First International Conference
Modelling and Development of Intelligent Systems
Sibiu - Romania, 22-25 October, 2009

Stable 3D Scene Restoration Using One Active PTZ Camera

Kiril Alexiev, Iva Nikolova, Georgi Zapryanov

Abstract

The active PTZ (Pan Tilt Zoom) camera is a key element of an intelligent surveillance system. The opportunity to control camera parameters significantly increases the abilities of these cameras as information sources. It is commonly regarded that a camera gives 2D presentation of 3D scene. The depth of the scene is irrevocably lost and only some image features may indirectly reveal the position of the objects in third dimension. The active camera can partially overcome this loss of information. The suitable control of camera parameters may be used for estimation of the depth of the observed objects. The paper discuses one of the methods for 3D scene restoration called "depth from defocus" and its inherited characteristics. All key points of the approach realization are described and commented. Experimental studies, using test patterns and real objects are presented to test its applicability.

1 Introduction

Many scientific and engineering applications require characterization of objects and phenomena occurring in our three-dimensional world. It is commonly regarded that the widespread digital cameras with CCD (Charge-Coupled Device) and CMOS (Complimentary Metal-Oxide Semiconductor) image sensors produce 2D presentation of the 3D environment. In that registration the location of objects such as angular coordinates in horizontal and vertical direction remains the same, but information about the distance to the objects is irrevocably lost. The restoration of the third dimension is, however, critically important for determining the actual spatial arrangement of objects, object tracking, understanding the spatial-temporal relationships between objects, evaluation of their behavior, and predicting future events. Scientists have long attempted to develop hardware and software tools for 3D recovery. Nowadays, there are professional CMOS video cameras, specially designed to capture video with depth information [1, 2], but unfortunately, they are too expensive and their resolution is a long way away from the quality of the usual CCD and CMOS cameras used today. Therefore, more research efforts are put into a software solution to the problem with standard video sensors.

Most of the currently available techniques on visual 3D recovering have focused on multisensor approach (stereo vision) and other algorithms that require multiple images, such as structure from motion, shape from shading, range from focus and depth from defocused images [4]. Depth estimation using frames from single camera is a difficult task, and it requires some prior knowledge about the scene and the global structure of the image. In this

paper, the problem of estimating distances to the objects in indoor scenes is discussed on the base of the well-known depth from defocus approach. This technique is the most attractive one with little hardware requirements, the small number of processed image frames and the absence of content-based image analysis. However, adequate spectral content and accurate information of the lens parameters of camera system must be ensured to get good estimates of the depth.

The automatic depth estimation requires several challenging problems to be successfully resolved during the camera frames analysis [9]: (i) image texture analysis; (ii) noise estimation for the processed part of the image frame; (iii) blur spot diameter determination; (iv) outlier detection and elimination; (v) distance/depth estimation.

The remainder of the paper comprises four sections. The second part gives the necessary mathematical information about depth from defocus approach. The third section deals with proposed in the paper solutions of the problems, cited above. Some experimental studies, results and concluding remarks are discussed in the last two sections.

2 Mathematical background of depth from defocus approach

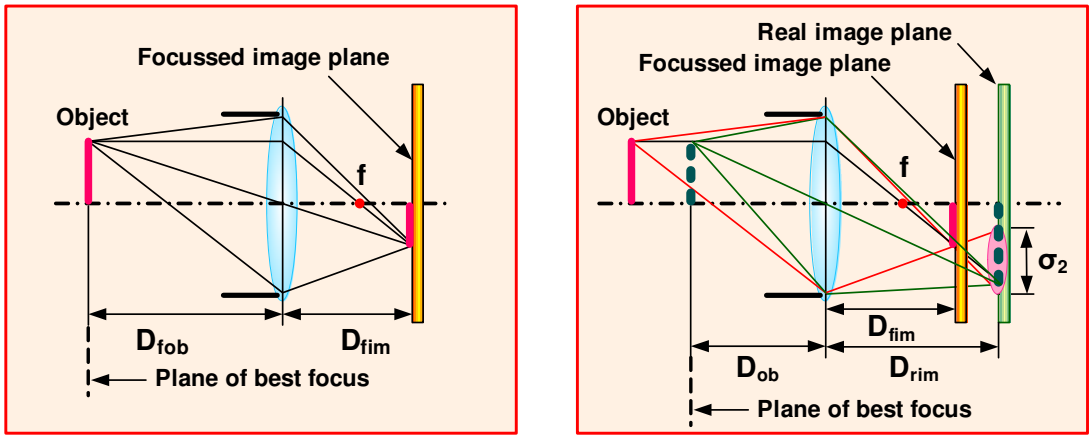
The defocus information in the image of an object formed by a camera system can be used to determine the distance (i.e. depth) to this object from the camera. The general principle of the methods for depth estimation by defocus exploits the physical effect produced by the modification of the focus length or the lens aperture, and the distance to an object on a received image. When a camera is focused on an object at a certain distance a clear (sharp) image is produced but other objects, both closer and farther than the focus distance, form spots more or less blurred according to their distance to the image plane (Figure 1). In case that the sensor is nearer or farther away from the lens than the corresponding lens focus length, the image becomes blurred due to the intersection of light rays either in front of, or behind, the sensor (image) plane. Another factor affecting the blur is lens aperture (iris). Decreasing a lens opening narrows the light rays passing through the lens and reduces defocus spot diameter. Practically, this means that the smallest lens opening will give the sharpest image for a scene of several objects at varying distances. When the aperture is relatively larger (i.e. the lens opening increases), the blur spot diameter becomes larger.

The methods proposed in the literature for depth estimation from blur [3]-[8] use different optical properties of the camera model. The most frequently used model with an intermediate level of complexity is thin lens model. It replaces the multi-lenses camera optic with a thin lens and the geometrical optics is used to derive some basic characteristics of focusing (Figure 1(a)). The Gaussian lens law postulates that:

$$\frac{1}{f} = \frac{1}{D_{fob}} + \frac{1}{D_{fim}} \tag{1}$$

where f is the focal length of the lens, D_{fob} is the distance from the object point to the lens center, and D_{fim} is the distance from the lens center to the plane on which the image of the observed object is in perfect focus. From Eq. 1 it follows that for a chosen focal length there is an infinite numbers of pairs (D_{fob}, D_{fim}), satisfying the equation. The pointed ambiguity shows that some restrictions have to be introduced to the camera model. These constrains stem from the realization of optical sensors. Choosing a suitable zoom setting, the user defines indirectly the scale parameter M – the ratio between the size L_{im} of the image of an object on the sensor matrix

and the actual size L_{ob} of the object. The scale uniquely defines the relation $M = \frac{L_{im}}{L_{ob}} = \frac{D_{fim}}{D_{fob}} = \frac{d_{ccd\ sensor}}{d_{FOV}}$. In the last expression the effective diagonal of the matrix is denoted with $d_{ccd\ sensor}$ and with d_{FOV} - the diagonal of field of view at distance D_{ob}. Furthermore, an additional valid constrain is: $D_{fob} + D_{fim} = D = const$.



(a) Focused object (b) Defocused object

FIGURE 1. Image formation process using thin lens camera model

Let σ_2 denote blur spot diameter (Figure 1 (b)), D_{rim} is the distance from the lens center to the plane of the taken image, B_2 is the diameter of the lens aperture, D_{fob} and D_{fim} are previously defined distances from the lens center to the object and to the plane of the focused image. All these parameters are related by the following equation:

$$\sigma_2 = \frac{B_2}{D_{fim}} abs(D_{rim} - D_{fim}) \tag{2}$$

The distance D_{fim} is expressed from the lens law:

$$D_{fim} = \frac{fD_{fob}}{D_{fob} - f} \tag{3}$$

Let consider that the real image plane is a focused image plane for an object, placed at distance D_{ob} (Figure 1(b)). Thus, Eq. 1 can be used again to express D_{rim}:

$$D_{rim} = \frac{fD_{ob}}{D_{ob} - f} \tag{4}$$

Substituting Eq. (3) and (4) into Eq. (2) gives:

$$\sigma_2 = \frac{B_2}{D_{fim}} abs\left(\frac{fD_{ob}}{D_{ob} - f} - \frac{fD_{fob}}{D_{fob} - f} \right) \tag{5}$$

According to Eq. 5, the diameter of the blur spot physically depends on the lens parameters (B_2 and f) and the depth D_{fob} of a scene point. Thus, focusing camera on different distances,

i.e. varying the focused distance D_{ob}, we obtain the functional dependency of blur spot diameter on D_{ob}, as it is shown on Figure 2, and therefore the distance D_{fob} can be calculated.

The ambiguity in determining the distance D_{fob} for a particular object point is due to the lack of function monotony (Figure 2). The object can be located at two different distances for one and the same value of the blur diameter. An example of the above statement is presented in Figure 2 where Point 1 corresponds to the distance of 2200 mm and Point 2 corresponds to the distance of 4700 mm - in this case, an enormous inaccuracy in the distance estimation for one and the same blur diameter can be detected. This uncertainty could be resolved by combining blur measurements from more than two images, obtained for different focal length settings of the camera.

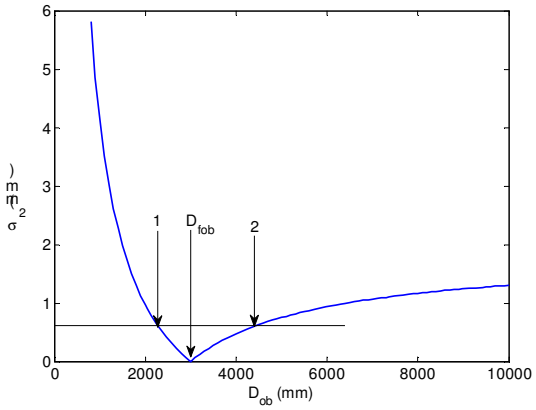


FIGURE 2. Dependency between blur spot diameter and distance to the object in case of a camera focused on distance of 3000 mm

3 Automatic depth estimation using "depth from defocus" approach

The generous algorithm of depth from defocus approach consists of the following steps: (1) Acquiring several image frames of the observed scenario with static camera under different focused distances (the camera settings remain the same, except for the focus setting). (2) Noise estimation. (3) Selection of the features/fields, depth of which will be evaluated. (4) Blur spot diameter determination. (5) Depth estimation of each features/fields in the image. These steps will be described in the following subsections.

3.1 Perceiving image frames

Two types of frame samples are perceived. The first sample consists of two or more frames, acquired under equal camera setting to estimate the noise in the images. The second sample consists of three or more frames taken at different focus settings of the camera. The minimal number of frames necessary for depth estimation is four (2+3=5 frames, but one of the frames in the first sample may used in the second sample too). The time interval for receiving such a quantity of frames usually does not exceed several hundreds of milliseconds. It is considered that this interval is small enough to regard the scene static. The assumption of static scene is not valid for the case of fast moving objects and an additional step for feature registration is required to be inserted into the general algorithm of depth recovery.

3.2 Noise estimation

Noise is the most important parameter, characterizing the quality of received images. The images with low level of the noise further better blur spot diameter estimation and thus enhance the accuracy of depth estimation. The International Standard IEC1146-1 regularizes the procedure of signal to noise ratio (SNR) for analogue cameras in laboratory conditions, but this approach is inappropriate for real time application. The SNR in image sensors can be determined by the ration of generated charge carriers (signal electrons) to the number of unwanted charge carriers (noise electrons). Let assume that the noise signal in an image pixel *(i,j)* is independent, identically distributed (iid) additive and stationary Gaussian with zero mean:

$$I_{i,j}(n) = S_{i,j}(n) + N_{i,j}(n), \tag{6}$$

where $S_{i,j}(n)$ is the useful signal amplitude from the *n*-th image frame, $N_{i,j}(n)$ is the corresponding noise signal and $I_{i,j}(n)$ is the received noisy signal. The intensity level of received signal is known and it is easy to be measured. The main problem is to evaluate the amplitude of the noise. The noise level may fluctuate given different conditions of work and have to be estimated without usage of calibrated sources of light. We propose to use the difference signal of two consecutive image frames with the same camera settings to estimate the noise level [13]:

$$\begin{aligned} &I_{i,j}(n) - I_{i,j}(n-1) = S_{i,j}(n) + N_{i,j}(n) - S_{i,j}(n-1) - N_{i,j}(n-1) = \\ &= S_{i,j}(n) - S_{i,j}(n-1) + N_{i,j}(n) - N_{i,j}(n-1) \end{aligned} \tag{7}$$

In the case of static scenes (static illumination and static objects), the difference signal will be mainly generated from noise. Even of the case of slight changes the difference of useful signal will be greatly depressed. The remaining image signal may be additionally removed by high pass filtering. At the same time (in the case of difference signal analysis) the noise variance will be doubled. Thus, the estimated level of noise variance will be:

$$\sigma_N^2 = \frac{\sigma_D^2}{2}, \tag{8}$$

where σ_D^2 is the estimated noise variance of the differential image.

If the noise is position dependent, the noise evaluation is performed for the pixels of the feature/field of interest.

3.3 Selection of features/fields

The selection of features/fields for depth evaluation is very important task, unresolved until now. It is clear that every pixel in the image plane corresponds to a point (area) from the scene with unique depth (at a given distance). The ambition to work with particular points can not be realized due to the lack of methods for blur estimation on a point.

All other suggested methods are based on multipoint analysis. These approaches have a serious drawback – there is not guarantee that all points correspond on one and the same depth.

Lines (edges, contours) in the image are the most commonly preferred features to be processed. Usually the choice of lines is validated by the fact that they determine the plane borders, and the scenes contain many lines and particularly straight lines. There are many well-developed relatively simple algorithms for line determination – Canny, Sobel, and etc. The very strong benefit of the line exploration is that the intensity change on them is assured.

Often the rectangular fields of different sizes of the image are analyzed. The study of intensity deviation in the field is mandatory for robust depth estimation. If the intensity deviation for all processed image frames doesn't exceed the noise variance for the same field, that field will not be useful in depth estimation.

3.4 Blur spot diameter estimation

The method used to estimate the diameter of the blur spot belongs to the so-called "early methods" of blur estimation. It relies on the analysis of lines detected in a camera image and it is well established with known pros and cons.

In this investigation the gradient analysis of image intensity in direction, orthogonal to the edge line, is used to estimate blur spot diameter. In many cases the gradient analysis is applied to a part of a line. The integration reduces the influence of additive Gaussian noise and improves the accuracy of the result. The brightness profile for different focus values – 1.3m, 1.4m, 1.5m, 1.6m, and 1.7m is depicted on Figure 3. The blur spot diameter estimate is received from the width of the brightness profile. In the case when the local template in the processed field disturbs the brightness profile the results are far from the true estimate.

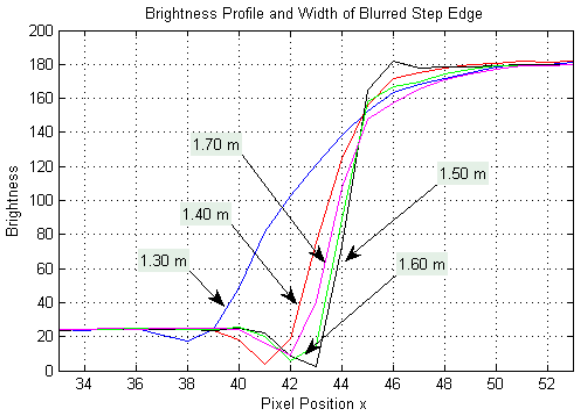


FIGURE 3. Brightness profiles on different focused distances

3.5 Automatic depth estimation using "depth from defocus" approach

It is proposed the object-to-camera distance evaluation to be performed by applying an optimization procedure. The nonlinear curve fitting task for depth estimation from defocus blur is determined in least squares sense: find the vector P of estimated parameters that minimizes the criterion $\min \sum_{i=1}^{m} \left(F(P, D_{ob}(i)) - \sigma_2(i)\right)^2$, where m is the number of the processed image frames, received on different camera focus settings. Here $F(P, D_{ob}(i))$ is the function from Eq. 5, calculating blur spot diameter for the used camera focus settings (distances $D_{ob}(i), \quad i = 1,...,m$) and focal length f. The blur spot diameters $\sigma_2(i)$ are measured for one and the same observed object in the processed image frames.

The parameter vector P consists of three elements: the real distance to the object D_{fob}, the iris diameter B_2 and the scaling coefficient M. The objective function is subject to constraints in the form of parameter bounds. The set of lower and upper bounds of the estimated parameters is determined by the admissible ranges of the camera parameters and the distance to the object. As can be seen from Figure 2 and Eq. 5, the objective function is nonlinear and its solution requires an iterative procedure to establish a direction of search the optimal value of the estimated distance to the object. This is achieved by the Levenberg–Marquardt algorithm [10-12], which interpolates between the Gauss–Newton algorithm and the method of gradient descent. The iterative minimization procedure starts, using an initial guess for the for the parameter vector P. The convergence of the algorithm to final solution - the global minimum, depends on the initial values

of these parameters, as well of the data obtained from measurements of the blur spot diameter. Usually the starting point for the estimated depth is chosen to be equal to the focused length of the camera, corresponding to the minimal blur in the received image frames.

Sometimes, the blur spot diameter cannot be properly estimated in the real scene images. This requires some additional blur estimates (a larger sample) to be taken and the outliers to be removed. Additional data are obtained, analyzing more images of the same object, taken at different focused distances. The outlier rejection is performed applying a simple procedure that detects the outliers by their relatively larger residuals. Then the optimization procedure restarts from the last "quasi-optimal" point.

4 Experimental results

Our experimental work has two goals: (i) to verify the applicability of the mathematical model to the practical camera system we use and to explore the dependency between the camera parameters and the scene characteristics and (ii) to test the evaluation accuracy of the recovered depths in a real scene. Two sets of experiments are conducted using Axis214 PTZ IP video surveillance camera.

(a) Test templates – scene 1 (b) Real scene 2 (c) Real scene 3

FIGURE 4. Experimental scenario

In the first group of experiments seven planar patterns, having two types of vertical edges – "inside" edges (belongs to the same plane of the pattern) and "outside" edges (formed on the transition from one pattern to another) with high contrast are placed at different, a priory known distances from the camera (Figure 4(a)). Three identical experiments were conducted: for shorter distances (1-4 m), for the middle distances (3-6 m) and for longer distances (4-7 m). The patterns (templates) are positioned at intervals of 50 cm. The camera is focused consecutively on each template under different zooms in the range of 6x-9x. The width of blur is calculated for the different camera parameter settings. The difference in pixel intensity is used in order to reduce the influence of the changes in illumination. The influence of the additive Gaussian noise is lowered by integrating up to a hundred points per line.

The second group of experiments concerned real partially structured scenes with many vertical lines (Figure 4(b,c)). The real distances to the object edges were measured in advance by laser distance meter Leica DISTO D3, with measurement accuracy of ±1 mm. The camera is focused consequently on different distances – from an initial selected position through 50 cm and under different zoom settings.

The received image frames were sequentially processed by several procedures: (i) utilization of a Canny algorithm for edge detection and localization; (ii) estimation of the blur spot diameter of the discovered edges; (iii) utilization of the Levenberg–Marquardt optimization procedure, using the blur estimates of the same edge in several frames as input data; (iv) object points distance calculation. Some of the results obtained during the experiments are shown in Table 1.

TABLE 1. Multiple depth recovery: accuracy evaluation.

Test templates – scene 1 Zoom 9x			**Real Scene 2 Zoom 6x**		**Real Scene 3 Zoom 6x**	
	Inside Edges	**Outside Edges**				
Real distance [m]	Estimated distance [m]	Estimated distance [m]	Real distance [m]	Estimated distance [m]	Real distance [m]	Estimated distance [m]
3.0	3.16	2.67	1.37	1.30	3.78	1.64
3.5	3.20	3.27	1.77	1.68	3.22	1.20
4.0	3.65	3.72	2.43	1.49	2.81	2.50
4.5	3.91	4.02	2.04	1.99	2.47	1.75
5.0	4.54	4.68	2.04	1.62	1.69	1.65
5.5	4.83	5.04	1.91	1.92	1.65	1.49
6.0	5.27	4.91	2.06	1.86		

* Focused distances: 1 – 3 m (real scene); 3 - 6 m (test templates)

5 Analysis of results and concluding remarks

In this paper a realization of an approach for computing distance to scene objects when multiple, defocused images are captured from active camera is proposed. The depth recovery task is presented as non-linear line fitting optimisation problem. The received at this early stage of evaluation results show that the proposed technique for estimating the distance to the object points is effective for the purposes of automatic depth perception. In some cases, independently of its easy implementation, it can yield to inaccurate results (see Table 1). The main sources of errors are: (i) improper calibration of camera parameters; (ii) lack of noise level estimation; (iii) failures in edge detection and localisation; (iv) inaccurate blur spot diameter estimate for an edge point. Furthermore, it should be noted that the experimental evaluations were conducted with conventional video surveillance PTZ camera, which is not specially designed for depth estimation purposes.

The thorough analysis of the main sources of errors and careful tuning of parameters of the used algorithms may limit the errors in the distance evaluation to a few percent. Unfortunately, we did not find a testbed for evaluation of depth recovery algorithms for a single PTZ video surveillance camera.

- Based on the performed experimental work with test patterns and real scene targets, the following conclusions and recommendations can be drawn:
- Estimating the scene depth from defocus using Levenberg–Marquardt algorithm requires at least three (better 5 or more) image frames, captured at different focused distances due to the number of the estimated parameters and the presence of outliers.
- The parameter of crucial importance on depth estimation procedure is the blur measure in defocused image frames. In most cases, the distance estimation errors for the 'inside' and the 'outside' edges of the test patterns are approximately equal (Table 1). However, the analysis of the edge intensity profile did not proved itself as reliable algorithm in real scenes, where the edges may have different local structure. The standard gradient operators fail to detect and localize edges when the blur scale, contrast and image noise level exceed some admissible threshold, and therefore the wrong results are received (Table 1, Real scene 3, Real distances 3.78 m and 3.22 m).
- It is necessary to recommend situating camera focus around or in the front of the object, rather than behind it, because the errors in blur spot diameter estimation on the steepest part of the function (Figure 2) lead to smaller errors in distance estimation.

Acknowledgement: This paper is supported by the Bulgarian Ministry of Education and Science under grants VU-MI-204/06.

References

[1] http://www.swissranger.ch
[2] http://www.3dvsystems.com
[3] Schechner, Y., Nahum K., "Depth from Defocus vs. Stereo: How Different Really Are They?," International Journal of Computer Vision, v.39 n.2, p.141-162, Sept. 2000.
[4] Namboodiri, V.P., Chaudhuri, S., "On defocus, diffusion and depth estimation," Pattern Recognition Letters, Volume 28, Issue 3, pp. 311-319, Feb. 2007.
[5] McCloskey, S. Langer, M. Siddiqi, K., "The Reverse Projection Correlation Principle for Depth from Defocus," In Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), pp. 607-614, 2006.
[6] Asada, N., Baba, M., Oda, A., "Depth from blur by zooming," Proc. Vision Interface, pp.165-172, 2001.
[7] Surya, G., Subbarao, M., "Depth from defocus by changing camera aperture: A spatial domain approach," Proc. CVPR, pp.61-67, 1993.
[8] Favaro, P, Mennucci, A., Soatto, S., Observing Shape from Defocused Images, International Journal of Computer Vision, v.52 n.1, p.25-43, April 2003.
[9] Nikolova, I., Zapryanov, G., Alexiev, K., "Detecting of Unique Image Features by Using Camera with Controllable Parameters," In Proceedings of the Fourth International Bulgarian-Greek Conference Computer Science'2008, Kavala, Greece, Volume 3, pp. 920-925, 2008.
[10] Levenberg, K., "A Method for the Solution of Certain Problems in Least-Squares," Quarterly Applied Math. 2, pp. 164-168, 1944.
[11] Marquardt, D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," SIAM Journal Applied Math., Vol. 11, pp. 431-441, 1963.
[12] More, J. J., "The Levenberg-Marquardt Algorithm: Implementation and Theory, "Numerical Analysis, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer Verlag, pp. 105-116, 1977.
[13] Lingfeng Chen, Xusheng Zhang, Jiaming Lin,Dingguo Sha, "Signal-to-noise ratio evaluation of a CCD camera", Optics & Laser Technology Journal, 41 (2009) 574–579.

KIRIL ALEXIEV
Institute for Parallel Processing – Bulgarian Academy of Sciences
Department of Mathematical Methods for Sensor Information Processing
25A Acad.G.Bonchev Str., Sofia, 1113
BULGARIA
E-mail: alexiev@bas.bg

IVA NIKOLOVA
Technical University of Sofia
Department of Computer Systems, Faculty of Computer Systems and Control
8 Kl. Ohridski Blvd. bl 1, Sofia 1000
BULGARIA
E-mail:inni@tu-sofia.bg

GEORGI ZAPRYANOV
Technical University of Sofia
Department of Computer Systems, Faculty of Computer Systems and Control
8 Kl. Ohridski Blvd. bl 1, Sofia 1000
BULGARIA
E-mail:gszap@tu-sofia.bg