UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

WYDZIAŁ TELEKOMUNIKACJI
I ELEKTROTECHNIKI

ZESZYTY NAUKOWE NR 258

TELEKOMUNIKACJA I ELEKTRONIKA 14

BYDGOSZCZ – 2011

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

ZESZYTY NAUKOWE NR 258

TELEKOMUNIKACJA I ELEKTRONIKA

14 2

BYDGOSZCZ – 2011

REDAKTOR NACZELNY
prof. dr hab. inż. Janusz Prusiński

REDAKTOR NACZELNY SERII
dr inż. Beata Marciniak

OPRACOWANIE TECHNICZNE
mgr inż. Daniel Morzyński

ISSN 1899-0088

Wydawnictwa Uczelniane Uniwersytetu Technologiczno-Przyrodniczego
ul. Ks. A. Kordeckiego 20, 85-225 Bydgoszcz, tel. 52 3749482, 3749426
e-mail: wydawucz@utp.edu.pl http://www.wu.utp.edu.pl/

Wyd. I. Nakład 80 egz. Ark. aut. 4,0. Ark. druk. 5,0.
Zakład Małej Poligrafii UTP Bydgoszcz, ul. Ks. A. Kordeckiego 20

Contents

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 258
TELEKOMUNKACJA I ELEKTRONIKA 14 (2011), 5-12

DESIGN AND TECHNOLOGY OF HYBRID MULTILAYER ELECTRONIC CIRCUITS

Dominik Jurków, Karol Malecha, Mateusz Czok,
Henryk Roguszczak, Michał Babiarz, Leszek Golonka

Faculty of Microsystem Electronics and Photonics
Wrocław University of Technology
50-372 Wrocław, Janiszewskiego 11/17, building C-2, Poland
dominik.jurkow@pwr.wroc.pl

Summary: The design and technology of hybrid electronic ceramic circuits with flip-chip and SMD (Surface Mounting Device) components are presented in the paper. The flip-chip audio amplifier and RF (Radio Frequency) transmitting and receiving modules are fabricated using LTCC (Low Temperature Co-fired Ceramics). X-Ray inspection is performed to analyze the solder bonding quality. The application of special underfill has increased the reliability of interconnections between flip-chip, SMD components and the ceramic multilayer substrate. The final structure of the audio amplifier module is encapsulated with ceramic housing.

Keywords: flip-chip, LTCC, X-Ray, thick-film, SIP

1. INTRODUCTION

The size of an electronic package is one of the most significant miniaturization factors. The size of a package is limited by the interconnection pitch. The flip-chip technology [1, 2], which is an advanced form of SMT (Surface Mounting Technology), is one of the most significant assembly developments. The method increases device reliability and productivity and simultaneously reduces costs. The flip-chip is a bare silicon chip which consists of solder pads placed on the whole bottom surface. In contrast to the common use of the wire-bonding technique, the flip-chip assembly method enables the bonding of all interconnections in one step and bonds devices with I/O placed on all bottom chip areas. Moreover, flip-chip technology allows the attainment of very high interconnections reliability and electric performance (low serial resistance, capacitance and inductance). However, the wire bonding method is still most often used in low density interconnection bonding <700 I/O on a single chip. The flip-chip technique is the main technique of higher interconnection density assembly. One of the main problems in assembly technique is mismatch between thermal expansion coefficients of chip and substrate. Epoxy laminates are the most common substrates in electronic industry. However, thermal expansion coefficient mismatch between epoxy and silicon is quite high. Therefore, interconnection reliability is decreased. The problem can be reduced by using

organic underfill which decreases thermal expansion coefficients mismatch between both materials. Moreover, the thermal expansion coefficient of new glass ceramics (Low Temperature Cofired Ceramics - LTCC) substrates is fitted to silicon. Therefore, the LTCC technology is compatible with the flip-chip technique.

The low Temperature Cofired Ceramic technique was developed around 30 years ago. The method was applied in the fabrication of hybrid multilayer electronic circuits [3-5]. The main advantages of LTCC are good long term stability and mechanical properties, compatibility with standard thick film technology and low cost. The interconnection density can be increased by using buried screen printed passive components. The standard LTCC process [6-7] consists of the following steps: design, via and shape cutting [8,9], thick film deposition and lamination [10,11] and cofiring. The ability to form 3D shapes and compatibility with the thick film technique enables us to fabricate various sensors [12-16], actuators [17, 18], microsystems [19,20] and wireless transmission devices [21,22] using LTCC technology.

The design and technology and inspection of hybrid electronic ceramic circuits with flip-chip and SMD components is presented in this paper.

2. RF MODULES ELECTICTRICAL CIRCUIT DESIGN

The electrical circuit of RF transmitter integrated with the LTCC module is presented in Figure 1. The electronic circuit consists of ATmega88 microcontroller, monolithic temperature and humidity sensor SHT11, radio frequency transmitter (RFM02), fixed-output voltage regulator (LM2980IM5-5.0) and additional passive components.

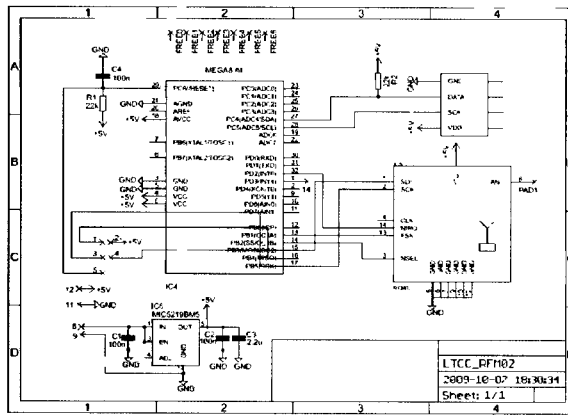


Fig. 1. Electrical circuit of RF transmitter module.

The electrical circuit of RF receiver integrated with LTCC module is presented in Figure 2. The electronic circuit consists of ATmega88 microcontroller, radio frequency receiver (RFM01), fixed-output voltage regulator (LM2980IM5-5.0) and additional passive components. Both modules were assembled to additional electronic circuits. The additional circuits consist of power supplies, LCD displays and RF antennas.

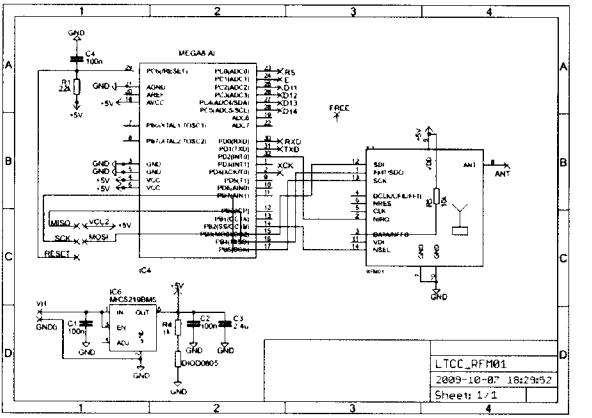


Fig. 2. Electrical circuit of RF receiver module.

3. TECHNOLOGY OF RF MODULE

The LTCC multilayer module consisted of two DP951 P2 green tapes: TOP, BOT, and two DP951 C2 tapes M1 and M2. The thickness of DP951 P2 and DP951 C2 tapes after firing was approximately 165 μm and 50 μm, respectively. The TOP and BOT layers consisted of surface conductive lines and solder pads, the M1 and M2 provided better mechanical properties of the ceramic module and masked the electric trucks. Therefore M1 and M2 layers were used as solder masks. The LTCC tapes were laser cut to proper dimensions and via holes were drilled. The silver paste DP 6141 was used to fill the vias. The silver paste DP 6145 was used to deposit electrical trucks. The silver palladium paste was used to deposit solder pads. The screen printing method was used for via filling and deposition of electric trucks and solder pads. After screen printing all ceramic tapes were stacked together in the proper order and laminated using an isostatic press with a pressure of 20 MPa at 70°C for 10 minutes. Finally, the LTCC multilayer module was co-fired in air in a two-step firing profile recommended by DuPont (Tmax = 850°C). The final fired structure with solder mask is presented in Figure 3. Substrates were assembled with SMD components. The final receiving and transmitting modules are shown in Figure 2a and 2b, respectively. The additional electronic circuits which consist of power supplies, LCD displays and RF antennas were fabricated with standard PCB (Printed Circuit Board) and SMT techniques. Power consumption of the devices was significantly decreased with the use of ATMEGA sleep modes and turning down unused microcontroller peripherals. The average current consumptions of transmitting and receiving modules were approximately 440 μA and 14 mA, respectively. The transmitting range was equal to 60 m in the open area.

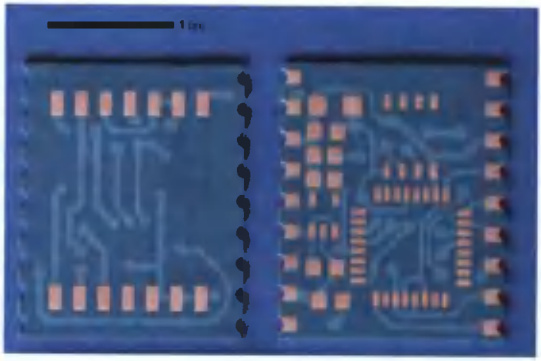
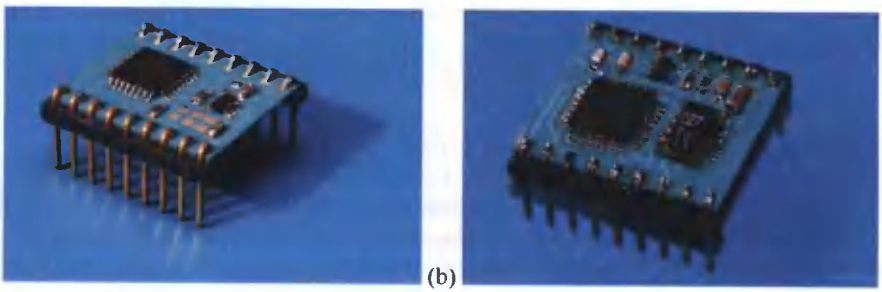


Fig. 3. Final fired multilayer substrate.



(a) (b)

Fig. 4. Final device, (a) receiver, (c) transmitter.

4. FLIP-CHIP MODULE ELECTICTRICAL CIRCUIT DESIGN

The electronic circuit of the flip-chip stereo audio power amplifier integrated with the LTCC substrate is shown in Figure 5. The circuit consists of a flip-chip stereo audio power amplifier TS4984, four negative feedback resistors and four capacitors. Each channel of the TS4984 device can supply 1.2 W (8 Ω load at 5V) [23]. The amplifier is suited for mobile phones, LCD monitors, portable audio devices, and PDA computers, etc.

The gain (k) of each channel is given by equation 1. The gain is set by R_1-R_4 resistors. The gain of 4.5 times was designed.

$$k = \frac{R_1}{R_2} = \frac{R_3}{R_4} \tag{1}$$

Bypass capacitors C_1 and C_2 provide power supply filtering. Input capacitors C_3 and C_4 separate the DC voltage from the flip-chip input terminations. The low-pass filter is set by R_2, R_4 resistors and input capacitors C_3, C_4. The cut-off frequency is given by equation 2. The cut-off frequency was set to 72 Hz.

$$f = \frac{1}{2\pi R_2 C_3} = \frac{1}{2\pi R_4 C_4} \tag{2}$$

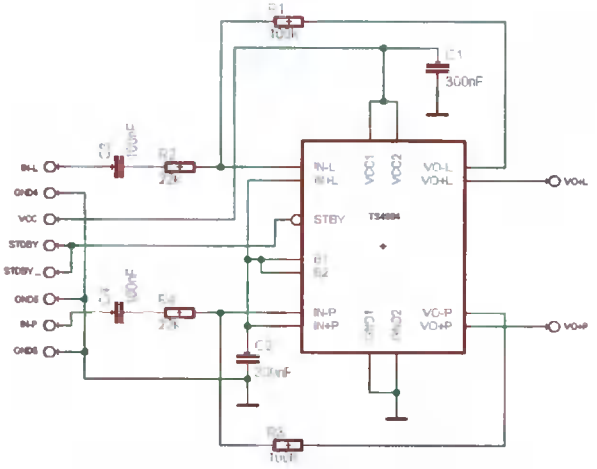


Fig. 5. Electrical circuit of stereo audio power amplifier based on flip-chip module TS4984

5. TECHNOLOGY AND POST INSPECTION OF THE FLIP-CHIP MODULE

The LTCC multilayer module consisted of four DP951 P2 green tapes: TOP, BOT, L1 and L2. The thickness after firing of each DP951 P2 tape was approximately 165 μm. The TOP layer consisted of surface conductive trucks and solder pads, the L1 and L2 layers consisted of inner conductive trucks and the BOT layer provided better mechanical properties of the ceramic module. The LTCC tapes were laser cut to proper dimensions. All conductive lines were made of silver palladium paste (DP 6146). Afterwards, vias were punched and filled with silver DP6161 paste with via filler. All ceramic tapes were stacked together in the proper order and laminated using an isostatic press with a pressure of 20 MPa at 70°C for 10 minutes. Finally, the LTCC multilayer module was co-fired in air in a two-step firing profile recommended by DuPont (T_{max} = 850°C). After firing the SMD and flip-chip, devices were assembled on top of the module and soldered with a precise flip-chip system (FC300 CAMMAX PRECIMA LTD.). The chip was encapsulated with underfill. The fired LTCC structure with assembled components is presented in Figure 6a. After the assembly of the SMD and the flip-chip components the ceramic module was housed. The package consisted of six DP 951 P2 LTCC tapes. The green ceramic tapes were laminated using the thermo-compression method (20 MPa at 70°C for 10 minutes). Next, the package geometry was formed with the CNC (Computer Numerical Control) microdrill and fired in a furnace in typical thermal profile. The top part of the package and final packaged LTCC module are shown in Figures 6b and 6c, respectively.

Fig. 6. Final device: a) assembled fired LTCC substrate, b) top housing part, c) housed structure

The alignment of the bare chip with solder pads after firing and the assembly processes was examined by a non-destructive X-ray tomography at ITR Warsaw. An X-ray image of the LTCC multilayer module with flip-chip is shown in Figure 7.

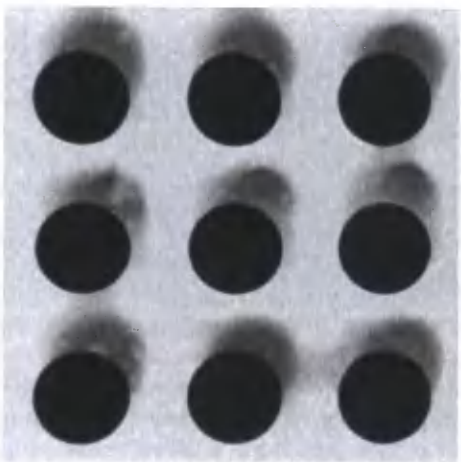


Fig. 7. X-ray image of the LTCC multilayer module

6. SUMMARY

The electrical circuit and layout of the RF transmitter/receiver and acoustic amplifier flip-chip modules integrated with LTCC substrates were developed.

Surface and buried electrical tracks were deposited using the screen printing method.

The multilayer LTCC RF modules were fabricated and tested positively . Their transmission range and power consumption were estimated to 60 m. and ca. 1.45 mW, respectively.

The semiconductor bare amplifier chip was encapsulated with underfill. Moreover, all electronic components were covered with ceramic housing.

The good alignment of the bare chip with solder pads after firing and the assembly processes was confirmed by X-Ray tomography.

The possibility of integration of surface electronic components (SMD, buried and surface screen printed passives and conductive tracks and flip chip devices) and package in one substrate enables us to fabricate sophisticated system-in-package (SIP) devices.

BIBLIOGRAPHY

[1] R. Tummala, 2001. Fundamentals of Microsystems Packaging, McGRAW-HILL, ISBN 0-07-137169-9.

[2] C. Harper, 2000. Electronic packaging and interconnection handbook, McGRAW-HILL, ISBN 0-07-134745-3.

[3] T. Gupta, 2003. Handbook of Thick- and Thin-Film Hybrid Microelectronics, Wiley-interscience, ISBN 0-471-27229-9.

[4] F. Barlow III, A. Elshabini, 2007. Ceramic Interconnect Technology Handbook, CRC Press/Taylor & Francis, ISBN 0-849-33557-4.

[5] L. Golonka, 2006. Technology and applications of Low Temperature Cofired Ceramic (LTCC) based sensors and Microsystems, Bulletin of the Polish Academy of Sciences, vol. 54, pp. 223-233.

[6] M.R. Gongora-Rubio, P. Espinoza-Vallejos, L. Sola-Laguna, J.J. Santiago-Aviles, 2001. Overview of low temperature co-fired ceramics tape technology for meso-system technology (MsST), Sensors and Actuators A, vol. 89, pp. 222-241.

[7] K. A. Peterson, K. D. Patel, C. K. Ho, S. B. Rohde, C. D. Nordquist, C. A. Walker, B. D. Wroblewski, M. Okandan, 2005. Novel microsystem applications with new techniques in Low-Temperature Co-Fired Ceramics, Int. J. Appl. Ceram. Technol., vol. 2, pp. 345-363.

[8] J. Kita, A. Dziedzic, L. Golonka, T. Zawada, 2002. Laser treatment of LTCC for 3D structures and elements fabrication, Microelectronics International, vol. 19, pp. 14-18.

[9] K. Nowak, H. Baker, D. Hall, 2006. Cold processing of green state LTCC with a CO_2 laser, Applied Physics A, Materials Science and Processing, vol. 84, pp. 267-270.

[10] M. A. Piwonski, A. Roosen, 1999. Low pressure lamination of ceramic green tapes by gluing at room temperature. Journal of the European Ceramic Society, vol. 19, pp. 263-270.

[11] H. Birol, T. Maeder, P. Ryser, 2006. Processing of Graphite-Based Sacrificial Layer for Microfabrication of Low Temperature Co-Fired Ceramics (LTCC), Sensors Actuators A, vol. 130-131, pp. 560–567.

[12] M.G.H. Meijerink , E. Nieuwkoop, E.P. Veninga, M.H.H. Meuwissen, M.W.W.J. Tijdink, 2005 . Capacitive pressure sensor in post-processing on LTCC substrates, Sensors and Actuators A, vol. 123-124, pp. 234-239.

[13] M. Gongora-Rubio, L.M. Sola-Laguna, P.J. Moffett, J.J. Santiago-Aviles, 1999. The utilization of low temperature co-fired ceramics_LTCC-ML technology for meso-scale EMS, a simple thermistor based flow sensor, Sensors and Actuators, vol. 73, pp. 215-221.

[14] M.H.H. Meuwissen, E.P. Veninga, M.W.W.J. Tijdink, M.G.H. Meijerink, 2006. Design study of a capacitive pressure sensor in non-silicon materials, Proceedings of the Institution of Mechanical Engineers, vol. 220.

[15] M.G.H. Meijerink , E. Nieuwkoop, E.P. Veninga, M.H.H. Meuwissen, M.W.W.J. Tijdink, 2005. Capacitive pressure sensor in post-processing on LTCC substrates, Sensors and Actuators A, vol. 123-124, pp. 234-239.

[16] W. Smetana, M. Unger, 2008. Design and characterization of a humidity sensor realized in LTCC-technology, Microsyst. Technol., DOI 10.1007/s00542-007-0465-3.

[17] H. Klumbies, U. Partsch, A. Goldberg, S. Gebhardt, U. Keitel, and H. Neubert, 2009. Actuators to be integrated in Low Temperature Cofired Ceramics (LTCC) microfluidic systems", Proc. 32nd IEEE International Spring Seminar on Electronics Technology, 13-17 May 2009, Brno (Czech Republic), pp. 1-4.
[18] E. Heinonen, J. Juuti, H. Jantunen, 2007. Characteristics of piezoelectric cantilevers embedded in LTCC", Journal of the European Ceramic Society, vol. 27, pp. 4135--4138.
[19] T. Thelemann, H. Thust, M. Hintz, 2002. Using LTCC for microsystems, Microelectronics International, vol. 19, pp. 19-23.
[20] L.J. Golonka, H. Roguszczak, T. Zawada, J. Radojewski, I. Grabowska, M. Chudy, A. Dybko, Z. Brzozka and D. Stadnik, 2005. LTCC based microfluidic system with optical detection, Sensors and Actuators B, vol. 111-112, pp. 396-402.
[21] A. Sutono, A. Pham, J. Laskar, W. R. Smith, "Development of three dimensional ceramic-based MCM inductors for hybrid RF/microwave applications", IEEE Radio Frequency Integrated Circuits Symposium, pp. 175-178, 1999
[22] A. Sutono, A. Pham, J. Laskar, W.R. Smith, 1999. RF/microwave characterization of multilayer ceramic-based MCM technology, IEEE Transactions on Advanced Packaging, vol. 22, pp. 326-331.
[23] Datasheet of TS4984 flip-chip Stereo Audio Power Amplifier.

The authors wish to thank the Polish Ministry of Science and Higher Education (grant no. R02 017 02) for financial support. Fellowship of Dominik Jurków co-financed by European Union within European Social Fund. The Foundation for Polish Science (FNP) is also acknowledged for financial support for Karol Malecha.

PROJEKTOWANIE I TECHNOLOGIA WIELOWARSTWOWYCH HYBRYDOWYCH UKŁADÓW ELEKTRONICZNYCH

Streszczenie

W pracy zaprezentowano projekt oraz realizację dwóch przykładowych hybrydowych układów elektronicznych: wzmacniacza audio opartego na elemencie typu flip-chip oraz nadajnika i odbiornika RF opartych na elementach do montażu powierzchniowego (SMD). Przedstawione układy zostały wykonane przy wykorzystaniu technologii bazującej na niskotemperaturowej ceramice współwypalanej (LTCC). Niezawodność połączeń lutowanych pomiędzy elementem typu flip-chip a polami kontaktowymi umieszczonymi na podłożu LTCC zbadano za pomocą metody rentgenowskiej. Poprawę niezawodności w przypadku układu z elementem typu flip-chip uzyskano stosując wypełnienie organiczne pomiędzy chipem a podłożem. Gotowy układ wzmacniacza audio zamknięto w specjalnie przygotowanej obudowie ceramicznej.

Słowa kluczowe: flip-chip, niskotemperaturowa ceramika współwypalana (LTCC), tomografia rentgenowska, warstwa gruba, SIP

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 258
TELEKOMUNKACJA I ELEKTRONIKA 14 (2011), 13-31

STEREO MATCHING AND DISPARITY CALCULATION BASED ON DISCRETE ORTHOGONAL MOMENTS OF CHEBYSHEV, LEGENDRE AND ZERNIKE

Tomasz Andrysiak

Institute of Telecommunications,
Faculty of Telecommunications and Electrical Engineering
University of Technology and Life Sciences
Al. S. Kaliskiego 7, 85-789 Bydgoszcz, Poland
tomasz.andrysiak@utp.edu.pl

Summary: In the article we present various theoretical and experimental approaches to the problem of stereo matching and disparity estimation. We propose to calculate stereo disparity in the moments space, but we also present numerical and correlation based methods. In order to calculate disparity vector we decided to use discrete orthogonal moments of Chebyshev, Legendre and Zernike. In our research of stereo disparity estimation all of these moments were tested and compared. Experimental results confirm effectiveness of the presented methods of determining stereo disparity and stereo matching for machine vision applications.

Keywords: orthogonal moments, stereo matching, stereo disparity

1. INTRODUCTION

One of the main research fields in machine and robotics vision is 3D scene perception based on techniques of measuring shapes, positions and relations between 3D objects that are visible in the scene. There are many known methods of retrieving information about the scene basing on 3D perception [2-4,6]. Those methods are based on disparity of stereoscopic scene elements and on information from the common part of stereoimages. After extracting the pair of corresponding points in two stereoimages, which are related to the same point in the scene, we can define the difference between coordinates of those points. Then basing on such differences, it is possible to create depth map for the visible scene by means of simple trigonometric transformations [18].

Recently, discrete orthogonal moments have gained much attention and have been successfully used in many applications of computer vision (e.g. pattern recognition) [1,9,14,16,17]. Therefore, in our research we decided to take advantage of discrete orthogonal moments' properties, in particular those introduced by Chebyshev, Legendre and Zernike, in order to calculate stereo disparity.

In the article, we are concerned with the images acquired from the axe-parallel robotics vision system. In such a system the optical axes of both cameras are parallel to each other and the image planes of stereoscopic pair are situated within the same

distance from the centre of the scene coordinates system. Only the common scene area covered by both cameras is further analysed (even though it is only a part of each of the image). In section 2, issues related to stereo-image acquisition in robotics vision system are described in details. In section 3, the discrete orthogonal moments of Chebyshev, Legendre and Zernike are introduced. In section 4, the method of displacement vector calculation in order to determine the margins and common part of stereo images is presented. Then, in Section 5, three approaches to calculate stereo disparity are described. Experimental results, discussion and conclusion are given in the next sections.

2. ACQUISITION OF THE STEREO IMAGES

Hereby we present the model of the used stereo camera system for robotics vision. Moreover, the principles of epipolar geometry and the method of camera system calibration are presented [5,28].

2.1. MODEL OF THE CAMERA SYSTEM

The model of the camera system represents geometrical and physical parameters of the cameras and transfers cameras' coordinates systems onto predefined coordinates system of the visible scene. A description of the model is necessary for correct measuring of the scene objects' shapes, sizes, position, direction and relations within the scene space [20].

Differences in localisation and directions of the cameras and, therefore in cameras' coordinates systems, determine various models of camera systems as presented in Fig. 1. In axe-parallel camera system the coordinates of the point $A^S(X,Y,Z)$ of the left camera $O^{KL}XYZ$ and the right camera $O^{KP}XYZ$ related to the scene O^SXYZ are given.

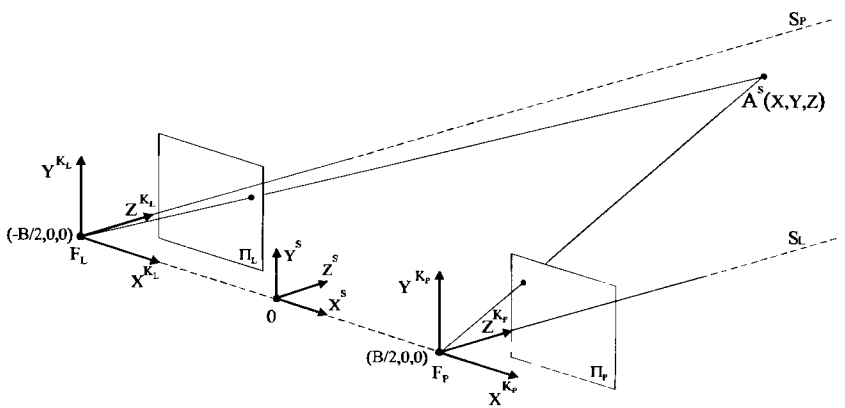


Fig. 1. The axe-parallel model of the stereo cameras vision system

$$\begin{bmatrix} X^{K_L} \\ Y^{K_L} \\ Z^{K_L} \end{bmatrix} = \begin{bmatrix} X^S - \frac{B}{2} \\ Y^S \\ Z^S \end{bmatrix}, \qquad \begin{bmatrix} X^{K_P} \\ Y^{K_P} \\ Z^{K_P} \end{bmatrix} = \begin{bmatrix} X^S + \frac{B}{2} \\ Y^S \\ Z^S \end{bmatrix} \tag{1}$$

The mapping of the point $A^S(X,Y,Z)$ onto the image planes Π_L (left) and Π_P (right) can be written as:

$$\begin{bmatrix} x^{\Pi_L} \\ y^{\Pi_L} \end{bmatrix} = \begin{bmatrix} f\frac{X^S - \frac{B}{2}}{Z^S} \\ f\frac{Y^S}{Z^S} \end{bmatrix}, \qquad \begin{bmatrix} x^{\Pi_P} \\ y^{\Pi_P} \end{bmatrix} = \begin{bmatrix} f\frac{X^S + \frac{B}{2}}{Z^S} \\ f\frac{Y^S}{Z^S} \end{bmatrix}. \tag{2}$$

Then, we can define the coordinates of the point $A^S(X,Y,Z)$ in the scene as:

$$\begin{bmatrix} X^S \\ Y^S \\ Z^S \end{bmatrix} = \begin{bmatrix} \frac{B\left(x^{\Pi_L} + x^{\Pi_P}\right)}{2d} \\ \frac{By^{\Pi_L}}{d} \\ \frac{Bf}{d} \end{bmatrix}, \tag{3}$$

where d is defined as the stereoscopic discrepancy as the difference between projection coordinates of point $A^S(X,Y,Z)$ onto image planes Π_P and Π_L, such as:

$$d = x^{\Pi_P} - x^{\Pi_L}. \tag{4}$$

Stereoscopic images acquired from the axe-parallel camera system can contain errors after the matching process if the chosen image elements do not have the correspondence elements on the second stereo-pair image. The solution for such a problem is to find the translation vector between stereo-images and defining the common part of the left and right image [20]. An example of a stereo-image pair acquired and calibrated in the described system is presented in Figure 2.

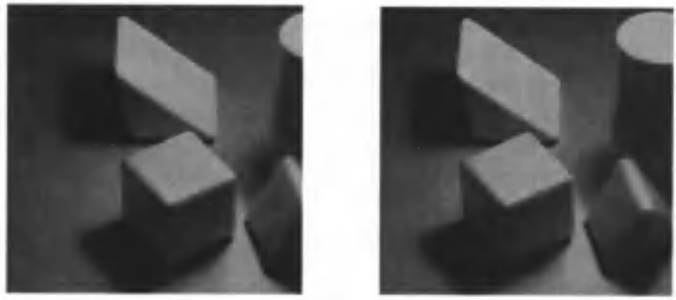


Fig. 2. The stereo-image „blocks" acquired in the axe-parallel camera vision system

2.2. EPIPOLAR GEOMETRY

An important characteristic of stereoscopic systems is the epipolar geometry, meaning that the projections of any point in the scene $A^S(X,Y,Z)$ onto the image planes Π_P and Π_L are localised on the corresponding epipolar lines. Each point placed on the epipolar line of the right image has its corresponding point on the analogical epipolar line of the left image. This relation is true only if we consider the common area of both cameras vision space [4].

The line between the focuses of the left and right cameras in the stereoscopic system is called a base-line, and the length of this line is called a base. The system consisting of the base-line and any point in the scene $A^S(X,Y,Z)$ determines unambiguously the epipolar plane. The intersection of the epipolar plane with the image planes Π_P and Π_L defines the epipolar lines. The number of detected epipolar lines is connected with the resolution of the camera system.

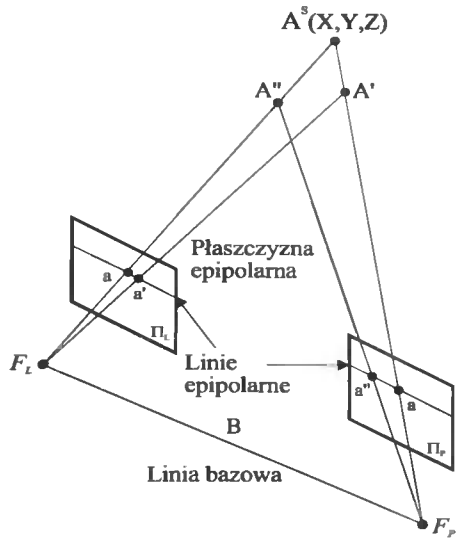


Fig. 3. Epipolar geometry in the stereoscopic camera system

2.3. CALIBRATION OF CAMERA SYSTEM

In order to perform the proper acquisition of stereoscopic images referring to the visible scene it is crucial that the camera system is properly calibrated. The process of calibration of the stereo-system is supposed to define relations between coordinates of the point in the scene $A^S(X,Y,Z)$, known coordinates of that point projection $a^{\Pi_L}(x,y)$ and $a^{\Pi_P}(x,y)$ onto the image planes Π_L and Π_P, and the geometrical and optical parameters of the camera system. Such parameters are the position and direction of cameras in relation to the defined coordinates system of the scene described by elements of the rotation matrix ΔR and the translation vector ΔT. Those parameters can determine the position and direction of the cameras in relation to the predefined coordinates system of the scene [27]. We have:

$$\begin{bmatrix} X^{K_L} \\ Y^{K_L} \\ Z^{K_L} \end{bmatrix} = \Delta R^{K_L} \begin{bmatrix} X^S \\ Y^S \\ Z^S \end{bmatrix} + \Delta T^{K_L}, \quad \begin{bmatrix} X^{K_P} \\ Y^{K_P} \\ Z^{K_P} \end{bmatrix} = \Delta R^{K_P} \begin{bmatrix} X^S \\ Y^S \\ Z^S \end{bmatrix} + \Delta T^{K_P}. \quad (5)$$

The translation vector describes the localization of centre points of the camera systems coordinates. The rotation matrix is the orthogonal matrix and its elements are the Euler angles that characterize orientation of camera coordinates systems axis.
The optical parameters are:

- scaling ratios k_x and k_y of the axis Ox^E and Oy^E,
- parameters p_x and p_y of the coordinates systems translations $O^E xy$ and $O^\Pi xy$,
- position of the center point of the coordinates system $O^E xy$,
- length f of camera focus,
- geometrical distortions in the camera optical system $\vartheta_x^\Pi, \vartheta_y^\Pi$.

Then:

$$\begin{bmatrix} x^\Pi \\ y^\Pi \end{bmatrix} = \begin{bmatrix} k_x x^E + p_x + \vartheta_x^\Pi(x,y) \\ k_y y^E + p_y + \vartheta_y^\Pi(x,y) \end{bmatrix}. \quad (6)$$

Description of all the geometrical and optical parameters of the stereoscopic camera system (for left and right camera separately) is the condition for proper measuring of the 3D objects' shapes, position and orientation [10].

The calibration problem is the problem of finding the values of geometrical and optical parameters for a considered camera system in order to define position and orientation of cameras in relation to the scene coordinates system basing on the images acquired from those cameras [26]. The solution of the calibration problem consists of:

- finding optical parameters and the values of distortion introduced by the lenses of both cameras,
- determining the correspondence list for features in the stereo-images,
- calculating the values of geometrical parameters (location and cameras orientation based on the obtained correspondence map).

Moreover, the feedback loop should be taken into account while analysing the camera calibration process. Then, the calculated parameters of position and orientation of the cameras can be tuned recursively until the desired parameters for a specified camera model are achieved. In result of the camera calibration the optical and geometrical parameters characterizing the acquired stereo-images are achieved [11].

3. DISCRETE ORTHOGONAL MOMENTS

In this section, three types of discrete orthogonal moments are presented: Chebyshev and Legendre moments which are realized in the cartesian coordinates system, and Zernike moments which are realized in the polar coordinates system. Due to orthogonalization of their base-functions these moments are characterized by fast algorithms of realization and by simple reconstruction formulas. Hence, these moments are calculated using Geometric moments, Central moments and scaled Geometric moments [7,15,23,24].

3.1. GEOMETRIC MOMENTS

The Geometric moments (GM) m_{ij} of order $i+j$ of a digital image $I(x,y)$ (left or right image of stereopair) of the size $N \times N$ are defined as [13,19]:

$$m_{ij} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x^i y^j I(x,y), \tag{7}$$

where $i, j = 0,1,...,N-1$.

The translation invariant Central moments M_{ij} are obtained by placing origin at the centroid of the image.

$$M_{ij} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (x-\bar{x})^i (y-\bar{y})^j I(x,y), \tag{8}$$

where

$$\bar{x} = \frac{m_{10}}{m_{00}}, \qquad \bar{y} = \frac{m_{01}}{m_{00}}. \tag{9}$$

Then, the scale invariant Central moments C_{ij} are defined as:

$$C_{ij} = \frac{M_{ij}}{\alpha^{(i+j+2)/2}}, \tag{10}$$

where $\alpha = M_{00}$.

Finally, the scale invariant Radial-Geometric moments R_{ij} are defined as:

$$R_{ij} = \frac{\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}\left(\hat{x}^2+\hat{y}^2\right)^{\frac{1}{2}}(\hat{x})^i(\hat{y})^j I(x,y)}{\alpha^{(i+j+3)/2}}, \tag{11}$$

where $\hat{x} = x - \bar{x}$ and $\hat{y} = y - \bar{y}$.

3.2. CHEBYSHEV MOMENTS

The Chebyshev moments (TM) of order $m+n$ of an image $I(x,y)$ (left or right image of stereopair) of the size $N \times N$ are defined using the scaled orthogonal Chebyshev polynomials $t_n(x)$, as [14,23]:

$$TM_{mn} = \frac{1}{\rho(m,N)\rho(n,N)}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} t_m(x)t_n(y)I(x,y), \tag{12}$$

Where $m,n = 0,1,...,N-1$ and $\{t_n(x)\}$ are a set of discrete orthogonal polynomials satisfying the following condition:

$$\sum_{n=0}^{N-1} t_m(x)t_n(x) = \rho(n,N)\delta_{mn}, \tag{13}$$

Where δ_{mn} is the Kronecker function:

$$\delta_{mn} = \begin{cases} 1 & if \quad m-n, \\ 0 & otherwise \end{cases}. \tag{14}$$

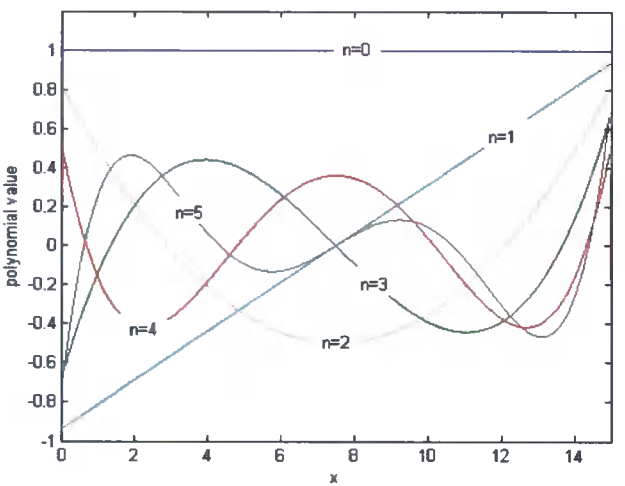


Fig. 4. Chebyshev polynomials $t_n(x)$ of order $n = 0,1,...,5$

The Chebyshev polynomials satisfy the property of orthogonality (13), with:

$$\rho(n,N)=\frac{(N-1)\left(N-2^2\right)...\left(N-n^2\right)}{(2n+1)N^{2n-1}} \tag{15}$$

and have the following recurrence relation:

$$(n+1)t_n(x)-(2n+1)(2x-N+1)t_n(x)+n\left(1-n^2N^{-2}\right)t_{n-1}(x)=0, \tag{16}$$

where $n=2,3,...,N-1$ and $x=0,1,...,N-1$ with the initial conditions:

$$t_0(x)=1, \quad t_1(x)=(2x-N+1)/N. \tag{17}$$

The inverse formula of Chebyshev moments is given by the following equation:

$$I(x,y)=\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}TM_{mn}t_m(x)t_n(y). \tag{18}$$

The Chebyshev moments of the same image may be expressed in terms of geometric moments as follows [14]:

$$TM_{mn}=A_mA_n\sum_{k=0}^{m}C_k(m,N)\sum_{l=0}^{n}C_l(n,N)\sum_{i=0}^{k}\sum_{j=0}^{l}s_k^{(i)}s_l^{(j)}m_{ij}, \tag{19}$$

where

$$A_m=\frac{1}{N^n\rho(m,n)}, \tag{20}$$

and

$$C_k(n,N)=(-1)^{n-k}\frac{n!}{k!}\binom{N-1-k}{n-k}\binom{n+k}{n}, \tag{21}$$

and $s_k^{(i)}$ are the Stirling number of the first kind, which satisfies:

$$\frac{x!}{(x-k)!}=\sum_{i=0}^{k}s_k^{(i)}x^i. \tag{22}$$

The explicit expressions of the Chebyshev moments in terms of geometric moments up to the first order are as follows:

$$TM_{00}=\frac{m_{00}}{N^2},$$

$$TM_{10}=\frac{6m_{10}+3(1-N)m_{00}}{N\left(N^2-1\right)}, \tag{23}$$

$$TM_{10}=\frac{6m_{01}+3(1-N)m_{00}}{N\left(N^2-1\right)}.$$

3.3. EGENDRE MOMENTS

The Legendre moments (*LM*) of order $r+s$ of discrete image $I(x,y)$ are defined as [12,21]:

$$LM_{rs} = \frac{(2r+1)(2s+1)}{N^2} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} P_r(\zeta_x) P_s(\zeta_y) I(x,y), \tag{24}$$

where $r,s = 0,1,...,N-1$ and the image coordinate transformation is given by:

$$\xi_x = \frac{2x-N+1}{N-1}, \quad \xi_y = \frac{2y-N+1}{N-1}, \tag{25}$$

where $P_s(\xi_y)$ is the Legendre polynomial of degree s and ζ_x, ζ_y are the normalized coordinates.

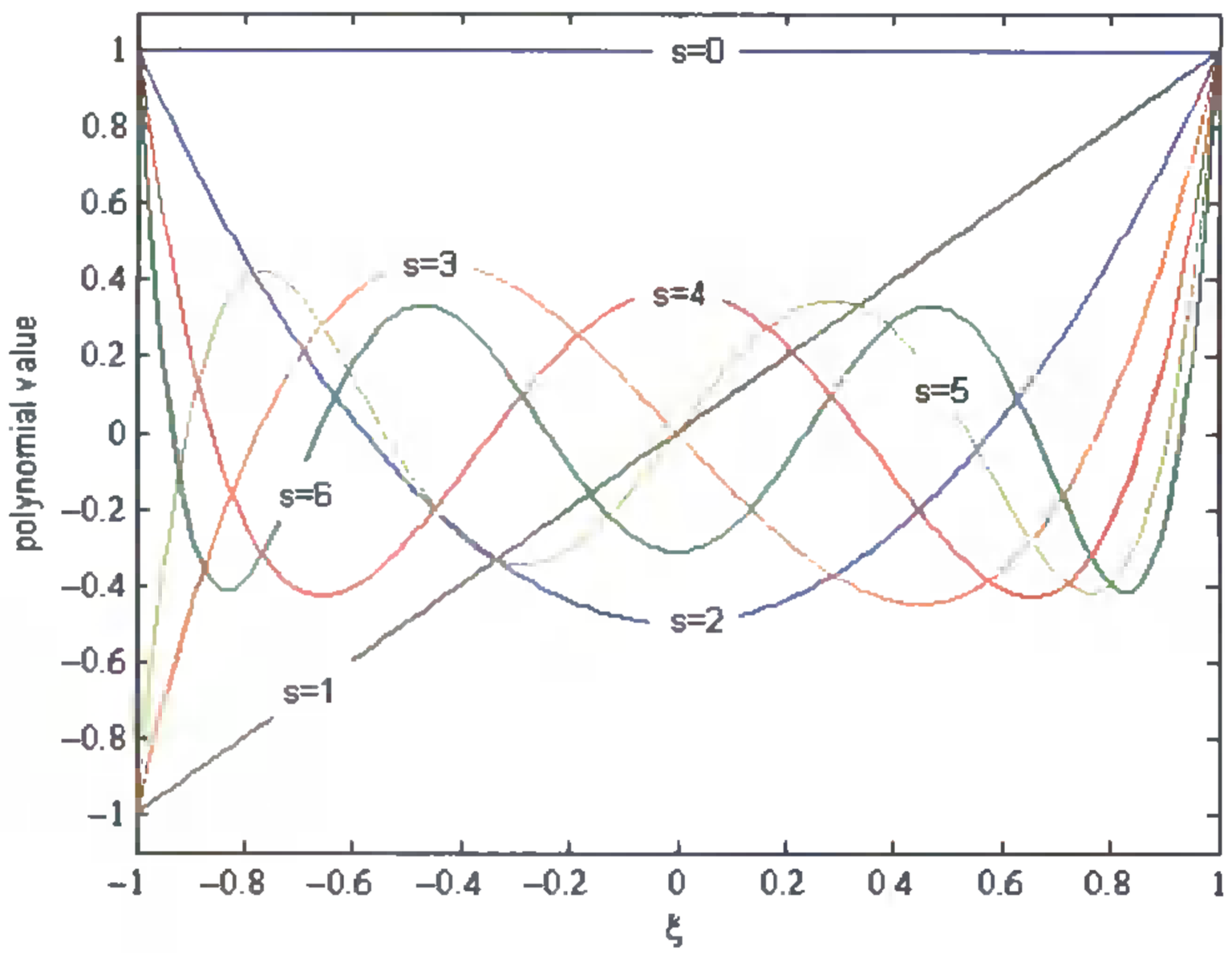


Fig. 5. Legendre polynomials $P_s(\xi_y)$ of order $s = 0,1,\ldots,6$

The Legendre polynomials $P_s(\xi_x)$ are a complete orthogonal basis set within the interval $[-1,1]$, for an order s and satisfying the following condition:

$$\int_{-1}^{1} P_r(\xi_x) P_s(\xi_x) d\xi_x = \frac{2}{2r+1} \delta_{rs}. \tag{26}$$

The Legendre polynomials are defined by:

$$P_s(\xi_x) = \frac{1}{2^s s!} \frac{d^s}{d\xi_x^s} \left(\xi_x^2 - 1\right)^s. \tag{27}$$

The inverse moment transform (only Legendre moments of order $\leq N$) which follows from the orthogonality of Legendre polynomials in the discrete domain, can be approximated by [12]:

$$I(x,y) \approx \sum_{r=0}^{N-1} \sum_{s=0}^{r} LM_{r-s,s} P_{r-s}(\xi_x) P_s(\xi_y). \tag{28}$$

The Legendre moments can be computed by Geometric moments as follows:

$$LM_{rs} = \frac{(2r+1)(2s+1)}{N^2} \sum_{k=0}^{r} \sum_{l=0}^{s} B_{rk} B_{sl} m_{kl}, \tag{29}$$

where

$$B_{rk} = \begin{cases} 0 & if \quad r-k = odd, \\ \frac{1}{2^k}(-1)^p \binom{r}{p} \binom{2r-2p}{r} & if \quad r-k = even. \end{cases} \tag{30}$$

and $p = \frac{r-k}{2}$.

3.4. ZERNIKE MOMENTS

Zernike moments (ZM) are the projection of the image $I(x,y)$ on the orthogonal basis $V_{pq}(x,y)$. The Zernike moments of order p with repetition q are defined as follows [8,25]:

$$ZM_{pq} = \frac{p+1}{\pi} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I(x,y) V_{pq}^{*}(x,y). \tag{31}$$

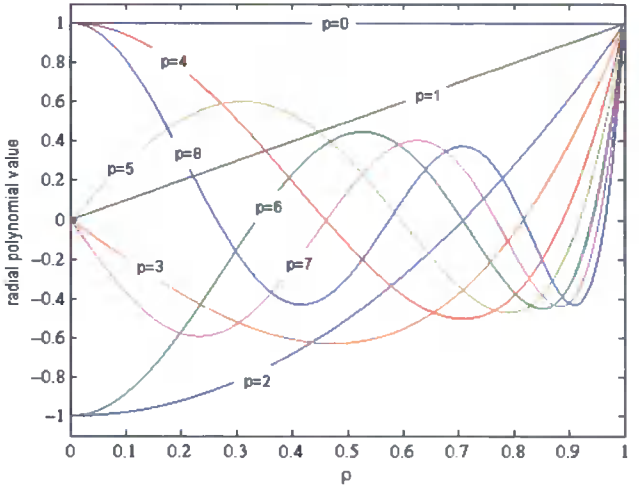


Fig. 6. Zernike polynomials of order $p = 0,1...8$ and $q = 0$ or 1

The Zernike polynomials:

$$V_{pq}(x,y) = R_{pq}(\rho)\exp(jq\theta) \tag{32}$$

are a complete set of complex valued functions orthogonal on the unit disk D: $x^2 + y^2 \le 1$, where $p \ge 0$, and $p - |q|$ is even positive integer.

The polar coordinates (ρ,θ) in the image domain are related to the Cartesian coordinates (x,y) by:

$$\rho = \sqrt{x^2 + y^2}\ , \quad \theta = \arctan(y/x)\ . \tag{33}$$

The Zernike polynomials $V_{pq}(x,y)$ are orthogonal basis set and satisfy the following condition:

$$\iint_D V_{pq}^*(x,y)V_{p'q'}(x,y)dxdy = \frac{\pi}{n+1}\delta_{pp'}\delta_{qq'}\ . \tag{34}$$

The radial polynomial $R_{pq}(\rho)$ is given by:

$$R_{pq}(\rho) = \sum_{l=0}^{(p-|q|)/2} (-1)^l \frac{(p-l)!}{l!\left(\frac{p+|q|}{2}-l\right)!\left(\frac{p-|q|}{2}-l\right)!}\rho^{p-2l}\ . \tag{35}$$

If Zernike moments of order $\le N$ are given, then the image intensity $I(x,y)$ can be approximated by [22]:

$$I(x,y) = \sum_{p=0}^{N-1}\sum_{q=0}^{p} ZM_{pq}V_{pq}(x,y)\ . \tag{36}$$

4. STEREO DISPLACEMENT SEARCH

The characteristic feature of the presented robotics vision stereo system is that axe-parallel cameras' visual plane contain different elements of the registered scene. It means that not all the elements on the left and right image have their counterparts. The lack of such a relation eliminates such elements from analysis of stereo images. However, extraction of such image areas, called margins of left and right stereo image, unambiguously characterize common (overlapping) parts of stereo images (the most important parts for stereometry image analysis). The problem of determining the margins and the common part of stereopair images is a problem of appointing the displacement vector between those images acquired from the left and right camera. The value of that vector depends on geometrical and optical parameters of the camera system.

In order to find the displacement vector d_t between stereopair images we perform calculation of the discrete orthogonal moments of the right I_P and left I_L stereo

image. We search for the displacement vector d_t by determining the minimum from the set of values:

$$\{U(0),U(1),...,U(d_{t_{\max}})\} \tag{37}$$

calculated accordingly to (38) characterizing adequately subtraction of reconstructed images I_L and I_P regarding to maximal displacement vector $d_{t_{\max}}$

$$U(d_t)=\sum_{x=0}^{N-d_t}\sum_{y=0}^{N-1}\left|I_L(x+d_t,y)-I_P(x,y)\right|+\sum_{x=0}^{d_t-1}\sum_{y=0}^{N-1}I_L(x,y)+\sum_{x=N-d_t}^{N-1}\sum_{y=0}^{N-1}I_P(x,y), \tag{38}$$

where $I_L(x+d_t,y)$ and $I_P(x,y)$ can be calculated on the basis of reconstructed intensity function values, for the Chebyshev (18), Legrande (28) and Zernike (36) moments, respectively.

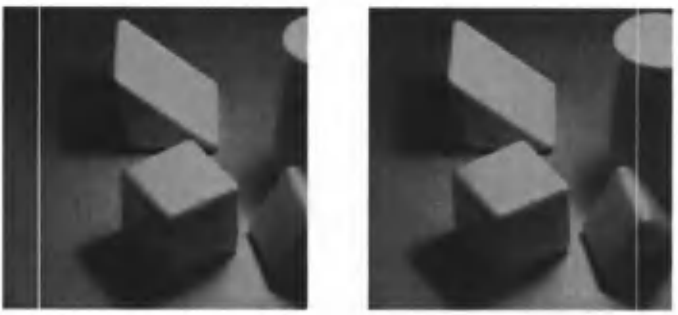


Fig. 7. The stereoscopic image „blocks" with the marked margins

5. MATCHING AND STREO DISPARITY CALCULATION IN MOMENTS SPACE

5.1. DISPARITY ESTIMATION USING NUMERICAL ANALYSIS

Stereo disparity d_x describes the difference in absolute localizations of the corresponding points in the epipolar lines. By analysing corresponding points of the left $I_L(x+d_t+d_x,y_e)$ and righ $I_P(x,y_e)$ images of stereopair we can define the following relation of their intensity function:

$$I_P(x,y_e)\cong I_L(x+d_t+d_x,y_e). \tag{39}$$

By utilizing the reconstruction formulas (18), (28) and (36) of the intensity function on the basis of the calculated moments for the stereopair images and by simplifying the factor dependent on y_e (characterizing the epipolar lines) we can denote:

– for the Chebyshev moments:

$$\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left(TM_{mn}^{(P)}t_m(x)-TM_{mn}^{(L)}t_m(x+d_t+d_x)\right)\cong 0, \tag{40}$$

– for the Legendre moments:

$$\sum_{r=0}^{N-1}\sum_{s=0}^{r}\left(LM_{r-s,s}^{(P)}P_{r-s}\left(\frac{2x-N+1}{N-1}\right)-LM_{r-s,s}^{(L)}P_{r-s}\left(\frac{2(x+d_t+d_x)-N+1}{N-1}\right)\right)\cong 0, \tag{41}$$

– for the Zernike moments:

$$\sum_{p=0}^{N-1}\sum_{q=0}^{p}\left(ZM_{pq}^{(P)}V_{pq}(x,y_e)-ZM_{pq}^{(L)}V_{pq}(x+d_t+d_x,y_e)\right)\cong 0. \tag{42}$$

Solving equations (40-42) in order to determine the stereo disparity in the analytical way is not possible. Therefore, in practice numerical methods have to be used. Each equation can be written as:

$$F(d_x;x)=0 \tag{43}$$

or for the Zernike moments (42) as:

$$F(d_x;x,y_e)=0. \tag{44}$$

We use the known iterative method of Newton-Rahpson to calculate, with the desired accuracy, stereoscopic disparity d_x which is the argument of the function for given parameters x and y_e on the basis of the following formula:

$$d_x^{(i+1)}=d_x^{(i)}-\frac{F\left(d_x^{(i)}\right)}{F'\left(d_x^{(i)}\right)}, \tag{45}$$

where F' is a derivative of F (45) in the i-th iteration.

The starting parameter of the method is in our case $d_x^{(0)}=0$. The achieved values of d_x are verified by the following conditions and properties: d_x is an integer number from the $\langle -d_{\max}, d_{\max}\rangle$, where $d_{\max}$ is a maximal possible value of the stereo disparity for a given pair of stereo images.

5.2. STEREO MATCHING BASED ON CORRELATION OF MOMENTS

In order to determine the stereo disparity d_x of the common part of the stereopair, the corresponding points on the epipolar lines have to be found in the process of stereo matching. In practice, in this approach, we search for the correlation between reconstructed intensity functions (18), (28) and (36) of the images I_L and I_P in the regions bounded by the window function:

$$W_z(x,y_e)=\left\{u,v\left|\begin{array}{l} x-\frac{z}{2}\le u\le x+\frac{z}{2} \\ y_e-\frac{z}{2}\le v\le y_e+\frac{z}{2}\end{array}\right.\right\}. \tag{46}$$

where z characterizes the size of the window function, and (u,v) are the coordinates describing the localization of window $W_x(x,y_e)$.

The correlation matching process is realized in the common part of stereopair (determined by the vector d_t) according to the following steps:

- for each point of the right image $I_P(x,y_e)$ choose its neighbourhood by the window function $W_x(x,y_e)$, where (x,y_e) is the centre of the window W_x,
- for all the points from the linear neighbourhood of the left image $I_L(x)$ choose its neighbourhood by the window function $W_x(x+d_x,y_e)$, where d_x characterizes the stereo disparity interval and d_x is within $\langle -d_{\max}, d_{\max}\rangle$,
- for the determined points and their neighbourhoods search for the minimum of the function C_{SSD}:

$$C_{SSD}(d_x)=\sum_{(u,v)\in W_z(x,y)}\left[I_P(u,v_e)-I_L(u+d_x,v_e)\right]^2. \tag{47}$$

The minimum of the C_{SSD} function determines the value of stereo disparity d_x for the matched points of the left and right stereopair image.

5.3. STEREO MATCHING BASED ON THE SIMILARITY OF VECTORS IN THE MOMENTS SPACE

An alternative approach to correlation matching method based on function (48) minimum search can be a similarity search in the feature vector space according to:

$$\Psi(d_x)=\min_{d_x}\left\|\lambda_x^{(L)}-\lambda_{x+d_x}^{(P)}\right\|, \tag{48}$$

where $\lambda_x^{(L)}$ is a vector consisting of moments values $\phi_i^{(L)}$ of the intensity function in a given window W_x with the centre in point (x,y_e) on the left image of stereopair given by:

$$\lambda_x^{(L)}=\left[\phi_1^{(L)},\phi_2^{(L)},\ldots,\phi_i^{(L)},\ldots\phi_M^{(L)}\right]_{\phi_i\in W_z(x,y_e)} \tag{49}$$

And $\lambda_{x+d}^{(P)}$ is a vector consisting of moments values $\phi_i^{(P)}$ of the intensity function in a given window W_x with the centre in point $(x+d,y_e)$ on the right image of stereopair given by:

$$\lambda_{x+d_x}^{(P)} = \left[\phi_1^{(P)}, \phi_2^{(P)}, ..., \phi_i^{(P)}, ... \phi_M^{(P)}\right]_{\phi_i \in W_z(x+d_x, y_e)}. \tag{50}$$

The similarity measure between the vectors calculated in the moment space according to equations (12), (24) and (31) is calculated on the basis of the Euclidean distance. The obtained minimum of the similarity measure $\Psi(d_x)$ determines the stereo disparity d_x between points (x, y_e) on the left image and $(x+d_x, y_e)$ on the right stereo image.

6. EXPERIMENTAL RESULTS

In the experiments we used stereo images of resolution 512×512 and 256 greyscale levels. All the images were acquired by the well-calibrated axe-parallel camera system. Sample test images are presented in Fig. 2.

Before disparity estimation we calculate the displacement vector d_t between each image stereopair (in other words, we find the common part of stereo images). This stage is based on search of the global minimum of the function $U(d_t)$ in the interval $< 0; d_{t_{\max}} >$, where $d_{t_{\max}}$ is set as ¼ of the image resolution. The ideal displacement vector for the presented sample image blocks is 52 and was equal to the one obtained by the Chebyshev reconstruction formula (18). The results achieved for other moments were characterized by small errors in the case of the Zernike moments (36), and large errors for the Legendre moments (28). Such a situation was caused by imprecise approximations in the reconstruction formulas and by the larger sensitivity of Legendre moments on intensity function deformations in the process of stereoscopic projections. Moreover, the errors were caused by the high orders of the used moments.

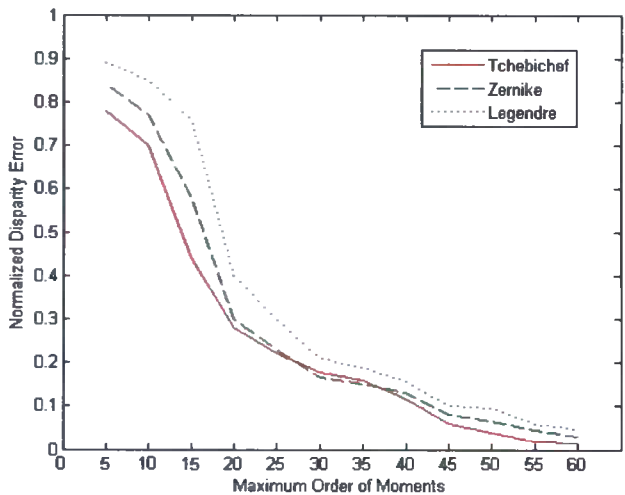


Fig. 8. Stereo disparity calculation error on the basis of the correlation of moments

Stereo disparity d_x search by numerically solving the numerical equations requires the estimation of derivative of function F (e.g. by differential quotient). Moreover, it requires the verification of the obtained results on the basis of demanded intervals of their changes. In the experiments we found out that we can often achieve the proper values of stereo disparity d_x after a small number of iterations if the starting value is set to $d_x^{(0)} = 0$. Even though, in general, this method is not computationally effective, it could be used as a verification tool for results obtained by other methods of stereo disparity calculation.

In order to verify the proposed methods of stereo disparity calculation (sections 5.2-5.3), we use the normalized disparity error, given by:

$$NDE = \frac{1}{N \times N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \frac{\left| d_x(x, y_e) - \hat{d}(x, y_e) \right|}{d_{x_{\max}}}, \tag{51}$$

where $d_x(x, y_e)$ is the calculated stereo disparity in the specified point (x, y_e) and $\hat{d}(x, y_e)$ is the ideal stereo disparity calculated by other methods (e.g. correlation matching) and $d_{x_{\max}}$ is the maximal stereo disparity for the given image.

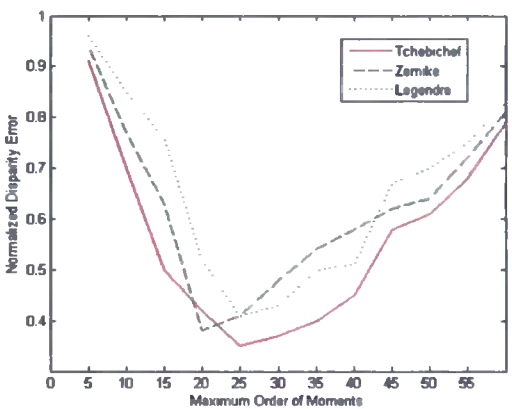


Fig. 9. Stereo disparity calculation error on the basis of the similarity of vectors in the feature space

Due to such formula (51), the normalized stereo disparity error values are varying within the interval $<0;1>$. In the Fig. 8. we presented the influence of the number of the used moments on the NDE. It decreases with the larger number of the used moments. Such situation reflects the impact of moment order on the reconstruction precision of the image intensity function (eq. 18, 28, 36 for different moments).

The results of stereo disparity estimation method implementation on the basis of the vectors similarity in the moments space and an interesting phenomenon of moments order adjustment are shown in the Fig. 9. The obtained value of NDE is optimal for the moments of order 20-25. Then, with the higher order of moments the results become worse, which is caused by the increased distance between vectors (48) in the Euclidean space.

7. SUMMARY

In the article, we presented the idea and implementation of using discrete orthogonal moments of Chebyshev, Legendre and Zernike in the process of matching and stereo disparity estimation. In order to optimalize those procedures, in the first stage we extracted the common parts of stereo images (which is important for matching) and the margins of stereo images.

In the article, three approaches to the problem of stereo disparity estimation were presented and tested. In the first method, we performed the estimation of stereo disparity d_x by numerically solving equations (40-42). The second approach was based on the correlation analysis of the reconstruction of image intensity function on the basis of discrete orthogonal moments. In the third approach, the problem of stereo disparity estimation was solved by similarity search in the vector space for the calculated moments characterizing the corresponding points of stereo images.

In the described methods we used the discrete orthogonal Chebyshev, Legendre and Zernike moments. After experiments we concluded that Chebyshev and Zernike were the most appropriate for stereo estimation moments, respectively. Much worse results were achieved by Legendre moments.

In the further work, we experiment with other discrete orthogonal moments applied to stereo disparity estimation task. We also try to analytically estimate the influence of moments values on stereo disparity d_x .Such a value of d_x could be used for optimalization and enhancement of the stereopair matching process.

BIBLIOGRAPHY

[1] T. Arif, Z. Shaaban, L. Krekor, S. Baba, 2005: Object Classification via Geometrical, Zernike and Legendre Moments, Journal of Theoretical and Applied Information Technology, vol. 7, no. 1, pp. 31-37.

[2] T. Andrysiak, M. Choraś, 2005: Multiresolution Matching and Disparity Calculation of Stereo Images in Frequency Domain, Information Extraction and Processing, no. 23 (99).

[3] T. Andrysiak, M. Choraś, 2005: Stereo Matching for Robotics Vision, Proc. of International Conference PELINCEC.

[4] S. T. Barnard, W. B. Thompson, 1980: Disparity analysis of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2, pp. 33-340.

[5] S.T. Barnard, A. Fischler, 1982: Computational stereo, ACM Computing Surveys, vol 14, no. 4, pp. 553-572.

[6] M.Z. Brown, D. Burschka, G.D. Hager, 2003: Advances in computational stereo. TPAMI, vol. 25, no.8, pp. 993-1008.

[7] T.S. Chihara, 1978: An Introduction to Orthogonal Polynomials. New York: Gordon and Breach.

[8] C.W. Chong, P. Raveendran, R. Mukundan, 2003: A comparative analysis of algorithms for fast computation of Zernike moments, Pattern Recognition., vol. 36, pp. 731-742.

[9] J. Fluseer, B. Zitova, T. Suk, 2009: Moments and Moment Invariants in Pattern Recognition, Wiley Knowledge For Generations.
[10] M.H. Han, S. Rhee, 1992: Camera calibration for three-dimensional measurement, Pattern Recognition, vol. 25, no. 2, pp. 155-164.
[11] Z. Hong, J. Yang, 1993: An algorithm for camera calibration using a three dimensional reference point, Pattern Recognition, vol. 26, no. 11, pp. 1655-1660.
[12] K.M. Hosny, 2007: Efficient computation of Legendre moments for gray level images, International Journal of Image and Graphics, vol. 7, no. 4, pp. 735-747.
[13] S.X. Liao, M. Pawlak, 1996: On image analysis by moments, IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 18, pp. 254-266.
[14] R. Mukundan, 2001: Image Analysis by Chebyshev Moments, IEEE Transactions on Image Processing, vol. 10, no. 9, 1357-1364.
[15] R. Mukundan, 2004: Some Computational Aspects of Discrete Orthogonal Moments, IEEE Transactions on Image Processing, vol. 13, no. 8, pp. 1055-1059.
[16] R. Mukundan, 2004: A New Class of Rotational Invariants Using Discrete Orthogonal Moments, IASTED Conference Signal and Image Processing, pp. 80-84.
[17] R. Mukundan, S.H. Ong, 2001: Lee P.A., Image Analysis by Chebyshev moments, IEEE Transactions on Image Processing, vol. 10, no. 9, 1357-1364.
[18] R. Mukundan, A. Pang, N. Khee, 2002: Stereo Image Analysis: A New Approach Using Orthogonal Moments, Proceedings of Asian Technology Conference in Mathematics, 513-522.
[19] R. Mukundan, K.R. Ramakrishnan, 1998: Moment Functions in Image Analysis – Theory and Applications, World Scientific.
[20] W. Mokrzycki, 1994: Constraction of 3D deph map from binocular stereo, Proceedings of 5th International School on Computer Vision & Graphics Microcomputer.
[21] Y. Pew-Thian, R. Paramesran, 2005: An efficient method for the computation of Legendre moments, vol.27, no. 12, pp. 1996-2002.
[22] H. Sun-Kyoo, K. Whoi-Yul, 2006: A novel approach to the fast computation of Zernike moments, Pattern Recognition vol. 39 pp. 2065-2076.
[23] M.R. Teague, 1980: Image analyis via the general theory of moments, J.Opt. Soc. Amer., Transactions on Pattern Analysis and Machine Intelligence vol. 70, no. 8, pp. 920-930.
[24] Ch. Teh, R.T. Chin, 1988: On Image Analysis by the Methods of Moments, IEEE , vol. 10, no. 4, 496-512.
[25] A. Wallin A, 1995: Complete sets of complex Zernike moment invariants and the role of pseudo invariants, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, pp. 1106-1110.
[26] G.Q. Wei, S.D. Ma, 1994: Implicit and explicit camera calibration: theory and experiments, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.16, no. 5, pp. 469-480.
[27] J. Weng, P. Cohen, M. Herniou, 1992: Camera calibration with distortion models and accuracy evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 10, pp. 965-980.
[28] U.R.Dhond,J.K. Aggarwal, 1989: Structure from stereo – a review, IEEE Transactions Systems Man and Cybern. vol. 19, no. 6, pp.1489-151.

PASOWANIE I OBLICZANIE NIEZGODNOŚCI STEREOSKOPOWEJ NA PODSTAWIE DYSKRETNYCH MOMENTÓW ORTOGONALNYCH CHEBYSHEVA, LEGENDRE'A I ZERNIKE'A

Streszczenie

W artykule przedstawiono teoretyczne i eksperymentalne podejścia do problemu pasowania i oceny niezgodności stereoskopowej. Zaproponowano realizacje obliczeń niezgodności stereoskopowej w przestrzeni momentów otrogonalnych, jak również przedstawiono podstawy do obliczeń numerycznych i metod opartych na korelacji. W celu obliczania wektora niezgodności zdecydowano się na użycie dyskretnych momentów ortogonalnych Chebysheva, Legendre'a i Zernike'a. W procesie badawczym oceny niezgodności stereoskopowej wszystkie proponowane momenty były testowane i porównywane. Wyniki badań potwierdzają skuteczność prezentowanych metod określania niezgodności i pasowania stereoskopowego dla zastosowań widzenia maszynowego.

Słowa kluczowe: momenty ortogonalne, pasowanie stereo, niezgodność stereoskopowa

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 258
TELEKOMUNKACJA I ELEKTRONIKA 14 (2011), 33-45

ANALYSIS OF LOSS AND SINGLE-MODALITY OF THERMALLY DIFFUSED AREAS IN SPLICED JOINTS OF SINGLE MODE TELECOMMUNICATION FIBERS

Marek Ratuszek, Zbigniew Zakrzewski,
Jacek Majewski, Małgorzata Ratuszek

Institute of Telecommunications,
Faculty of Telecommunications and Electrical Engineering
University of Technology and Life Sciences
Al. S. Kaliskiego 7, 85-789 Bydgoszcz, Poland

Summary: In this work diffusion processes in thermally connected cylindrical fibers with weakly guiding and circular cross-section, that is telecommunication fibers, have been presented. There have been discussed diffusion distributions of the core dopant of fibers spliced in $t \approx 2000°C$. Gaussian approximations of the core dopant concentration distribution and refractive index in the connecting area of single mode telecommunication fibers have been presented. Theoretical analysis of propagation and loss characteristics for thermally diffused expanded-core (TEC) of single mode telecommunication fibers has been performed, as well. It has been shown that because Gaussian profile approximates well diffusion distribution, this thermally diffused core area (splice) remains single-modal.

Keywords: single mode optical fibers, thermally diffused expanded-core, dopant distribution, refractive index profile, Gaussian approximation

1. INTRODUCTION

Optimization of spliced joints of telecommunication fibers of different types involves thermal diffusion of the connected fibers core dopant (most often GeO_2) in such a way that the mode field radiuses in the thermally diffused expanded-core (TEC) – the splice intermediate area, will equalize [5,7,12,13] – fig. 1. The loss of such area is small if it remains single-modal.

The purpose of this work is an analysis of diffusion processes within the splice, their approximation by Gaussian distributions, and evaluation of diffusion optimal times, as well as mainly theoretical analysis of losses occurring in the transit area resulting from mismatch of mode fields and TEC area dimensions, and finally comparison of the analysis results with experimental data. Second goal of this work is to proof, that TEC area after diffusion process remains single-modal.

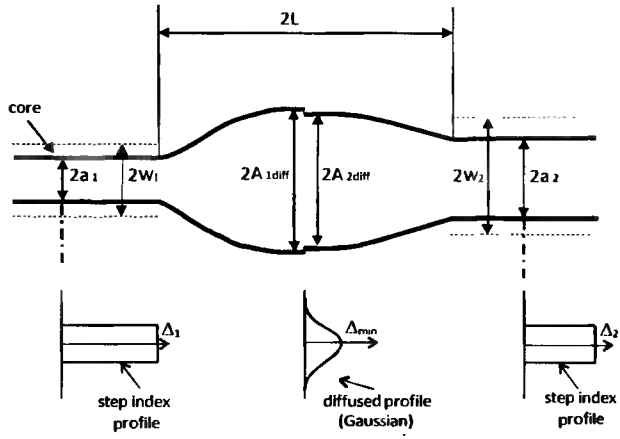


Fig. 1. Schematic illustration of thermally diffused expanded-core of the splicing connected fibers, where: a – core radius, W – mode field radius before diffusion, A_{diff} – core radius after diffusion, $2L$ – length of the TEC area

2. DIFFUSION PROCESSES IN THE TRANSIT AREA OF THE CONNECTED FIBER

In fig. 2, 3 calculations of diffusion distributions are presented, their Gaussian approximation of GeO_2 concentration in SiO_2 for five splice times (in seconds): $t_1 = 2$ s, $t_2 = 3$ s, $t_3 = 5$ s, $t_4 = 7$ s, $t_5 = 10$ s , and fusion splicing temperature $t = 2000°C$.

For distributions from fig.2 diffusion from layer with limited thickness was assumed, the core thickness was $h = 2a = 8\mu m$, dopant (GeO_2) concentration (before diffusion) $N_0 = 6.79 \times 10^{26}\,m^{-3}$ which corresponds to the refractive index coefficient in the core $n_r = 1.45149$, refractive index coefficient in the cladding (undoped SiO_2) was $n_p = 1.44680$ – n_r and n_p were calculated with the use of Sellmeier formula [4]. The above presented parameters can correspond to single mode telecommunication fibers with step refractive index profile SI SMF of the type G.652 [8] with $\Delta[\%] = \frac{n_r^2 - n_p^2}{2n_r^2} 100\% = 0.32\%$ and numerical aperture $NA = 0.116$ (the above presented parameters were calculated for $\lambda = 1.31\mu m$). For calculations a diffusion coefficient $D = 5 \times 10^{-13}\,m^2/s$ was accepted [7,13].

For distributions from fig. 3 also diffusion from layer with limited thickness was assumed, but the core thickness was $h = 2a = 6\,\mu m$, dopant concentration (before diffusion) $N_0 = 1.27 \times 10^{27}\,m^{-3}$ which corresponds to the refractive index coefficient in the core $n_r = 1.45569$, refractive index coefficient in the cladding (undoped SiO_2) was $n_p = 1.44680$ – n_r and n_p were calculated with the use of Sellmeier formula [4]. The above presented parameters can correspond to single mode telecommunication fibers

with step refractive index profile SI SMF of the type G.653 or G.655 [9,10] with $\Delta[\%] = \frac{n_r^2 - n_p^2}{2n_r^2} 100\% = 0.61\%$ and numerical aperture $NA = 0.161$ (the above presented parameters were calculated for $\lambda = 1.31\,\mu m$). For calculations, in this case, diffusion coefficient $D = 10^{-12}\,m^2/s$ [7,13] was accepted as along with higher dopant concentrations an increase of the diffusion coefficient can be observed [2].

Diffusion distributions were calculated with the use of an expression of diffusion from a layer with limited thickness $h = 2a$ – the core diameter:

$$N(r,t) = \frac{N_0}{2} \cdot \left(\mathrm{erf}\left(\frac{r+h}{2 \cdot \sqrt{D \cdot t}} \right) - \mathrm{erf}\left(\frac{r}{2 \cdot \sqrt{D \cdot t}} \right) \right) \quad (1)$$

where: N_0 – core dopant concentration before diffusion,
h – layer thickness (core diameter),
D – diffusion coefficient,
t – diffusion time,
r – distance from the core center.

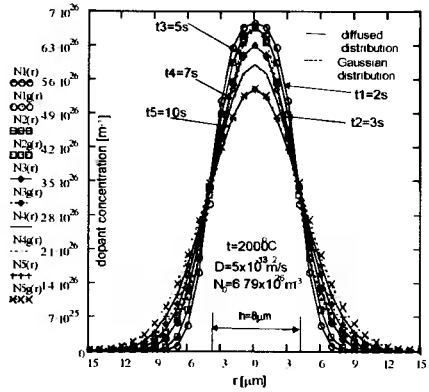


Fig. 2. Dopant diffusion distributions and their Gaussian approximation for SI SMF of G.652 type

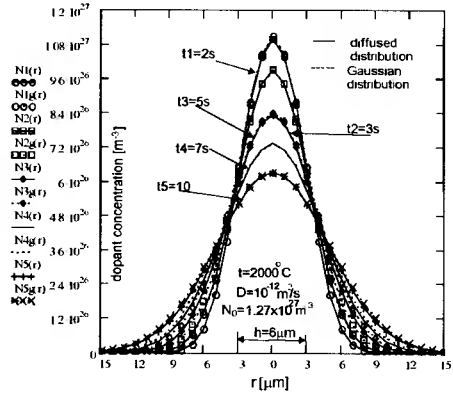


Fig. 3. Dopant diffusion distributions and their Gaussian approximation for SI SMF of G.653 or G.655 type

Gaussian distributions calculated using an expression:

$$N(r,t) = N_0(0,t) \exp\left(\frac{-r^2}{A_{diff}^2} \right) \quad (2)$$

where: $N_0(0,t)$ – concentration in the core center decreasing over time, $A_{diff}^2 = a^2 + 4Dt$ was defined in the work [11], here A_{diff} is the core diameter after diffusion where the dopant concentration decreases by e times.

Good consistence of diffusion and Gaussian distributions was established, it was the better with the longer diffusion time and greater diffusion coefficient – fig.2 and fig. 3.

2.1. Refractive index profiles

Dopant distribution after diffusion is consistent with Gaussian distribution which leads to Gaussian distribution of the refractive index:

$$n^2(r) = n_p^2 + (n_r^2 - n_p^2)\exp\left(\frac{-r^2}{A_{diff}^2}\right) \tag{3}$$

where: n_r is the value of the refractive index in the core center ($r = 0$) after diffusion.

Refractive index profiles for five diffusion times (splicing), calculated according to expression (3) and corresponding to the dopant distributions – fig. 2 and fig. 3 are presented in fig.4 and fig.5.

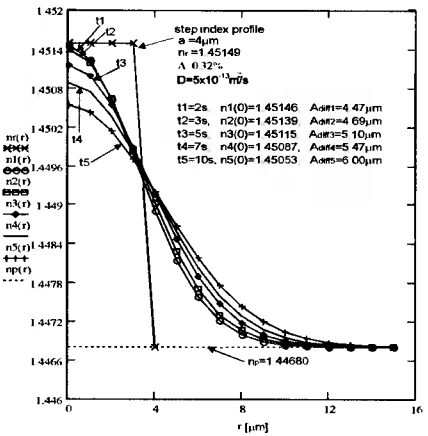


Fig. 4. The core refractive index profiles in SI SMF of the type G.652 for five diffusion (splicing) times, $\lambda = 1.31\,\mu m$

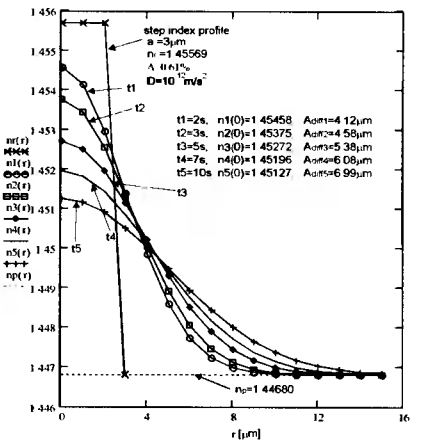


Fig. 5. The core refractive index profiles in SI SMF of the type G.653 or G.655 for five diffusion (splicing) times, $\lambda = 1.31\,\mu m$

2.2. Matching of mode field radiuses

If the mode fields of the connected fibers are not totally equalized then, apart from transmission losses of TEC area there also occur losses A due to mismatch of the mode fields. With the assumption that the fundamental mode field distribution LP_{01} can be approximated with the Gaussian distribution – which is true for step and Gaussian refractive index profiles [1,4,12] we receive:

$$A = -10\log\left[\left(2W_{diff1}W_{diff2}/(W_{diff1}^2 + W_{diff2}^2)\right)^2\right] \tag{4}$$

where: W_{diff1}, W_{diff2} mode field radiuses of the connected fibers in the connection place (here after diffusion , though the formula is of general character), in TEC.

Splicing connecting practice of fibers with significantly different parameters e.g. G.652 and G.655 indicates that it is possible to achieve loss of these fibers lower than 0.08 dB [7] using dopant diffusion within the splice (TEC).

It is optimal, however, to base upon calculations, thus with the above given diffusion coefficients and thermal connecting temperature it is advisable to match optimal splicing times after diffusion on the basis of the core radius values, mode fields radiuses and TEC transmission parameters.

3. ANALYSIS OF THE TRANSIT AREA TRANSMISSION PROPERTIES OF THE SPLICED FIBERS

Even with an assumption of an ideal matching of W_{diff1} and W_{diff2} there will occur the TEC transmission losses $-10\log[T_f(L,\gamma_{max})]$ [11] – fig. 1.

Dependency curves $-10\log[T_f(L,\gamma_{max})]$ show the maximum for the TEC for characteristic length $L_{c\cdot}$:

$$L_{c\cdot} = \left[\gamma_{max}(\gamma_{max}-1)\right]^{1/2}\frac{\pi n_r W}{\lambda} \tag{5}$$

where: $2L$ length of the TEC area fig. 1,

$\gamma_{max} = A_{diff}/a$,

W - mode field radius before diffusion.

Loss of the transit area is of course lower than the maximum if its length $L > L_{c\cdot}$ or $L < L_{c\cdot}$, and significantly smaller $L >> L_{c\cdot}$ or $L << L_{c\cdot}$ [3,11].

The fact that loss is small if $L >> L_{c\cdot}$ or $L << L_{c\cdot}$ causes that the transmission losses in TEC area can be described by two different processes – transmission models. When $L << L_{c\cdot}$ the transit area is an equivalent of the gap between connected fibers [3,11], when $L >> L_{c\cdot}$ a phase front transformer model is used [11]. When the TEC area length is $L >> L_{c\cdot}$ or $L << L_{c\cdot}$ then the transmission loss is negligibly small which is the most desirable case for optimization of thermal fibers connecting process e.g. fusion splicing. The choice of temperature and thermal connecting times in such a way that good matching of the mode field radiuses and simultaneously the smallest losses connected with TEC dimensions are due to occur finishes the process of thermal connecting optimization.

Loss TEC area in gap model does not depend on γ_{max}. Thus, optimization of losses due to different W is not practically possible, which does not remain in consistence with experimental data [6]. So the model is useless in case of optimization by thermal diffusion method. Therefore, it is necessary to conduct the diffusion process so that the TEC area length will be $L >> L_{c\cdot}$. Loss of the TEC area in the function L for a gap model and joints G.652-G.652 with parameters accepted in the work, are presented in fig. 6.

Loss of the TEC area in function L, for a phase-front transformer for G.652-G.652 and G.655-G.652 joints models with parameters accepted in this work, and five diffusion times are shown in fig.6 and fig.7 respectively. The results presented in fig. 6 and indicate that neglecting mode field matching, the shortest diffusion time the smaller TEC areas loss. $L_{c\cdot}$ can be defined approximately from the intersection of curve presented on fig. 6.

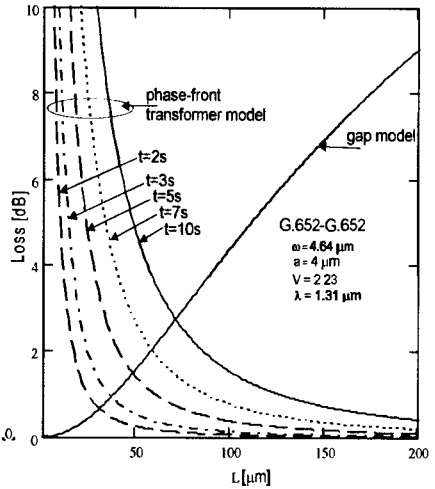


Fig. 6. Loss of the TEC in the function L for joining the same type of fibers (652-G.652) for the gap model and the phase-front transformer model $\lambda = 1.31\,\mu m$

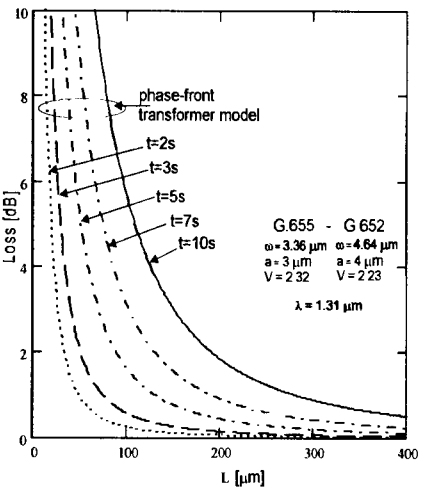


Fig. 7. Loss of TEC areas in the function of length L for joining fibers of the types G.655, G.653 & G.652 – phase front transformer model $\lambda = 1.31\,\mu m$

Length $2L$ of the TEC area during optimization of the splicing process (splicer Ericsson FSU 925RTC) was evaluated on the basis of thermoluminescence of the spliced fibers. Fig.8 presents a picture of thermoluminescence intensity profile, along the cores, of the fusion spliced fibers G.652 & G.655. The splicing temperature in the center was $t \approx 2000°C$, splicing time $t_2 = 3\,s$. The temperature was decreasing from the splice center until it reached room temperature, where the thermoluminescence line changes into a horizontal one from the left to the right side of the splice. The distance which can be identified as $2L$ is 675μm and is considerably longer than L_c.

Total losses of the TEC area are a sum of losses due to mismatch of mode field radiuses and resulting from transmission losses. In table 1 total losses $A_\Sigma = -10\log\left[\left(2W_{diff1}W_{diff2}/\left(W_{diff1}^2 + W_{diff2}^2\right)\right)^2\right] - 10\log T_f$ [7] for the assumed by the authors diffusion times and connecting G.655-G.652 have been presented. Results presented in table 1 are correct when splice area after diffusion (TEC area) remain single-modal.

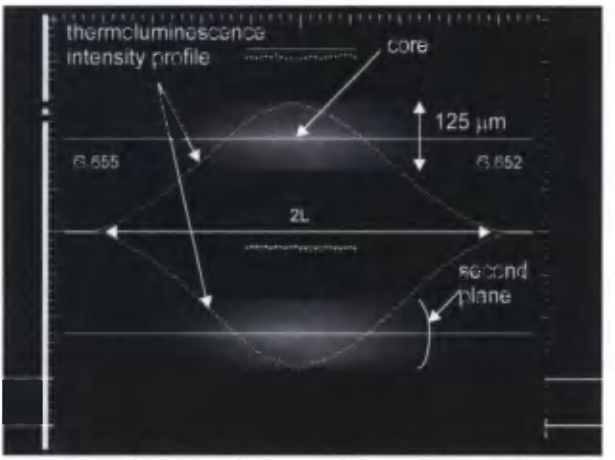


Fig. 8. Exemplary thermoluminescence intensity profile of fusion spliced fibers G.652 & G.655

Table 1. Values A_{Σ}[dB] for $\lambda = 1.31$ μm and $\lambda = 1.55$ μm

Diff. Time [s]	A_{Σ}[dB] $\lambda = 1.31$ μm	A_{Σ}[dB] $\lambda = 1.55$ μm
2	0.070	0.084
3	0.058	0.074
5	0.163	0.220
7	0.362	0.486
10	0.768	1.027

4. ANALYSIS OF SINGLE-MODALITY OF THERMALLY DIFFUSED AREAS

4.1. STEP INDEX PROFILES – PROFILES OF EQUAL VOLUME

Expression (6),

$$n^2(R) = n_r^2[1 - 2\Delta f(R)] \tag{6}$$

in general case, describes profile of refractive index of waveguide fiber with circular section – fig. 9.

The same distribution for optical fiber with constant value of cladding refractive index can be expressed as given below [12]:

$$n^2(R) = \begin{cases} n_p^2 + \left(n_r^2 - n_p^2\right)\left[1 - 2\Delta f(R)\right] & \text{for } 0 \le R < 1 \\ n_p^2 & \text{for } 1 \le R < \infty \end{cases} \tag{6'}$$

where: $R = \dfrac{r}{a}$ – normalized radius,

a – core radius,

r – distance from core center,

n_r – value of core refractive index (for $r = 0$),

n_p – value of cladding refractive index,

$\Delta = \dfrac{n_r^2 - n_p^2}{2n_r^2}$,

$f(R)$ – function of changes of refractive index.

The function $[1 - f(R)]$ describes profile form. For all profiles with equal values of Δ and n_p, the so-called effective volume of profile Ω [12]

$$\Omega = 2\pi a^2 \int_0^\infty \left[1 - f(R)\right] R dR \tag{7}$$

can be calculated, where a is a core radius or characteristic dimension of profile.

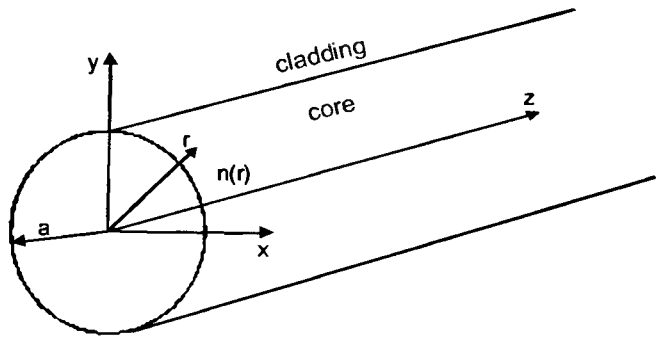


Fig. 9. Cross section of waveguide fiber with circle symmetry: fiber sizes in directions r and z are unlimited, core radius – a, refractive index profile – $n(r)$, and here symmetry axis is identical with z axis

Profiles with equal Δ have equal effective volume, but, in general case, different a. For our analysis, as the base, let us choose the profile of refractive index, and let us calculate Ω for it. The power profile (8) when limited to the core, that is for $q = \infty$, behaves as a step profile

$$n^2(R) = \begin{cases} n_r^2(1 - 2\Delta R^q) & \text{for } 0 \le R < 1 \\ n_p^2 = n_r^2(1 - 2\Delta) & \text{for } 1 \le R < \infty \end{cases} \tag{8}$$

Therefore for the step profile we obtain

$$\begin{aligned} f(R) &= 0 \quad \text{for} \quad 0 \le R < 1 \\ f(R) &= 1 \quad \text{for} \quad 1 \le R < \infty \end{aligned} \tag{9}$$

Hence, the effective volume of the step profile is given by

$$\Omega = 2\pi a_S^2 \int_0^1 R dR = \pi a_S^2 \tag{10}$$

where $a_S = a_{seff} = a$ is a core radius for optical fiber with the step profile of refractive index (it is equal to the effective radius).

Meanwhile, for the power profile limited to the core $2a_{peff}$ (fig.3), the effective volume can be expressed as

$$\Omega = 2\pi a_{peff}^2 \int_0^1 (1 - R^q) R dR = \pi a_{peff}^2 \frac{q}{q+2} \tag{11}$$

where $R = \frac{r}{a_{peff}}$.

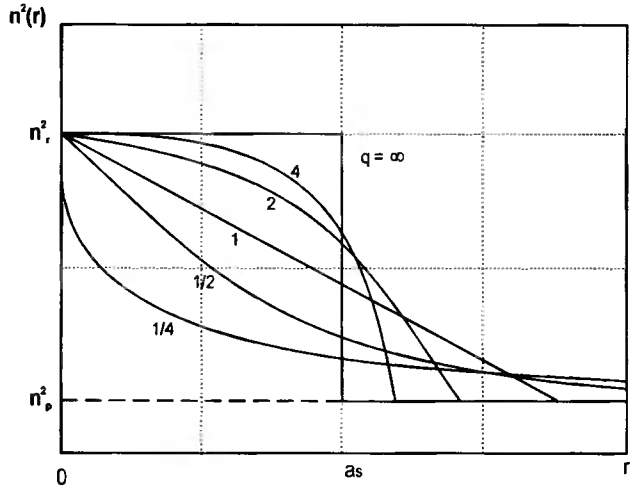


Fig. 10. Refractive index changes for step profile and power profiles under assumption of equal effective volumes

From (10) and (11), we get

$$a_{peff} = a\left[\frac{q+2}{q}\right]^{1/2} \tag{12}$$

The refractive index changes for the step profile and power profiles provided that equal effective volumes are presented in fig.10.

4.2. Cut off frequency for profiles with equal volumes

Cut off frequency V_c' for any form of profile with equal volume can be expressed by a_{eff} and V_c [12]

$$V_c' = \left(\frac{a_S}{a_{eff}}\right) V_c \tag{13}$$

where V_C is cut off frequency, which describes single-modal threshold for any form of profile, a_{eff} is core radius or characteristic dimension of profile with equal volume.

For the power profile of equal volume, this expression assumes the following form

$$V_C' = \left(\frac{a_S}{a_{peff}}\right) V_C = \left(\frac{q}{q+2}\right) V_C \tag{14}$$

For the step profile $q = \infty$ and $V_C' = V_C = 2.405$. V_C' is in practice independent of the profile form. For example, for power profile with changes of q from 1 to ∞, with good accuracy (5.1%) [12], it can be assumed that $V_C' = 2.405$.

Generally, for different profile form of refractive index with the same volume, we get (in comparison to step profile):

$$2\pi a_{eff}^2 \int_0^{\infty} [1 - f(R)] R dR = 2\pi a_S^2 \int_0^1 R dR \tag{15}$$

and

$$\frac{a_{eff}}{a_S} = \left[\frac{\int_0^1 R dR}{\int_0^{\infty} [1 - f(R)] R dR} \right]^{1/2} \tag{16}$$

where $R = \dfrac{r}{a_{eff}}$.

For power profile $f(R) = R^q$, and for Gaussian profile $f(R) = 1 - \exp\left(-R^2\right)$.

Because $\int_0^{\infty} \exp\left(-R^2\right) R dR = 0.5$, then for Gaussian profile, that approximates very well the diffusion distribution of dopant, we get (16), where $a_{Cieff} = a_S$. It means that the cut off frequency for the step profile is the same as for the Gaussian profile with the same volume (with equal Δ), because $V_C = \dfrac{a_{eff}}{a_S} V_C' = \dfrac{a_{Cieff}}{a_S} V_C' = V_C' = 2.405$.

So, if for the step profile the optical fiber remains single-modal, then, for the Gaussian profile of equal volume with the same Δ (fig.11), it remains also single-modal. It also mean that the optical fiber area with the step profile after diffusion e.g. after performing the operation of thermal connecting, remains single-modal. It is so because step index optical fibers remain single-modal if quantity of dopant in border of step profile remains the same [11].

Core radius after diffusion, which is a_{Cieff}, can be calculated using the method presented below.

It is assumed, that after diffusion process the distribution of dopant (as well as of the refractive index) remains the step one, and the quantity of dopant across the connection length L remains the same.

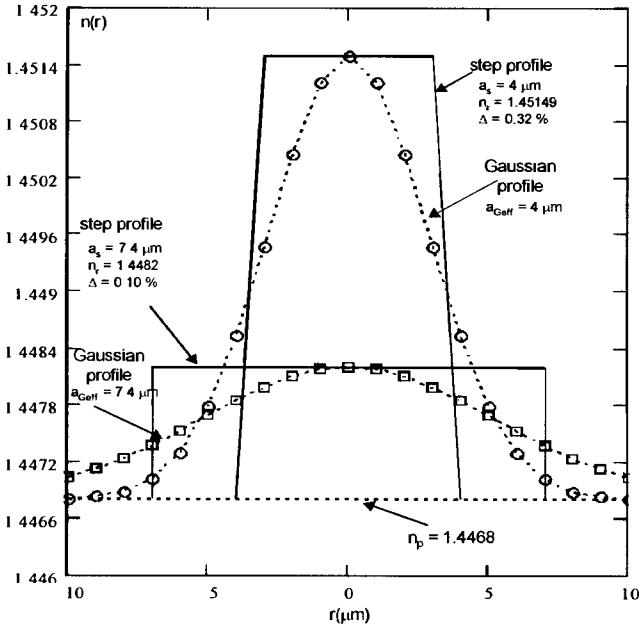


Fig. 11. Examples of profiles with equal volume before and after diffusion: step and the corresponding Gaussian ones

Therefore, we have

$$\pi a_{S0}^2 \cdot L \cdot N_0 = \pi a_S^2 \cdot L \cdot N_1 \tag{17}$$

where: L – length of thermal connection,
a_{S0} – core radius before diffusion,
a_S – core radius after diffusion,
N_0 – dopant concentration before diffusion,
N_1 – dopant concentration after diffusion.

Moreover, core radius, for step and diffusive (Gaussian) distribution, assumes the following form

$$a_S = a_{Geff} = a_{S0}\sqrt{\frac{N_0}{N_1}} \tag{18}$$

The form of the function $N_1(r)$ follows from the diffusion distributions [12]. N_1 in (18) is defined as $N_1(r)$ for $r = 0$. For $r = a_{Geff} \rightarrow \dfrac{N_1(a_{Geff})}{N_1(0)} = \dfrac{1}{e}$.

5. CONCLUSIONS

In the process of thermal fiber connecting there occurs the core dopant diffusion. Good consistence of the dopant diffusion and Gaussian distributions was found out, the longer the diffusion time i.e. thermal connecting the better the consistence. Thus, the dopant diffusion distributions and in effect the refractive index profile can be approximated with Gaussian distributions. Gaussian field distribution approximation which significantly simplifies calculation of losses due to mismatch of mode field radiuses of the spliced fibers is especially useful in these cases of refractive index profiles for which there is no analytic scalar wave equation solution. Gaussian refractive index profile is such an example. This form of profile is of great importance from the practical point of view as, first, it reflects dopant diffusion processes between the core and cladding in the fiber production process, second, it can be used for approximation of refractive index in the transit areas of thermally connected telecommunication fibers of different types. In so connected fibers it is possible, thanks to proper dopant diffusion, to achieve a thermally expanded core transit area (TEC) and join fibers with different values and refractive index coefficients almost without any losses. The above carried out theoretical analysis of the TEC transmission area properties showed that it is possible to achieve loss below 0.08dB for a thermal joint (splice) of fibers such as: G.652 & G.655, whose parameters differ significantly. Theoretical loss of TEC area is small if it remains single-modal. In the work, for step, Gaussian and power refractive index profiles, for which fundamental mode LP01 distribution can be approximated by Gaussian function, the so-called equal volume profiles were given. For these profile forms, it has been shown that the TEC area remain single-modal independently of changing values of core radius e.g. after diffusion. It has been shown that because Gaussian profile approximates well diffusion distribution, this thermally diffused core area, which has constant quantity of dopant in the diffusion area, remains single-modal.

BIBLIOGRAPHY

[1] Anderssen R.S., de Hoog F., Sammut R.A., 1981. Modes of Gaussian profile optical fibres, Optical and Quantum Electronics, Vol. 13, ss. 217-224.
[2] Jost W., 1960. Diffusion in solids, liquids, gases, Academic Press, New York.
[3] Kihara M., Matsumoto M., Haibara T., Tomita S., 1996. Characteristics of thermally expanded core fiber, Journal of Lightwave Technology, Vol. 14, No.10, 2209-2214.
[4] Majewski A., 1991. Teoria i projektowanie światłowodów, WNT, Warszawa.
[5] Ratuszek M., 2005. Analysis of reflectometric measurements losses of spliced single mode telecommunication fibers with significantly different parameters, Optica Applicata, Vol. XXXV, No. 2, 347-363.
[6] Ratuszek M., Majewski J., Zakrzewski Z., Zalewski J., 1999. Examination of Spliced Telecommunication Fibers of the NZDS-SMF Type Adjusted for Wavelength Division Multiplexing, Optica Applicata, Vol. XXIX, No. 1-2, pp. 73-85.
[7] Ratuszek M., Zakrzewski Z., Majewski J., 2009. Characteristics of Thermally Diffused Transit Areas of Single-Mode Telecommunication Fibers, Journal of Lightwave Technology, Vol. 27, ss. 3050-3056.

[8] Recommendation ITU-T G.652, 2003. Characteristics of a single-mode optical fibre cable.
[9] Recommendation ITU-T G.653, 2003. Characteristics of dispersion-shifted single-mode optical fibre and cable.
[10] Recommendation ITU-T G.655, 1997. Characteristics of a non-zero dispersion-shifted single-mode optical fibre and cable.
[11] Shiraishi K., Yanagi T., Kawakami S., 1993. Light-propagation characteristics in thermally diffused expanded core fibers, Journal of Lightwave Technology, Vol. 11, No.10, 1584-1591.
[12] Snyder A.W., Love J.D., 1983. Optical Waveguide Theory, Chapman & Hall, London.
[13] Yablon A.D., 2005. Optical Fiber Fusion Splicing, Springer Verlag, Berlin Heidelberg New York.

ANALIZA TŁUMIENIA I JEDNOMODOWOŚCI TERMICZNIE ROZDYFUNDOWANYCH OBSZARÓW W SPAWANYCH POŁĄCZENIACH JEDNOMODOWYCH ŚWIATŁOWODÓW TELEKOMUNIKACYJNYCH

Streszczenie

W pracy przedstawiono procesy dyfuzji zachodzące w termicznie łączonych jednomodowych włóknach o słabym prowadzeniu i przekroju kołowym, tj. jednomodowych światłowodach telekomunikacyjnych. Przeanalizowano dyfuzyjne rozkłady koncentracji domieszek rdzeniowych w tychże światłowodach spawanych w temperaturze $t \approx 2000°C$. Zaprezentowano gaussowskie aproksymacje rozkładów koncentracji domieszek rdzeniowych oraz współczynnika załamania w obszarze łączenia włókien. Przeprowadzono teoretyczną analizę propagacji i tłumienia światła w termicznie rozdyfundowanych obszarach rdzeni (TEC) jednomodowych światłowodów telekomunikacyjnych. Wykazano, że ponieważ profil gaussowski dobrze aproksymuje rozkład dyfuzyjny domieszki rdzeniowej to rozdyfundowany obszar rdzenia (spawu) pozostaje jednomodowy po dyfuzji – procesie termicznego łączenia.

Słowa kluczowe: światłowody jednomodowe, termicznie rozdyfundowany rdzeń, rozkład domieszki, profil współczynnika załamania, aproksymacja gaussowska

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 258
TELEKOMUNKACJA I ELEKTRONIKA 14 (2011), 47-60

WĄBRZEŹNO CITY BROADBAND IP NETWORK

Sławomir Bujnowski, Łukasz Zabłudowski

Institute of Telecommunications,
Faculty of Telecommunications and Electrical Engineering
University of Technology and Life Sciences
Al. S. Kaliskiego 7, 85-789 Bydgoszcz, Poland
slawomir.bujnowski@utp.edu.pl; lukasz.zabludowski@utp.edu.pl

Summary: In the paper the designing project (plan) of Wąbrzeźno City broadband IP optical network has been presented. The extended version of network plan constitute technical part of network Feasibility Study, that will be implemented in Wąbrzeźno and be financed from European Regional Development Funds. The network plan presented in the paper contains both topological structure of fiber optic network as well as the active equipment for the network. In the project described in the paper it has been suggested to use Modular Cable System – MCS for passive infrastructure and Metro Ethernet technology for active equipment. The presented solution provides low cost of construction (CAPEX), ease of implementation of the network and low operating cost (OPEX). Moreover the parameters of installed Metro Ethernet switches in the network guarantee the scalability of the network for at least 10 years.

Keywords: broadband, IP network, active infrastructure, passive infrastructure, backbone network, access network.

1. INTRODUCTION

Modern networks based on IP protocol are not currently used only for browsing websites and transfer information to global Internet, as it used to be in the initial stage of IP network development, but are also used as an efficient communications systems transferring different types of signals, including audio and video. Therefore, when designing (planning) a broadband IP network of Wąbrzeźno City, the network structure that have been considered then, should be opened for new users services and traffic volume growth during the next few years. When designing a network, it is essential to define very particular optimization problem. In his paper Richard van Slyke [8] has classified potential network design problem according to the list presented in table 1.

Table 1. List of potential network design problems.

	Given	Determine	Objective
1.	Traffic requirements, network topology, routing of traffic	Capacity of network transmission channels	Optimize tradeoff between channel costs and network performance
2.	Traffic requirements, network topology, capacity of network transmission channels	Routing of traffic in network	Minimize traffic delay
3.	Traffic requirements, network topology	Capacity of network transmission channels, routing of network traffic	Optimize tradeoff between channel costs and network performance
4.	**Traffic requirements**	**Network topology, routing of traffic, capacity of network transmission channels**	**Optimize tradeoff between channel costs, network performance and reliability**
5.	Terminal locations, traffic requirements	Location of multiplexers, concentrators, and/or routers	Minimize channel costs
6.	Terminal locations, traffic requirements, location of multiplexers, concentrators, and/or routers	Assignment of terminals to multiplexers, concentrators, and/or routers	Minimize channel costs

The Wąbrzeźno City broadband IP network required to be designed as a completely new network, both in passive and active network parts, so according to the table 1 it is defined as No 4 problem. In that case, when a completely new network is planned, the major task, is to find a compromise solution which comply several aspects impacting on the investment network cost (CAPEX), operating costs (OPEX), as well as users revenue. The aspects that should be taken into consideration during network planning process includes [8]:

1. The definition of all current and future services which planned network will offer and the classes of users for those services;
2. The designation of traffic volume generated by internet users located in areas served by the network;
3. The designation of expected level of the planned network scalability;
4. The designation of the proposed network topology, that depends on the geographic range of the area in which the network is implemented;
5. The designation of the level of network reliability, measured by network availability;
6. The definition of the network security level, including in particular secure access to active network resources;
7. The designation of maximal acceptable level of capex and opex costs for planned network guaranteeing fulfilling others network aspects defined above.

In fact, designing Wąbrzeźno City IP network, only some of mention above aspects have been considered. The aspects considered during Wąbrzeźno City network planning process are listed below:

- designed network should connect defined locations in Wąbrzeźno City;
- traffic generated from the locations should be transferring toward (from) global Internet;

- the locations should be connected with the use of fibers cables;
- in every location an active equipment (Ethernet switch or IP router), should be installed for generated traffic transfer,
- implemented network should minimize the total investment cost (CAPEX and OPEX costs);
- planned network structure should maximize network reliability;
- network should be scalable in time period not shorter than 10 years.

Taking under consideration the defined above network planning aspects, it is possible to formulate the optimization network designing problem, as follows. Design new network that:

1. minimize total investment cost;
2. maximize the network reliability;
3. maximize total network performance (i.e. throughput);
4. active equipment installed in the network should fulfill scalability for at least 10 years of its life.

As, defined above, optimization problem is very difficult for solving, so it has been reformulated as an optimization problem with constraints. The analyzed optimization problem for Wąbrzeźno City network, has been defined as follows:

- maximize the total network performance (throughput),
- under constraints:
  - the investment cost should not exceed defined value;
  - the network should operate in the proper way in the case even when one link or node will be damaged (defined network reliability level);
  - the active equipment should be scalable for at least 10 years.

When designing, completely new IP network, either on local (city) or on regional level, two key elements, that affect on the network structure and active equipment efficiency, should be considered. These key elements are:

1. User expectations (demands) of access lines bandwidth growth ratio. Nielsen law [7, 9] says that the users expectation of access line bandwidth grows exponentially, with the growth rate of about 50% every year by an average;
2. The total traffic volume growth in the global ip network. Based on cisco reports [2,3,4], it follows that the growth rate of traffic volume is exponential with cagr (compound annual growth rate), in period of the years from 2007 to 2012, of around 40%.

The backbone of Wąbrzeźno City Broadband IP Network has been planned using ring topology. The total throughput of the inner link between BN_1, BN_2 and BN_3 nodes is 10 Gb/s. The access nodes are connected to the backbone nodes on the basis of star topology. The total throughput of the access links is 1 Gb/s. Considering above the planned network keeps relatively high quality to price ratio.

The presented in this paper Wąbrzeźno City network plan constitute technical part of network Feasibility Study, that will be implemented in Wąbrzeźno and be financed in 75% from European Regional Development Funds.

2. SERVICES OF WĄBRZEŹNO CITY BROADBAND IP NETWORK

As it has been shown in the introductory section, one of important aspect affecting the overall network design is the range of services offered to network users. Modern IP networks offer services to any class of the users, i.e. for residential ones, for users required special needs (security or emergency ones), for local government units, for users that use telemedicine applications, for business users. Below the list of possible services in Wąbrzeźno City network is presented.

1. Services for residential users include:
   a) Internet access services;
   b) Peer-to-Peer services (P2P).
   c) additional services:
      i. VoIP telephony;
      ii. IP television;
      iii. Video on Demand;
      iv. local community portals;
      v. monitoring services;
   d) services offered to residential users by local government, i.e. e-government services for the citizens of the city, educational institutions, business;
   e) home office;
   f) teleeductation;
      i. e-learning: access to selected text and visual content, displayed on Internet portals, virtual language courses and e-learning;
      ii. e-school.

2. Services for users required special needs:
   a) central control of traffic lights to improve traffic in the city and the region as well as provide actual information about traffic
   b) dynamic information about departure times of public transport;
   c) certain municipal tasks telemetry;
   d) shared 112 number for communication in emergency situations;
   e) numbers for communication in undefined situations, such as 311 number in the U.S. or the 115 number in Germany;
   f) monitoring of selected areas of the city and schools based on digital cameras and IP video surveillance system;
   g) monitoring and surveillance of various alarms and monitoring center;
   h) Emergency Services Management Center;
   i) integration of emergency services in crisis situations.

3. Services for local government units:
   a) communication services ensuring save data transmission between local government units (LGUs) on the basis of VPN network;
   b) VoIP telephony communications among different LGU's;
   c) management of LGU's subordinate infrastructure (lighting, traffic lights etc.);
   d) public access to Internet, Internet kiosks and hotspots.

4. Services for business users:
   a) pictures of interesting sites of the city, transferring to global Internet;
   b) tourist information and accommodation infrastructure;
   c) Home Office;
   d) Teleworking;
   e) Video-conference, including telepresence;
   f) Corporate VPN's or VLAN's networks.

5. Telemedicine services:
   a) remote hospital consultations on special cases of the diseases;
   b) patients monitoring, and remote diagnostics, ie. telecardiology, teleradiology etc;
   c) teleconferences, including videoconferences and telepresence;
   d) access to patients data bases;
   e) VPN networks for communication among health units (i.e. hospitals).

For the above services listings, the following assumptions for access links has been considered:

a) group number 1 – 6 Mb/s symmetrical;
b) group number 2 – 20 Mb/s symmetrical;
c) group number 3 – 20 Mb/s symmetrical;
d) group number 4 – 6 Mb/s symmetrical;
e) group number 5 – 20 Mb/s symmetrical;

To define network topology, as well as to estimate necessary efficiency of an active equipment, installed in the designed network, it is needed to know the network nodes to which traffic, generated by various services, is directed. First of all, it is necessary to define whether the traffic is oriented towards other nodes installed in the planned network, or it is routed to Internet eXchange Point (IXP), that join the planned network with global Internet. The nodes to which traffic generated is directed in the regional or city (local) network are depicted in table 2.

Table 2. Service types and transfer direction of traffic generated by these services.

Type of service	Transfer direction of traffic	Type of traffic
Peer to Peer (P2P)	Global Internet (IXP)	Symmetrical
Video for PC, ie. YouTube. etc.	Partly regional network, mainly global Internet (IXP)	Asymmetrical
IP TV, VoD	Servers in regional network	Asymmetrical
WWW	Global Internet (IXP)	Asymmetrical
Interactive games	Global Internet (IXP)	Symmetrical
Videoconference	Mainly – regional network	Symmetrical
VoIP	Mainly – regional network	Symmetrical
Data transmission VPN	Mainly – regional network	Symmetrical
Monitoring	City network – within the area of distribution node	Asymmetrical
Telemedicine	Mainly – city and regional network	Symmetrical
Services for security	Mainly – city and regional network	Symmetrical

Cisco reports [2,3,4] show that the largest part of global traffic volume makes P2P traffic. Moreover, large parts of traffic volume make also video for PC, interactive

games and WWW traffics. Table 2 shows that the significant part of entire traffic volume, generated in the city network is directed toward Internet eXchange Points (IXP), so planned Wąbrzeźno City network should be called, "traffic oriented network" [10]. "Traffic oriented network" is meant the network that has been implemented in the way ensuring transfer of largest part of entire traffic volume, via the shortest paths. For city or regional networks the optimal "traffic oriented network" structure is star topology with IXP as the root. As in fact, city network is usually considered as the network included in regional one, making access network of it, so from table 2 it follows, that the largest part of traffic volume generated in city network is oriented toward global Internet via regional network (IXP consists connection point between city network and regional one).

The total traffic volume directed to IXP, consist of Peer to Peer traffic, Web traffic, traffic for interactive games, video for PC's traffic, all oriented towards global Internet, as well as video for TV traffic (such as IP TV, Video on Demand), videoconference traffic, Voice over IP traffic, e-health services and hospital applications traffic, all oriented towards regional network. As all this part of traffic are oriented to IXP, so the total volume of traffic transferred via IXP reaches more than 95% of entire one [2,3,4]. The only traffic that is oriented towards other nodes in Wąbrzeźno City network, is the traffic generated by monitoring systems installed in Wąbrzeźno.

Table 2 also shows the nature of the traffic transferred in city network in relation to a particular class of service. And so, the service, IP TV, VoD, video for PCs, generate asymmetric traffic, for which volume transmitted down the network, i.e. to the users, is much larger than the volume of the traffic directed up the network (to IXP). The traffic generated by monitoring systems is also, asymmetrical, but in difference to IP TV, Web, or video for PC traffic, the significant volume of this traffic is transferred from aggregation points (such as monitoring cameras) to the monitoring center, located locally, in the city.

The information contained in Table 2 will be used both, for determination of the traffic management way in Wąbrzeźno City network and for designation the optimal parameters of switching devices (i.e. Metro Ethernet switches or IP routers) installed in this network.

3. PASSIVE INFASTRUCTURE OF WĄBRZEŹNO CITY NETWORK

Topological structure of Wąbrzeźno City network passive fiber optic physical layer is shown in figure 1. This structure shows layout of all build, under the ongoing project, fiber optic cables, both for backbone and access network. In table 3 Wąbrzeźno City network node locations and characteristic of those nodes has been depicted. As the fibers are planned to be put along the road the length of the fibers has been calculated on the basis of google maps indications.

Fig. 1. Planned optical fiber routes for Wąbrzeźno City network.

Figure 2 in turns, shows a topology structure of network nodes and links connections between the nodes depicted in figure 1, that was created on the basis of fibers cable physical layout. The diagram presented in figure 2, has been used for a mapping of physical topology depicted in figure 1 into logical network topology shown in figure 4. In planned Wąbrzeźno City IP network it has been assumed that 3 nodes (BN_1, BN_2, BN_3) function as the distribution/aggregation nodes and the other ones as the access nodes.

Table 3. Node location and their characteristics.

No.	Node name	Node type	Location
1	Ampol	Access (A1)	Mikołaja z Ryńska st.
2	GPZ	Access (A2)	Matejki st.
3	Boiler	Access (A3)	Pruszyńskiego st.
4	Estate	Access (A4)	Kętrzyńskiego st.
5	Police	Access (A5)	Wolności st.
6	Marketplace	Access (A6)	Marketplace
7	The Cort/prosecutor	Access (A7)	Wolności st.
8	Fire dept.	Access (A8)	Matejki st.
9	Hospital	Access (A9)	Wolności st.
10	Municipal office	Access (A10)	Mickiewicza st.
11	Tax office	Access (A11)	Rataja st.
12	Node 1 - MZECIK	Access (A12)	Tysiąclecia st.
13	Node 2 – TBS	Access (A13)	Łabędzia st.
14	Node 3 – City Council	Access (A14)	Wolności st.
15	Node 4 – WDK	Backbone (BN3)	Wolności st.
16	Node 5 – Meat Proccessing	Backbone (BN2)	Kętrzyńskiego st.
17	Node 6 – TBS	Access (A15)	750-lecia Wąbrzeźna st.
18	Node 7 - MZECIK	Access (A16)	Żeglarska st.
19	Water Tower	Backbone (BN1)	Królowej Jadwigi st.
20	KPSI	IXP	Królowej Jadwigi st.
21	KPSI	IXP	Wolności st.

Distribution/aggregation nodes of Wąbrzeźno City network plays the role of backbone layer network nodes connected together by the fiber links. The backbone layer of every network should be characterized by high level of network reliability. So, over designing process of physical topology Wąbrzeźno City network structure, it has been assumed that distribution/aggregation nodes will be connected with double fiber rings. Double ring topology structures are used very often as the topology of backbone layer, as on the one hand this topology ensures high level of network reliability but on the other one, simplifies the management process of traffic transfer in the network [6]. Nowadays, in optical mesh networks, ring structures are used in wider term, creating so called p-cycles [5]. As our planned network is considered as fast recovery network the p-cycles provides the desired goal. P-cycles are always directed. When one of the links on the cycle fails, similar to a ring with multiplex-section protection, all wavelengths on the span are redirected in the reverse direction around the cycle that are used for network recovery in the event of link failure [1].

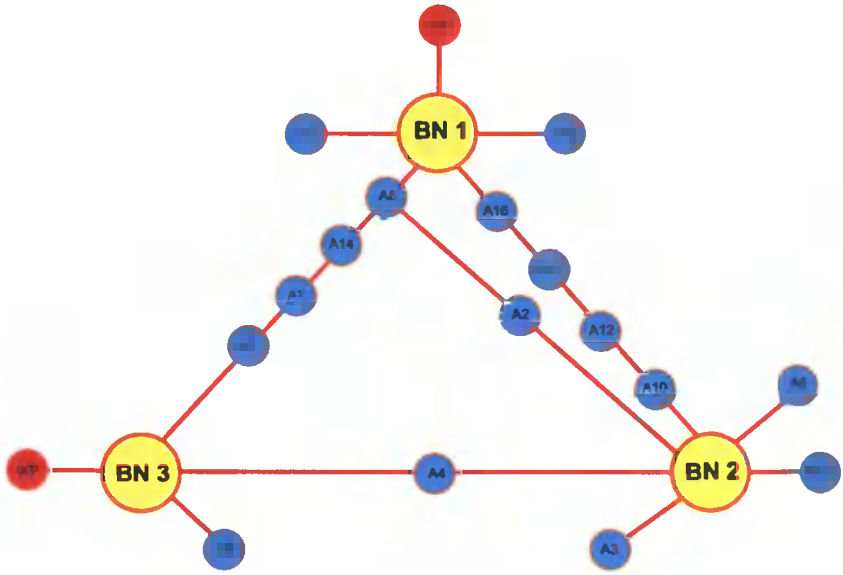


Fig. 2. Physical connection scheme following from network structure shown in figure 1.

The way of increasing reliability level of traffic transfer in the ring structure uses one of well known methods of link or path protection. The main principle of protection lies in fact that all selected pairs of nodes, that communicate with each other, are connected by two disjoint routes. In double fiber optical ring, each of these paths uses single optical fiber. One path is called working path and the second one, protecting path, They are implemented with the use of disjoint rings. Traffic transmitted from any two nodes in backbone goes through dedicated fiber with a total throughput of 10 Gb/s. As in Wąbrzeźno City network three backbone nodes are connected without any optical system like xWDM, so communication between any two nodes, in working or protecting paths, are always done via the third node (either in working or protecting path).

On the basis of fiber rings layout depicted in figure 2, real physical network structure of Wąbrzeźno City has been modified to "traffic oriented network". Modified Wąbrzeźno City physical network infrastructure includes two fibers rings, that join three backbone nodes denoted as BN_1, BN_2, BN_3, and 16 links joining access nodes to backbone ones. The figure 2 is supposed to visualize the idea of the physical connections. Nevertheless the access nodes which lays between backbone nodes do not mediate the same path with backbone nodes what will be explained in next section. The modified physical fiber cables structure of Wąbrzeźno City IP network is implemented in the form of double fibers ring connecting distribution/aggregation nodes (backbone layer), and branches (spurs), joining backbone nodes with access ones located in the city. The real structure of Wąbrzeźno City network topology, build on a physical basis of topological structure presented in figure 1, is shown in figure 3. From figure 3 outcomes that two backbone nodes BN_1 and BN_2 are connected to regional (K-PSI) network via IXP. Two connections of Wąbrzeźno City network with regional network has been assumed to increase the level of reliable transfer to the global Internet. Practically the connections to IXP points are realized via two 1 Gb/s links located in backbone nodes.

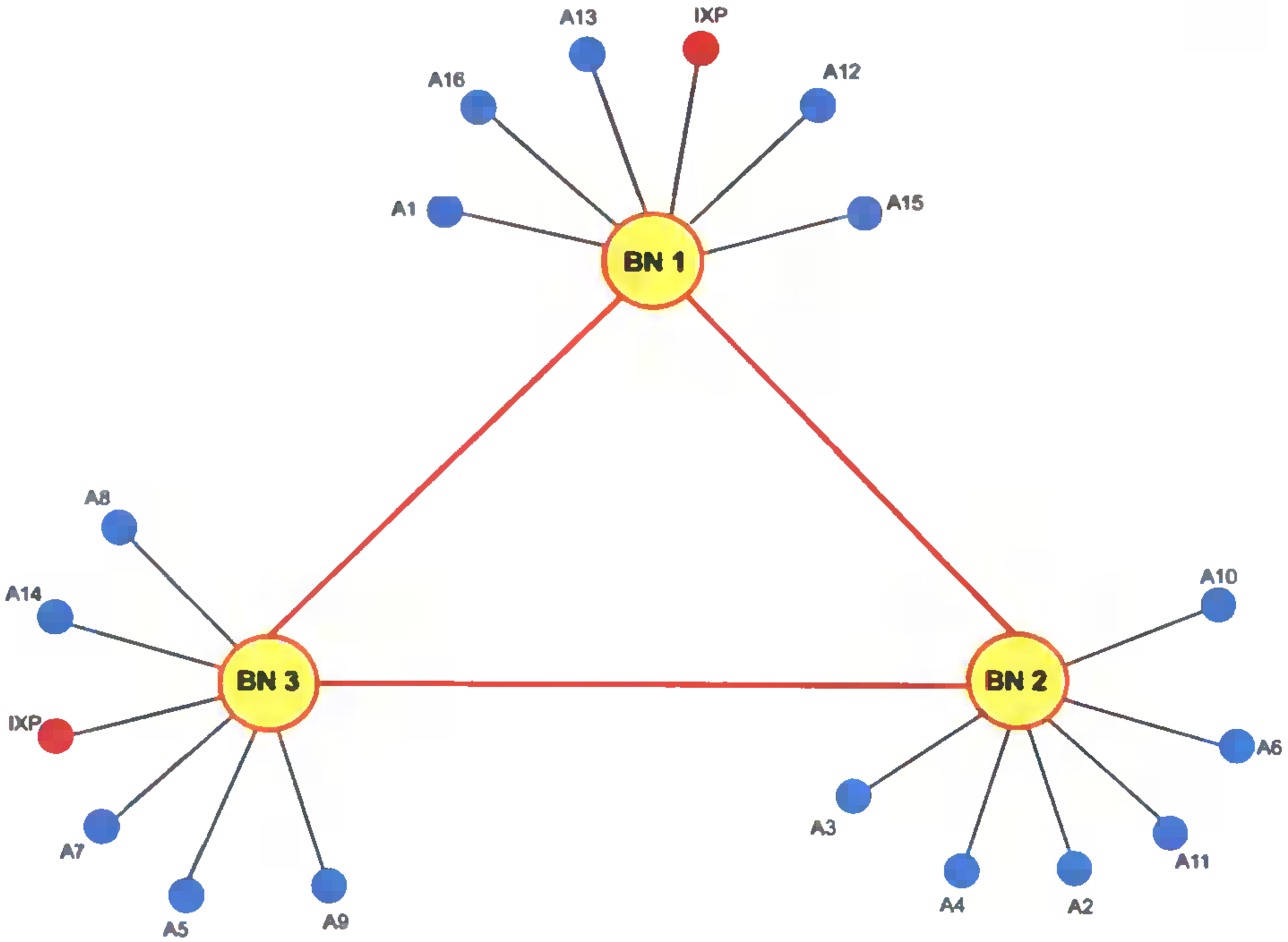


Fig. 3. Logical topological structure of Wąbrzeźno City IP network

To provide a wide range of scalability of the planned network, it has been assumed that the main nodes of the network (backbone ones), are to be connected to each other using links with transmission capacity of 10 Gb/s, while the access nodes has been joined to the backbone nodes using links with 1 Gb/s transmission capacity. Logically, the access nodes together with backbone one, forms a star topology (the optimal topology for "traffic oriented network") with backbone nodes as the main star nodes.

Functionally, both the backbone and access nodes of Wąbrzeźno City network meet mainly distribution (aggregation) functions, as their role is traffic transfer, to (from) main nodes BN_1, BN_2, BN_3, from (to) access nodes and further to (from) regional network and then to (from) global Internet. All access nodes lying both, on network ring or outside network ring are connected with dedicated fiber to only one backbone node. Admittedly, the star structure of access network does not guarantee high level of reliability, but it is the cheapest structure, that allow efficient traffic transfer. In fact, the damage of a fiber optic link will affect only the small group of users.

To be able to realize the topological structure of planned network (depicted in fig. 3) the suitable usage of optical fibers is needed. The main idea of fibers usage is shown below in figure 4. Access nodes located on the spurs of network structure are connected with dedicated fibers to the ring, and then using dedicated ring fibers, to backbone nodes. Access nodes located in the ring has been connected to backbone nodes using fibers in the ring. Connections of access nodes located in the ring are realized out as follows: in each relation (optical cable joining neighboring nodes) one pair of fibers (let us take the pair with number 1) is used for connection of backbone nodes. The other pair of fibers in analyzed relation can be used for connection of access nodes located in this relation with the nearest (and only one) backbone node. To connect each access node, the described idea requires to use separate pair of fibers. Obviously, similar situation is when connecting access nodes on spurs, as in the ring the fibers for joining spur nodes are chosen in the same way as the fibers for connection of access nodes lying in the ring.

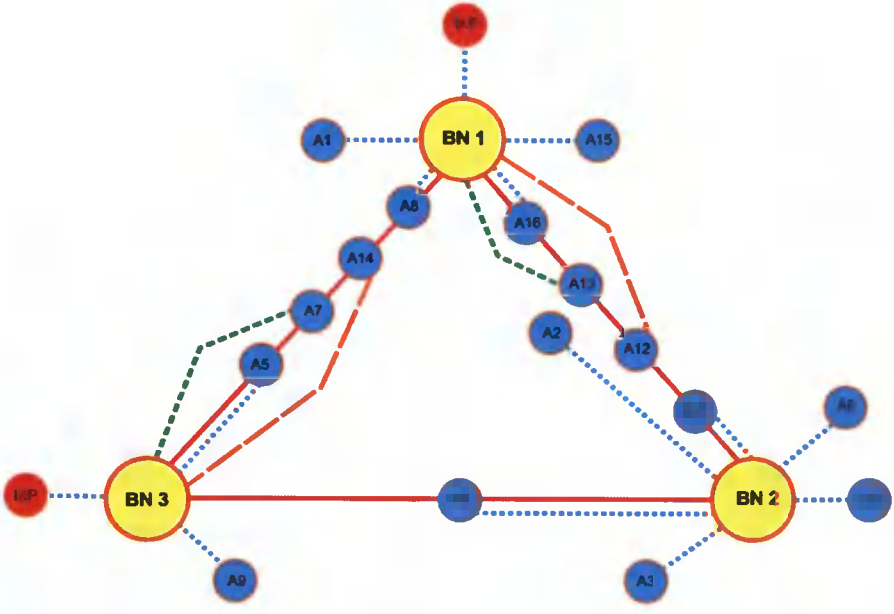


Fig. 4. Physical fibers connections for backbone and access network implementation.

To reduce the number of needed fibers, it has been assumed that the pair of fibers in the cable with the same number can be used repeatedly, however, two access nodes located in the network ring can be connected, with the same number of fibers pair, to

different backbone nodes. For example, in figure 4 it has been shown that the pair of fiber marked with a dotted line (let us assume fiber pair with i-number) connects, in the same cable relation, A_{16} with BN_1 and A_{10} with BN_2. The described above method of fibers reuse, gives topology required minimal number of fiber pairs usage. For network presented in figure 3, the construction of this network require only four pairs of fibers.

The Project of the passive physical network layer (fiber optic network) involves the construction of the network using Modular Cable System (MCS) technology with the usage of one pipe equipped with three micro pipes. The assumptions about the construction of passive infrastructure are as follows:

- The fibers optical cables are done based on the MCS system.
- Pipes used are suitable for dig in MCS cable directly in the ground. The pipes are equipped with 3 micro pipes. The fiber cable can be blown into each of micro pipe.
- In the first stage of project implementation to one of three micropipe fiber cable is blown. The cable contains 96 fibers on the basis of 8 micro cables with 12 fibers in each micro cable.
- To extend the network transmission capacity in the future, at each 1 km distance cables are placed into bunker.

The total length of fiber infrastructure in Wąbrzeźno city is about 12 km long. To estimate Wąbrzeźno City network investment cost, it has been assumed (information were obtained from different telecommunication companies), that the average cost of 1 km fiber cable implementation is around 110 thousand PLN. So, the total cost of Wąbrzeźno City network passive infrastructure (fiber cabling) was estimated on 1,3 millions PLN.

4. ACTIVE INFRASTRUCTURE OF WĄBRZEŹNO CITY NETWORK

As, it has been discussed in the previous parts of this paper, in order to improve reliability of Wąbrzeźno City backbone network the dual ring topology has been proposed for it. Dual ring network topology connecting backbone nodes ensure connectivity of backbone even in the event of network failure. However, access nodes joined to the backbone nodes create the configuration of star topology that is relatively inexpensive to implement and easy to locate of any failure in the network. From planned network physical layer description follow that each access node is connected to one backbone node with one pair of fibers only. It is clear that such connections do not give any redundancy, as it is in the case of backbone nodes, so the connection between each access node and backbone node results in a total cut of services offered by the network. However, access layer star topology significantly reduces the CAPEX and OPEX costs and possess big impact on the later stage of proposed network development and exploitation.

Before the necessary choice of the equipment is possible it is essential to evaluate the total volume of traffic generated in Wąbrzeźno City network. The traffic in this network has been calculated in accordance with the procedure outlined in the paper [8]. For traffic evaluation the following assumptions has been taken:

1. Around 20% of all households in Wąbrzeźno will be connected (via ISP networks) to Wąbrzeźno City network. Average capacity of user access line will be 6 Mb/s;

2. Around 15% of total number of SME companies (Small and Medium Enterprises) will be connected to planned network. Average capacity of business user access line will be 20 Mb/s;
3. All educational institutions, government, and local government agencies, health care, police department, fire department, municipal and county offices will be connected to designed network. Average capacity of access line for institutional users will be 10 Mb/s;
4. Overbooking factor is assumed as:
   - 20 for residential users,
   - 10 for business user;
   - 6 for institutional ones.

Taking into account the assumptions presented above, total volume of traffic generated in Wąbrzeźno City has been estimated equal to 500 Mb/s.

As an active equipment for the network the following options has been considered:
1. optical CWDM system and routers or switches;
2. routers IP with path protection;
3. Metro Ethernet switches with path protection.

Considering in detail three possible active equipment options, the one that has been suggested is the option number 3, i.e. Metro Ethernet switches with path protection. This option is both, cost (allows to build the network with minimal cost) and bandwidth effective (ensures high throughput). The details requirements for Metro Ethernet switches installed in Wąbrzeźno City network are given below:

- Two neighboring backbone nodes connected (on double fiber rings) to the third one are using 2 optical links with capacity of 10 Gb/s. Such high capacity of backbone connections will allow to operate Wąbrzeźno City network of over next few years, ensuring network scalability of at least 10 years and guaranteeing the appropriate level of services offered in the network. Switches ports of 10 Gb/s capacity should be equipped with optical interfaces ensuring proper transmission on 10 km range. Every Metro Ethernet switch installed in backbone layer node additionally should support MPLS protocol, as the protection functions in the backbone will be provided by MPLS protocol. MPLS protocol ensures traffic switching into protecting path in the case of working path failure.
- Each backbone node is equipped with 24 optical ports with 1 Gb/s capacity, through which these nodes will be connected to lower level access nodes, located throughout the city. Similarly to the specification of backbone ports, also 1 Gb/s access ports should provide efficient transmission of 10 km distance.
- In access nodes the switches are equipped with one 1 Gb/s optical port (for connection with backbone node) and at least 6 wire ports (for user connection) with capacity 1 Gb/s.
- It has been assumed that active equipments of backbone nodes are installed mostly in the indoor cabinets. The cabinets are equipped with power supply unit, air condition device and UPS module. The outdoor cabinets are only located at the police station, hospital, fire department and City Hall, as there was no possible to install indoor ones. Outdoor cabinets are also equipped with air condition devices, power supply unit as well as UPS.

In designed Wąbrzeźno City network the backbone nodes have been located in three main locations: Water Tower, Meat Processor and WDK. It should be noted however that described network would operate in future as Carrier of Carriers network type, where the users (both residential and business) will be joined to the network via local ISP networks. So, in the planned here network do not exist access lines connecting end users, as in fact this city network is "last mile before" network.

It has been suggested that the Wąbrzeźno City network should be connected to the global Internet via Kuyavia-Pomerania Regional Broadband (K-PSI) network. The Internet Exchange Point with K-PSI network will be installed in two backbone nodes located in a Water Tower and WDK. It has been assumed also that connection between Wąbrzeźno City network and K-PSI regional network would be implemented for every node through two links with of 1 Gb/s each (all together 4 Gb/s).

The total cost of active infrastructure installed in Wąbrzeźno City network has been estimated on 700 thousand PLN.

5. CONCLUSIONS

The paper is devoted to the designing project (plan) of Wąbrzeźno City network. In the paper, the topological structure of fiber optic network as well as the final solution for active layer of the network has been presented. For construction of passive infrastructure the technology of Modular Cable System – MCS has been suggested. Within this technology the fairly wide range scalability and cost effective network has been provided.

For the construction of the city network active layer the Metro Ethernet technology is used. Metro Ethernet technology provides a low cost of construction (CAPEX), ease of implementation of the network and low operating cost (OPEX). The parameters of installed network switches (switches) Metro Ethernet guarantee the scalability of the network for at least 10 years. The project assume also, that the city network in Wąbrzeźno cooperate with the regional K-PSI network (Wąbrzeźno City network and K-PSI network exchange traffic via common IXP point).

Total investment cost of Wąbrzeźno City IP network is estimated on 2 millions PLN.

REFERENCES

[1] Christian Mauz. p-cycle Protection in Wavelength Routed Networks, Communication Technology Laboratory.

[2] Cisco Visual Networking Index – Forecast and Methodology, 2007–2012, June 2008.

[3] Cisco Visual Networking Index - Forecast and Methodology, 2008–2013, June 2009.

[4] Cisco Visual Networking Index - Forecast and Methodology, 2009–2014, June 2010.

[5] W.D. Grover, 2003. Mesh-Based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking, Prentice-Hall, New Jersey.

[6] Ming-Jun Li, et al., 2005. Transparent Optical Protection Ring Architectures and Applications. Journal of Lightwave Technology, vol. 23, no. 10.
[7] Nielsen Jakob, 2008. Nielsen's Law of Internet Bandwidth. Alertbox for April 5, 1998, retrieved on.
[8] R. Van Slyke, 2001. Network Planning and Design, Polytechnic University Report.
[9] Study to Assess Broadband Bandwidth Usage and Key Trends in Europe, February 2008. Produced independently by Ventura Team LLP for the Fibre to the Home Council Europe.
[10] A. Zabłudowski, 2009. Broadband IP Network Development. Academy Publishing House EXIT, Warsaw.

SZEROKOPASMOWA SIEĆ IP DLA MIASTA WĄBRZEŹNO

Streszczenie

W pracy przedstawiony został projekt szerokopasmowej sieci IP dla miasta Wąbrzeźna. Rozszerzona wersja projektu stanowi część techniczną Studium Wykonalności zadania, które będzie wdrażane w mieście Wąbrzeźno. Zadanie to będzie współfinansowane z funduszy Regionalnego Programu Operacyjnego. Prezentowany w pracy projekt sieci zawiera zarówno opis części pasywnej (topologia kabli optycznych), jak i części aktywnej. Dla budowy części pasywnej sieci sugeruje się wykorzystanie systemu mikrokabli. Z kolei, dla budowy części aktywnej sieci sugeruje się wykorzystanie przełączników (switchy) Metro Ethernet. Zaprezentowane w pracy rozwiązanie zapewnia niski koszt inwestycyjny (CAPEX) budowy sieci, łatwość implementacji oraz niskie koszty operacyjne (OPEX). Ponadto także parametry wydajnościowe zastosowanych w sieci przełączników Metro Ethernet zapewniają skalowalność sieci przez okres co najmniej 10 lat.

Słowa kluczowe: szerokopasmowy, sieci IP, aktywna infrastruktura, pasywna infrastruktura, sieć szkieletowa, sieci dostępowe.

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 258
TELEKOMUNKACJA I ELEKTRONIKA 14 (2011), 61-67

EYETRACKING – METHODOLOGY AND APPLICATION

Arkadiusz Rajs[1], Mariusz Aleksiewicz[1], Agnieszka Banaszak[2]

[1] Institute of Telecommunications,
Faculty of Telecommunications and Electrical Engineering
University of Technology and Life Sciences
Al. S. Kaliskiego 7, 85-789 Bydgoszcz, Poland
arajs@utp.edu.pl; mariusz.aleksiewicz@utp.edu.pl

[2] Institute of Physics, Faculty of Mathematics,
Physics and Technique Science,
Kazimierz Wielki University,
85-052 Bydgoszcz, Powstańców Wlkp. 2,
agnb@ukw.edu.pl

Summary: Eyetracking gives great capability of computer's systems control and study of usability applications. In this paper we show construction of eyetracker and range of applications.

Keywords: eyetracking, usability applications, heatmap.

1. INTRODUCTION

At present a much attention is focused on the development of alternative methods of communication and computer control. One of the noninvasive and contact less method is the eye tracking method. The important feature of this method is no interference in natural human behavior and a broad range of application. The goal of this paper is the presentation of the eye tracker operation principles and possibilities of its applications.

2. EYETRACKER – CONSTRUCTION, PRINCIPLES AND OPERATION IMAGE PROCESSING

The construction of the eye tracking device is based on the processing of the image from camera situated on the front of the computer operator and few infrared light sources, which illuminate the observer eyes [2]. In the first stage of operation is the eyes localization, determination of the sight lines and the distance between the centers of eyeballs. The eyes must be localized on horizontal line and the observer face should be directed to the camera objective. These operations can be realized on many ways as for example the way using brightness gradient. Next, the image of both eyes recorded by

camera operating with frequency of 16-60 Hz is processed in order to determine the coordinates of eye pupil. In this purpose is used a phenomenon of the Purkinje reflection. Usually is used so called first Purkinje reflection it means the reflection of the light wave from the front surface of the eye cornea. This reflection is visible in the form of white point in the pupil vicinity (big black circle) and allows on detailed analysis of its movements. The image of the Purkinje reflection is shown in Fig. 1. The characteristic feature of this reflection is its stability during the eyeballs movements [1].

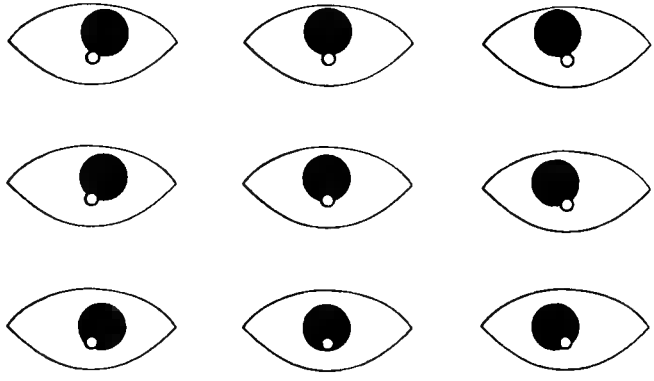


Fig. 1. Localization of pupil in relation to Purkinje reflection – camera view

On the basis of standing item-the Purkinje reflections should be determined, on the image from camera, the coordinates of crossection of the line from eye crossed the computer monitor. The coordinates should be determined separately for each eye assuming that the point of view for both eye is the same. It is shown in Fig. 2.

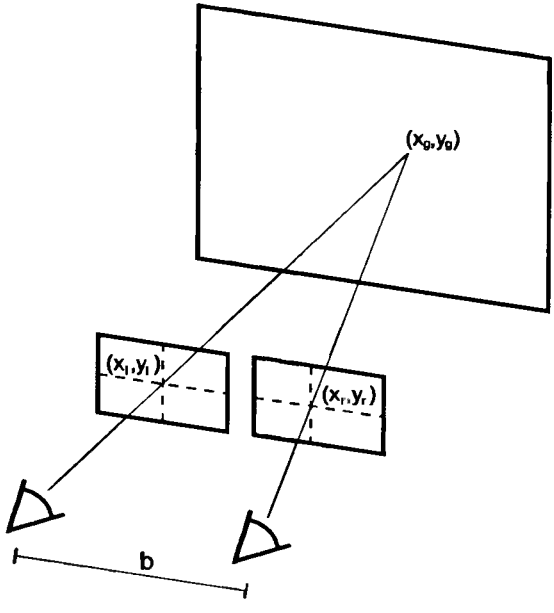


Fig. 2. Basic principle of gaze lines define

On the basis of the determined coordinates of left eye (x_l,y_l), right eye (x_r,y_r) and the distance b between ayes it is possible to determine the coordinates (x,y) of the point on which the sight is directed according to following equations:

$$x_g = \left(1 - \frac{b}{x_l - x_r + b}\right) x_h + \left(\frac{b}{x_l - x_r + b}\right)\left(\frac{x_l + x_r}{2}\right)$$
$$y_g = \left(1 - \frac{b}{x_l - x_r + b}\right) y_h + \left(\frac{b}{x_l - x_r + b}\right)\left(\frac{y_l + y_r}{2}\right) \quad (1)$$

The main problem of the devices proper operation is the observer movements. The movements of the whole body or of the head have negative influence on the proper determination of the eye pupil positions. By using of the additional algorithms it is possible to minimize these unfavorable factors [1].

3. EYETRACKING – THE RANGE OF APPLICATIONS

During the studies the eye tracker very precisely tracks the eye movement of the analyzed person. In the effect we can get two types of information's. The first information is the precise localization of the points (fixations) on which the sight of the analyzed person stopped for certain time (fixation time in general is the range of 0.15-1.5 s). Secondly we analyze the so called sakkad, it means the path on which the sight of the analyzed person is moved between fixations. The sakkad is performed in average 4 to 6 times per second with duration from 0.03 to 0.06 s. On that basis we are able to make (formulate) conclusions on type of the cognitive processes of the analyzed person.

The properly prepared software collects all information concerning the eyes positions at certain time and time duration of the sight stop on chosen position on the screen. The whole path on which the sight moves is also registered.

In the result the suitable processed data give the possibility of the generation of so called heatmaps and the focus maps i.e. the images on which, using suitable colors, are presented the average results of the sight localizations of the analyzed persons. The heatmaps essentially present the heat distribution of the focus points of chosen objects, which are easier observed than other negligible objects. In the case of leader of eyetracker producers- Tobii company- the software allows on the generation of two type of the heatmaps: gray-scale heatmaps (more intense focus point corresponds to certain level of transparency) and color heatmaps. Bright red color usually corresponds to point on which the sight is most frequently focused and lightgreen color corresponds to points at which the sight is rarely focused.

The next results of the studies by using eyetracker are, so called gazeplots. The gazeplots present the direction of eye scanning path during subsequent fixations. At is was shown in Fig.4 the single circles illustrate the subsequent fixations (the circle diameter is proportional to the duration of fixation). The lines which connect the circles correspond to direction of the eye scanning path. This process is called the eye sakkad movement.

Fig. 3. Heatmap shows regions which attract especial attention of user

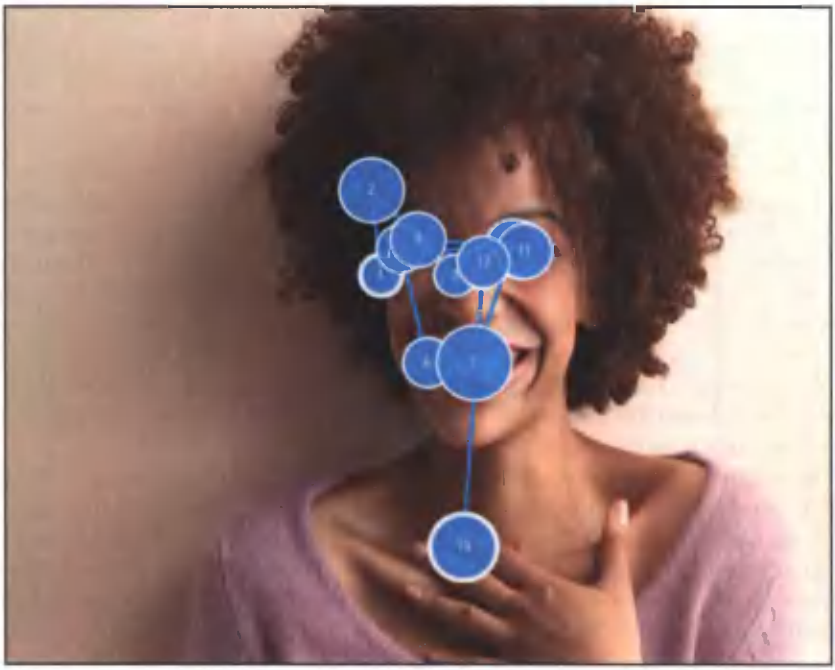


Fig. 4. Gazeplots show order and length of particular fixations

Additional forms of the eyetracking analysis are so called areas of interest (AOI). The investigator is able to determine the crucial, from his point of view, elements and to verify their popularity among respondents (how long and when the elements were observed). It is possible to perform the analysis of direction and duration of fixation in frame of given area. Moreover this method offers the possibility of comparison of the results between different areas what can lead to very interesting and valuable conclusions. Finally we can get the consistent information about number of peoples, which are especially interested in chosen AOI and how long continued the everage fixation.

The data obtained from eyetrack experiment allow to determine the time the persons spend on observation of the interesting us elements on the screen (especially prepared presentation as for example videos or websites); what is the sequence of observation; what disturb the observers. Additionally we can also to determine which element immediately catches the eye and which are ignored.

The biggest popularity the problem of eyetracking achieved in group of neuromarketing specialists. Neuromarketing is interested in problem of the using of measurement tools used in psychophysical studies and in the studies of human behavior in the optimalization process of marketing stimuli. Neuromarketing supports the studies of psychophysical processes and that basis checks the reaction of consumers on, for example. products and advertisement.

Nonregular eye movements correlate with difficulties of finding chosen element and longer fixation are connected with more intense stimuli processing. These data allow to have deeper insight into conscious and unconscious processes during interaction with consumer.

The second, also very big area of eyetracking applications is the support of software (desktop type and websites) usability analysis. The usability is the science which concerns to ergonomics of the interactive devices and their application. Generally speaking usability covers the problems of the convenience of the using of different devices.

The creators bandy himself in products improving and in applying of new technologies. It is, unfortunately, the proper operation of the given software is screened by other contents (not necessary in explicit way).

Usability or webusability is the proper way of the golden mean finding between the number of functions and convenience of using them. The final recipients will be not happy from the ergonomic WWW formula if they will not able to perform desired action. From the other hand they will be not satisfied if they will have theoretical possibility to performed desired tasks but in very complicated way that practically is unusefull. For this goal the eyetracking studies of softwares and WWW services are performed. The obtained data allows to propose new models or the products graphical structure before it emerges on the market.

The both, the neuromarketing and usability studies are the areas which have only business character. The other eyetracker application was proposed by scientist, who are interested in the support of the daily life of disabled peoples. It was discovered that studies of sight position of the person sitting in front of the computer monitor (classical eyetracking) can be a basis for elaboration of special software to control mouse by eyes. This idea is based on analogical assumption which are applied in PDA type devices where special stylus is used. At very beginning always the calibration is performed i.e. adjusting eyetracker operation conditions to person in front of the monitor.

The transfer of eye activity on text field activates virtual keyboard QWERTY, which then allows of the content introduction. In summary, the correlation of the sight of the person in front of monitor with operation of the sight controlled mouse allows for used different devices.

4. CONCLUSIONS

Undoubtedly from point of view of practical use of collected information, the best feature of the studies with eyetracker use is a potential objective determination of the consumer activity and perception as function of chosen time unit. The properly constructed software is able to support the studies of usability of the created software market realized in Internet on WWW Websites and also to support the advertisement realized by TV Advertisement Agencies, billboards etc.

The eyetracking has also certain limitations. It may happen that due device construction (the necessity of eye illumination by infrared, camera position etc.), some analyzed persons during the procedure must dress off the glasses or to take of the contact lenses. Additionally, it necessary for respondent to have stable position during the studies procedure and sight must be in the range of camera operation. All these elements can have a negative influence on the activity of stimuli and in consequence the whole procedure can be unbelievable. In the context of eyetracker use as a transducer element between the disabled human sight and the fulfill of all condition can be problematic.

It is also important to mention that the necessity of averaging of the obtained data have an influence on proper interpretation and proper conclusions formulation. The stimuli (advertisement movie, static images) which are presented to the respondent influence them not uniformly. Each of them reacts on chosen presentation or Website in different way. In the consequence the data of honestly performed studies on the big group of peoples must be averaged and extreme opinion removed.

An interesting point of view on eyetracking seems to be the determination of the emotional activity for chosen moment of studies. In the order to achieve that goal the studies cam be enriched to record electroencelograph measurements. Because the unequivocal interpretation of such a record is very difficult and is characterized by big error therefore the problem is undertaken more and more frequently by specialist from the field of neuropsychology.

BIBLIOGRAPHY

[1] Duchowski, A. T., 2007. Eye Tracking Methodology: Theory and Practice, Springer-Verlag, 360 p.

[2] Yoshinobu Ebisawa, Shin-ichi Satoh, 1993. Effectiveness of Pupil Area Detection Technique Using Two Light Sources and Image Difference Method., Engineering in Medicine and Biology Society. Proceedings of the 15th Annual International Conference of the IEEE, pp. 1268-1269.

[3] Tadeusiewicz R., Korohoda P., 1997. Komputerowa analiza i przetwarzanie obrazów. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków.

[4] Lee N., Broderick A.J., Chamberlain L. 2007. What is 'neuromarketing'? A discussion and agenda for future research. International Journal of Psychophysiology 63, pp. 199--204.

[5] van Elk, et al. 2008. You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy. NeuroImage, doi. 10.1016/ j.neuroimage.2008.07.057.

METODOLOGIA I ZASTOSOWANIE EYETRACKINGU

Streszczenie

Eyetracking znajduje bardzo szerokie zastosowanie w systemach komputerowych do sterowania komputerem oraz w badaniach nad użytecznością aplikacji. W niniejszej pracy przedstawiono budowę eyetrackera oraz scharakteryzowano obszary zastosowań.

Słowa kluczowe: eyetracking, badanie użyteczności, mapy ciepła

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 258
TELEKOMUNKACJA I ELEKTRONIKA 14 (2011), 69-78

SOLUTION IMPLEMENTATION BASED ON MODIFIED KALMAN FILTER FOR PURPOSE OF BUS ARRIVAL TIME PREDICTION

Damian Ledziński[1], Michał Jezierski[2], Beata Marciniak[1],
Tomasz Marciniak[1]

[1] Institute of Telecommunications,
Faculty of Telecommunications and Electrical Engineering
University of Technology and Life Sciences
Al. S. Kaliskiego 7, 85-789 Bydgoszcz, Poland
[Damian.Ledzinski, BeataMarciniak,Tomasz.Marciniak]@utp.edu.pl

[2] Electronics Company SIMS, Bydgoszcz
ambinanitelo@gmail.com

Summary : This paper describes use of Kalman's filter for prediction of time of arrival of bus. Kalman filter is recursive algorithm determining the minimum-variance estimate of the state vector of dynamic system, based on the measurement of inputs and outputs of the system. Three prediction algorithms used: difference algorithm, traditional Kalman filter and Kalman filter with changing weights of input data. Authors studied the bus arrival time predictions. Used for this purpose data send by radio from vehicles to prediction server. The smallest average prediction error obtained for the Kalman filter with variable weights.

Keywords: Kalman filter, time prediction

1. INTRODUCTION

In the era of technological development especially in large cities, is increasing need for effective and fast movement. Very often, public transportation is used for this purpose. Thus, there can be observed an increasing demand for accurate information on the transportation possibilities. The traditional schedule fixed in advance is not sufficient any more, travelers expect information in real time, along with the changing traffic conditions in the city.

The possibilities, of presenting such data are also increasing, including electronic boards at bus stops, offices and shopping centers and dynamically updated web pages being browsed by the mobile phones. However, there is a problem of storage data on the vehicle position and its processing in order to be able to predict traffic delays.

The bus arrival time prediction problem already was discussed in the literature. Usefulness of different types of input information and computing methods has been studied [1,2,3,4,5,6,7,8,9,10]. In this paper the authors present a system which public transportation vehicles, send the information about the bus arrival at the next stops via radio. This information is collected and processed by the prediction server (Fig. 1). The

authors present and compare three prediction algorithms: difference algorithm, Kalman filter and Kalman filter with changing weights of input data. The Kalman filter algorithm has been used in many subjects including: weather prediction [11], asset pricing [12], navigation [13] and visual tracking [14].

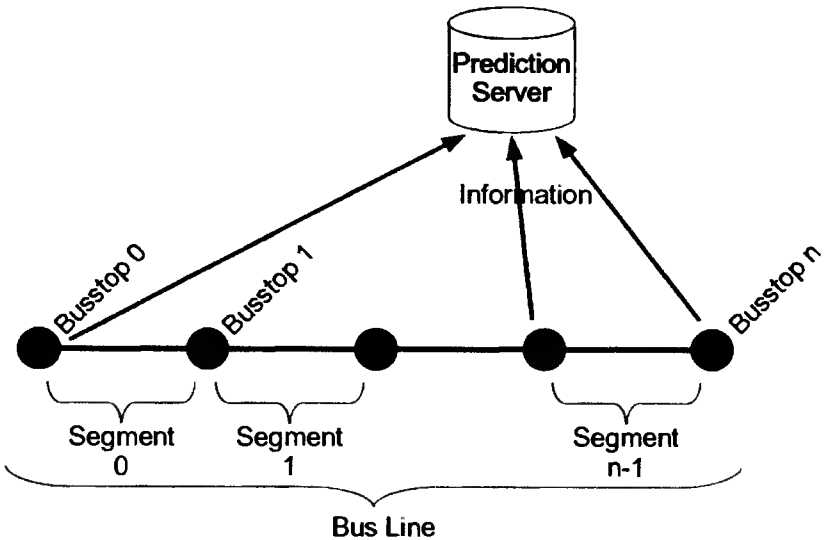


Fig. 1. Bus line schematic

Each bus line can be described by a graph where a stops are nodes and segments are the edges. If each distance is assigned with the time in which is defeated, we receive a weighted graph. The sum of the lengths of these edges describes the travel time for whole route. Theoretical travel time is regulated by schedule.

1.1. DIFFERENCE ALGORYTHM

The simplest to be used is the differential algorithm. It checks delay of bus at the stops, defined as the difference between real and scheduled time of departure from the bus stop.

$$\Delta t(B_k) = t_R(B_k) - t_s(B_k) \tag{1}$$

where:

$\Delta t(B_k)$ – delay on bus stop B_k,

$t_R(B_k)$ – real time of departure from the bus stop B_k,

$t_s(B_k)$ – scheduled time of departure from the bus stop B_k.

The basic idea of the difference algorithm is that, if the vehicle is late at bus stop B_k, by the time Δt, then at the next stop B_{k+1} it is also a late by the time Δt

$$\Delta t(B_{k+1}) = \Delta t(B_k) \tag{2}$$

Estimated time of arrival of the bus at the next stop, is the scheduled time corrected by the delay from the previous stop

$$t_x(B_{k+1}) = t_S(B_{k+1}) + \Delta t(B_k) \tag{3}$$

where $t_x(B_{k+1})$ is estimated arrival time from bus stop B_{k+1}, so

$$t_x(B_{k+1}) = t_S(B_{k+1}) + t_R(B_k) - t_S(B_k) \tag{4}$$

Difference algorithm was used here as a reference point for determination of efficiency of the algorithms presented further in this paper.

2. RESEARCH METHODOLOGY

For all the algorithms presented in this paper test method was the same. Data for the tests was taken from an existing system. Those were the times of departures from the bus stops one of the most frequented used lines in one of the Polish cities. These times were collected one month. Totally, there were more than 50,000 departures.

For the study, the authors developed a simulator, implemented in Java programming language (see Fig. 5). It introduced the data to be used for computing of prediction (the real times of departure from previous stops, or previous rides of the tested distance). Then, as result calculated the difference between predicted time and real travel time (prediction error).

2.1. INTRODUCTION TO THE KALMAN FILTER

The considered model consists of two linear stochastic difference equations:

$$x_n = x_{n-1} + w_{n-1} \tag{5}$$

$$y_n = x_n + v_n \tag{6}$$

Our aim is to estimate unknown (unobservable) variable x, whose evolution in time is described by the first equation (5). Quantity x is assumed to be 'true' value and unknown to us. All of the values $y_0,...,y_n$ are known and indicate measurements of x in discrete points of time $0,...,n$. Because of uncertainty of measurement process, the second equation (6) contains v_n summand. We assume here, that in each time step, ie. for each n, process noise w_n and measurement noise v_n are both white and Gaussian (v_n and w_n are uncorrelated random variables)

$$w_n \sim N(0,Q) \tag{7}$$

$$v_n \sim N(0,R) \tag{8}$$

In the following paper we considered the filter to be stationary, which means that variances Q and R are fixed. We will use the following notation:

$\hat{x}_n^-$ – a priori estimation taken in step n with $n-1$ given observations, ie. $\hat{x}_n^- = \varepsilon\left[x_n \mid y_0, \ldots, y_{n-1}\right]$,

$\hat{x}_n$ – a posteriori estimation – process state estimation with n given observations, ie. $\hat{x}_k = \varepsilon\left[x_n \mid y_0, \ldots, y_{n-1}, y_n\right]$,

$e_n^- = x_n - \hat{x}_n^-$ – a priori estimation error,

$e_n = x_n - \hat{x}_n$ – a posteriori estimation error.

Kalman filter is a tool used to find an optimal estimate of x which minimizes the a posteriori estimation error P_n, ie. minimizes mean-squared error given by the formula

$$P_n = \varepsilon\left[e_n^2\right] = \varepsilon\left[\left(x_n - \hat{x}_n\right)^2\right]. \tag{9}$$

Under above assumptions, estimation $\hat{x}_n$ of system state x_n which minimizes mean-squared error (9), is given by linear combination

$$\hat{x}_n = \hat{x}_n^- + K_n\left(y_n - \hat{x}_n^-\right), \tag{10}$$

where

$$K_n = \frac{P_n^-}{P_n^- + R}, \tag{11}$$

$$P_n = \left(1 - K_n\right) P_n^-, \tag{12}$$

$$P_n^- = P_{n-1} + Q, \tag{13}$$

$$\hat{x}_n^- = \hat{x}_{n-1}. \tag{14}$$

Reader interested in detailed proof of above equations may refer to original R. Kalman's article [15]. A theoretical introduction to Kalman filter theory can be found in many papers including [16,17,18].

Roughly speaking, Kalman filter is the way to compute an optimal (ie. with minimal P_n) estimate from two given values. It is a kind of weighted mean which uses process and measurement uncertainties Q and R instead of standard weights. In our case for every n, Kalman filter algorithm estimates $\hat{x}_n$, basing on $\hat{x}_n^-$ and y_n (see Fig. 2). From equation (10) it follows, that when Q is significantly greater than R then $\hat{x}_n \approx y_n$. Analogously, if R is significantly greater than Q, then $\hat{x}_n \approx \hat{x}_n^-$.

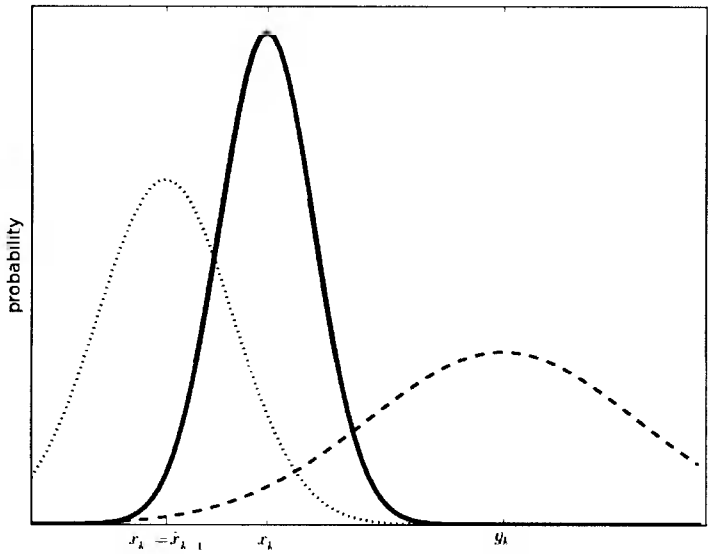


Fig. 2. Estimation taken in k-th step under assumption $Q < R$.

Due to the recursive nature of the estimation algorithm, the process can be organized into two stages: time update step (a priori estimation) and the measurement update step (a posteriori estimation). The above two-step approach is shown in the following diagram (Fig. 3), where left side presents a posteriori estimation, and right side a priori estimation (indices n are omitted for simplicity)

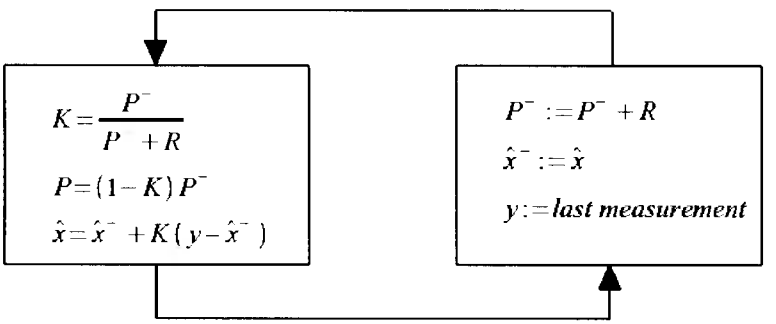


Fig. 3. Time update and measurement update steps

2.2. USAGE OF THE KALMAN FILTER

In order to improve efficiency (minimize prediction error), we decided to use the Kalman filter. The input data of filter were previous travel times on tested distance. That is, travel times of this distance immediately before the tested one.

In each distance the server uses an independent filter. When the vehicle has traveled the distance, the filter performs the next iteration. This iteration was calculated

from the new input data. The result of this iteration was the prediction of travel time the next vehicle on the same distance. Moreover, a filter at each iteration did not receive a single travel time, it received an average weighted from the last three times of traveling over a given distance.

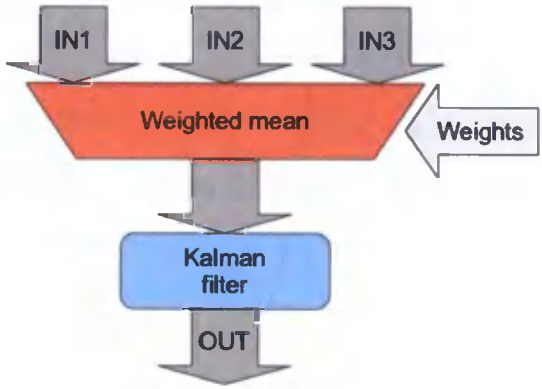


Fig. 4. Filter construction

$$\Delta t_x(V_m) := F\left(\frac{\sum_{i=1}^{3} w_i \cdot t_D(V_{m-i})}{\sum_{i=1}^{3} w_i}\right) \tag{15}$$

where:

$F(x)$ – The Kalman filter iteration with input data x,
$\Delta t_x(V_m)$ – estimated segment traveling time by vehicle V_m,
w_i – weight for i-th parameter,
$t_D(V_{m-i})$ – travel time over a distance by earlier vehicle.

Looking for the optimal solution, assumed weight was 1, 0.6, 0.3.

2.3. KALMAN FILTER WITH CHANGING WEIGHTS

To improve results, we tested the previous algorithm with different values of weights w_1, w_2, w_3. Our goal was to find the optimal values of weights. The best weights were those with the lowest prediction error. Such a test was performed for each distance separately.

```
private final double Q, R, M, F;
private double p = 1.0;
private double xhat = 1.0;

public double predict(double inputValue) {
  double x_hat_k_minus;
  double p_k_minus = Math.pow(F, 2) p + Q;
  double K = p_k_minus * M / (p_k_minus * Math.pow(M, 2) + R);
  double x_hat_k = x_hat_k_minus+ K*inputValue-M*x_hat_k_minus);
  double p_k = (1 - K) * p_k_minus;
  p = p_k;
  xhat = x_hat_k;
  return x_hat_k;
}
```

Fig. 5. Implementation of the Kalman filter in Java programing language

Results (most optimum weight for each section) are shown in Table 1. Notice that there are different values for each distance. Inspired by this observation, we concluded that weight could be periodically modified by the simulations were carried out on historical data from some period.

Table 1. Optimal values of weights on each segment

Distance	w_1	w_2	w_3
0	1	0,7	0,7
1	1	1	0,8
2	1	0,4	0,4
3	1	0,4	0,4
4	1	1	1
5	1	0,4	0,4
6	1	1	1
7	1	1	1
8	1	0,6	0,6
9	1	0,1	0,1
10	1	0,8	0,7
11	1	0,8	0,1
12	1	1	1
13	1	0,5	0,5
average	**1**	**0,69**	**0,62**

3. RESULTS

Applying discussed in the previous chapter the methods of determining the time prediction obtained the following results (table 2). For better clarity, derived calculations have been rounded to 2 decimal places. The smallest average prediction error was obtained for the Kalman filter with variable weights. Some individual time prediction errors received for the differential algorithm had little value (within 0 and 1 in Table 2). However, in this algorithm all errors are cumulative, making it impossible to obtain satisfactory results.

Table 2. The average prediction error for each distance for different algorithms

Distance	Diference Algorithm	Kalman Filter	Kalman Filter with changing weights
0	7,61	19,63	5,22
1	7,63	15,65	6,74
2	16,27	15,03	14,29
3	59,15	13,10	10,56
4	17,51	34,24	18,80
5	24,85	20,20	24,67
6	55,97	27,63	7,42
7	35,36	15,75	20,09
8	40,06	59,51	40,33
9	59,63	19,78	50,79
10	18,84	40,36	11,65
11	37,97	45,01	33,44
12	38,64	31,77	26,70
13	33,77	23,64	19,55
average	**32,38**	**27,24**	**20,73**

4. CONCLUSIONS

As shown in the table 2, the Kalman filter with variable weights has the best effectiveness. Kalman filter with fixed weights produces better results than a difference algorithm.

The main advantages of the difference algorithm are its relatively good performance and simplicity. The disadvantage is that it requires constant predefined schedule. It cannot be used in situations when we do not have a regular schedule. If the schedule is poorly prepared, that is , the travel times of the distances are poorly matched, the algorithm will produce more mistakes. Delayed drivers will probably try to make up for the delay, whereas, the difference algorithm assumes the delay to be constant.

The advantage of using a Kalman filter is better efficiency, especially with many vehicles on the line. If at any point in the city there is a traffic jam, the filter quickly adopts to the situation (fast response to changing time of travel through a given distance). In addition, it does not require a fixed schedule. This filter will not work correctly when the line is loaded less than other bus line (eg, a few rides during the day).The filter requires historical data. So, when the program starts and there is no data collected, will not work properly. To make this possible, it passes several iterations and it can takes whole day.

Kalman filter with changing weights proved to be the best although it needed performance of periodic simulations for different values of weights, which requires much computing power. And it is its biggest disadvantage. Moreover, it requires storage of lots of historical data.

The described method can be applied in other areas of technology, wherever we can use the Kalman filter and collect historical data.

In the future, we will test the Kalman filter having more parameters than just the last tree travel times. Perhaps, expected results would be given by a simulation using the optional values of fixed parameters of Kalman filter.

BIBLIOGRAPHY

[1] Kajan E., 2002. Information technology encyclopedia and acronyms. Springer, Berlin Heidelberg New York.
[2] Broy M., 2002. Software engineering – From auxiliary to key technologies. In: Broy M., Denert E. (eds). Software Pioneers. Springer, Berlin Heidelberg New York.
[3] Che M, Grellmann W, Seidler S, 1997, Appl Polym Sci 64:1079-1090.
[4] Ross D.W., 1977. Lysosomes and storage diseases. MA Thesis, Columbia University, New York.
[5] Padmanaban, R.P.S., Divakar, K., Vanajakshi, L., Subramanian, S.C., 2010. Development of a real-time bus arrival prediction system for Indian traffic conditions. In: Intelligent Transport Systems, IET. vol. 4, issue 3, pp. 189-200.
[6] Hao Chu, Yun Cai, Xiaoguang Yang, 2007. Research on Bus Arrival Time Prediction Based on Multi-Source Traffic Information. In: Telecommunications, ITST '07. 7th International Conference on ITS, pp. 1-5.
[7] Jian Zhang, Ling Yan, Yin Han, Jing-Jing Zhang, 2000. Study on the Prediction Model of Bus Arrival Time. In: Management and Service Science,. MASS '09. pp. 1-3.
[8] Jeong, R., Rilett, R., 2004. Bus arrival time prediction using artificial neural network model. In: Intelligent Transportation Systems,. Proceedings. The 7th International IEEE Conference, pp. 988-993.
[9] Latos P., Dubalski B., Marciniak T., Marciniak B., 2010. Demand forecasting for spare parts. Zesz. Nauk. UTP w Bydgoszczy, Telekomunikacja i Elektronika 13, pp. 103-114.
[10] Padmanaban, R.P.S., Vanajakshi, L., Subramanian, S.C., 2009. Automated Delay Identification for Bus Travel Time Prediction towards APTS Applications, Emerging Trends in Engineering and Technology (ICETET), 2nd International Conference, pp. 564 569.
[11] Galanis G., Louka P., Katsafados P., Kallos G., Pytharoulis I., 2006. Applications of Kalman filters based on non-linear functions to numerical weather predictions, Ann. Geophys., vol. 24, pp. 1-10.
[12] Bidarkota P., Dupoyet B., 2006. Asset Pricing with Incomplete Information In a Discrete Time Pure Exchange Economy; Florida International University, Department of Economics.
[13] Brock L.D., Schmidt G.T., 1970. General Questions on Kalman Filtering in Navigation Systems, Chapter 10 of Theory and Applications of Kalman Filtering C.T. Leondes, Editor, NATO AGARD.
[14] Funk, N., 2003. A Study of the Kalman Filter applied to Visual Tracking; University of Alberta.
[15] Kalman R.E., 1960. A New Approach to Linear Filtering and Prediction Problems, Transaction of the ASME – Journal of Basic Engineering, pp. 35-45.

[16] Jacobs O.L.R., 1993. Introduction to Control Theory; 2nd Edition. Oxford University Press.
[17] Welch G., Bishop G., 2006. An Introduction to the Kalman Filter; Department of Computer Science at the University of North Carolina at Chapel Hill Tech.
[18] Maybeck, Peter S., 1979. Stochastic Models, Estimation, and Control, vol. 1; Academic Press Inc.

IMPLEMENTACJA FILTRU KALMANA DO PROGNOZOWANIA CZASU PRZYBYCIA AUTOBUSÓW

Streszczenie

W pracy przedstawiono zastosowanie filtru Kalmana do prognozowania czasu przybycia autobusów. Filtr Klamana to algorytm rekurencyjnego wyznaczania minimalno-wariancyjnej estymaty wektora stanu układu dynamicznego, na podstawie pomiaru wejść i wyjść tego układu. Zbadano trzy algorytmy predykcji: algorytm różnicowy, tradycyjny filtr Kalmana oraz filtr Kalmana ze zmiennymi współczynnikami. Autorzy badali odchylenie od prognozowanego czasu przyjazdu autobusów. Używano do tego celu danych przesyłanych drogą radiową z autobusów do serwera predykcji. Najlepsze wyniki uzyskano dla filtru Kalmana ze zmiennymi współczynnikami.

Słowa Kluczowe: filtr Kalmana, prognozowanie czasu

ISSN 1899-0088