Research on Real Aperture Methods for Determining Distance to Objects in a Scene using Defocus Cue

I. Nikolova, M. Karamihalev, G. Zapryanov

Key Words: *Real aperture methods; depth recovery; image processing; computer vision.*

Abstract*. This paper deals with the challenging task of acquiring the three coordinates for the point of interest in the field observed by a camera. This problem is known as scene depth recovery. The present work discusses image analysis techniques relying on depth cues for determining the distance to objects in a scene. Two fundamentally different depth estimation methods, grounded on measuring the degree of image blur and depth map recovery are examined. Both of them are real aperture imaging methods and are based on a limited depth of field of the camera optics. The objective of the present research investigation and conducted experiments is to verify the effectiveness of the evaluation methods in providing reliable depth estimation of real scenes from digital still camera images. For the purposes of the comparative method of analysis, application software and procedure for determining the geometric parameters of the experimental camera, needed in the calculations ,are designed and developed.*

1. Introduction

The problem of recovering the three-dimensional information of a scene observed by a camera plays an important role in many fields such as entertainment, security systems, industrial automation, medical imaging and biometrics, information extraction, etc., and forms the basis for the development of various applications like structure recovery, image enhancement, object recognition, classification and tracking, rendering of novel images, navigation, etc.

This paper deals with the challenging task of acquiring the three coordinates for the point of interest in the field observed by a camera. This problem is known as *scene depth recovery*. The primary consideration in depth estimation from images has been mostly on the utilization of various cues - geometric and/or photometric in order to infer the depth. In depth estimation by geometric cues it is the geometric relationship among various objects in the scene that encode the depth [3]. In stereo based vision two images of a scene are taken from two different viewpoints and measuring the relative disparity between the locations of the objects in the images, the depth in the scene can be recovered. Structure from motion follows a similar principle where a moving object is captured and the relative movement offers an estimate of depth in the scene. While geometry offers a strong cue, it is possible to obtain depth by considering photometric cues as well.

In depth estimation by photometric cues one uses the depth based variations in shading, texture and focus. In shape from shading, given an image of a scene with a known light source and given a reflectance model one uses the variation of shading to estimate the shape in the scene. In shape from texture, instead of assuming that the reflectance of the scene is known or constant, one assumes that the deformation of individual texture elements in the scene is due to projective deformation caused by the variation in orientation and shape in the scene and this is used to recover the depth in the scene. These photometric cues due to the assumptions are fairly limiting in nature. Photometric cues that do not impose limiting assumptions on the scene are those generated by the camera lens and affected by the camera parameters - focus/defocus depth cues. In the area of depth recovery from focus/defocus there have been two main approaches: *depth from focus* (DFF) [7] and *depth from defocus* (DFD) [5,6,10,12], which rely on the real aperture imaging model, discussed in [1].

The objective of this paper is to explore the use of defocus cue in depth estimation and to investigate the practical applicability of two defocus-based depth recovery methods using real scenes, captured by a single digital still camera. Experimental results demonstrate the effectiveness of the methods in providing a reliable estimation of the depth of a scene, and also outline their advantages and drawbacks.

2. Real Aperture Methods for Depth Recovery Using Defocus Cue

Depth form defocus methods rely on the fact that a real lens blurs the observed scene before the camera captures it. The amount of blurring depends on the actual lens (the quality of the optical camera system; the current settings of the camera focus and aperture size), but also on the distance of the observed object to the lens. The basic problem addressed in the DFD methodology is the measurement of the relative defocus between the observations. The relative change in blur is assessed on the base of analysis of a few (usually 2-3) images, produced with different camera settings: focus or aperture size (f-stop). Such analysis aims at determining characteristic quantities of the degrees of defocus blur of the objects presented in the scene that can be used to estimate distance. The sensitivity of this approach depends on the relative change of the defocus blur with distance. The term *focal gradient* has been introduced in [8] to describe this value. Further it has been shown that the focal gradient is inversely proportional to the squared distance to the object and this fact explains the general loss of accuracy at farther distances.

Despite of their variety, the DFD methods can be informally grouped in two groups, depending on the nature of the scenes: (***i***) *methods applicable to scenes with sharp edges*; (***ii***) *methods applicable to arbitrary scenes*.The first group of methods relies on analysis of scenes, composed primarily of sharp-edge objects. The blur is measured using a simple metric based on the width of the edge [8]. The second group includes methods that try approximating the point spread function (PSF) of the camera, which describes the image intensity caused by a single point light source. The size of the camera PSF is a measure of the amount of the defocus. The techniques involved here might be (non-strictly) assigned to the following classes: Methods based on Fourier analysis in the frequency domain, Stochastic methods with Markov random fields to model the parameters of defocus blur [9], Coded aperture methods based on synthetic lens aperture shape usually achieved by using a customized diaphragm [13], Methods, operating in the spatial domain and based on direct raster operations on the image pixel luminance values [12,14].

In the focus of the presented study is the assessment of the practical applicability of two representative DFD estimation methods in the spatial domain for arbitrary scenes: *S-transform* and *Iterative Optimization*.

2.1. S-transform Method for Depth Estimation

The S-transform approach for depth estimation is based on the concept for determining the depth from defocus by measuring the amount of blur in a small region of an image. The solution for the depth of the object is approached by means of finding out the blur circle radius as a function of the relation between variably blurred (observed) images. For this purpose, it is proposed, the optical camera system to be modeled by a convolution operation as shown in *figure 1a*. The *S-transform* is defined as a spatial-domain convolution/deconvolution operation [11]. For the special case of image defocus analysis, the two-dimensional camera image is represented as a composition of planar regions and approximated by a cubic polynomial. Based on this model the following deconvolution formula is derived

$$(1)\quad s(x,y) = i(x,y) - \frac{\sigma_p^2}{4}\nabla^2 i(x,y)$$

where $i(x,y)$ is the blurred image of the scene, σ_p - the spread parameter of the camera (the blur parameter), and ∇^2 is the Laplacian operator.

Modelled in terms of the paraxial optics the blur parameter σ_ρ is a function of the camera parameters (focal length of the lens, f; distance between the lens and the image detector, s; and aperture diameter D) at which the analysed images are captured and the distance to the object, u, and can be expressed as:

$$(2)\quad \pm\sigma_p = mu^{-1} + c$$

where

$$m = -\frac{Ds_0}{2\sqrt{2}},\ c = \frac{Ds_0}{2\sqrt{2}}\left[\frac{1}{f} - \frac{1}{s}\right].$$

The sign of σ_ρ depends of the direction of defocusing observed on the image detector. With s_0 is denoted a pivotal value of s, used for normalizing the images towards magnification.

In the general case σ_ρ and the object function $s(x, y)$ are unknown. Hence, in order to obtain the value of the blur parameter, two blurred images $i_1(x, y)$ and $i_2(x, y)$ of the same scene are required. The images are acquired with different camera parameter settings (m_1, c_1) and (m_2, c_2) corresponding to blur parameters σ_1 and σ_2. After certain transformations using Eq.(1) and Eq.(2), it can be shown that in case only the aperture size is altered between image acquisitions, the spread parameter σ_2 of the second image can be calculated using the equation

$$\sigma_2 = \sqrt{\frac{\pm G}{\alpha^2 - 1}}$$

where

$$\alpha = \frac{m_1}{m_2} \text{ and } G = 16\frac{\iint_S (i_1(x,y) - i_2(x,y))^2\,dxdy}{\iint_S (\nabla^2 i(x,y))^2\,dxdy}.$$

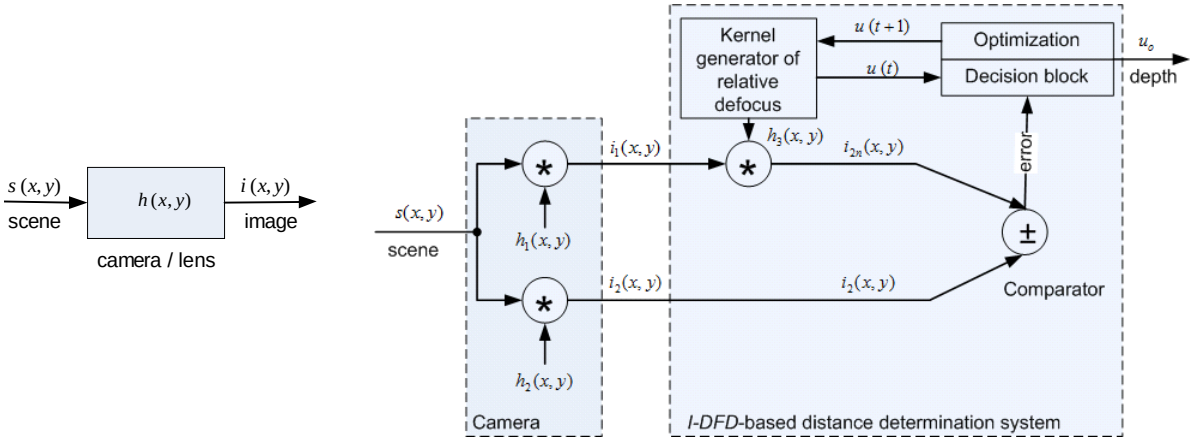
The $\nabla^2 i(x,y)$ is the mean value of the Laplacians of both images $i_1(x, y)$ and $i_2(x, y)$.

In order to enable the method of the S-transformation to be applied in practice to determining the distances to the objects in the scene, it is necessary to have the parameters of the camera at wich the analysed images are acquired. Most frequently they are not known in advance or only some of them are known (e.g., focal length, aperture), but they can be obtained experimentally through a calibration procedure).

2.2. Iterative Optimization Method for Depth Estimation

The method of iterative optimization (I-DFD) is based on measuring the relative defocus between two images $i_1(x, y)$ and $i_2(x, y)$ acquired under differing focal gradients, and using this measure to imply a particular depth. The goal of the iterative optimization procedure is to find the optimal value of the relative defocus without solving the inverse convolution as opposed to the S-transform method [4].

The basic principle of I-DFD is illustrated in *figure 1b*. It is grounded on the idea of how much more the sharp image would have to be defocused in order to be as blurry as the blurry image. This is equivalent to blurring image $i_1(x, y)$ with a new defocus PSF, $h_3(x, y)$ as shown in *figure 1b*. The purpose of the relative defocus kernel generator is to generate a unique relative defocus PSF (h_3) for any particular distance $u(t)$. The image comparator has to determine how well matched image $i_{2n}(x, y)$ and image $i_2(x, y)$ are, where image i_{2n} is derived from image i_1 after convolution with the generated kernel h_{23}. The result of this comparison, in the form of error estimate, is submitted to the decision block. The task of this block is to select the next combination $(u(t), h_3)$ to try and to produce the new value of the distance $u(t + 1)$, needed for the next iteration. This process is repeated until the comparator determines that



a b

Figure 1. (a) Convolution model of an optical system and (b) The mechanism of *I-DFD* method

the two images are a good match. The relative defocus PSF used to obtain this match is associated with a unique depth. After that the decision block terminates the search and announces the current value of the distance $u\ (t)$ as the final value u_0. The accuracy of the whole system depends of how well the relative defocus h_3, matches the camera PSFs and correlates to the desired depth. The camera PSFs $h_1\,(x,\,y)$ and $h_2\,(x,y\ )$ may be described by one of the models of defocus upheld in literature - *geometric (pillbox)* or *Gaussian* parametric models [4, 8, 12]. Depending on this choice, further specific variants of I-DFD method, named I-DFD-R (for *regularized shaping*) and I-DFD-G (for *Gaussian shaping*) are distinguished to model the relative defocus and determine the kernel h_3 respectively.

3. Implementation of the Assessed Depth Recovery Methods

The process of depth recovery using defocus techniques described above has been implemented in three primary stages realized by program modules.

3.1. Image Pre-processing Module

The purpose of the image pre-processing module is to align the original digital images obtained directly from the experimental camera, in a form suitable for their analysis by the camera calibration and/or distance determination procedures. The functional block diagram of the module is presented in *figure 2*, where $I^*_1, \ldots, I^*_n$ denote the input images. The main features of the module and the algorithms involved in it are as follows: (***i***) *Noise reduction:* Wienner filter (W), Adaptive Median filter (M) or Gaussian filter; (***ii***) *Scaling*: an optional operation for resizing the image dimensions if needed; (***iii***) *Feature detection and Tracking*: Kanade-Lucas-Tomasi (KLT) Feature Tracker [2] used for precise alignment of the images necessary to compensate for camera movements or vibrations during acquisition of the image series. (***iv***) *Sorting*: an optional functionality that performs sorting of the images according to their relative defocus level.

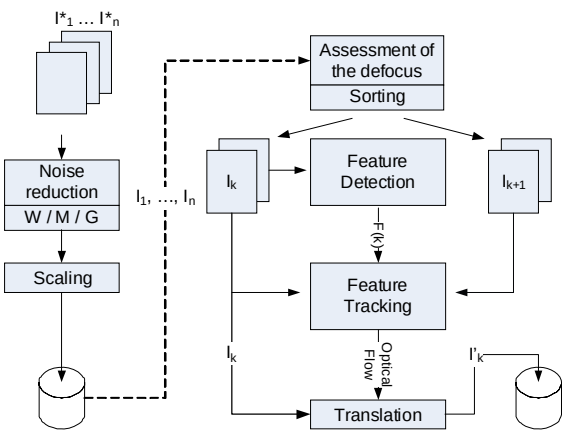


Figure 2. Functional block diagram of the image pre-processing module

3.2. Digital Camera Calibration Module

The camera calibration module is used to indirectly determine the geometric parameters of the experimental camera which are significant and/or represent the input data for the algorithms for determining distance. These parameters are represented by their equivalent coefficients *m* and *c* in the relation shown with Eq. 2. A pair of *m* and *c* can be assigned to each fixed set of camera settings (focus, aperture, zoom). The tabulated values of *m* and *c* (for the range of settings) form the calibration characteristic for a given camera. The functional block diagram of the camera calibration module is illustrated in *figure 3*. The operation of the module is based on the analysis of series of images, $I^*_1, \ldots, I^*_n$ of a calibration pattern (step edge), located at different, apriori known distances from the camera, in order to produce the camera calibration characteristics. The camera settings at which a given calibration image is captured are stored using a data serialization format according to certain syntactic rules.

The calibration process passes through the following two basic steps:

- ***Step 1***. Calculation of the line spread function (LSF) from the linear transient characteristic (edge spread function) of the camera. Two LSF models are considered: *pillbox* and *Gaussian*. The calculation process requires pre-selecting rect-

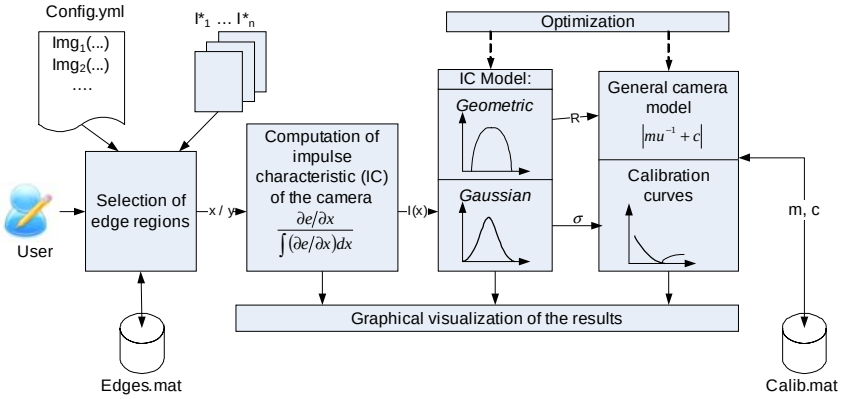


Figure 3. Functional block diagram of the camera calibration module

angular image regions, so that the reference edge is located approximately in the middle part of the region. The size of the region in a direction perpendicular to the direction of the edge should be large enough (it is considered to be 5:2 in favor of the size, perpendicular to the edge direction) to fully cover the transition zone. Moreover, an inaccuracy of the order of 1° in determination of the orientation of the edge leads to an error of approx. 1%.

- ***Step 2***. Determination of the parameters of the general camera model (Eq. 2) applying the model of regression on the values of the defocus blur parameter σ_ρ and the function

$$(3)\quad \left|m(u+e)^{-1}+c\right|$$

where e is a parameter introduced for tracking the systematic error due to the unknown position of the first (front) focal plane of the camera lens.

3.3. Region Selection and Distance Determination Module

The module for selection of regions of interest (ROI) of the scene and determining the distance allows choosing of suitable image sectors for use in the distance determination algorithms. The selected regions should contain planar sections of the scene. The size of the regions should be larger than the largest possible relative defocus kernel. The functioning of the module relies on the usage of the camera calibration curves obtained from the calibration module and offers two types of region selection modes: (*i*) **Manual region selection**, where any number of image regions (rectangular and corresponding to planar zone of the scene) can be entered and removed, resized and moved by the user. (*ii*) **Automatic region selection**. The automatic determination of regions is based on a preset size, number and minimum distance between central points of two adjacent regions. The region selection is performed on the base of the differences in the level of defocusing of a certain region of the scene in an image pair captured under different camera settings. As a measure of a given image region is considered the dispersion of a measure of the defocus level. The obtained values are sorted in descending order and the first *N* are taken. The size of the automatically selected (square) regions their number *N* and the minimum distance between the central points of the neighboring regions can be adjusted by the user.

4. Experimental Evaluation, Results and Discussions

Experimental tests conducted in the context of the discussed methods can be conditionally divided into two groups: (***i***) experiments to establish the parameters of the digital camera (camera calibration); (***ii***) experiments to test the practical usefulness of the methods for depth recovery based on defocus analysis.

Experiments are conducted using conventional digital still camera Olympus E-P2 with lens „Olympus OM-System Zuiko“, 50 mm, f-stop 1:1.4. This lens has a very high sharpness, especially for higher aperture values. Initially, experiments for determining the parameters of the experimental camera were conducted under the following conditions: (***i***) a calibration pat-

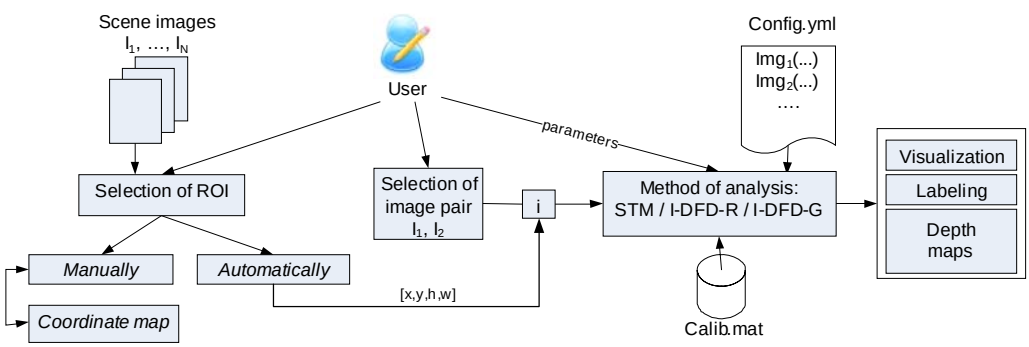


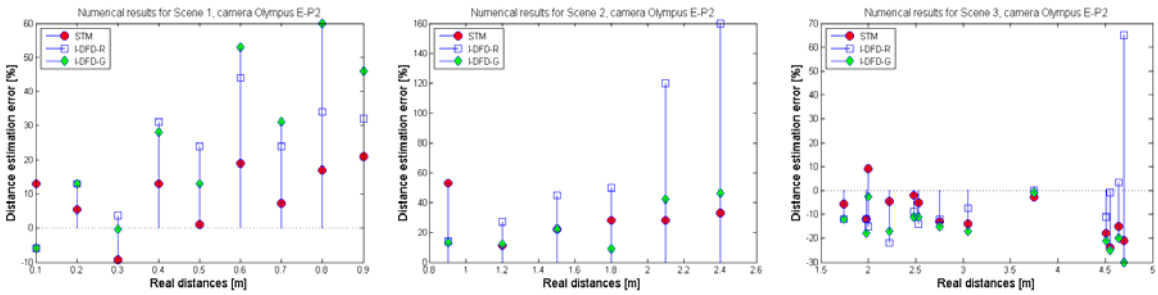
Figure 4. Functional block diagram of the distance determination module

tern of step edge was placed in the front of the camera such that the edge was located near to the center of camera's field of view; (***ii***) the camera was focused to 3.6 m and to infinity and 2x96 images of the pattern were acquired for distances from 3.6 m to 0.30 m by step of 30 cm; (***iii***) for each distance the aperture's size of the camera was changed from 1.4 to 16 and 8 images were acquired. The image data were processed by the calibration module and as result the calibration curves of the experimental camera were received.

During the second group of experiments, several simple and more complex test scenes are composed to assess the accuracy of the investigated DFD methods. The real distances to selected objects from the scenes were previously measured by laser distance meter Leica Disto™D3 with measurement precision ± 1 mm, in order to be used as ground truth data to determine the evaluation accuracy of the investigated DFD methods. Both methods (STM and I-DFD with its variations I-DFD-R and I-DFD-G) are compared, and the results obtained for three different scenes (*figure 6*) are summarized and plotted in *figure 7* in the form of relative distance estimation. In *figure 8* are presented the reconstructed depth maps for Scene 1 with each of the three approaches. As it can be seen from the plots, the resultant depth measures obtained were however not very accurate and the uncertainly in distance determination is higher, especially for complex scenes - under- or overestimation of the distance (*figure 7*).The methods are more accurate for nearby objects than for distant objects, because the blur varies linearly with inverse distance. The S-transform method is restricted to isolated objects; presence of other objects nearby (within a distance of about twice the spread parameter of the object) affects depth estimation (*figure 6 b* and *figure 7 b*) increasing significantly the measurement error. Also, the effective range of this method depends on the constants m and c in Eq. (2) and the image quality in terms of spatial and gray level resolution.Nevertheless, the STM is more accurate as opposed to I-DFD and its two modifications.

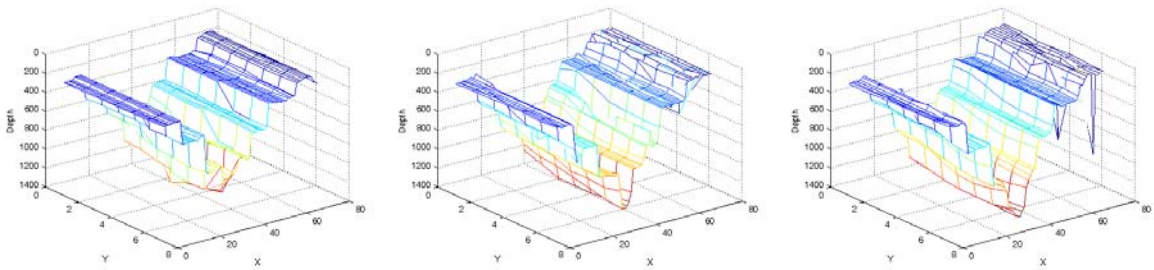
(a) *Experimental Scene 1* (b) *Experimental Scene 2* (c) *Experimental Scene 3*

Figure 6. Selected test scenes acquired with camera Olympus E-P2



(a) Scene 1 (b) Scene 2 (c) Scene 3

Figure 7. Relative error of distance estimation for STM, I-DFD-R and I-DFD-G methods



(a) STM I-DFD-R I-DFD G

Figure 8. Recovered depth maps of experimental scene 1

5. Conclusions

In this paper, two fundamentally different methods of determining the distance to objects in a given scene depth map recovery are examined. Both methods are grounded on measuring the degree of image blur and rely on a limited depth of field of the optical system of the camera.

The methods are implemented in software developed by the authors. Also, a procedure for determining the geometric parameters of the experimental camera, needed in the calculations, is implemented. Various experiments with capturing and analysis of real scene images are carried out with the help of the developed software suite to verify the practical applicability of the investigated methods.

With a view to the results obtained for simple and complex scenes, it can be concluded that the investigated methods characterize with low accuracy when applied to real scene images with many and overlapping objects. However, given some of their advantages, the methods based on the degree of defocus can serve as an initial step in various hybrid approaches (for example, in combination with stereo approaches). Theoretic models and experimental results acknowledge that the discussed methods are effective only in the steep area of the focal gradient of the optical system. This practically means that they are not suitable for determining distance to remote objects (except when telephoto lenses are used). Therefore, optical systems designs with greater depth of field are not responding well to these methods.

6. References

1. Alexiev, K, I. Nikolova, G. Zapryanov. 3D Scenes Recovery through an Active Camera Based on Blur Assessment of the Resulting Image. - *Journal on „Information Technologies and Control“*, Year VI, No. 3-4, 2008, ISSN 1312-2622.

2. Bouguet, Jean-Yves. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm. Intel Corporation Microprocessor Research Labs.

3. Cyganek, B., J. Siebert. An Introduction to 3D Computer Vision Techniques and Algorithms, John Wiley & Sons, 2009.

4. Ens, J. and P. Lawrence. An Investigation of Methods for Determining Depth from Focus. - *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, 1993, 97-108.

5. Favaro, P., S. Soatto, S. A Geometric Approach to Shape from Defocus. -*IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (3), 2005, 406-417.

6. Leroy, Jean-Vincent, Simon, Thierry, Deschenes, François. Real Time Monocular Depth from Defocus. Proceeding ICISP ’08 Proceedings of the 3rd International Conference on Image and Signal Processing, Springer-Verlag Berlin, Heidelberg, 2008.

7. Matsuyama, T, N. Asada, H. Fujiwara. Edge and Depth from Focus. - *International Journal of Computer Vision*, 28 (2), 1998, 153-163.

8. Pentland, A. A New Sense for Depth of Field. Proceedings of the 9-th International Joint Conference on Artificial Intelligence - IJCAI, 1985, 988-994.

9. Rajagopalan, A. N. and S. Chaudhuri. A Recursive Algorithm for Maximum-likelihood-based Identification of Blur from Multiple Observations. - *IEEE Trans. Image Processing*, 7 (7), 1998, 1075-1079.

10. Soon-Yong Park. An Image-based Calibration Technique of Spatial Domain Depth-from-defocus. - *Pattern Recognition Letters* 27 (12), 2006, 1318-1324.

11. Subbarao, M. Spatial-Domain Convolution/DeconvolutionTransfrom. Tech. Report No. 91.07.03. Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY, 11794-2350.

12. Subbarao, M. and G. Surya. Depth from Defocus: A Spatial Domain Approach. - *Journal on Computer Vision*, 13 (3), 1994, 271-294.

13. Zhou, Changyin, Lin, Stephen, Nayar, Shree K. Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring. - *International Journal on Computer Vision*, 93, 2011, 53-72.

14. Ziou, Djemel, Francois Deschenes, Depth from Defocus Estimation in Spatial Domain. - *Computer Vision and Image Understanding*, 81, February 2001, Issue 2, 143-165, ISSN 1077-3142.

Manuscript received on 13.10.2011

Iva Nikolova *is an assistant professor of computer systems engineering at Technical University of Sofia. She received a MSc. degree in Computer Systems and Technologies from the Department of Computer Systems at the Faculty of Computer Systems and Control. The major fields of her professional and scientific research interests include digital image processing, high performance computer architectures and parallel programming.*

Contacts:
Technical University of Sofia, Sofia, Bulgaria
e-mail: inni@tu-sofia.bg

Mavrodi Karamihalev *received a MSc. degree in Computer System and Technologies from the Department of Computer Systems at the Faculty of Computer Systems and Control, Technical University of Sofia. He is currently working as a service engineer at Siemens Enterprise Communications. His professional interests are concerned with VoIP communication systems, IP networking, platform virtualization, solution troubleshooting lab operation. His scientific interests are related with digital photography and algorithm design and analysis.*

Contacts:
Technical University of Sofia, Sofia, Bulgaria
e-mail: karamihalev@abv.bg

Georgi Zapryanov *is an assistant professor of computer systems engineering at Technical University of Sofia. He received a MSc. degree in Computer Systems and Technologies from the Department of Computer Systems at the Faculty of Computer Systems and Control. His professional and scientific research interests include algorithm design and analysis, digital photography, image sensors and digital image processing.*

Contacts:
Technical University of Sofia, Sofia, Bulgaria
e-mail: gszap@sofia.bg