, shorter lead times, adaptation to the changing situation in short term and additional, this are benefits which organizations would like to receive. They can predict problems before they occur and provide appropriate detention (corrective) appliance, IMS can be very useful to support the expected level of competitiveness. [3] Department of Industrial Engineering of Faculty Mechanical Engineering in University of Žilina participate in research and development Intelligent Manufacturing System ZIMS. ZIMS is a working name of Zilina Intelligent Manufacturing System. This concept was created at the initiative of prof. Ing. Milan Gregor, Phd., prof. Ing. Branislav Mičieta, Phd. and prof. Ing. Štefan Medvecký, Phd. Their ambition is to create in Žilina together with the University of Žilina, UKAI and CEIT a future intelligent manufacturing system. ZIMS wouldn´t represent letter for letter IMS. This working place will focus on directed research in this area. The main vision is to create in University of Žilina a prototype of intelligent manufacturing enterprise. Part of this concept is the application of information systems (SCADA/HMI, MES, ERP etc.) to the existing model of modular production flow system represented by FESTO FMS 500 (Figure 2.) [5]

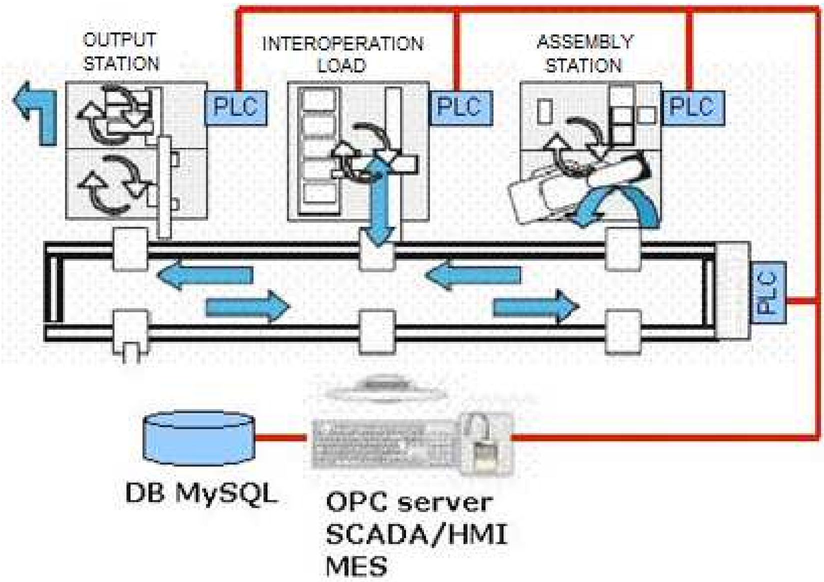


Fig. 2. Model of FESTO FMS 500 [5]

Seeing that FMS 500 is designed primarily as a model of automated assembly line without the ability to respond to the information flow from higher layer control system, it was necessary to change its control algorithms. In developing the control algorithms should be particularly attentive to:

1. Specification and analysis of key determinants of production, which are necessary to the production system to react flexibility.
2. The method of algorithms design and their implementation into the control system.

The first and second point will vary depending on the specific applications, but the methodology of creation of control programs in automated equipment is largely changed from current status. Creating an IMS will require much greater cooperation between the different layers of control. Current technologies however allow full this cooperation. As solution to this situation are offering several models and interdependence among different layers of control.

4. Conclusion

In company is today a lot of information technology. Because of this companies began considering of idea of integration all levels into enterprise-wide information system. Thereby they can avoid the possible shortcomings in their mutual communication. Standard represent the present and future between information systems. Seeing that companies have usually two-level hierarchy of IT have started implementation of MES. These systems introduce interlayer which plays a very important role and ensures common integration of information. Although this issue is still dealing with major organizations that determine standards and direction in this area, companies don’t use the all potential of MES solutions. Many discussions of future orientation of the development bring new areas of research directions, which can bring competitiveness to the companies. Nowadays is discussing about possibilities of common collaboration MES and IMS.

Acknowledgement

This paper was made about research work support: VEGA no. 1/0241/10.

References

[1] ANSI/ISA-95.00.01:2000, Enterprise-Control System Integration - Part I: Models an Terminology.

[2] Introduction. 2011. Available on the Internet: <http://www.ims.org/introduction>.

[3] Benyoucef, L., Grabot, B., Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management. Springer-Verlag London 2010. ISBN 978-1-84996-118-9.

[4] Kletti, J. 2007. Manufacturing Execution Systems - MES. Springer - Verlag Berlin Heidelberg New York, 2007, ISBN 978-3-540-49743-1.

[5] MACEK, P. - ŠIMÁK, V. - ROFÁR, J. - MIČIETA, B.: Aplikácie nových automatizačných technológií v laboratóriu automatizácie a simulácie procesov vo výrobe. In: Acta Mechanica Slovaca. - ISSN 1335-2393. - Roč. 12, č. 1-A (2008), s. 317-320.

[6] MESA International: White Paper Number 1-6. 1997. Available to members of the MESA.

[7] MEYER, H., FUCHS, THIEL, K.,. 2009. Manufacturing Execution Systems. Optimal Design, Planning and Deployment. The McGraw-Hill Companies, Inc. 2009, 243 s. ISBN 978-0-07-162602-6.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Towards Developing Personal Diet Assistant for Diabetics

*Miroslav Rusin
*University of Žilina, Faculty of Management Science and Informatics, Department of Informatics, Univerzitna 1, 01026 Žilina, Slovakia, Miroslav.Rusin@fri.uniza.sk

Abstract. Diabetes Mellitus is a disease that causes pain to the individuals and the whole society as well. Telemedicine is one of many efforts to help with this condition. In this paper, we propose a telemedicine system focused on diabetes management, especially on diet. Diet has a key role in a life of a diabetic and should not be neglected.

Keywords: diabetes mellitus, telemedicine, diet.

1. Introduction

Information and communication technologies play vital role in the area of medicine. They help in diagnosing and treatment of various diseases [1]. Diabetes is a chronic disease that deals with abnormal level of blood glucose. This is caused by insufficient production of insulin; a body is not capable to absorb glucose. Diabetes is incurable but can be managed by maintaining a healthy diet, exercising daily and monitoring blood glucose several times per day. On the other hand, neglecting this condition can lead to severe complications such as retinopathy (blindness), neuropathy (leg amputation), nephropathy (kidney failure), cardiovascular problems, etc.

Situation concerning diabetes disease is grave as the number of deaths related to it is among top 10 in the world [2]. Another fact that contributes to the importance is the rising number of diabetics, estimations say that it will rise from 194 million in 2003 to 334 million in 2025 [3]. Society suffers from this as well. European countries allocate 5-10% of their health care cost for diabetes [4].

It is no surprise that there have been many initiatives to help in treating this condition. Telemedicine is one of them. Telemedicine means medical care on distance; it utilizes information and communication technologies to overcome geographical barriers and distances. Access to information supports decision making and this way a patient’s well being.

A diabetic needs to understand nutrition and its impact on his/her health status. There already exist Internet-based applications that help people with nutrition in Slovakia (http://www.florastranky.sk/e-dennik). A man can enter some meals and obtain information about the nutrition values contained there and comparison to the values that are expected for his/her target group. This way nutrition information is provided but this area requires more sophisticated approach including menu recommendation, personalisation or prediction. For these reasons we propose system called Personal diet assistant for diabetes (PDAD).

There are many telemedicine diabetic systems that consider nutrition but only in terms of gathering data and evaluating them [5]. Methods for diet assessment [6], eating behaviour profiles creation [7] were also introduced. However, the focus was on diet menu generation and meal planning. We provide some existing approaches to this problem.

Automated meal planner was proposed by Bulka et al. [8]. A new concept, glycemic index, is introduced in praxis. Glycemic index describes the extent to which a product raises blood glucose after eating. Now, it is not only important how much food a patient eats but also in what combinations as different combination causes different glucose level alteration. It utilizes genetic algorithm to find optimal meal plan. The best solution is optimal combination of products with suggested amounts. This planner is intended only for educational purpose as the model does not consider all the factors that influence blood glucose level and model parameters for all specific patients are not always easily determined.

Solution for diet-recommendation problem is proposed in [9]. This approach is based on fuzzy logic as there are uncertainties associated with diabetes patient's everyday life, linguistic uncertainties and uncertainties with the experts' opinions. The evaluation results confirmed that with this approach they are capable of generating individualized and sensible diet suggestions. Although they consider personalization, they did not provide any algorithms for learning from patients' choices in dynamic way as they may choose to ignore prepared menu and alter it. The authors also rely on general knowledge from expert's in the area of nutrition and do not test in some way what consequences can the generated menu have on an individual.

2. Future Health Service Model

During the system design process it is advantageous to consider existing models and trends. U.S. National Institute of Health described a future health model that served as basis for 6-P's paradigm [10]. This paradigm depicts the desired properties of new health systems (see Fig. 1). Two fundamental questions that are laid are: "What kind of health care decisions should be made?" "How health care decisions should be made?". The answer lies in 6 P-properties. *Prevention* will focus on individuals as well as population. *Prediction* risk models and *Pre-emptive* treatment to specific diseases might prevent life-threatening situations. *Personalization* is of essence as each individual is different and health care provision must adapt to the specific case. Objective of *Pervasive* care is to be available to a patient, independent of place or time. *Participatory* means that a patient can share information and decision making with health care providers.

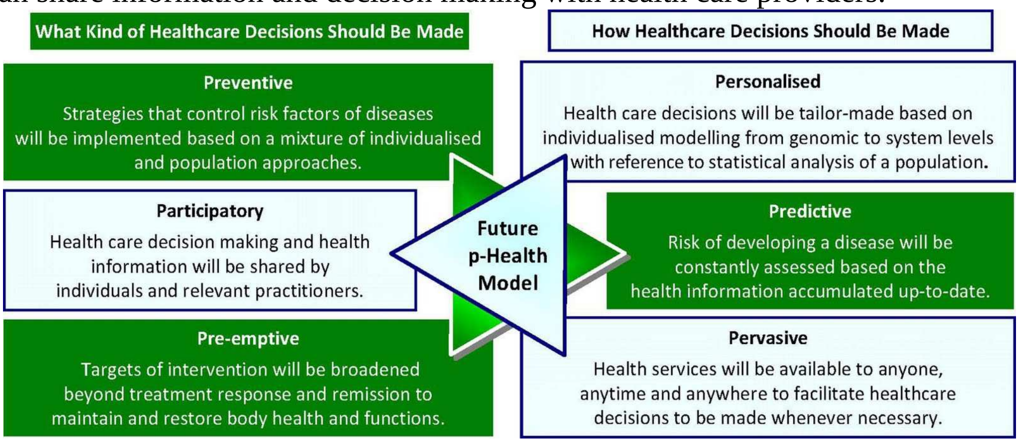


Fig. 1 Future health model [10]

We use this 6-P's paradigm to further describe the proposed system. The stated properties will translate themselves in our system in this way:

- Personalised – there will be a basic profile filled out by a user including, e.g. sex, age, weight, etc. Furthermore, the system will be able to expand it. It will contain a block of artificial intelligence able to learn about its user from a dietary point of view as much as possible. The learning data will be provided by a user feedback on his/her changes to a diet and subjective evaluation of the diet personalisation measure and objective evaluation of its effect, i.e. weight and blood glucose level. We want to exploit the existing data about an individual stored by our system in the process of individualized menu planning.
- Participatory – access to patient health data will be shared by both a patient and a physician with supervision
- Pervasive - this self-management assistant will be available as mobile and web application

- Preventive – maintaining healthy weight, as recommended by WHO, can prevent or delay diabetes. Furthermore health lifestyle advices will be offered on the web page of national diabetic organization (http://zds.sk/) that we cooperate with.
- Predictive - before we offer a meal plan, we will try to predict the possible outcome based on the past records and obtained knowledge about the user
- Pre-emptive - the application is intended to be put in praxis and used on daily basis by patients whose health condition is dependent on following certain diet

3. Project of System – Personal Diet Assistant for Diabetes

The telemedicine system for nutrition was first described in [11] where basic formalization of the problem and desired system properties were given. We proposed an automated tool focused on diet and its impact in the life of a diabetic. Here we want to provide more detailed specification along with the development outline. After the certain time period of using, it will offer menus adjusted to the user specific situation but at the same time consider eating habits of Slovaks.

To accomplish this functionality we can split the process of its development into several steps. These components will be developed:

1. Food database. It should consider relationships between various types of food to allow their combining or substitution.
2. Patient module with related data storage. Developing methods for recording patient`s history and reactions on particular meal which could help in the process of predicting a future development of the health state.
3. Algorithms for suggesting menus. Theses algorithms should consider a man’s current health condition, preferences, dislikes, allergies and long-term health plans.
4. Algorithms for learning eating behaviour
5. Algorithms for prediction a future development of a patient’s health condition taking into account various factors: their history recorded in the system, their eating habits, physical activities, age etc.
6. Algorithms for system reliability analysis.
7. All supporting methods and modules in parallel.

After finishing the implementation phase, our intention will be to test the system capability. Firstly, we will perform some tests using computation methods and ask the experts for their opinion and assessment. Then we would like to offer it to patients under supervision of nurses and doctors.

A brief description of the data processing in our system follows next (see Fig. 2). There will be two types of system users: a patient and a physician.

In a patient’s point of view, the user creates his/her own profile by entering some input values, i.e. sex, age, weight, carbohydrate intake (see Patient module in Fig. 2). These data are fed to a menu generating mechanism (see Menu generator in Fig. 2) that provides menu for a certain period of time. We can accomplish personalised menus via 2 ways: a user has the opportunity to fill out a questionnaire regarding his/her likes, dislikes and we will consider them in the algorithms or the system will be able to track past changes to the menu and learn from them. These changes are anticipated as it is not very likely that a user will behave exactly according to the given recommendations. It will be possible to replace some meals in a menu in favour of the preferred ones (see Menu module in Fig. 2). Menu generating mechanism should adapt to the patient’s personal goals, i.e. losing/gaining weight. An interface for diet database management will be available for a dietician (see Diet module in Fig. 2). Diet is usually given in a form of advices, i.e. what food to avoid, what nutrition values should daily meals contain and in what proportions, etc. So we need to map this into usable inputs.

First step of the work will be to design and implement food database (see Food DB in Fig.2). There are tables stating nutrition values for each food item. Taking into account these data, it is up to a diabetic to design his/her menu. Firstly there will be the nutrition calculator based on patient’s choice of meals. The plan is to implement this module into an existing web page of national diabetic organization (http://zds.sk/).

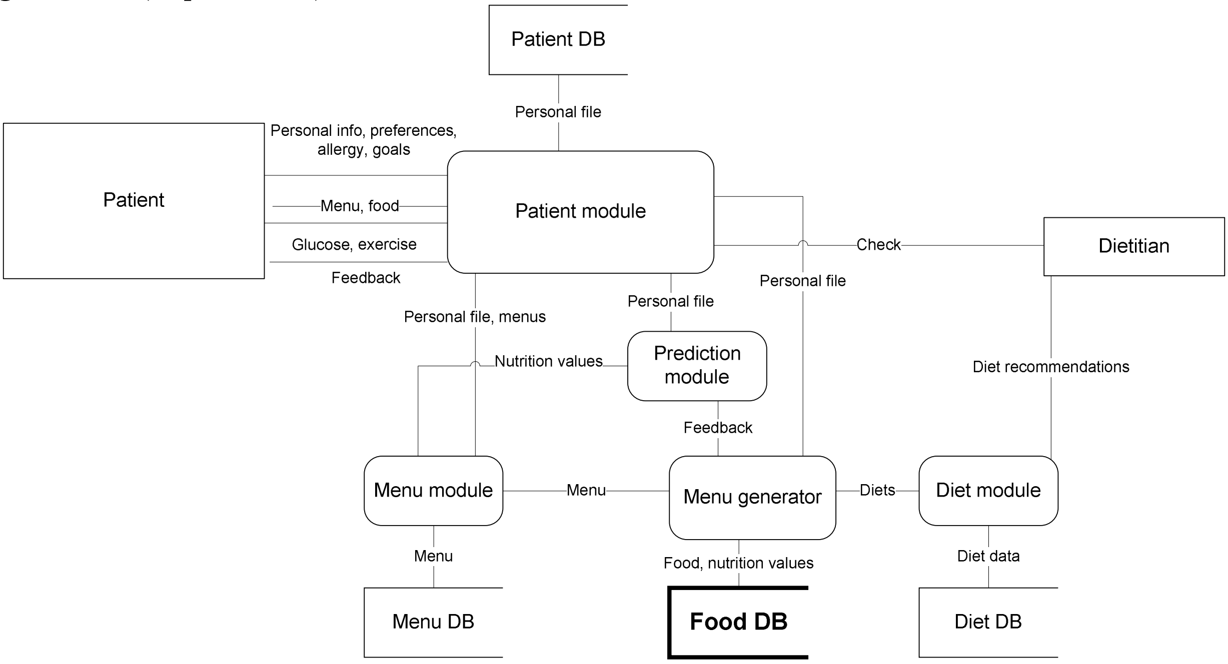


Fig. 2 Data flow diagram of the proposed system

4. Conclusion

Diabetes Mellitus is a disease that causes pain to the individuals and the whole society as well. Telemedicine is one of many efforts to help with this condition. In this paper, we propose a telemedicine system focused on diabetes management, especially on diet. Diet has a key role in a life of a diabetic and should not be neglected. We provide the problem formalisation and future plans are presented. We believe that by process of personalisation a patient will follow the diet that will lead to his/her well being.

References

[1] LEVASHNKO V., et al. *Nomogram for assistance in treatment of prostate cancer*. Proc. of the 5th International workshop on Digital Technologies (DT 2008), 20-21 November, Zilina, Slovakia, 2008.
[2] http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf
[3] WILD, S., et al. *Global Prevalence of Diabetes.* In *Diabetes Care*, 2004.
[4] http://www.idf.org/webdata/docs/DiabetesWorkshopProceedingsJune2004.pdf
[5] TATARA, N., et al. *A Review of Mobile Terminal-Based Applications for Self-Management of Patients with Diabetes*. International Conference on eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED '09., pp.166-175, 1-7 Feb. 2009
[6] MEI-HUI, W. et al. *Property and application of fuzzy ontology for dietary assessment*. 2010 IEEE International Conference on Fuzzy Systems (FUZZ), pp.1-8, 18-23 July 2010
[7] LAUBSCHER, K, et al. *Toward a probabilistic model of eating behavior for patients with type 1 diabetes*. Systems and Information Engineering Design Symposium (SIEDS), 2010 IEEE , pp.147-152, 23-23 April 2010
[8] BULKA, J., et al. *Automatic meal planning using artificial intelligence algorithms in computer aided diabetes therapy*. Proc. of 4th International Conference on Autonomous Robots and Agents, 2009. ICARA 2009, pp.393-397, 10-12 Feb. 2009
[9] CHANG-SHING, L., et al. *A Type-2 Fuzzy Ontology and Its Application to Personal Diabetic-Diet Recommendation.* IEEE Transactions on Fuzzy Systems, vol. 18, no. 2, April 2010
[10] ZHANG, Y. T., et al. *Editorial Note on Bio, Medical, and Health Informatics.* IEEE Transactions on Information Technology in Biomedicine, vol.14, no.3, pp.543-545, May 2010
[11] RUSIN, M. *Personal diet assistant for diabetes.* Winter School MICT, Šachtičky Slovakia, 3-8 Jan. 2011 [accorded]

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

To Performance Modelling of Virtual Computers

*Dan Slováček, **Peter Hanuliak
*European polytechnic institute, Kunovice, Czech Republic, slovacek@edukomplex.cz
**Polytechnic institute, Dubnica nad Vahom, Slovakia, phanuliak@googlemail.com

Abstract. The paper describes development of the new analytical model for the study of the basic parameters of the virtual computers. The suggested model considers for every node of the computer network one part for the own node´s activities (communication functions) and another one for the modelling of the node's channel for data transmission. We supposed a using of multiprocessor system as modern node's communication processor in order to model both overheads (the own node's activities and the node's transmission latency). Such analytical model includes the real non exponential nature of the input to the individual transmission channels. The achieved results of the developed model will be compared with the results of the common used analytical models to estimate the magnitude of improvement.

Keyword: Virtual computer, performance modelling, network of workstations, queuing theory

1. Introduction

For the contemporary technical level of the reachable computer means (desktop computers, minicomputers, supercomputer etc.) is dominate using of various typical forms of the connected processors or computers (Virtual computer). In this sense recent trends are using more than one core or processor (SMP workstation, modern graphical cards) in order to achieve higher performance of virtual system. This trends are using also in high performance computing (HPC) through using networks of workstations (NOW, cluster) as a cheaper alternative of virtual parallel computer. A massive using of connected virtual computers (NOW, cluster) builds various forms of Grid system (Metacomputer).

2. The architectures of the virtual systems

We can divide realised virtual systems to the two following groups

- model with shared memory [1, 4, 6, 10]. The basic system properties are given through the existence of some kind of the common shared memory,
- model with distributed memory [1, 3, 6, 9, 11]. This group covers the field of various forms of computer networks (NOW, Grid),
- hybrid model as a mix of the both previous models.

The basic question is to model and analyse behaviour of all of these models. A crucial question is modelling of used communication system.

3. Modelling of communication system

The communication model for the whole transport system of the virtual system corresponds to the mutual model connections of the used communication computer according the chosen topology. The common model of servicing transport network consists from U-nodes where the i-th node can create j-servicing transmission lines to the next nodes of the transport network. Each modes of transmission line has the average servicing time $1/\mu$ seconds (exponential servicing time distribution). If we define the individual computer network nodes as graph nodes and their mutual communication lines as graph edges we get in common for the transport system with U-communication computers oriented graph with U-nodes according the Fig. 1. where

- $\gamma_1, \gamma_2, \ldots, \gamma_u$ represent the total intensity of input data stream to the given node. It is given as Poisson input stream with the intensity λ demand in time unit,
- r_{ij} are given as the relation probability from node i to the neighbouring connected nodes j,
- $\beta_1, \beta_2, \ldots, \beta_u$ correspond to total extern output stream of data units from the individual nodes.

This model corresponds in queuing theory to the model of open servicing network. The adjective "open" characterize the extern input and output data stream to the servicing transport network [2,4, 7]. In common they are the open Markov servicing networks, in which the demand are mixed together at their output from one queuing theory system to another connected queuing theory system in a random way to that time as the are leaving the network. To the given i-th node the demand stream enter extern, with the independent Poisson arrival distribution and the total intensity γ_i demands in seconds. After servicing at i-th node the demand goes to the next j-th node with the probability r_{ij}.

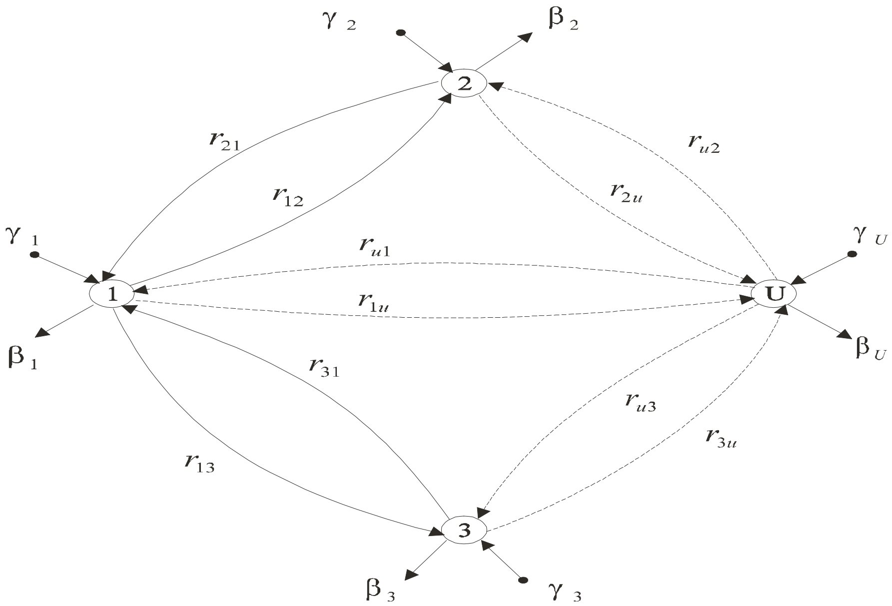


Fig.1. Basic model of the communication system.

4. Analysis methods of the communication system

To the analysis of computer network have been developed two basic differentiate principles

- analytical methods using the queuing theory [2, 4, 5, 7]
- discrete simulation [12].

The approach on the basis of mathematical queuing theory application is a very effective and practical tool for the analysis mainly of the large complex distributed computer networks. The discrete simulation can give the very interesting sight to the behaviour analyze of the smaller computer networks but it is fully unusable to the analysis of massive virtual computer. It is very useful for the analysis mainly in these cases in which they do not exist any analytical methods.

4.1. The analytical approach

For the mathematical model of the virtual computer it is necessary to specify

- the statistical character of the input demand stream,
- the servicing network and the servicing times,
- way of servicing (FIFO, LIFO, servicing with priorities etc.),
- number of buffers for storing the transmitted data unities.

For the presented communication model we derived the whole delay as

$$T = \frac{1}{\gamma}\left(\sum_{j=1}^{u} \lambda_{ij} \cdot T_{ij}\right) \qquad (1)$$

where the defined parameters are

- γ - the whole extern input flow to the communication system,
- λ_{ij} - the whole input flow to the j-th transmission channel at i-th node,
- u_i -the number of transmission channels at i-th node,
- T_{ij} - the average servicing time of the j-th queue of the transmission channel at i-th node.

This basic model decomposes the whole communication system to the individual M/M/1 systems according the Fig. 2.

Fig. 2. Model of one transmission channel of the i-th node.

To improve more precise communication model we suggest to derive whole delay as

$$T = \frac{1}{\gamma}\left[\sum_{i=1}^{U}\left(\lambda_i \cdot T_i + \sum_{j=1}^{u_i} \lambda_{ij} \cdot T_{ij}\right)\right] \qquad (2)$$

where new parameters are defined as

- λ_i - the whole number of incoming messages to the i-th node, that is the sum both of external and internal message inputs to the i-th node
- T_i -the average servicing time in the message queue (the waiting in a queue and servicing time) in the i-th node.

5. Conclusions

To improve the mentioned problems we are going to suggests the behaviour analysis of the virtual computers the improved analytical model, which extends the previous communication model in two areas

1) it considers also the delays caused through the activities of communication processor and through awaiting this service,
2) it takes into account correction factor to take into account also the influence of real non exponential nature of the inter arrival time of inputs to the transmission channels.

These corrections contribute to precise behaviour analysis of actual virtual computer for the typical communication activities and for the variable input loads.

References

[1] ABDERAZEK, A. B., *Multicore systems on-chip – Practical Software/Hardware design*, 200 pp., Imperial College Press, August 2010.

[2] DATTATREYA, G. R., *Performance analysis of queuing and computer network*, 472 pp., University of Texas, Dallas, USA, 2008.

[3] FOSTER, I., KESSELMAN, C., *The Grid 2, - Blueprint for a New Computing Infrastructure (Second Edition)*, Morgan Kaufmann, 748 pp., USA, 2003.

[4] GELENBE, E.., *Analysis and synthesis of computer systems*, 324 pages, Imperial College Press, April 2010

[5] JOHN L, K.., EECKHOUT, L., *Performance evaluation and benchmarking*, CRC Press, 2005.

[6] HANULIAK. I., *Parallel architectures – multiprocessors, computer networks*, 187 pp., Ed.: Book Centre, Žilina, 1997.

[7] HANULIAK, I., *Parallel computers and algorithms*, 327 pp., Publ.: ELFA Košice, 1999.

[8] HANULIAK, J.., *To performance evaluation of distributed parallel algorithms*, Kybernetes, West Yorkshire, Volume 34, No. 9/10, pp 1633-1650, United Kingdom, 2005.

[9] HANULIAK, P.., *Comparison of HPC and GRID systems*, In Proc.: Veda 2006, FRI ŽU, Žilina, pp. 140-146, 2006

[10] KIRK, D. B., HWU W. W., *Programming massively parallel processors*, Morgan Kaufmann, 280 pages, 2010

[11] PATERSON, D. A., HENNESSY, J. L., *Computer Organisation and Design*, 4 - th Edition, 912 pp., Morgan Kaufmann, 2009

[12] KOSTIN, A., ILUSHECHKINA, L., *Modelling and simulation of distributed systems*, 440 pages, Imperial College Press, Jun 2010.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Virtual Reality Simulators in Education

*Ondrej Števka, **Monika Welterová, ***Miroslav Bariak, PhD.
* University of Žilina, Faculty of Applied Mechanics, Department of Applied Mechanics, Univerzitná 1, 01026 Žilina, Slovakia, ondrej.stevka@fstroj.uniza.sk
**University of Žilina, Faculty of Operation and Economics of Transport and Communications, Department of Railway Transport, Univerzitná 1, 01026 Žilina, Slovakia, monika.welterova@fpedas.uniza.sk
***111, s.r.o., Dolné Rudiny 3, 010 01 Žilina, Slovak Republic, Miroslav.Bariak@111.sk

Abstract. Educational simulations are one of the technological approaches available to facilitate the building of skills such as critical thinking. They do not represent a new phenomenon as they have been a part of education for decades. In that course of time, they have proven as effective in supporting traditional educational technology approaches to instructional design, delivery and facilitation. At the moment, there are many educational simulators in education process at the University of Žilina. There are also several virtual reality simulators there. Since late 2008, there is an important project of virtual reality simulator to be mentioned located in the premises of the department of Water Transport.

Keywords: educational simulations, virtual reality, stereoscopic projection.

1. Introduction

This The University of Žilina as a modern university provides a full range of technological, economic, management, and also a limited range of humanistic and natural science education at under-graduate, graduate and post-graduate levels. During its existence the University has become a reputable institution within the university educational system of the Slovak Republic.

There are several specialized devices and laboratories around the University of Žilina used for educational simulations. A good example is the laboratory of railway transport. Total length of the rails installed in this simulator is around 100 meters. This laboratory is actively used for education for last over ten years and allows the students to come in touch with real railway transport controller devices and consoles installed around Slovak Republic. This simulator is not based on virtual reality. It advantages very high simulation immersivity and good reproduction of reality. Although there are disadvantages such as fixed simulation scenarios and excessive physical space needs for its installation.

2. Simulators in Education

The education evolved addressing new perspectives on pedagogy and the demands of society. The simulation is used quite commonly for educational purposes. It is an important point in education, where the students come in close touch with practice. The technology that is currently being developed in the computer game and simulation domain has sparked much of the interest about that potential. This new technology represents a powerful set of tools for educational technology that can change the way instructional designers create experiences as well as the way instructors facilitate those experiences. Educational simulations are one of the technological approaches available to education to facilitate the building of skills such as critical thinking. Traditionally, simulations are something that bridges the gap between the typical classroom setting and the real world where actual practice occurs. They have been used to assist in the capacity for students to understand and use information to solve problems that are actually relevant to a real context.

A simulation is a model of events, items or processes that do or could exist [12]. There are many different examples of models that are used to convey information. Verbal models present statements about the world while visual models can present graphic representations of abstract concepts. A simulation is another kind of model that is differentiated from the others by its dynamic nature. It represents an operating model of a system. It allows an observer to view not only a single point in time in the model but also how it changes under different parameters [11]. It is not meant to be a complete representation of an event but rather an abstraction that focuses on a specific aspect of that event.

3. Virtual Reality Simulators in the University of Žilina

There are two important virtual reality simulators in the University of Žilina. The first of them - Flight and navigation procedures trainer is held by Air Transport Department located at the Žilina Airport. It's located in university premises in a specially equipped room. The simulation itself is performed by six standard industrial nineteen inch PCs based on Unix and MS-Windows platforms. There are four PCs for video processing, one for mathematical model calculations and one for data I/O operations. Data from pilot console are acquired by I/O cards of this device and passed to mathematical PC. The simulation output consists of three projection screens in the front and two LCD screens in the console. The I/O computer is also used as an instructor console. The simulation is displayed using DLP projectors and LCD screens in 2D projection.

The simulator is in use already for more than seven years. It is certified for pilot training as defined in JAR-STD 3A regulation and it is suitable for IR(A) - device handling qualification, and MCC - multi crew co-operation. The types of airplanes that can be simulated are Beechcraft B200, Piper Seneca 5 and Piper Archer 5. Model cabin of simulator is 95% identical to the cabin of real Beechcraft B200 airplane. This simulator has been delivered as a closed-source project, therefore no further modifications to its functions are allowed by Žilina University staff.

There is a newly installed virtual reality simulator in the premises of the department of Water Transport. This simulator is a gift or better to say a heritage from the state navigation office. The installation in premises of Žilina University was done in late 2008. Nowadays the simulator is in testing and stabilization phase. Unfortunately this is an unfinished project at the moment. The project is dated to the year 1999, when a simulator had been ordered by the state navigation office. The author of the simulator is VUJE Inc. - an engineering company that performs design, supply, implementation, research and training activities, particularly in the field of nuclear and conventional power generation. The project was stopped in year 2003 in its second construction phase out of four. That is the reason, why there are only two simulation scenarios available at the moment - 10km of Danube river for inland navigation in vicinity of Bratislava city and the famous ship-locks in Gabčíkovo. The simulator can simulate only one ship - the tug-boat with the cargo. It simulates the water flow based on average flow in specific profile and on the relief of the bottom. Simulations of wind, fog and icerung are not possible at the moment. The ship-locks in Gabčíkovo scenario can be operated via the instructor console of the simulator. The project of the simulator is functional, but unfinished. The department of Water Transport is trying to find a way how to finance further growth of simulator functionality, how to add more virtual reality scenarios, new maps, new ships.

This simulator after it is finished will be an open project. That means that students, young researches and other third parties will be able to deliver new functionalities to the simulator. For example, the signal processing at the I/O part of the simulator could be upgraded by cooperation with Faculty of Electrical Engineering, the map scenarios could raise in cooperation with the faculty of Civil Engineering. Special emergency scenarios could be issued in cooperation with Faculty of Special Engineering, the simulator engine itself can be further developed in cooperation with Faculty of Management Science and Informatics, the graphical interface could be upgraded or redesigned in cooperation with the Faculty of Science and the mechanical design or a real movement simulation could be brought by cooperation with the Faculty of mechanical engineering.

At the moment, there is a research project running at the department of Water Transport in cooperation VUJE Inc. The department staff together with postgraduate students and third-party companies is taking part in this project in order to add new simulation scenarios, additional functions and 3D stereoscopic projection to the existing simulator engine. It is expected, that the department of the Water Transport would be able to educate not only university students, but also persons concerned from extra university setting.

4. The future of the virtual reality simulators

One of the most important trends in virtual reality projection is the 3D stereoscopic projection. Recently several affordable solutions of 3D projection were issued. Fortunately there's also an interface for obsolete graphical engines (as those used in both mentioned simulators) to be able to project simulation scenarios in 3D stereoscopy. Therefore an upgrade to 3D stereoscopic projection can be expected soon. As both simulators utilize NVidia graphics, there is an easy way for 3D projection upgrade using NVidia Vision 3D solution, or adding secondary projectors and polarized filters.

A dual-projector polarized projection relies on two projectors to deliver video to the screen, while each projector delivers a unique perspective for a specific eye. Each projector lens is attached to its own polarized filter. The viewer has to wear glasses, but the glasses contain no electronic parts and simply use passive polarized filters. The filter over the right eye will block out the polarized video that is intended for the left eye and the filter over the left eye will block out the polarized video intended for the right eye. This way, each eye only sees its intended perspective, even though both perspectives are displayed on the same screen. The most significant advantage of the dual-projector polarized method for stereoscopic 3D are these polarized glasses, that can be purchased for less than an Euro per piece. Considering the downside - the up-front costs are much higher in this case. A dual-projector system requires two independent projectors per each projection screen. On top of that, there's a need for polarized-compatible screen and polarized filters for each projector. The total cost of such setup can be kept just under 3.000 Euro per each projection screen. The cost is on the other hand still comparable to a single 50" 3D-ready monitor, which requires at least one pair of expensive glasses to operate. There are several other disadvantages, too. The polarized filters are never 100% perfect at blocking all of the light from one of the projectors, so if there is something dark beside something bright, an eye might notice light intended for the other eye. This issue is known as the crosstalk. This usually isn't much of an issue, but it can be quite noticeable in some scenarios. Another slight disadvantage of such setup is that most filters polarize the light across a plane, so the more a viewer tilts his head, the more crosstalk he sees.

Another 3D solution is the 3D Vision projection called also The Alternate-Frame Sequencing. Nvidia's 3D Vision solution uses a method called Alternate-Frame Sequencing. Alternate-Frame Sequencing works by alternately displaying a frame of video for each eye. First, a frame of video for the left eye is shown, and then a frame of video for the right eye. This changes back and forth, 120 times each second. The key to making this system work is LCD shutter glasses. These glasses alternatively block each eye at the same frequency (60 times a second for each eye) in order to allow only the intended frame of video to be seen by the targeted eye. A rate of 60 frames of video per second is what is used to be seen on conventional LCD TVs. At this speed, the viewer shouldn't be able to perceive any strobing or flickering. The advantage of a 3D Vision projector over a polarized dual-projection system is the startup cost. It is just a fraction of any other large-screen 3D Vision setup. A good 3D Vision projector can be found for 1.000 Euro including the 3D Vision glasses kit, and only a single projector is required for each projection screen. Some ghosting might be seen on LCD screens in this case if very bright objects are displayed, but with digital light processing (DLP) projectors, the refresh rate is fast enough to prevent this issue. There are still downsides to take into consideration. The main problem with alternate-frame sequencing solutions is the high cost of the glasses. For example, Nvidia's 3D Vision kit, including a single pair of 3D glasses and an IR emitter required to synchronize the glasses to the proper frame of video, is about 200,- Euro. Each extra pair of glasses after that will typically cost 150,- Euro each.

5. Conclusion

New methods in education are addressing current demands of society. Therefore the simulation is nowadays used quite commonly for education as it allows the students to come in close touch with practice. Simulations are used to assist the students to understand and to use information to solve problems that are actually relevant to a real context. The stereoscopic projection is one of the most important trends in virtual reality simulator design. Several affordable solutions of 3D projection were recently issued including interfaces for easy upgrade of older graphical engines. The upgrade to 3D stereoscopic projection can be therefore expected very soon. New technologies in simulator designs represent a set of tools for educational technology that can change the way instructional designers create experiences as well as the way instructors facilitate such experiences.

References

[1] ANDERSON, R. J. Security Engineering: A Guide to Building Dependable Distributed Systems, John Wiley & Sons, Inc., 2001

[2] WEINGART, S. H. Physical Security Devices for Computer Subsystems: A Survey of Attacks and Defenses, in Cryptographic Hardware and Embedded Systems CHES 2000, LNCS 1965, Springer, 2000

[3] ANDERSON R.,KUHN M. Tamper resistance – a cautionary note, The Second USENIX Workshop on Electronic Commerce Proceedings, USENIX Association, Oakland, California, November 18–21, 1996

[4] CHIANG, W.H. 3D-Groundwater Modeling with PMWIN: A Simulation System for Modeling Groundwater Flow and Transport Processe. 2^{nd} ed. Springer, 2008

[5] ALDRICH, C. Learning by doing: A comprehensive guide to simulations, computer games, and pedagogy in e-learning and other educational experiences. San Francisco: Pfeiffer, (2005)

[6] SAETTLER, P. *The evolution of American educational technology.* 3^{rd} ed. Greenwich, Connecticut: L. Erlbaum Associates, (2004)

[7] WANG, Z., CHUI, C., CAI, Y. & ANG, C.. *Multidimensional volume visualization for pc-based microsurgical simulation system*. Paper presented at the SIGGRAPH, Los Angleles, (2004).

[8] Literature resources of Department of Railway Transport and Air Transport Department, University of Žilina

[9] The guide to 3D projection, http://www.nvidia.com

[10] DLP projection survey, http://www.dlp.com

[11] REENBLAT, C.S. Gaming-simulation-rationale, design, and applications: A text with parallel readings for social scientists, educators, and community workers, Sage Publications, 1975, ISBN 0470325003

[12] PAULSEN M.B., FELDMAN K.A. *Taking teaching seriously: Meeting the challenge of instructional improvement*. Graduate School of Education and Human Development, George Washington University (1995)

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

The Analysis of Kinematic Parameters in Biomechanical Systems Using Virtual Instrumentation

*Stanislav Štofan , *Dušan Koniar, *Libor Hargaš, *Miroslav Hrianka
*University of Žilina, Faculty of Electrical Engineering, Department of Mechatronics and Electronics, Univerzitná 8215/1, 01026 Žilina, Slovakia,
{stanislav.stofan, dusan.koniar libor.hargas, miroslav.hrianka}@fel.uniza.sk

Abstract. This paper describes generally the kinematic analysis of biomechanical systems using image analysis. The object of our interest is microscopic biomechanical system represented by respiratory epithelium cilium. Due to this microscopic size, the solution is not based on conventional kinematic analysis using various sensors for kinematic parameters, such as accelerometers and others. Ciliary cells play an important role in the elimination of inhaled particles from the respiratory tract. The effective function of system depends on kinematic parameters of every one cell but also on their interrelationship and synchronization. Our solution for kinematic analysis is based on three processes (image acquisition, image preprocessing and extraction of kinematic parameters). We describe the specifications of all these processes. The one of important kinematic parameter is frequency of object motion. We designed the algorithms for measurement of beating frequency of cilium and also the verification of these algorithms.

Keywords: Ciliary cell, high-speed video acquisition, beat frequency, virtual instrument

1. Introduction

Development in area of information technology (SW's + HW's equipments) brings the requisite to review existing ways and means of resolving problems in the number of technical activities. The achievements are searching for new solutions of problems by improvement and efficiency of existing solutions. Existing medical diagnostic methods are no exception. The image processing has enormous significance in this area. It gets more accurate and more relevant information for medical experts. This will allow a more objective diagnosis, supports their decision and thus minimize the errors.

We focus in our work on the possibilities of using virtual instrumentation for appropriate use of medical experts. These software tools are replacing existing equipment. It allows more flexibility and further adjustment of the required parameters and change functionality. This can be achieved by adjusting algorithms.

Human respiratory tract is exposed to foreign particles in the region. It has also created some of the defense mechanisms how these particles can be excluded. The respiratory mucous membrane is created from glandular epithelium (covering the layer producing mucus) and ciliated epithelium of various types. They move mucus with foreign particles out of the respiratory apparatus. Each epithelium cell contains around 200 cilia. The size of cells is 6 micrometers. They beat with frequency up to 30 Hz. [1]

Periodical movement of cilia consists of two phases. The fast and effective stroke (ESD) is the first phase. The time of this phase is 15% of the movement period. The tip of cilia moves mucus out of the respiratory system during this movement. This movement is in a plane perpendicular to the surface of cells. The second phase is called recovery stroke. The cilium returns to the initial position during this phase. The movement is more slow in a plane inclined to the cell surface. This phase takes a 75% of the all movement period. [1]

The size of ciliary cells is the reason why we can not use the conventional methods for kinematic analysis (based on physical sensors). The kinematic properties of ciliary cell can be use for diagnostic and prediction of respiratory diseases. The tool will allow to investigate the kinematics of biomechanical micro objects by digital image processing. The modern medical applications are also based on very powerful hardware. The important features of these systems must be also high reliable. The computer based system is created to support medical experts in their decision. But final verdict is still on the doctor, not on the machine. [2]

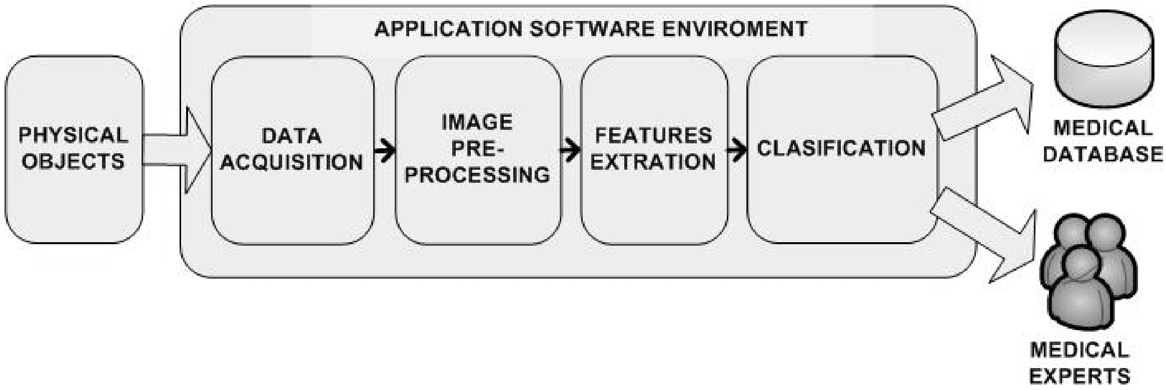


Fig. 1. Simplified scheme of modern medical diagnostic system.

The idea of modern diagnostic systems is composed of several processes (Fig. 1). Similar system can be used for diagnosis of ciliary cells. The data acquisition converts the motion of real objects from microscope linked with camera to digital image. This is first modified by algorithms for digital filtering. It makes image enhancement, noise reduction or contrast adjustment. This step is followed by feature extraction. The group of various algorithms acquires specified properties from data (color, shape, beat frequency, texture). The last step in system is classification. The algorithm for classification is based on set of features typical for the group and thus it helps experts. The results can be saved to database and they will be use in next cases or they could be distributed to other medical department or central database.

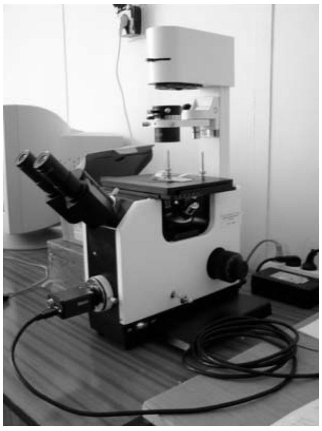
2. Virtual instrumentation in process of kinematic analysis

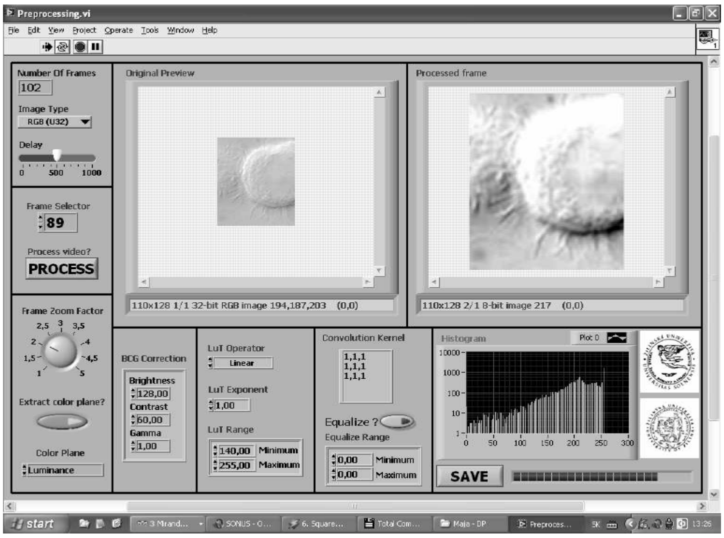
Virtual instrument is a software analogy of real meter with all functions. We use universal measuring PC card and so we can simulate big number of expensive meters. User can freely modify design, controls and indicators of virtual meter. Virtual instrument is characterized with high degree of flexibility. One of development systems for virtual instrumentation is National Instruments LabVIEW. LabVIEW is graphical development system containing many tools and functions for data and signals measurement, analysis and presentation. Vision Development module contains powerful tools of image processing and analysis and standalone application Vision Assistant. Script for each virtual instrument is called Block Diagram, using icons and color data links script is different like text source and suitable for fast and easy editing, debugging or repairing. User graphical interface (called Front Panel) is automatically generated for each Block Diagram. [3][4]

We can create a powerful tool for medical diagnostic support. Using LabVIEW we create a virtual instruments they can store the observed image from microscope´s camera, we can create virtual instruments they can adjust this video sequence or image and also we can create the algorithms which can analyze the kinematic parameters (beat frequency, stability of this frequency, description of trajectory and more).[2][5][6]

3. Video sequence acquisition

The samples of cells come from voluntary donors, but also from patients. They are observed by inverse light microscopy. The resolution of this group of microscopes is 1micrometer. This value is sufficient to examine the kinetic properties of cilia. The electron microscopy has better parameters but it is more expensive and less readily available. It is more suitable for analysis the structure of cells.





a) b)

Fig. 2.a The light microscope linked with the camera

Fig. 2.b The front panel form user application created in LabVIEW environment – preprocessing application

Inverse light microscope, in our case (Fig. 2a), is linked with acquisition camera. The beat frequency of object is important parameter for the choice of camera´s frame rate. In case of observation the ciliary cell, when we estimated the beat frequency up to 30 Hz [2], it is suitable use the high speed camera. Nowadays, high frame rate cameras with high resolution are available. But this brings the big value of data and it requires a special electrical interface to transport high transport rate (Camera link, GigaE, IEEE 1394a,b), the fast and big requires for memory (RAM) and also the sufficient volume of storage medium. The choice of suitable resolution can reduce this volume of data. The volume of data is the reason why we do not make the kinematic analysis in real time. Video sequence analysis is first process step and the next step is the preprocessing of video data. The data are ready for the kinematic analysis now.

RESOLUTION	FRAME RATE	RAW GRAY SCALE BIT RATE
1280x1024	500 fps	5.243 Gbps
640x480	500fps	1.23 Gbps
640x480	120 fps	295 Mbps

Tab.1. The relationship between resolution, frame rate and bit rate

The important part of video acquisition process is using of suitable light source. This source could be provided with sufficiency of optical power and the suitable wavelength of light. We designed the solution with automatic regulation loop and LED lamp as an optical source. The output power of LED lamp is regulated by PWM regulator. The regulator is controlled of main acquisition application in LabVIEW via HW card line. The main application calculates the histogram in actually image frame and then set the optimal value of duty cycle in PWM regulator.

4. Conclusion

Due to fast digital camera, system contains intelligent illumination dimming hardware automatically regulated through measurement card. Regulating parameter for dimmer (PWM duty cycle) is computed from image features, histogram distribution and intensity relations. Dimming helps system to preserve optimal acquisition light conditions for accurate image/sequence processing and eliminates abnormal heat generation in microscope condenser when using high-power lamp. This system can be used in sophisticated measurements in many educational, research and industrial applications where moving objects of investigation can’t be equipped with sensors of kinematics parameters.

Described system for diagnosis should fully support a medical expert in diagnosis for respiratory defects. The kinetic analysis based on image processing is also important tool for medics. It provides numbers of quality information.

It is the next challenge for us to implement the others algorithms for features extraction. Time and spatial correlation relationships between every cell are also important in this area. In the future we want to analyze 2-D correlation map between subimages.

Acknowledgement

Results of these experiments are parts of project Centre of Experimental and Clinical Respirology II (ITMS Code 26220120034) and grant VEGA No. 1/0320/10

„We supported the research activities in Slovakia/Project is co-financed from EU sources“

References

[1] JAVORKA, K. ET AL.: *Medical Physiology*, Martin, Osveta, ISBN 80-8063-023-2 (in Slovak), 2001.

[2] HARGAŠ L., KONIAR D., HRIANKA M., PRÍKOPOVÁ A.: *Kinetics Analysis of Respiratory Epithelium by Virtual Instrumentation*. Sensors & Transducers Journal, pp. 11-18, ISSN 1726- 5479, January 2008.

[3] NI Vision Concepts Manual

[4] JURIŠICA, L. - SUROVÍK, T.: *Vizuálne systémy ako prostriedok pre vyhodnocovanie pohybu mechatronického systému*. In: Acta Mechanica Slovaca, Modelovanie mechanických a mechatronických sústav MMaMS 2008, Červený Kláštor, Slovensko, pp. 375-380, ISSN 1335-2393, 14.- 16. 10. 2008 (2008).

[5] YI W. J., PARK K. S., LEE CH. H., RHEE CH.S., NAM S.W. Directional Disorder of Ciliary Metachronal Wawes Using Two-Dimensional Correlation Map. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*. Vol. 49, No. 3, March 2002.

[6] M.A. CHILVERS, C. O’CALLAGHAN: *Analysis of ciliary beat pattern and beat frequency using digital high speed imaging: comparision with the photomultiplier and photodiode methods*, BMJ Publishing Group Ltd & British Thoracic Society, Thorax 2000, 55:314-37(April).

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Application of the RFID Technology in the Field of Gun Management for the Police of the Czech Republic

*Jiří Švub
*VŠB-TU Ostrava, Faculty of Mining and Geology, Institute of Economics and Control Systems, International RFID Laboratory,
17. listopadu 15/2172, Ostrava-Poruba, 708 33, Czech Republic, {jiri.svub}@vsb.cz

Abstract. This document makes a short report from the ongoing project of application the ultra high frequency (UHF) radiofrequency identification (RFID) technologies into the field of gun management. This project is managed by the ILAB RFID – International RFID Laboratory at the VŠB-Technical University Ostrava. Objectives of the research are to realize essential tests and necessary analyses that are required for the proper design of the newly created gun management system for the Police of the Czech Republic. Current tasks lie in the analysis of the processes that will be automated and in the definition of the optimal parameters of the UHF transponder, its location and the method of attachment to the body of the gun.

Keywords: RFID, tag, gun management, testing, analysis

1. Introduction

The constantly developed identification methods and information technologies of today's world implies more frequent use of the application of radiofrequency identification. Its applications occur in various fields that require very accurate and also user-friendly and automatic identification method.

1.1. Current situation

One of such solutions is a potentially interesting application of the UHF RFID technology in the handguns management systems. Nowadays the registration of firearms is based on registration numbers that are placed on the body of firearms. In the process of identification is the number read and registered by the operator. This operation alone does not take too much time, but the sequential identification of more weapons is already relatively time-demanding. The manual identification also involves a significant risk of human factor errors.

1.2. Intentions of the Project

Identification based on the RFID technology allows us to accelerate significantly this process and also increase the accuracy and reliability, since the possibility to load a wrong registration number is practically eliminated in the correct use of the technology. Guns are marked by a new identification element – the RFID tag – on the body of the handgun. After that we are able to read the registration number of the weapon in a distance of up to several meters, even through some kinds of materials (fabric, plastic, leather, etc.) without direct visibility. All information about the weapon - previously recorded on paper or in static databases - can be transferred into electronic form, enriched with dynamic data acquired from the RFID system while the weapon moves by the read points during its use. These data can then be available online to all authorized users.

2. Current tasks

As this project is still ongoing I had to divide this article into two parts. The first part deals with the tasks that are currently done or are worked on. These tasks are followed by the future plans that could not start before the first ones are finished.

2.1. Analysis of user requirements

We cooperate on this project with one of the divisions of the Police of the Czech Republic – The Intervention Unit of Ostrava. This task was done with the help of the members of this unit. First of all the user requirements where collected. The user requirements consisted for example of the demands about the locations of the identification areas, ergonomic gun tagging and readpoints design. The user requirements have to be analyzed after the collection. The analysis resulted in various important findings.

2.2. Hardware selection

We had to select appropriate hardware platform for this kind of application. The RFID frequency band had to be chosen as well as the specific readers and tags. We had to deal with problems based on the material of the gun bodies that are mainly made of metal. Finally we found a solution of this problem and we selected a few tags that are able to operate on the body of a handgun.

2.3. Location of the tag

Another problem that had to be solved was to find a best location of the tag on the body of the weapon. The project team consisted of laboratory staff and policemen found few possible locations of the tag on the guns body. Each location had its pros and cons. After subsequently performed and analyzed tests we decided for the specific location of the tag that offered best performance parameters. Some of the analyses that have been performed were analysis of the directional characteristics of the tagged weapon, analysis of the influence of the close proximity of two handguns to the readability, test of the influence of placing a handgun in the common plastic box, etc.

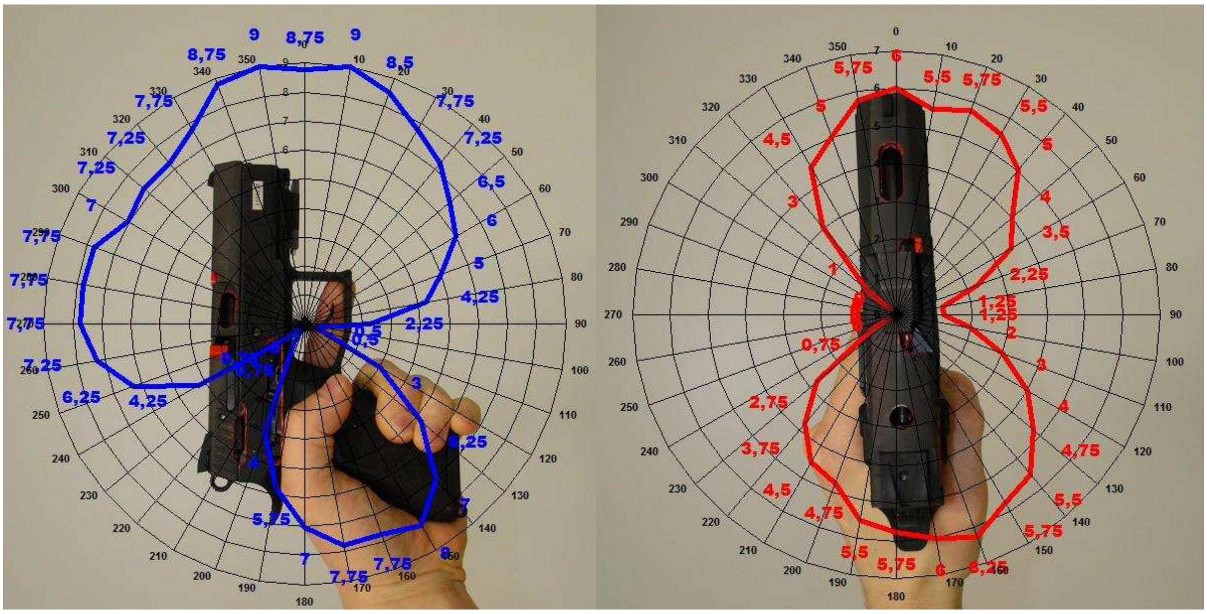


Fig. 1. Directional characteristics of the tagged weapon in the medial and the transversal plane

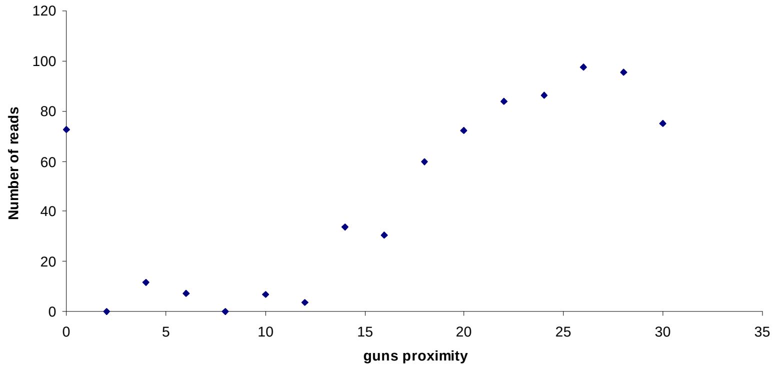


Fig. 2. Influence of the close proximity of two handguns to the readability

3. Subsequent tasks

These days we continue our work by building on the previous tasks. We started to cooperate with external manufacturer of tags that would make a special tag right according to parameters specified by our lab. This tag should be able to have even better performance than already tested serially manufactured tags. After the custom design tag arrives the body of gun has to be mechanically adapted for the final integration of the tag. The tagged weapon will have to undertake laboratory tests to obtain important experimental parameters as well as the tests in vivo in the real environment of the Intervention Unit center.

4. Major benefits of the solution

- Elimination of errors resulting from wrong identification of the gun.
- Reducing the cost of administration associated with registering and evaluating the use of a weapon.
- Contactless identification of a handgun without direct visibility (in the case, in baggage, etc.).
- Simple access to reports about the history of use of identified guns.

5. Conclusion

Although much work has been already done in this project, there are various tasks that have to be accomplished. We believe the whole project will be finished in less then a year and then much more results that are currently excluded of this small report will be published.

References

[1] DOBKIN, Daniel M. *The RF in RFID : Passive UHF RFID in Practice.* Elsevier, 2008.

[2] FETTE, Bruce, et al. *RF & WIRELESS TECHNOLOGIES : know it all.* Elsevier, 2008.

[3] LAHIRI, SANDIP. *RFID Sourcebook.* IBM Press, 2006.

[4] MORADPOUR, S., BHUPTANI, M. *RFID Field Guide: Deploying Radio Frequency Identification Systems.* Prentice Hall. 2005.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Visualization of Large Multivariate Data Sets Using Parallel Coordinates

*Ľuboš Takáč
*University of Žilina, Faculty of Management Science and Informatics, Department of Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, lubos.takac@fri.uniza.sk

Abstract. Data visualization is very useful method of data processing and data mining. This paper deals with an approach of visualization multivariate large data sets using parallel coordinates. In the following text there is detailed description of the process of creating parallel coordinates graph which can be used for exploring large multivariate data sets.

Keywords: data visualization, data mining, data processing, large multivariate data sets, parallel coordinates, computer graphics.

1. Introduction

One of the best ways how to transfer a lot of information from a computer to a human is to visualize it. Imagine the difference between the verbal descriptions of the picture which we have never seen and real visual form of this picture. Everyone will agree that it is better to see it. Also when we are looking at the picture we automatically see essential features which are sometimes not clearly recognizable from non-visual description. As an example check Fig. 1 and try to understand these data.

	A	B	C
1	0,0000	4,9950	4,3400
2	0,0030	4,9950	4,3400
3	0,0060	4,9950	4,3400
4	0,0080	4,9950	4,3400
5	0,0110	4,9950	4,3400
6	0,0140	4,9950	4,3400
7	0,0170	4,9950	4,3400
8	0,0190	4,9950	4,3200
9	0,0220	4,9850	4,3100
10	0,0250	4,9950	4,2800
11	0,0280	4,9950	4,2900
12	0,0310	5,0000	4,3000

Fig. 1. The figure shows the sequences of real numbers.

By reading values from Fig. 1 you cannot easy understand data. Also there is a lot of analytical methods for describing real value data. But by properly selected method of visualization you can see the main character of data. (See Fig. 2)

If you want to explore or understand big data sets, visualization is an appropriate method because of following reasons: you can see dataset globally that means all data in one picture; in a good visualization you do not or poorly see insignificantly elements and significant features are highlighted; Main purpose of visual view of data is to see some significant characteristics or anomalies which can be further examined to gain some additional information about raw data.

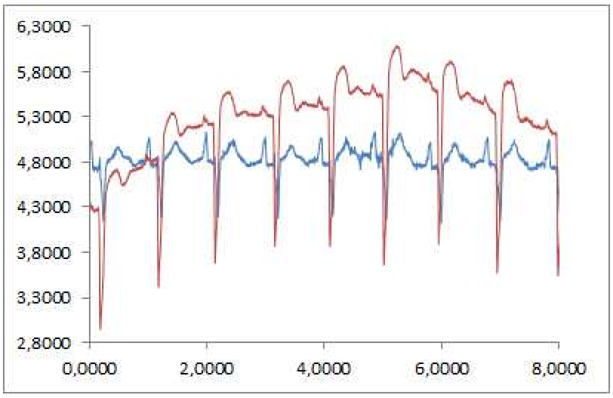


Fig. 2. The figure show simple visualization of dataset from Fig. 1. (An ECG diagram)

2. Parallel Coordinates

Parallel coordinates (PAC), also known as parallel axes, were invented by Maurice d'Ocagne in 1885. Then they were independently rediscovered and popularized by Alfred Inselberg in 1959 [3]. The major advantage of visualizing by PAC is possibility to visualize multivariate data. In basic graph of function we can see at most three dimensions. I prefer only two because of simple construction of graph and better display possibilities. Using PAC we elegantly solve this problem. So what is the difference between PAC and graph of function? In ordinary graph of function there is all axes perpendicular to each other. In PAC are all axes parallel. (See Fig. 3)

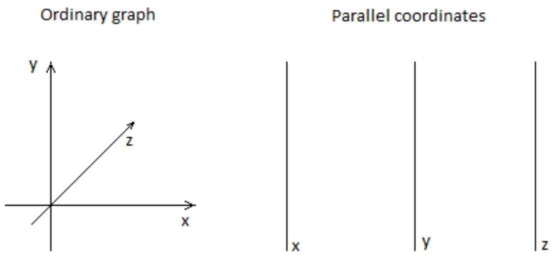


Fig. 3. Main difference between ordinary graph of function and parallel coordinates is in position of axes.

If we would like to visualize multivariate (multidimensional) data using parallel coordinates, we simple add consecutively appropriate number of axes.

Data records are represented in PAC by polylines. For example 2D point a[x=1, y=4] is represented by line which connect axes in corresponding values, see Fig. 4, top left image. It is interesting that point is represented as line and line is represented as point; see Fig. 4, top center image. There are also some another examples of 2D functions on Fig. 4 visualized by PAC.

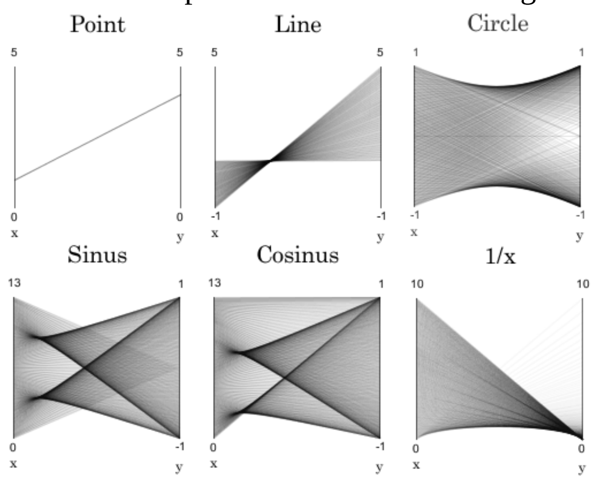


Fig. 4. Examples of some 2D functions visualized using parallel coordinates.

3. Visualizing Large Data Sets using Parallel Coordinates

3.1. How to Visualize using PAC

When we would like to visualize large datasets using PAC we have several problems. Imagine that you would like to display PAC graph on monitor with resolution for example 1024x768 and you have data set containing 10 thousands records roughly uniformly distributed over axes. By simply painting on canvas you do not see anything, only one big blur. (See Fig. 5, left image) Also there is another problem that if you paint a few of records, they may overlap each other, so you can lose some information. Similarly there is difference between overlapping two same records and hundreds and by simple painting you will not see that difference.

These problems can be solved by preprocessing data before visualizing what can be difficult and sometimes it requires knowledge of character of data. Another simpler approach is to paint data records sophisticated by alpha compositing.

Alpha compositing is a computer graphics painting method which use alpha channel to define each color. Alpha channel is something like transparency factor of color. If you paint an object with color which is not opaque, the resulting color will depend on background color too. So using this method to visualize large data set via PAC we avoid mentioned problems in previous paragraphs. (See Fig. 5, right image) Alpha channel can be used also if we would like to positively or negatively discriminate some records. I prefer to optimize alpha parameter interactively when you see visualization.

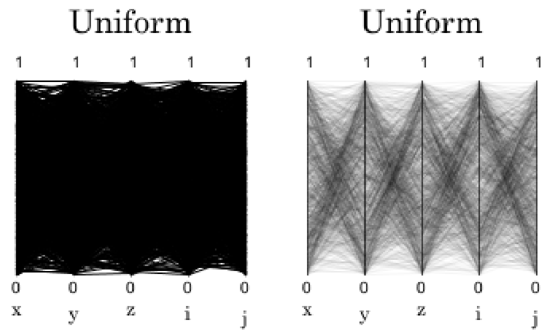


Fig. 5. Visualizing the same randomly generated multivariate data set, by opaque colors (left image) and using alpha compositing technique (right image).

3.2. What to Visualize using PAC

Appropriate data for visualizing using PAC are multivariate data with numerical character. The best is a set of objects where each object is characterized by the same numerical attributes. For example records of people with their age, height, weight, salary etc. By visualizing it using PAC we can track dependencies between attributes if they are, people trends or distributions per attributes. With non-numerical attributes there are problems because it is ambiguous how to sort data or how to group it to appropriate form. It is possible to use this kind of data but a good knowledge of the characteristics of data is necessary to transform it to appropriate numerical form to visualize it using PAC.

3.3. Software Tool

I developed an interactive software tool for visualizing large multivariate data sets using PAC based on mentioned principles. Software tool offers data import from text file, options for changing colors and alpha value, interactive marking or excluding data record based on attributes values, see Fig. 6. Records count is limited to hundreds thousand at resolution of image 1920x1080. Generated images are antialiased and in high quality, so interactive options can be slow when you visualize a great number of records. Software tool is developed in Java using Java 2D API.

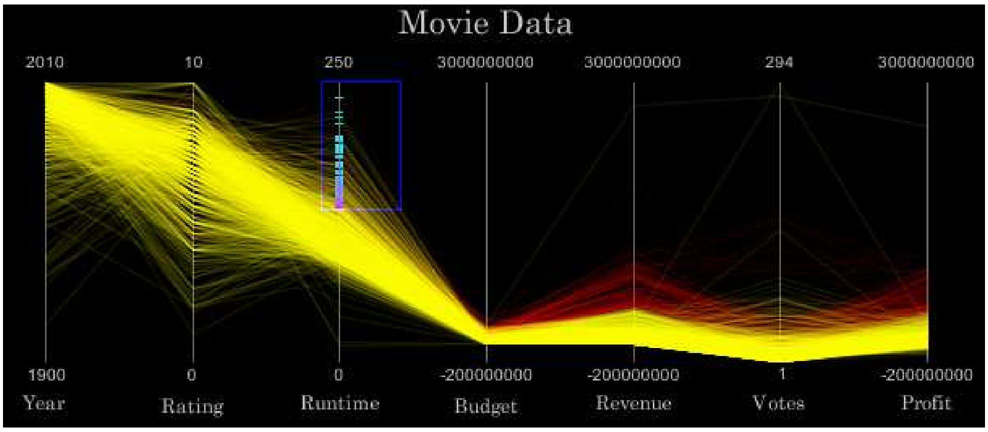


Fig. 6. Demonstration of developed software tool. You can interactively mark or exclude sets by mouse based on attributes values. Visualized data sets come from IMDb (Internet Movie Database).

My plan for the future is to use this tool to analyze multivariate medical data to see or find main characteristics and dependencies between them. There is also a lot of multivariate data from practice for examination and exploration via PAC. All the figures 4-6 were generate by this software tool. If you want to use it do not hesitate to contact me.

4. Conclusion

This paper provides detailed description how to visualize multivariate large data sets using PAC. At the beginning philosophy and importance of data visualization is explained following with description of PAC visualization method with some examples. Next section deals with concrete approach for creating PAC graph for large multivariate data sets and problems associated with it. All mentioned theory is proved by my own developed software tool with some additional functions for selections and manipulation with data sets. Visualizing large multivariate data sets using PAC using mentioned approach or developed software tool can help decision makers and data analyst to gain some added information to do better decision.

References

[1] TAKÁČ, Ľ. *Data processing over very large databases.* Winter School MICT, Šachtičky Slovakia, 3-8 Jan. 2011

[2] INSELBERG, A. *Homepage*. http://www.math.tau.ac.il/~aiisreal

[3] TRICAUD, S., KARA, N., SADDÉ, P. *Visualizing network activity using parallel coordinates.* System Sciences (HICSS), 2011 44th Hawaii International Conference on, vol., no., pp. 1-8, 4-7 Jan 2011

[4] ZÁBOVSKÝ, M., ZÁBOVSKÁ, K., *Big data.* Proceedings of 12th International Conference System Integration 2010, 6-7 Sept. 2010

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Recognition of the European Traffic Signs

*Štefan Toth
*University of Žilina, Faculty of Management Science and Informatics, Department of Software Technologies, Univerzitná 8215/1, 010 26 Žilina, Slovakia, stefan.toth@fri.uniza.sk

Abstract. This article proposes a traffic signs recognition system for the vertical European traffic signs by a Vienna convention. It describes methods and approaches for the traffic signs detection and recognition. Artificial neural networks are used for the recognition as a classifier to recognize shape and type of the traffic signs. After recognition of the sign is finished it is necessary to track the sign while it does not get out from the image. The proposed recognition system for traffic signs is developed to real time recognition, from the image captured by video camera from a moving vehicle.

Keywords: traffic sign recognition, detection, tracking, neural network.

1. Introduction

The traffic signs are one of the most important elements in a transport infrastructure. They have a great importance – govern the right of way, inform and warn road users through their commands, prohibitions and restrictions meaning. They occur in a various shapes and colors to catch eye for the driver's attention.

There are many applications for the traffic signs recognition (TSR). For example, advanced driver assistance system (e.g. warning if the driver exceeds speed limits) [1], or collecting traffic signs with their localization via GPS to databases for navigation systems of maps or for local authorities [2].

The first work in this area can be traced back to the late 1960's, however only in the 1990's, when the idea of autonomous intelligent navigation was popularized, significant advances were made [3].

Since then, many publications on a given subject have been published, in which authors presented many various methods for detection and classification of traffic signs. In many of those, they used classical approaches for detection based on *thresholding and color segmentation* using various *color spaces* (RGB, HSL, HSI, CIECAM97) [4] [5] [6], or *based on shapes* in greyscale image [7], or combination of both [8]. Nunn et al. [9] even used for their work 3D modeling. Next possible approaches use machine learning algorithms, either for detection, classification, or both. Possible algorithms are artificial *neural networks* [10], *support vector machine* (SVM) [11], *boosting* [12]. Other algorithms are *template matching* [13], *genetic algorithm* and so on.

My contribution deals with detection of the traffic signs by using segmentation based on color and recognition of detected signs through the neural networks. The procedure of the proposed TSR system is described in the following chapters.

2. Traffic Signs

According to Marxwell [14], there are two main traffic signs systems – European based on the *Vienna convention* and American based on *MUTCD* (*Manual on Uniform Traffic Control Devices*). This work deals with a small set of the traffic signs used on the European roads, mainly in Slovakia.

3. Traffic Signs Recognition

The procedure of detection and recognition proposed is illustrated at Fig. 1. The procedure stages will be described in details in next subchapters.

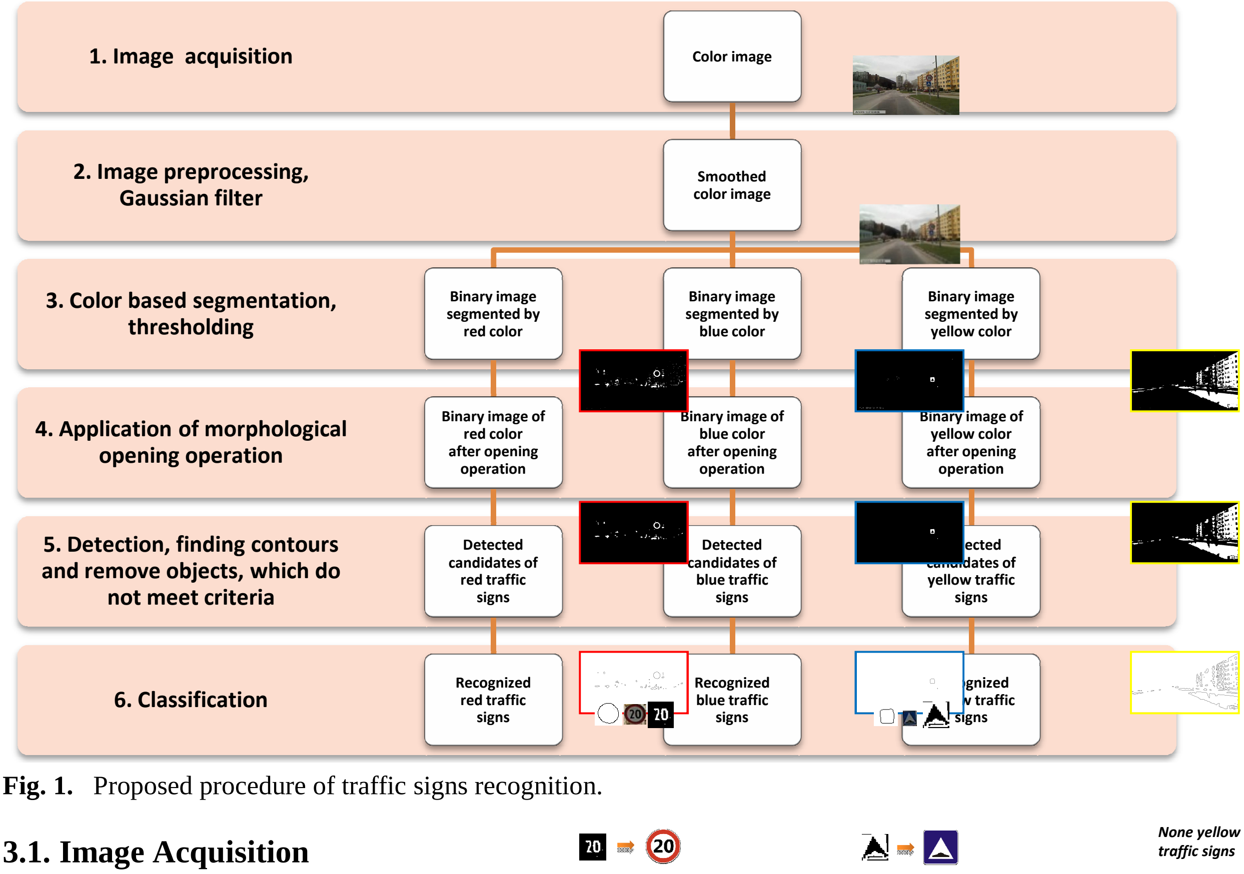


Fig. 1. Proposed procedure of traffic signs recognition.

3.1. Image Acquisition

The color image is acquired by using a video source like video camera or video file. The image is represented by RGB color space. I have chosen the resolution of images acquired from camera to be 640 x 360, which also appears to be sufficient for the real-time recognition.

3.2. Image Preprocessing

The next stage after image acquiring is preprocessing, that means preparation of the image to detection and recognition, for example by removing the noise from the image. It is necessary if the video source is a compressed video file or camera. Here I used smoothing using a Gaussian filter.

3.3. Color Based Segmentation and Thresholding

In order to detect candidates I decided to use a segmentation using color information. The advantage of this approach is low computational complexity unlike the segmentation based on shapes. I converted the image from RGB to CIELAB color space for easier processing. Now from the new image represented by CIELAB model I got simple three binary images using global thresholding by red, blue and yellow color. By means of it the current system can detect three types of traffic signs by color, e.g. red, blue and yellow traffic signs. Other colors like green or brown are not recognized by the current system, but they can be simple implemented later.

3.4. Application of Morphological Opening Operation

In the next stage I used a morphological opening operation. It is a complex operation consisting of basic morphological transformations of erosion and dilation. This operation removes noise and small objects and enlarges bigger objects.

3.5. Detection of Candidates

After processing of all 3 binary images I used *algorithm for finding contours*. This algorithm finds contours of objects on the binary images. Each contour represents continuous area, which has some properties such as height and *width*, *bounding rectangle* of contour, size of the *area* or *perimeter*. In addition *convex hull,* the *smallest circumscribed rectangle* and characteristics like *oblongness*, *squareness, roundness,* and *convexity* [15] can be computed. I used these properties to remove contours, which the most likely seem not to be traffic signs. I tried to set the values of these properties by experimentation. However, there was a problem of finding the balanced values of mentioned properties in order to reject the largest possible number of false detection of signs and vice versa, to accept candidates that are actually traffic signs. After a long analysis and observations I have found a compromise values that meet the parameters for the test sample of different images. Using these methods many false candidates were eliminated, although there are enough false candidates accepted. They will be eliminated by the classification in the next stage.

3.6. Classification of Candidates

The output of the classification for one candidate is a class i.e. recognition of traffic sign. I *used artificial neural networks* (ANN) for classification. I have proposed overall 15 ANN based on color and shape. Type of all used ANNs is a multilayer perceptron with one hidden layer.

Each candidate to be considered traffic sign is initially classified according to shape. The output of this first neural network is the decision, which shape is the candidate (circle, triangle, square etc.) or that “it is not a traffic sign”. If the candidate is accepted, i.e. it has not been classified as "it is not a traffic sign”, the process of classification will be continued by the next neural network to recognize the name of the traffic sign.

4. Tracking of Traffic Signs

The procedure of traffic signs detection and recognition proposed above works only in one still image. The aim of this work is the recognition of traffic signs from a moving vehicle that is from sequence of still images. Therefore it is necessary to track the traffic signs and to determine that one sign in the first image is the same sign in the next image. After the sign gets out from the image the tracking of the sign is finished.

Tracking of the traffic signs was realized by *Lucas-Kanade algorithm*. It compares two consecutive images, the first (previous) at the time t and the second (current) image at the time of $t + \Delta t$. New positions of tracking points in previous image are calculated into new points in the current image.

5. Tests and Results

At the beginning, I had to get the input data for classifiers to learn the neural networks. Therefore it was necessary to drive several routes by vehicle with a web camera placed in front of the windscreen of the vehicle. After the video files were captured I annotated detected the objects on each frame. Thus I collected 25,958 objects include traffic signs and also not traffic signs. Overall I have received 75 different types of traffic signs.

For the testing I have taken new routes, one from Partizánske to Žilina and the second route around the city of Žilina. The results of the test were 70% success detection of the traffic signs. The average processing speed of one frame with tracking was 25 milliseconds (range was from 10 to 50 milliseconds, depending on the complexity of the scene). This time is very good for the real time recognition.

These results of detection are not so sufficient. The most signs that were not detected were blue pedestrian crossings. Detection excludes these signs because they were too small and the camera was very far from them. However, the biggest problem was caused by light conditions, because the recognition uses segmentation based on color. Colors are keys, so it is a very big problem if the signs are faded or have bad colors. Likewise, reflections cause problems if there is too much blur, which in turn occurs in low light conditions. Furthermore, rotation and errors on signs - painted, missing parts or otherwise damaged or obscured by tree branches or other objects.

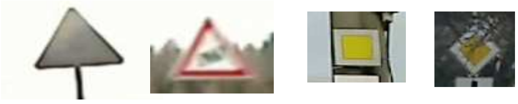


Fig. 2. Examples not recognized traffic signs.

6. Conclusion

The result of my work is an application for traffic signs recognition. The system is currently trained to recognize *75 different types of traffic signs* and can be further extended by learning new signs.

The overall success of traffic signs detection and recognition is *70%*. Success of the neural networks is approximately *85%*. The average processing speed of one image stands at 25 milliseconds that is a satisfactory result for the real time recognition.

I would like to improve the detection by trying another way of segmentation and subsequently selection of better features of candidates to input the classifier.

References

[1] Traffic sign recognition. Wikipedia, the free encyclopedia. [Online]. http://en.wikipedia.org/wiki/Traffic_sign_recognition

[2] TOTH, Š. *Rozpoznávanie dopravných značiek a ich použitie v mapových aplikáciách*, in GIS Ostrava 2011.

[3] RUTA, A., LI., Y. *Towards Real-Time Traffic Sign Recognition by Class-Specific Discriminative Features*, 2007.

[4] PACHECO, L., BATLLE, J., CUFI, X. *A new approach to real time traffic sign recognition based on colour information*, 1994.

[5] BROGGI, A., CERRI, P., MEDICI, P., PORTA, P. P., Ghisio, G. *Real Time Road Signs Recognition*, 2007.

[6] FOLTÁN, S. *Car color recognition from CCTV camera image* in Theoretical and applied aspects of cybernetics : International scientific conference of students and young scientists. Kyjev: Bukrek, 2011, pp. 307-310.

[7] Chiung-Yao Fang, Sei-Wang Chen, Chiou-Shann Fuh *Road-sign detection and tracking*, 2003.

[8] CARDARELLI, E., MEDICI, P., PORTA, P. P., GHISIO, G. *Road signs shapes detection based on Sobel phase analysis*, 2009.

[9] NUNN, C., KUMMERT, A., MULLER-SCHNEIDERS, S. *A novel region of interest selection approach for traffic sign recognition based on 3D modelling*, 2008.

[10] ZHENGHE, S., BO, Z., ZHONGXIANG, Z., MENG, W., ENRONG, M. *Research on Recognition Methods for Traffic Signs*, 2008.

[11] SHI, M., WU, H., FLEYEH, H. *Support vector machines for traffic signs recognition*, 2008.

[12] BAHLMANN, C., ZHU, Y., RAMESH, V., PELLKOFER, M., KOEHLER, T. *A system for traffic sign detection, tracking, and recognition using color, shape, and motion information*, 2005.

[13] WANG, Y., SHI, M., WU, T. *A Method of Fast and Robust for Traffic Sign Recognition*, 2009.

[14] MAXWELL, L. *History of Traffic Signs* in *The human factors of transport signs*. Boca Raton, Florida: CRC Press LLC, 2004.

[15] DOBEŠ, M. *Zpracování obrazu a algoritmy v C#*. Praha: BEN - technická literatúra, 2008.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Subjective Assessment of MPEG Compression Standards Using DSIS Method

*Miroslav Uhrina, *Martin Vaculík
*University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitná 1, 01026 Žilina, Slovakia, {miroslav.uhrina, martin.vaculik}@fel.uniza.sk

Abstract. This article focuses on the subjective evaluation of the well-known compression standards MPEG-2, MPEG-4 Part 2 (AVS) and MPEG-4 Part 10 (AVC/H.264) using DSIS method. In the first part a short characteristic about the MPEG standards is written. In the second part the subjective assessment and DSIS method are described. The last part of this article deals with the measurements and results.

Keywords: MPEG, subjective assessment, DSIS.

1. Introduction

In recent years the field of a multimedia technology has rapidly increased. Many new compression techniques and standards are being developed, most of them are based on the MPEG technology. The compression and the transmission link imperfection are the most common factors that influence the video quality. Due to this reason, the video quality evaluation has become an important role.

2. MPEG compression standards

MPEG-2 is the compression standard which was approved in 1994 and whose video coding scheme is a refinement of MPEG-1. The most important application of MPEG-2 is broadcast digital television, but it also specifies the format of movies and other programs that are distributed on DVDs and similar disks. MPEG-2 is suitable for coding both progressive and interlaced video. MPEG-2 also defines the Profiles and Levels. A profile describes a degree of functionality whereas a Level describes resolution and bitrates. But not all Levels are supported at all Profiles [1], [2], [3].

MPEG-4 Part 2 (Visual) improves on the popular MPEG-2 standard both in terms of compression efficiency and flexibility. It achieves this in two main ways, by making use of more advanced compression algorithms and by providing an extensive set of “tools” for coding digital media. Some of the key features that distinguish MPEG-4 Visual from previous coding standards include: efficient compression of progressive and interlaced video sequences, coding of video objects, support for effective transmission over networks, coding of still “texture”, coding of animated visual objects such as 2D and 3D polygonal meshes, animated faces and animated human bodies, coding for specialist applications such as “studio” quality video [4].

MPEG-4 part 10 (H.264/AVC) is the latest standard, designed for a wide range of applications, ranging from mobile video to HDTV. The key advantages of this standard are: up to 50% bit rate saving, high quality video, error resilience, network friendliness. The new features include smaller block sizes, more flexible prediction both temporally (inter-frame) and spatially (intra-frame), an inloop deblocking filter to reduce the visibility of the characteristic blocking artefacts. MPEG-4 Part 10 also defines Profiles and Levels but its organization is much simpler than in MPEG-4 Part 2 [3], [5].

3. Subjective quality assessment

The subjective assessment consists of the use of human observers who score the video quality. It is the most reliable way how to determine the video quality and cannot be replaced with objective testing but it is time consuming method and human resources are needed. The well-known subjective methods are:

- Double-Stimulus Impairment Scale Method – DSIS,
- Double-Stimulus Continuous Quality-Scale Method – DSCQS,
- Single Stimulus Continuous Quality Evaluation – SSCQE,
- Simultaneous Double Stimulus for Continuous Evaluation – SDSCE.

By evaluation at least 15 observers should be used. They should be non-expert that means that they are not directly concerned with video quality as part of their normal work and are not experienced assessors. The number of assessors needed depends upon the sensitivity and reliability of the test procedure adopted and upon the anticipated size of the effect sought.

The assessors should be at first carefully introduced to:

- the method of assessment,
- the types of impairment or quality factors likely to occur,
- the grading scale,
- the sequence,
- the timing.

The source signal provides the reference picture directly and the input for the system under test. It should be of optimum. The absence of defects in the reference part of the presentation pair is crucial to obtain stable results.

In the figure 1 the presentation structure of test session is shown.

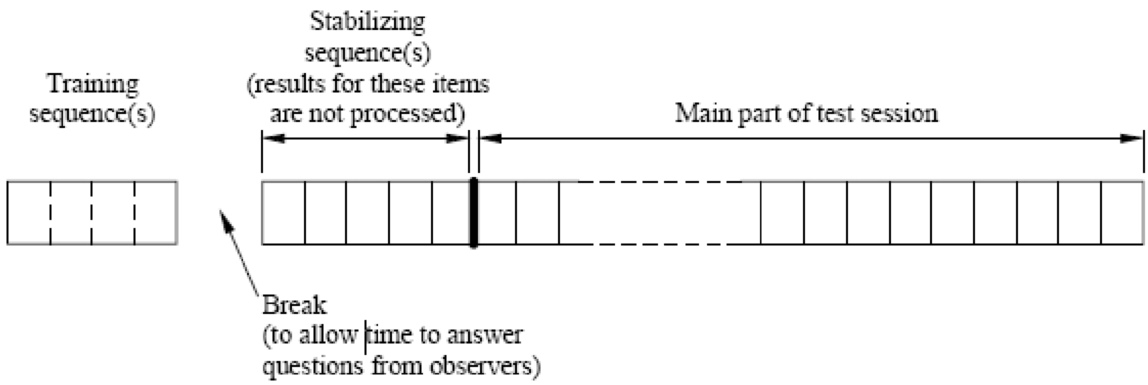


Fig. 1. Presentation structure of test session.

Training sequences demonstrating the range and the type of the impairments to be assessed should be used with illustrating pictures other than those used in the test, but of comparable sensitivity.

A session should last up to half an hour. At the beginning of the first session, some sequences (from three to five) should be introduced to stabilize the observers´ opinion. The data issued from these presentations are not taken into account in the results of the test. A random order should be used for the presentations.

Finally, the calculation of the mean score is performed:

$$\overline{u_{jkr}} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr} \ . \qquad (1)$$

where: u_{ijkrs}: score of observer i for test condition j, sequence k, repetition r,
N: number of observers [6], [7].

3.1. Double-Stimulus Impairment Scale Method – DSIS

By this method the unimpaired (reference) sequence is first presented to the assessor (shows the video material in at the highest quality) and then the same sequence impaired (the test one) as is shown in the figure 2.

Reference sequence	Test sequence	Vote

Fig. 2. Presentation structure of test material.

Following this, the assessor is asked to vote on the second, keeping in mind the first. The five-grade impairment scale should be used:

- 5 imperceptible,
- 4 perceptible, but not annoying,
- 3 slightly annoying,
- 2 annoying,
- 1 very annoying.

In sessions the assessor is presented with a series of sequences in random order and with random impairments covering all required combinations. At the end of the series of sessions, the mean score for each test condition and test picture is calculated [6], [7].

4. Measurements

In our experiments two test sequences were used – one with slow motion (called the “Train” sequence) and one with dynamics scene (called the “Football” sequence). Both sequences were in the resolution 720x576 with 25 fps (frames per second). The length of these sequences were 220 frames, i.e. 8,8 seconds. Both sequences were downloaded in the uncompressed format (.yuv) from [8] and used as the reference sequences. Afterwards they were encoded by different MPEG compression standards (MPEG-2, MPEG-4 Part 2, MPEG-4 Part 10) using the FFMPEG program version SVN-r24872 [9]. The parameters of the encoded sequences were set to Main Profile, Main Level for MPEG-2 standard; Main Profile, Level 3 for MPEG-4 Part 2 standard; Main Profile, Level 3.1 for MPEG-4 Part 10 standard. The target bitrates were in range from 3 Mbit/s to 15 Mbit/s, changed in 2 Mbit/s step. For the subjective evaluation the DSIS method was chosen. 19 assessors (16 men and 3 women) evaluated the test sequences. Their age was in the range from 20 to 71 years, the mean age was 27 years. After the testing the mean opinion score for each sequence, compression standard and bitrate using the formula 1 was calculated.

The figure 3 shows the results of the performed tests of the “Train” sequence and the figure 4 shows the results of the performed tests of the “Football” sequence.

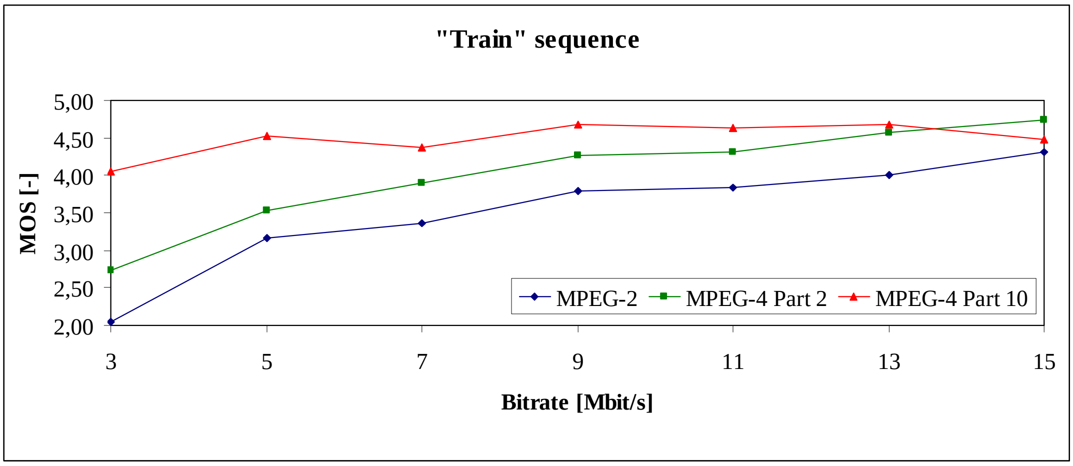


Fig. 3. Results of the “Train” sequence using DSIS method.

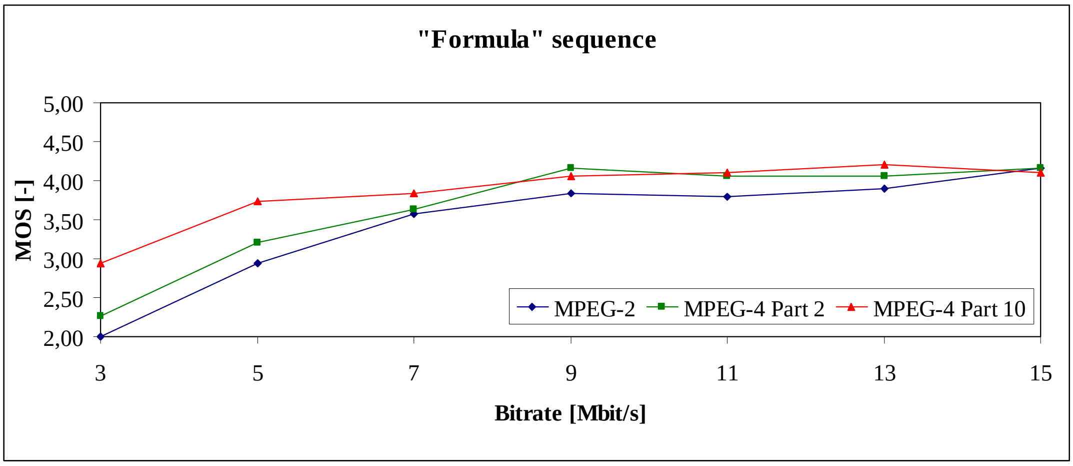


Fig. 4. Results of the “Football” sequence using DSIS method..

5. Conclusion

In this article the subjective evaluation of the well-known MPEG compression standards (MPEG-2, MPEG-4 Part 2 and MPEG-4 Part 10) using the DSIS method was done. According to the graphs, in both sequences H.264/MPEG-4 Part 10 compression standard can be considered as the best one. In the “Train” sequence with slow motion is the difference between the H.264 standard and the rest ones bigger than in the dynamic sequence. It can be due to the fact that assessors perceived the difference of the video quality between the compression standards better in the scene with slow motion.

References

[1] WATKINSON J. *The MPEG Handbook (second edition).* Focal Press, 2004. 435 pages. ISBN 0-240-80578-X.

[2] RICHARDSON E.G. *Video Coding Design*. John Wiley and Sons Ltd., 2002. 299 pages. ISBN 0-470-84783-2.

[3] WINKLER S. *Digital Video Quality: Vision Models and Metrics*. John Wiley and Sons Ltd., 2005. 175 pages. ISBN 0-470-02404-6.

[4] RICHARDSON E.G. *H.264 and MPEG-4 Video Compression*. John Wiley and Sons Ltd., 2003. 281 pages. ISBN 0-470-84837-5.

[5] Scientific Atlanta, *MPEG-4 Part 10 AVC (H.264) video encoding*. Scientific-Atlanta, June 2005. 19 pages. Part Number 7007887 Rev B.

[6] RECOMMENDATION ITU-R BT.500-11, *Methodology for the subjective assessment of the quality of television pictures*.

[7] WU H.R., RAO K.R. *Digital Video Image Quality and Perceptual Coding*. Taylor and Francis Group LLC, 2006. 594 pages. ISBN 0-8247-2770-0.

[8] http://media.xiph.org/vqeg/TestSeqences/ThumbNails/df.

[9] http://www.ffmpeg.org/download.html.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Methods of Multi-criteria Scheduling Jobs in Grid Environment

*Michal Ulbricht
*University of Žilina, Faculty of Management Science and Informatics, Department of Informatics,
Univerzitná 8215/1, 010 26 Žilina, Slovakia, Michal.Ulbricht@fri.uniza.sk

Abstract. In this paper are presented some of methods for multi-criteria scheduling jobs in grid environment. In a commercial sphere a contractor often asks how long does it take to complete the task and how much does it cost. Assuming that faster calculation would be more expensive it is quite logical question. Therefore multi-criteria methods are very needed in grid computing. Some of presented methods belong to the family of heuristics and other to the family of exact methods. There are also presented future plans how to choose suitable approach and where some improvements should be done.

Keywords: grid environment, multi-objective genetic algorithms, multi-objective simulated annealing, multi-criteria mathematical model.

1. Introduction

Problem of solving complex tasks whether in commercial sphere or in the science is where to find computer with sufficient performance. Those tasks usually require large number of processing cycles and involve huge amount of data. Their computation on a single computer would be very time consuming or not possible to be finished in required time. Those tasks that are beyond the processing limits of individual computers are suitable for grid systems. One of the projects using grids is Oxford University's Centre for Computational Drug Discovery's project which utilizes more than one million computers to look for a cancer cure. People around the world donate a few CPU cycles from their computers through the "screensaver time" in order to help analyze 3,5 billion molecules for cancer-fighting potential. Project GIMPS (Great Internet Mersenne Prime Search) focuses to search for Mersenne prime numbers. Thanks to volunteers who use freely available computer software, project has big performance available. Another example might be the project SETI (Search for Extraterrestrial Intelligence) which uses computation potential contributed by users around the world looking for signs of extraterrestrial life by analyzing signals coming from outer space.

2. Grid Systems

Probably the most appropriate definition of grid system can be found in books by authors I. Foster and C. Kesselman - The Grid: Blueprint for a New Computing Infrastructure [1] where they introduce grid computing as hardware and software infrastructure providing reliable, continuous, consistent, pervasive and inexpensive access to computational resources. This definition was in 2002 refined and more key features that must grid meet were added:

- Coordinates resources not subject to centralized management
- Uses standard, open and general protocols and interfaces
- Provides non-trivial quality of services. This means that the quality of services provided by grid is greater than the quality of services that could provide any of the resources individually

There are many options how the grid environment can be designed. A Typical method is when the global scheduler is the highest in the hierarchy and makes decisions based on information from the grid information service, which resource will be allocated to the task. Global scheduler then sends this task to the local scheduler, which decides the way the task will be scheduled in the resource in its autonomous domain. For simplicity of presentation, this environment will be simplified to only one scheduler (queue contains tasks that are waiting to be planned) and a basic set of entities (see Fig. 1.):

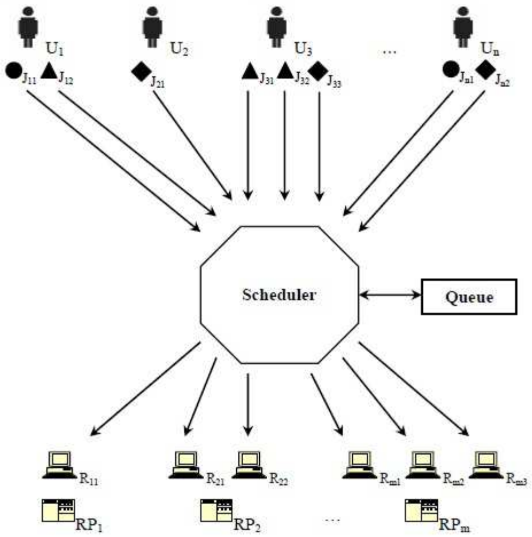


Fig. 1. Grid environment suitable for simpler presentation of problem

- $U = \{U_1, U_2, ... ,U_n\}$ - A set of users who assign jobs.
- $J = \{J_{11}, J_{12}, ... ,J_{ni}\}$ - A set of tasks assigned by users. The task is understood as the basic unit of planning in grid systems.
- $RP = \{RP_1, RP_2, ... ,RP_m\}$ - A set of resource providers who provide resources.
- $R = \{R_1, R_2, ... ,R_j\}$ - A set of available resources. Resource is the physical or virtual component of limited capacity[1] (resource with unlimited capacity does not make sense to plan) which task needs for its processing insight grid. Resources may be renewable or non-renewable depending on whether the task completion is again available to other tasks or not.

3. Multi-criteria scheduling

Scheduling problem occurs in all types of systems where it is necessary to organize and/or distribute the work among the entities of the system. One of the definitions of scheduling introduces M. Pinedo [2]: "Scheduling concerns the allocation of limited resources to tasks over time. It is the decision-making process that has as a goal the optimization of one or more objectives." Scheduling in grid systems can be divided as:

- Priori optimization - In this case, all contractor's criteria and preferences are aggregated into one composite objective function and solution is determined for this one problem as a whole. Result of this optimisation is one solution.
- Posterior optimization - this approach aim to provide the decision maker (contractor) set of Pareto optima, among which belongs the most satisfactory solution. The set of Pareto optima is suggested to the decision maker and he chooses one of them.

An essential part of the scheduling problem is an objective function that allows evaluation of solution profit according to selected preferences (preferences are added by contractor explicitly or preferences are discovered by system based on previous experience).

[1] Capacity of resource can be understood for example as CPU time, the size of free memory, network bandwidth, external devices and so on.

4. Exact methods

The very important part of this chapter is multi-criteria mathematical model. One model is in details described and explained in [3]. This model contains two criteria: completion time of task and execution cost. Every job executed by users (contractors) contains criteria preferences chosen by users themselves. Number of criteria can be easily increased. If the mathematical model is ready then methods of mathematical programming (methods of mathematical programming belong to family of priori optimization methods) should be used to get optimal solution. For this reason were tested some applications and one of the best was freeware Gurobi 3.0.1 [4]. Advantage of using exact methods is obtained optimal solution. Big disadvantage is solving time. To get optimal solution even for small scheduling problem it can take hours or days. Because of this disadvantage are exact methods not suitable for scheduling in grid systems. But they can be used as comparative methods to determine efficiency of much faster methods like heuristics.

5. Heuristics

There are a lot of multi-criteria scheduling algorithms. They can be divided into two classes (see chapter 3):

Algorithms in first class are trying to approximate optimal solution. They achieve it in much shorter time as exact methods. So the advantage is shorter calculation time but optimal solution is not guaranteed. Single objective algorithms provide one solution and use composite objective function. Examples of algorithms are:

- genetic algorithm
- simulated annealing
- tabu search
- hill climbing algorithm
- etc.

In second class algorithms are trying to approximate Pareto curve. Result is a Pareto set of solutions which are not affected of contractor`s preferences (see Fig. 2.). Then one of the solutions is chosen by contractor based on his preferences. Solutions (on Fig. 2.) without squares are approximation of Pareto curve - so they are results of multi-objective algorithm with two criteria (C1 and C2). Other solutions (with squares) are dominated.

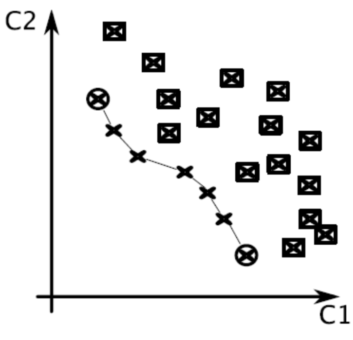


Fig. 2. Approximation of Pareto curve

Examples of multi-objective algorithms are:

- archived multi-objective simulated annealing (AMOSA)
- random weighted genetic algorithm (RWGA)
- improved Pareto evolutionary algorithm (SPEA2)
- niched Pareto genetic algorithm (NPGA)
- Pareto envelope-based selection algorithm (PESA)

- etc.

6. Conclusion and future plans

Exact methods are very important not for scheduling itself but as a tool to determine efficiency of other much faster heuristics. Once the model [3] was used to determine efficiency of SPEA2 [5] algorithm and proves as a useful tool. In future model will be used to determine efficiency of AMOSA. Both of those algorithms belong to family of posterior optimization algorithms. Future plans involve implementation of single-objective versions of those two algorithms. Single objective algorithms belong to family of priori optimization algorithms. Then they will be compared (single and multi objective algorithms) and confronted with optimal solutions obtained via mathematical model. Better approach for scheduling jobs in grid systems will be chosen (priori or posterior) and new algorithm or improvements in existing algorithms will be done.

References

[1] FOSTER, I., KESSELMAN, C. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers, ISBN: 1-55860-475-8, 1998.

[2] PINDEO, M. *Scheduling: Theory, Algorithms and Systems*. Springer, 3-rd edition, ISBN: 978-0-387-78934-7, July 2008.

[3] MURÍN, M., ULBRICHT, M. *Mathematical Model of Multicriteria Scheduling in Grid Environment and its Improvements, 8th Joint Conf. on Math. And Comp. Sci*., July 14-17, Komárno, Slovakia, In press, 2010.

[4] Gurobi optimization, URL: www.gurobi.com

[5] ULBRICHT, M., MURÍN, M. *Mathematical Model for Determination of Efficiency of Multi-Criteria Scheduling Algorithm* (SPEA2), *GCCP 2010: 6th international workshop on grid computing for complex problems*, November 8-10, Bratislava, Slovakia, ISBN 978-80-970145-3-7, 2010.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Contribution to Optimization SQL Query on the Base of Syntax

*Monika Vajsová, *Karol Matiaško,
*University of Zilina, Faculty of Management Sience and Informatics, Department of Informatics,
Univerzitna 2, 01026 Žilina, Slovakia, {Monika.Vajsova, Karol.Matiasko}@fri.uniza.sk

Abstract. Optimization of SQL queries consists not only in using right access methods, normalization of database model or building properly indices. Often is also important the way how is written the SQL query. In case of more complex queries mostly exist more variants how to write SQL query with the same results but with completely changed requirement. This effect we will demonstrate success stories examples.

Keywords: Optimization SQL, Database systems

1. Introduction

This paper aims at Optimization of SQL queries on the base of analyze of actual database query. General SQL query have been created not only in using right access methods, normalization of database model or building properly indices. Often is also important to analyze the way how is written the SQL query. In case of more complex queries mostly exist more variants how to write SQL query with the same results but with completely other cost or time requirements.

2. Syntax controlled processing

In this article we would like to concentrate attention on three useful kinds of preprocessing of SQL query. There are: avoiding unnecessary table, rewriting join statements and utilization an aggregation functions in condition

2.1. Unnecessary table

In data model we can find situation that two tables depends on the same table. If we don't need any data from the master table is useful not create joins in SQL query according the data model, but join detail tables directly and avoid using of master table. In this case we suppose to avoid the relationship which is not necessary for complete and correct result of the SQL query.

2.2. Join statement variants

If we have SQL query where we need to select data only from master table and condition is from detail table, it is useful use master query with subquery instead of one complex query. For the decision we have followed opportunities: use standard join, use standard join with subquery with Exists clause and use standard join with subquery with In clause. From this point of view is very important to identify right way of joining some subqueries.

2.3. Utilization Aggregation functions in condition

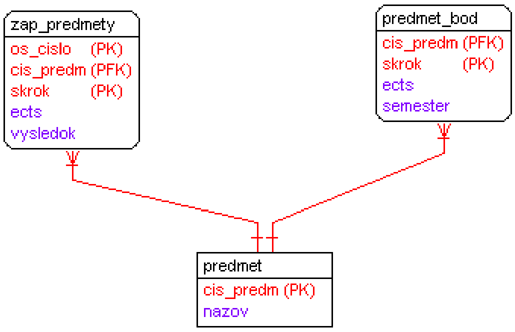
Using aggregation functions in conditions of SQL query has some other variants. In our solution we offer followed possibilities how to solve this problem: use correlated subquery, use condition in having clause , use ALL operator and uncorrelated subquery and use uncorrelated subquery.

3. Solutions

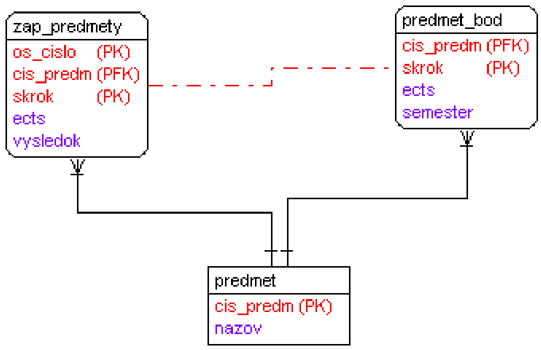
For the solution of described problems we choose some examples from our actual database.

3.1. Unnecessary table

The goal of the query is obtain subjects from the school year 2011, where the number of student’s credits doesn’t correspond to default number credits for this subject. In the variant A we use select that use all tables exactly according the relation ships in the data model. In the variant B we will exclude the table predmet because is not necessary, while both remaining tables contain the foreign key from the table predmet and this reduction will not change the result set of the query.



a)



b)

```
a) select zp.os_cislo, zp.skrok, zp.cis_predm,
zp.ects, pb.ects
from zap_predmety zp, predmet pr,
     predmet_bod pb
where zp.skrok = pb.skrok
  and pr.cis_predm = zp.cis_predm
  and pr.cis_predm = pb.cis_predm
  and zp.skrok = 2011
  and zp.ects <> pb.ects;
```

```
b) select zp.os_cislo, zp.skrok, zp.cis_predm,
          zp.ects, pb.ects
from zap_predmety zp, predmet_bod pb
where zp.skrok = pb.skrok
  and zp.cis_predm = pb.cis_predm
  and zp.skrok = 2011
  and zp.ects <> pb.ects
```

Oracle	**Number of rows**	**Cost**
a)	2284	266
b)	**489**	266
PostgreSQL	**Cost**	**Execution time**
a)	49.89	321.930
b)	**45.06**	**280.425**

Result: As we can see in the both used DBS the variant B is more efficient don’t use the unnecessary table. In this case play important role also the existing indexes for column cis_predm in both of detail tables. (This can be seen on the detail query execution plan.)

3.2. Join statement variants

For solving of this problem we prepared the task where is necessary to create the list of students that studied at least one subject in school year 2010. This task is more complex and can be resolved with and without subquery, because we need to see data from one table and the condition is in second table. Variant A is the standard approach using inner join. Variant B is with subquery Exists but we have forgotten the table zap_predmety in the main select. This often occurring mistake is repaired in the variant C. And the last variant is with using of subquery In.

a)
```
select st.os_cislo, st.rocnik
from student st join zap_predmety zp on
( st.os_cislo = zp.os_cislo)
where skrok = 2010;
```

b)
```
select st.os_cislo, st.rocnik
from student st
where exists
( select 'x' from zap_predmety zp
where st.os_cislo = zp.os_cislo
and skrok = 2010);
```

c)
```
select distinct st.os_cislo, st.rocnik
from student st join zap_predmety zp on
(st.os_cislo = zp.os_cislo)
where exists
( select 'x' from zap_predmety zp
where st.os_cislo = zp.os_cislo
and skrok = 2010);
```

d)
```
select st.os_cislo, st.rocnik
from student st
where os_cislo in
( select os_cislo from zap_predmety zp
where skrok = 2010);
```

Oracle	**Number of rows**	**Costs**
a.	10502	295
b.	14735	1009
c. EXISTS	**3320**	**295**
d. IN	**3320**	**295**
PostgreSQL	**Time**	**Costs**
a.	52.74	760.177
b.	22140.89	18829.607
c. EXISTS	101.48	4173.695
d. IN	**32.54**	**278.536**

Result: On this example we can see also the difference between DBS Oracle and Postgres. In both DBS is the using of right subquery better than normal Join in main select. In DBS Oracle are Exists and In equal, however in DBS Postgres using In clause lead to lower costs.

3.3. Utilization Aggregation functions in condition

On this example (Repeated study of subject) we will show the possibilities how to use aggregation function for search duplicities. Variant A is written in non effective way because the same table zp is used in main query and also in the subquery with aggregation function. Variant B is standard way how use aggregation function in the having clause. In this case is not necessary to use subquery. Variant C is correct way how to use subquery without using table zp in the main query and using operator ALL. In the last variant D is rewritten operator ALL to using aggregate function max. In DBS Oracle we can use nested aggregate functions in select, if you don't select anything else except this function. In DBS Postgres this variant must be written using nested selects.

```
a) select distinct o.meno, o.priezvisko
from os_udaje o join , student s,
zap_predmety z
where o.rod_cislo = s.rod_cislo
and s.os_cislo = z.os_cislo
and 1<(select count(*)
    from zap_predmety zp
    where z.os_cislo = zp.os_cislo
       and z.cis_predm = zp.cis_predm);
```

```
b) select distinct o.meno, o.priezvisko
from os_udaje o join student s using
(rod_cislo) join zap_predmety zp using
(os_cislo)
group by os_cislo, zp.cis_predm,
            o.meno, o.priezvisko
having count(*) >1;
```

```
c) select distinct o.meno, o.priezvisko
from os_udaje o, student s
where o.rod_cislo = s.rod_cislo
and 1< ALL
   ( select count(*) from zap_predmety z
     where z.os_cislo = s.os_cislo
     group by z.os_cislo, z.cis_predm);
```

d) *Postgres*

```
select distinct o.meno, o.priezvisko
from os_udaje o, student s
where o.rod_cislo = s.rod_cislo and
1< ( select max(tab.pocet)
      from ( select count(*) as pocet
                from zap_predmety z
                where z.os_cislo = s.os_cislo
               group by z.os_cislo, z.cis_predm )
          tab   );
```

d) *Oracle*

```
select distinct o.meno, o.priezvisko
from os_udaje o, student s
where o.rod_cislo = s.rod_cislo
and 1< ( select max(count(*))
             from zap_predmety z
             where z.os_cislo = s.os_cislo
             group by z.os_cislo, z.cis_predm);
```

Oracle	**Costs**	**Time**
a)	518K	01:43:44
b) Having	**1267**	00:00:16
c) All	10583	00:02:07
d) Max	10501	00:02:07
PostgreSQL		
a)	7324.859	2262
b) Having	5862.135	2262
c) All	**324.261**	*2689*
d) Max	3482.311	2262

Result: On this example is evident that using having leads to best execution time. Interesting is that despite of lower costs of variant c in DBS Postgres, the execution time is higher. Variant A is the less suitable way how to write the query as we have been expecting.

4. Conclusion

In this paper we have described some opportunities how to solve the SQL query simplifying or rewriting SQL statement. We have evaluated all suggested solutions in DB system Oracle and Postgres for ability to compare results in different database engines with very interesting results. We can notice that the same solution has very often different optimal result from the point of view execution time or costs of queries. From this reason is possible to recommend to each database designer to test more variants of solution directly under DB engine to find the best one.

References

[1] MATIAŠKO, K. a kol. *Databázové systémy a technológie*. STU Bratislava, 2009.

[2] BERNSTEIN, P. – NEWCOMER, E.: *Principles of Transaction Processing*, Morgan Kaufman, San Francisco, 1997.

[3] ATZENI, P.- CERI, S.- PARABOSCHI, S.- TORLONE, R.: *Database Systems – concepts, languages & architectures*, McGraw-Hill, England, 1999.

[4] ULLMAN, J.D. at all : *Database Systems The complete book*, Prentice Hall, 2006.

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Data Warehouse for Breeding Enterprise

*Eva Valentíková
*University of Žilina, Faculty of Management Science and Informatics, Department of Informatics, Univerzitná 8215/1, 01026 Žilina, Slovakia, Eva.Valentikova@fri.uniza.sk

Abstract. Processing of large databases and the ability to process a great amount of data are nowadays one of the fundamental demands of various firms and organizations. There is a growing interest in Business Intelligence tools that are used for building systems, which assist people with decision making and business management. This way the need for creating multidimensional databases becomes widely recognized. The main aim of this paper is to show effective design and usage of analytical databases.

Keywords: data warehouse, data transformation, integration services, DW architectures.

1. Introduction

Existing database systems are particularly appropriate in terms of consistency of data retention and achieve very good performance especially in transactional operations such insert, delete or update data in OLTP (online transactional processing) databases in real time. Usually data are located in many different sources and this is a problem for big analyzes taking data from more than one local system. We can say that big disadvantage of transactional systems is the scattered data in different heterogeneous systems unsuitable for time saving and efficient creation of complex analysis. In case when we need to make complex analysis over more than one system, we need to use another approach instead of the transactional one. One of the solutions of the problem described above is to design data warehouses specific for complex analysis. There are many problems in process of data warehouse design, which need to be resolved and will be discuss later in this paper.

2. BASIC DEFINITIONS

Data warehouse (DW) is database serves as the main storage space for historical data of the company. One of the base tasks of DW is the selection of data from transactional databases (or other sources) and their storage in central multidimensional database. Data stored in DW are accessible in case of temporary failure of one of the original sources and analyzes can be performed in every time irrespective to state of the OLTP sources. Main purpose of DW is to relieve OLTP from complex analytical queries for decision support and data mining. The main structure of the DW consists from few dimensions tables and one or more fact tables. There are two basic structures of data warehouse design according to count of fact table [3, 5]: *star schema*, where only fact table make relationship between all dimensions; *snow flake schema*, where dimensions are connected by more fact tables. The base definition of DW is from the "father of data warehousing" W.H. Inmon [3] and he says that: "A data warehouse is a Subject Oriented, Integrated, Non-volatile, and Time-variant collection of data in support of management's decisions." Mathematically we can say, DW schema (DWS) is a pair of dimensions D and fact relations F: DWS = (D, F), where $D = \{D1, D2,..., Dn\}$ is a set of dimensions and $F = \{F1, F2, ..., Fn\}$ is set of fact relations. As shown in figure Fig.1. the relationship between dimensions and fact relation is 1:M. Simultaneously Fig.1. shows designed DWS for breeding enterprise needs.

Data mart (DM) is a kind of data warehouse with small range. It contains information useful to the different departments of a company. Usually data marts are designed for smaller companies and their departments, or for the departments of bigger companies. Data are highly targeted there therefore installation time and costs may be considerably less in compare with initialization cost of DW.

Staging area (SA) is a kind of temporary data location between the data sources and DM/DW. Data from sources are placed here and operation such cleaning and transformation of data are performed. Cleaned and transform data are then transported to DW (or DM) prepared for the next processing.

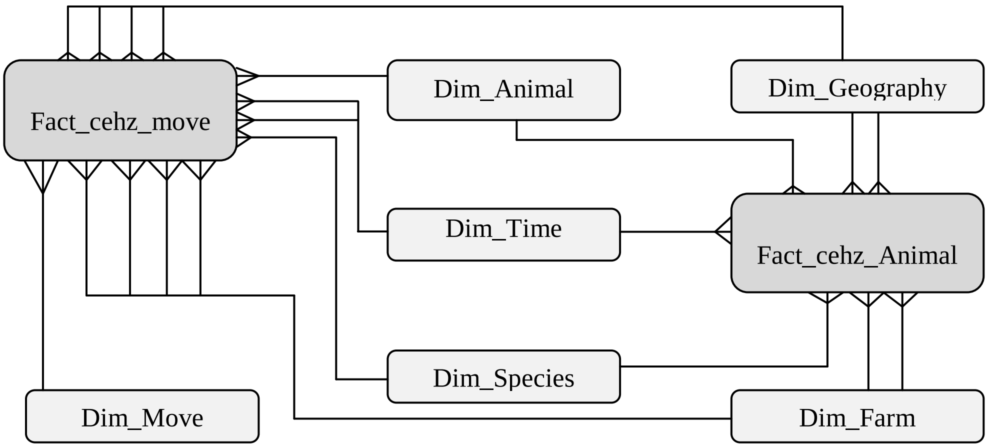


Fig. 1. Star schema for breeding data warehouse with two fact tables

3. PROBLEM SPECIFICATION

Since there are several interesting example of building the data warehouse environment, I would like to limit discussion of this paper to DW design aimed at breeding enterprise of Slovak Republic. Breeding enterprise stores data about central register of animals in Slovak Republic and provides complex analysis that can help in mapping and localizing infections and diseases. Monitoring of animal movements may help in allocation of risk pool of animals and appropriate actions may be taken as measure. These data may also be used for statistical and scientific purposes in other breeding or statistical organizations. The main requirement was primarily to reduce the burden on breeding workers and the streamlining of extensive analysis. Among the desired outcomes were included the following reports: number of livestock in various regions; number of animal births in each region (or farm) for a certain period; report of state of farms and the diversity of species there; ability to track the movements of animals between farms and regions; monitoring the animal movements reasons between farms; and many other analysis of a similar nature. An important task was also to detect the weaknesses and deficiencies of input data.

3.1. Data placement

Data entering the analysis are stored in Oracle 9.2 and MySQL database systems. Many historical data exists in .dbf files. Data stored in primary sources contains many duplicate and zero values, therefore analysis and their modification are required. This process is usually called data cleaning or data scrubbing. Given that the existing database captures data for more than 30 years, it is necessary to think about the differences in methodology of preparation and processing of data in different time periods.

4. Main tasks in DW design for breeding enterprise

Data warehouse design involves many particularly issues like environment selection, data transformation process (ETL process), dimensions and fact table design, hierarchy design [6] and so on. The selection of views to materialize is considered as one of the most important tasks of the DW design. Interesting work involving the view selection problem can be found in [1, 2, 4]. In order to limit the number of pages, only brief preview of breeding data warehouse design is given. To build data warehouse for breeding enterprise MS SQL Server environment and SSIS (SQL Server Integration Services) has been chosen. As mentioned before, the original data for breeding enterprise were stored in two different database systems and there exist few .mdf files containing historical data. Since it was not possible to access all data for testing because of security reasons, staging area (called CEHZ) in form of transactional database has been created. This database contains consolidated data from all existing systems and files. Schema of the staging area has been designed in compliance with original databases. Breeding enterprise DW has been designed as star schema with two fact tables and seven dimensions (Fig. 1.). Two fact tables were selected because of the required analysis nature, which can be divided into two different areas: 1) analysis about animal movement, 2) analysis about animal deployment and character. An interesting proposal in designed fact tables is that they contains only foreign keys attributes from designed dimensions and no facts can be found there. Such tables are called fact-less tables [5]. Although they do not contain numerical unit, it is still possible to perform operation such as COUNT (number of rows), SUM (amount), MIN (minimum), MAX (maximum) called also aggregated functions. In some cases, column with value 1 is added to fact-less table. Basically, this column does not add any information to fact table, but allows create more readable SQL queries.

5. MEASUREMENTS OF QUERY EXECUTIONS

Since there was not possibility to access original sources, query executions measures were performed with staging database CEHZ and with designed data warehouse CEHZ_DW. Consider a report where we need to know how many animals were moved to which district and which year.

$$Q_{CEHZ} = {}_{L}\gamma_{G}\left(\pi_{A}\left(t_1 \bowtie t_2 \bowtie t_3 ⟕ t_4 ⟕ t_5 ⟕ t_6 \bowtie t_{4_1} \bowtie t_{5_1} \bowtie t_{6_1} \bowtie t_{4_2} \bowtie t_{5_2} \bowtie t_{6_2} \bowtie t_7 \bowtie t_8\right)\right)$$

$$\text{where}$$

$$L = YEAR(t_3.date\_move), t_7.district\_name, t_8.species\_name \tag{1}$$

$$G = count(*), sum(\theta_1), sum(\theta_2)$$

$$A = t_3.date\_move, t_7.district\_name, t_8.species\_name, t_{6_1}.district\_id, t_{6_2}.district\_id$$

To each year, district and animal species we need to know additional information like how many animals come from their born district, how many come from another district and how many animals come from unknown (we have no information in database) district. We need to use 14 relations to perform mentioned query, while physically there are stored only eight of them. This means that some of the relations are used multiple times but with different meanings. It is need to use 10 joins and 3 outer joins to link the relations. Also aggregate functions as sum and count are needed, what means using the group by clause in query. Described query for CEHZ may be represented by extended relational algebra as (1), where t_i is i-th table used in query, $⟕$ sign represents left outer join, $\bowtie$ is sign of natural join, π means projection of columns, ${}_{L}\gamma_{G}$ is sign of expanded relational algebra represented GROUP BY clause in the query, where L is the set of grouping attributes and G represents the set of aggregate functions. Same query for the designed data warehouse with the same results (returning the same rows) may than looks like (2), where d_i represents dimensions of data warehouse and f_i represents fact tables.

Note that all dimensions d_i are in many-to-many relationship by fact table f_i. We can say, that Q_{CEHZ_DW} execution take only half of execution time in compare with execution time for Q_{CEHZ}.

$$Q_{CEHZ\_DW} = {}_{L}\gamma_{G}\left(\pi_{A}\left(f_1 \bowtie d_1 \bowtie d_2 \, o\!\bowtie d_3 \bowtie d_{3_1} \bowtie d_{3_2} \bowtie d_4 \bowtie d_5\right)\right)$$

$$\textit{where} \tag{2}$$

$$L = d_5.calendar\_year, d_{3_2}.district\_name, d_4.species\_name$$

$$G = count(*), sum(\theta_1), sum(\theta_2)$$

$$A = d_5.calendar\_year, d_{3_2}.district\_name, d_4.species\_name, d_3.district\_id, d_{3_1}.district\_id$$

Beside queries mentioned above there were tested many other queries and we can divide them to the next groups: **According query run-time:** *Short (TS):* less than 10 sec.; *Medium (TM):* 10 to 30 sec.; *Long-running (TL):* more than 30 sec. **According resulting dataset size:** *Small (SS):* less than 1000 values; *Medium (SM):* 1000 to 100 000 values; *Large (SL):* more than 100 000, up to several million values. The results are shown at *Fig.2.*.

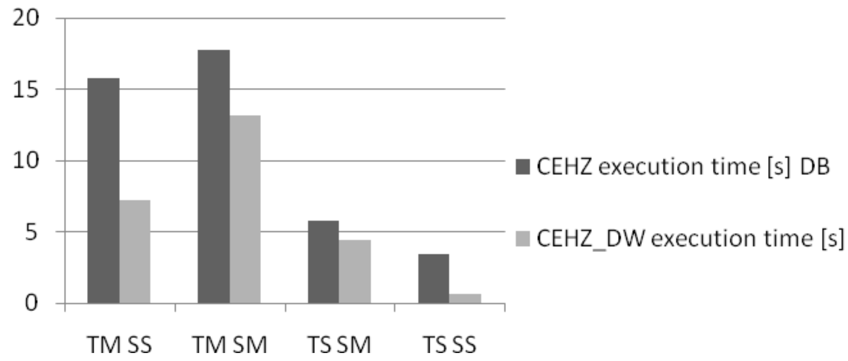


Fig. 2. Execution time for different query types

6. CONCLUSION

The results suggest that the data warehouse building is efficient for queries with medium to long run-times, which are returning small to medium datasets. These kinds of queries are very common in analytical data processing using aggregation, performed on data warehouses. It is irrelevant to produce queries returning big data sets when we want to make decisions build on historical data. Usually we want to summarize big data sets to smaller, readable and more illustrative data set. Of course, not for all queries have to be data warehouse advantageous. In ideal case, we know most of the required analyses and the data warehouse designer propose fact relations and dimensions for them. There are many tools for better performance of both transactional database and data warehouse, like are for example indexes, parallel or distributed databases.

References

[1] Bauer A., Wolfgang L., On Solving the View Selection Problem in Distributed Data Warehouse Architectures. *15th International Conference on Scientific and Statistical Database Management,* 2003 ISBN: 1099-3371/03

[2] Bauer A., Wolfgang L., On Solving the View Selection Problem in Distributed Data Warehouse Architectures. *15th International Conference on Scientific and Statistical Database Management,* 2003 ISBN: 1099-3371/03

[3] Inmon W.H., Building the Data Warehouse. *John Wiley & Sons, 2002 (3th edition) ISBN: 0-471-08130-2*

[4] Jin-Hyuk Y., In-Jeong Ch., Materialized View Selection in Data Warehouse, *International Journal of Information Processing Systems,* June 2006, ISSN: 1738-8899

[5] Ralph Kimball, Margy Ross; The Data Warehouse Toolkit Second Edition, The Complete Guide to Dimensional Modeling, ISBN 0-471-20024-7

[6] Defining and Configuring Dimensions, Attributes, and Hierarchies http://msdn.microsoft.com/cs-cz/library/ms174537(en-us).aspx (13.3.2011)

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

The Influence of the Dish Efficiency on the Quality of Satellite Signal

*Jacek Łukasz Wilk
*The Kielce University of Technology, Faculty of Electrical Engineering, Automatics and Computer Science, Al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland, jwilk@tu.kielce.pl

Abstract. One of the parameters to characterize the properties of satellite communication system is a dish efficiency, which relates to the power delivered to the antenna and the power radiated or dissipated within the antenna. A high efficiency antenna has most of the power supplied at the antenna's input radiated away. A low efficiency antenna absorbes most of the power as losses within it. This paper presents the influence of the antenna’s receiver efficiency on the quality of a satellite signal Polonia channel in Kielce city. TV Polonia broadcasts many of TVP’s domestic programs as well as local and national news from Polish communities all over the world.

Keywords: antenna efficiency, antenna gain, satellite system, SNG, VSAT.

1. Parameters

A very important parameter for measuring the performance of a satellite link is the satellite signal location – longitude, latitude of the selected location and its altitude above sea level. This is directly from the „visibility” of the satellite signal. As a place to pick up the signal has been chosen Kielce city. Latitude and longitude for the city are 20.62E and 50.87N, respectively. The city is characterized by significant differences in levels between 260 and 408 m above sea level. Therefore, the average altitude was 300 m above sea level.

1.1. Polarisation angle

In contrast to land-based antennas, the satellite antennas also requires a setting in the vertical plane. As a measure usually adopted in setting an azimuth – an angle between the direction of the south and the direction of receiving antenna in the horizontal and the elevation – an inclination of the angle in a vertical plane of the antenna. Therefore, knowledge of the geographical coordinates of a satellite dish installation is necessary to determine the appropriate measure of the antenna. SMW Link software v. 2.0 was used to determine the polarisation angle:

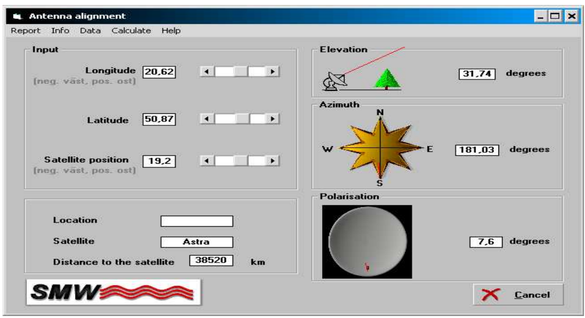


Fig. 1. Determination of the polarisation angle for Kielce city.

The polarisation angle for Kielce city of to receive programs from satellite Astra1KR is 7.6 °.

1.2. Antenna aperture

Antenna aperture can be visualised as the area of a circle constructed broadside to incoming radiation where all radiation passing within the circle is delivered by the antenna to a matched load. Generally, antenna gain is increased by directing radiation in a single direction, while necessarily reducing it in all other directions since power cannot be created by the antenna. Thus a larger aperture produces a higher gain and narrower beamwidth. Ground antennas and the antennas are usually a bundle of point and tend to be Very Small Aperture Terminal (VSAT) and the Satellite News Gathering (SNG). The size of the receiving antenna is selected according to the received signal power, which in turn depends on the distance the lower power signal and vice versa. This is due to the fact that a large-diameter antenna is characterized by a small angular width of the main beam, so that undesirable levels of radiation around the antenna are minimal. Antenna gain is directly proportional to aperture. Gain of the large-diameter antenna is thus very large, and thus the power transmitted may be small to achieve a given link budget. The inaccuracy of setting the main axis of the antenna beam towards the geostationary orbit should not cause a decrease in antenna gain by more than 1dB for any frequency. In Poland, it is recommended to use antennas with a diameter of at least 60 cm. The size of antenna was used 0.80x0.90 m (Model Triax 800 DAP).

1.3. Losses

Noise temperature of antenna depends mainly on the noise of the sky (the absorption of atmospheric and galactic noise) and the noise coming from the Earth (system noise, man-made noise, etc.). Satmaster + can automatically designate the antenna system noise. Satellite dish should not be permanently deformed, as well as it should not require readjusting when exposed to wind velocity of up 130 km/h. The antenna should maintain the direction of the satellite with an accuracy of 20% of the width of the beam, which corresponds to a reduction in the received signal to within ± 0.5 dB – the antenna beamwidth is defined for the decrease in the 3dB signal power. The main antenna axis deviation from the desired direction can be a result of wind velocity between 130 km/h and 160 km/h. Wind velocity up to 160 km/h should not cause permanent deterioration of antenna parameters: a reduction of gain, deformation characteristics of the directional signal or the increase of the attenuation of the orthogonal polarization signal. Satellite dish also should not be permanently damaged or deformed as a result of heavy rainfall or snow, and the crust of ice thickness up to 25mm. It was assumed that that the coupling loss of a satellite system was approximately 0.4 dB, whereas the antenna mispointing loss (including as a result of poor stability dish antenna) was 0.3 dB. Usually these parameters can be determined empirically. Noise interference between the antenna beams is considered the most serious source of interference in multibeam satellite systems. Interference effect appears as an increase in thermal noise. It is estimated that the multibeam satellite systems noise interference is 40% of total noise. The parameters affecting the interference are the effective isotropic radiated power (EIRP), frequency and location of the neighboring broadcasting satellite. The study incorporates impact of ASTRA 2A satellite (EIRP = 51dBW, bandwidth equals 33 MHz, position 28.2°E) and ASTRA 1E satellite (EIRP = 51dBW, bandwidth equals 26MHz, position 23.5°E). It was assumed a rain model in accordance with ITU-R.

2. Measurements

The signal of the TV Polonia channel is emitted via the Astra 1KR satellite. Weather conditions (including the suppression of atmospheric gases, rain, water vapor, as well as the impact of clouds and mists) may result in a temporary unavailability of the TV channel. Attenuation caused by atmospheric gases depends mainly on the length of the route which the electromagnetic wave is present in the atmosphere – is the result of wave absorption by oxygen and water vapor, as well as nitrogen (especially for wave frequencies above 100 GHz). Usually only oxygen and water molecules cause the effect of carrier suppression, while the other gases contained in the atmosphere may cause the attenuation of electromagnetic wave in a very dry air and above the frequency of 70 GHz. Therefore, the division shall be made to the suppression by dry air (oxygen and nitrogen) and water vapor. For frequencies below 10 GHz attenuation of water vapor and dry air is fairly low and therefore often overlooked. Losses are a result of absorption, displacement currents, as well as through the dispersal of the wave on the water droplets. Attenuation is caused by increase in rain for higher frequencies over long distances of a satellite link. In practice, the horizontally polarized waves are more suppressed than waves of vertical polarization.

Increase in the antenna efficiency in non-rainy and rainy weather between 60 and 73% causes the following results:

Increase in the antenna efficiency affects on the increase in the antenna gain.

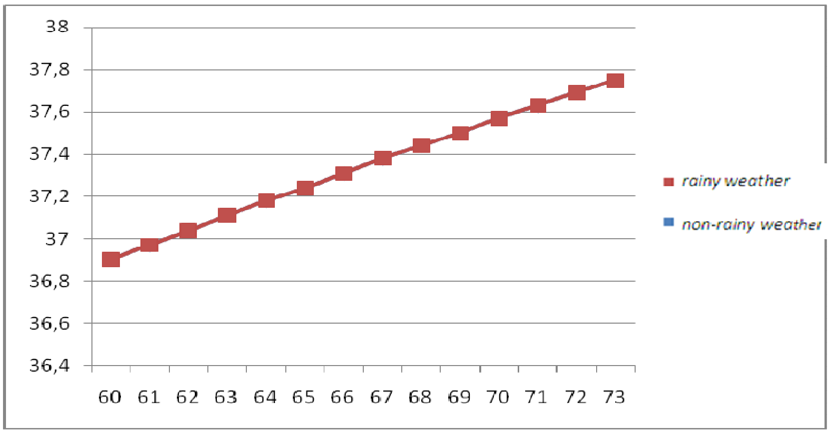


Fig. 2 The antenna efficiency [dB] vs the antenna gain [dBi].

Increase in the antenna efficiency affect on the decrease in the system noise temperature. In rainy day the system noise temperature is higher than in non-rainy weather.

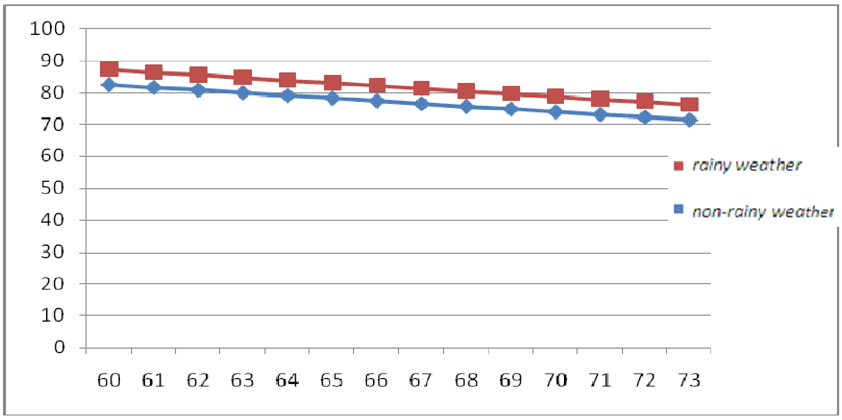


Fig. 3 The antenna efficiency [dB] vs the system noise temperature [K].

Increase in the antenna efficiency causes the increase in the carrier power at LNB output – in rainy weather the carrier power at LNB output is less than in non-rainy weather approximately about 0.1 dBW:

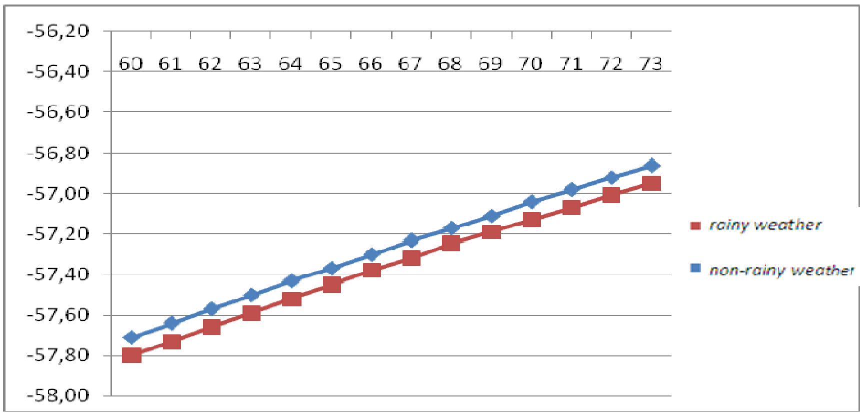


Fig. 4 The antenna efficiency [dB] vs the carrier power at LNB output [dBW].

3. Conclusions

The aim of this study was to analyze the ground dish efficiency and its influences on the quality of TV Polonia channel broadcasts via Astra 1KR satellite depending on the interferences of two adjacent satellites in geostationary orbits (Astra 1E, Astra 2A), as well as the technological parameters and disturbance of the reception quality of this TV channel in non-rainy and rainy weather in Kielce city. The study incorporates impact of the ASTRA 2A satellite (EIRP = 51dBW, bandwidth equals 33 MHz, position 28.2°E) and the ASTRA 1E satellite (EIRP = 51dBW, bandwidth equals 26MHz, position 23.5°E). In practice, the main source of noise are interference fading – due to the influence of reflections from the Earth's surface and multiway propagation in the atmosphere, rain attenuations – due to absorption and dissipation of energy wave in the rain, snow or hail, dropouts diffraction – increase in the attenuation as a results of shielding effect obstacles due to deflection (diffraction) of radio wave, power fading – due to unusual deflection of radio wave or keep it especially in tropospheric radio duct which is equivalent to growth attenuation in free space. In satellite systems the interferences are as result of other satellites, radio links, industrial disturbance (to a lesser extent) and radar jamming (there are at equal intervals of time), wide bandwidth and power amplifiers operating near saturation point. It was assumed a rain model in accordance with ITU-R. As we can see, the system noise temperature depends on the antenna efficiency – it decreases with an increase in the antenna efficiency from 82.43 K (DAP TRIAX antenna efficiency is 60%) to 71.51 K (η = 73%) in non-rainy weather and from 87.17 K (η = 60%) to 76.25 K (η = 73%) in rainy weather. With an increase in the antenna efficiency in the range from 60% to 73% increases the carrier power at LNB output from -57.71 to -56.86 dBW in non-rainy weather and from -57.80 to -56.95 dBW in rainy weather. The gain of a real antenna can be as high as 40-50 dB for very large dish antennas (although this is rare). Directivity can be as low as 1.76 dB for a real antenna, but can never theoretically be less than 0 dB.

References

[1]. BOBA, A. Szumy w systemach telekomunikacji satelitarnej, Kielce 2009.

[2]. BOGUCKI, J. *Anteny łączności satelitarnej*, Warszawa 2001.

[3]. KUŁAKOWSKI, P. *Analiza wpływu warunków terenowo-klimatycznych na pracę systemów radiokomunikacyjnych*, Kraków 2003.

[4]. WILK, J. Ł. *The influence of the antenna aperture on the quality of the satellite signal*, Wisła-Kopydło 2010.

[5]. WILK J. Ł. *Naturalne źródła szumów w transmisji satelitarnej*, Kielce 2010.

[6]. WILK J. Ł. *Wybrane zagadnienia dotyczące szumów w komunikacji satelitarnej*, Kielce 2010.

Netography

[1]. http://en.wikipedia.org/wiki/Antenna_aperture

[2]. http://en.wikipedia.org/wiki/Coupling_loss

[3]. http://sirius.cs.put.poznan.pl/~inf74839/materials/tc.doc

[4]. http://www.antenna-theory.com/basics/gain.php

[5]. http://www.aval.com.pl/aval,instalacja_satelitarna,62,2,12.html

[6]. http://www.broadbandinfo.com/time-warner-cable/digital-cable-tv/tv-polonia-polish-language-channel-on-time-warner-cable.html

[7]. http://www.satsig.net/azelhelp.htm

Transcom 2011, 27-29 June 2011
University of Žilina, Žilina, Slovak Republic

Automatic Web Image and Text Extraction

*Martina Zachariášová, *Róbert Hudec, *Miroslav Benčo, *Zuzana Dubcová, * Patrik Kamencay
*University of Žilina, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitná 8215/1, 01026 Žilina, Slovakia, {martina.zachariasova, robert.hudec, miroslav.benco, zuzana.dubcova, patrik.kamencay }@fel.uniza.sk

Abstract. In this paper, the system designed for the images automatic extraction and their textual comments available on the web is described. This system is called Filimage system. The problematic, the architecture system, and processing steps are described. The implementation consists of three sequential processing steps. The disassembly of textual and visual objects is the suggested solution allowing a separate processing. The contextual exploration method is applied for the textual extraction. Textual semantic analysis is based on hierarchical textual indices. Automatic Matching and Associating (AMA) strategy make up short multimedia documents from the full-length web documents [1,7].

Keywords: Semantic filtering, Filimage system, Automatic Matching-Associating (AMA)

1. Introduction

In this article, approach based on a semantic method to determine the relationship between digital images and textural segments in web documents by using a natural language processing NLP techniques is described.

A multimedia document consist of different components (texts, images, sounds, videos, etc.) in order to ensure the optimal use of any heterogeneous resources, therefore the goal was to establish automatically the correspondences between the images and their associated texts. It is necessary to provide a powerful system, implementing semantics resources and offering advanced features so as to produce short documents from downloaded web pages.

The most important reasons for which users need supports are:

- Information quantity is high point expanding on Internet;
- Information variety forms available electronically are also increasing [2].

Fig.1. The Web page contains several objects (texts, images, etc.).

2. Filimage system

The Filimage system allows an automatic processing of various components of Web pages. Figure 3 present the system architecture called Filimage. Automatic extraction takes place along two axes, one oriented towards the text and the other oriented image.

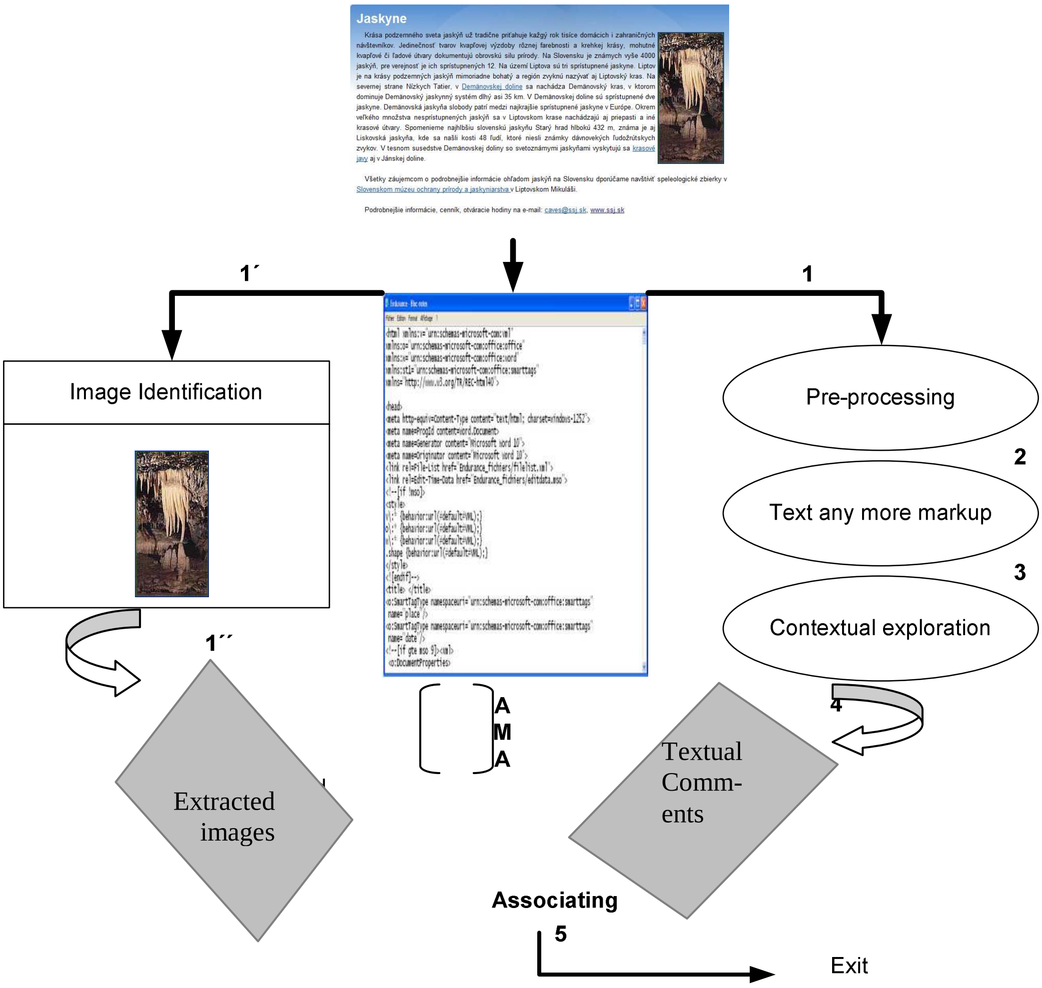


Fig.2. The Filimage architecture [1]

The key steps for its implementation are as follows:

1´´- shows the extraction module how the images are identified.

1- displays first text transition step is the pre-processing module.

2- it is in charge of the reading of the source code, and then returning a result to the last module.

3- it converts document content via data structure to a text any more mark up.

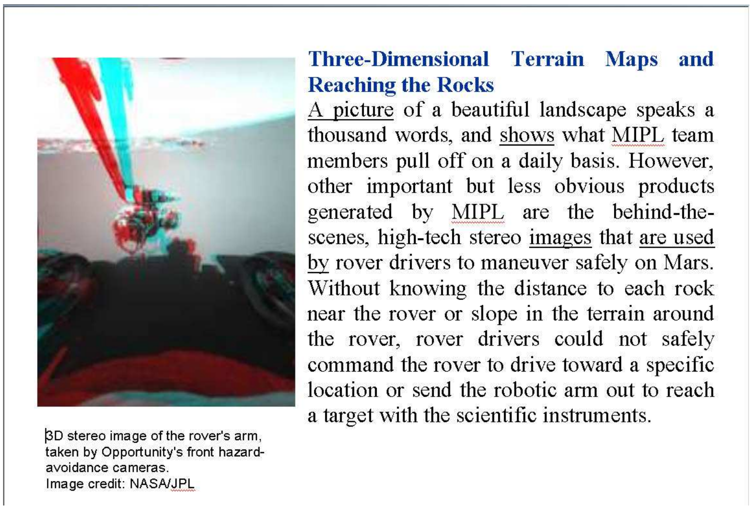
4- the textual segments are coming from semantic analysis.

5- a connection between the two modules allows automatic matching and associating operation.

After the document is unfolded, the objects such as images and texts are then stored into varied databases. The extracted textual comments are also stored [2].

2.1. Semantic filtering

This section is focused on text semantic analysis. Contextual exploration method is used to obtain the textual segments. It is the computational linguistic method, which is based on research of hierarchical textual indices. The consolidation of the linguistic resources is made through indices, rules, and databases. The linguistic expressions are analyzed in the context, which allows automatic exploitation of relevant markers. This is called CE – contextual exploration and enables extraction based on semantics, with a wide variety of a text's specific relationships involved (time, space, causality, definition, etc.). Different textual information types are considered from the diverse sources used in Language Processing (title, text, legend, source indication…). As the word usage in any language is full of ambiguity (sense of a word depends on context) there is a problem with correct localization of markers. To solve this problem the localization of markers is used in a context. The main aim of the analysis is to find the relationship between indices co-present in the same context. The texts semantic analysis is also a very helpful tool for understanding which types of indices are more precisely corresponding to visual component and must be taken as the relevant markers. An extracted segment related to an image can be displayed as its textual comment and markers are localized in the context, thanks to the released rules. A context analysis example can be seen on (Cf. Fig. 5) [2,6]


Three-Dimensional Terrain Maps and Reaching the Rocks

A picture of a beautiful landscape speaks a thousand words, and shows what MIPL team members pull off on a daily basis. However, other important but less obvious products generated by MIPL are the behind-the-scenes, high-tech stereo images that are used by rover drivers to maneuver safely on Mars. Without knowing the distance to each rock near the rover or slope in the terrain around the rover, rover drivers could not safely command the rover to drive toward a specific location or send the robotic arm out to reach a target with the scientific instruments.

3D stereo image of the rover's arm, taken by Opportunity's front hazard-avoidance cameras.
Image credit: NASA/JPL

Fig.3. We see in this preceding screen-capture, the context including the relevant image, it legend and the textual comments. The indices used by CE semantic analysis are underline [2].

The relevant indicator is "A picture" and the associated indices: "shows" is on the indicator's right. A rule allows localizing the relevant indicator "image" and the associated indices is: "are used by". An example of CE rule [2]:

Name of the rule: *RSBXXX;*
Released task: *Filimage*
Comment: *type scheme: table, argument: following table shows…*
Indicator Class
E1= *Creating space*
C1= *Indices Class*
C2= *Indices Class*
Condition1
Condition2
Attribute (tag)

2.2. Automatic matching and associating module (AMA)

In this part, the third module is described. The system AMA (Automatic matching and associating) allows solving the integration issues of a visual component (image) and a textual component (semantic content). This module is founded on resultant extraction and be used visualize combination result. Solution is CE semantic – based retrieving together joining the exempt picture of into rare sending result. The system deals of a table containing the components, and uses those components to design a new result web page [2]. After the extracted images are stored into the databases, the extracted textual segments are use for match with them. This process is necessary to the sequential organization of the document space. Linguistics markers are participating to held coherence between the textual components and images. The image is based on the context constraints [3,4].

3. Conclusion

This paper provides a short review about text automatic extraction system to the image information retrieval. System presented in the paper is available to provide an automatic extraction by using the AMA module, which is capable of summarizing any digital document. The system is based on a combination of automatic image identification and NLP techniques with using text semantics. The solution and realization under the postulation; such association provides further communication through the relevant and higher quality information was proposed. The World Wide Web contains lot of information - multimedia contents, that create a new requirements for effectiveness of access to information. Semantic entities are created from the images and their textual comments. The interaction activity enables a better cognitive understanding and the key point of Filimage system provides high level of integration objects from user point of view [1,3].

Acknowledgement

This paper has been supported by the Slovak Scientist project VEGA grant agency, project no. 1/0570/10, "Algorithms research for automatic analysis of multimedia data".

References

[1] Behnami, Sh. *Identification et Extraction Automatique des Formules Scientifiques et leurs Commentaires Textuels* , SETIT, International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, Sousse, Tunisie, 2004.

[2] Behnami, Sh. Filimage System: *Web's Images and Texts Automatic Extraction*, The World Scientific and Engineering Academic Society (WSEAS), Izmir, Turkey, 2004.

[3] Behnami, Sh. *Automatic Matching and Associating (A.M.A) for Extraction of E-document Components*, 3rd WSEAS International Conference on E-Activities, Advanced Web Technologies, E- Communities, Gis, Crete Island, Greece, 2004.

[4] Behnami, Sh. *Scientific Formulas Extraction Oriented Towards Web Summarizing File Cards*, 5th Int. WSEAS Conf., Transactions on computer, Issue 6, Vol. 3, Venice, Italy, 2004.

[5] Behnami, Sh. *Multimedia Medical Application by Filimage System: Automatic Extraction of Images Associated to Textual Comments*, 2nd IEEE on Information & Communication Technologies: from Theory to Application, ICTTA'06, Syria, 2006.

[6] Behnami; Sh. *Advanced E-access Content Filimage System: Synthesis File cards through Automatic Images-Captions Web-Pages Extraction*, Innovations in Information Technology, 2006 , vol., no., pp.1-4, Nov. 2006.

[7] Behnami, Sh. Filimage *: Système Multimédia de Résumé Automatique et Multilinguisme* , SETIT, 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, Sousse, Tunisie, 2005.

9 788055 403724

TRANSCOM 2011

Proceedings, Section 3

Published by University of Žilina
First Editions
Printed by EDIS-Žilina University publisher

ISBN 978-80-554-0372-4