Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision, Malta, September 15-17, 2005 (pp88-93)

PICASSO 2 – a System for Performance Evaluation of Image Processing Methods

I.V. GRIBKOV, P.P. KOLTSOV, N.V. KOTOVICH, A.A. KRAVCHENKO, A.S. KOUTSAEV, A.S. OSIPOV, A.V. ZAKHAROV
Scientific Research Institute for System Studies, Russian Academy of Sciences
(NIISI RAN)
Nakhimovskii pr. 36-1, Moscow, 117218
RUSSIA

Abstract: The new version of the system for a comparative study of image processing algorithms named PICASSO 2 is presented. We demonstrate its use in comparative evaluation of three energy minimizing methods (by Mumford-Shah, Geman-Reynolds and the method with piecewise-linear energy minimizing function). The testing includes quantitative and qualitative estimation of noise reduction property. Also three methods for segmenting an image into textured and nontextured regions are tested by using PICASSO 2.

Key-Words: image processing, energy minimization, image segmentation, image restoration, edge detection, textured and nontextured regions, ground truth, performance evaluation.

1 Introduction

One of the important tasks in computer vision research is the development of the methods which provide a comparative assessment of image processing algorithms, [1-3].

In order to obtain a tool for testing and evaluation of the image processing methods, we are developing the software system named PICASSO (PICture Algorithms Study SOftware). Originally it was designed to compare various edge detection algorithms on a set of artificial 2D images [4]. The first results of its practical usage allowed us to formulate the main goal of our further research: to create an adaptive system for real image segmentation on the basis of PICASSO. Reaching our goal requires the inclusion of new methods into the existing system in order to cover a wide range of image processing tasks, such as image restoration, edge detection, boundary improvement and texture analysis. Also the testing technique needs further revision. To meet these tasks, the new version of PICASSO system, named PICASSO 2, has been created.

In the present paper, we describe the main features of this new version and demonstrate its use in comparative evaluation of a certain class of image restoration algorithms as well as some algorithms intended for segmenting an image into textured and nontextured regions.

The paper is organized as follows. In Section 2 we briefly describe the new features of PICASSO 2. Section 3 contains a comparative analysis of energy minimizing methods used for image restoration. We mention that nowadays there exists a certain amount of classes of energy minimizing methods applicable not only for edge detection, but also for image restoration and boundary optimization. In [5] we introduced some results concerning the testing of two such classes by means of our system. We tested their edge detection and boundary improvement properties. Here we consider the image restoration capabilities of one class of the methods studied in [5]. Finally, in section 4 we test three methods of discriminating textured and nontextured regions of an image [10-11].

2 PICASSO 2 – Basic Features

The original version of PICASSO system has been described in [4] and as mentioned above, PICASSO 2 represents its further extension. Note that the core feature of both PICASSO and PICASSO 2 is modeling of typical situations in image processing. We have worked out a set of synthetic images. These images simulate a collection of situations which are difficult in some sense for the image processing methods. Now the image database of PICASSO 2 contains a set of grayscale images as well as a set of corresponding reference images (ground truth) and their brief descriptions. Also the system contains the special image editor, software implementation of the methods tested, noise generators, filling templates for background and objects.

After the selection of a method or a group of methods for testing, the following procedure is implemented in PICASSO 2. It consists of four stages, namely:

Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision, Malta, September 15-17, 2005 (pp88-93)

For the selected methods, forming a set of test images with given features
Optimization of the methods' parameters
Selection of performance metrics
Statistical processing of results

This approach to testing is widely known (see e.g. [9]) and already proved its efficiency.

3 Energy Minimizing Methods

In this section we test energy methods for image restoration which use functionals: a) Mumford-Shah's, b) Geman-Reynolds', c) functional with piecewise-linear energy minimizing function (we mark them as MS, GR and PL). Originally MS and GR methods were developed not only for image restoration but also for boundary extraction Formulas of these functionals were rather complicated [6, 7]. However, as shown in [8], the tasks of image restoration and boundary extraction by means of MS and GR methods can be split up into two steps. The first step deals with image restoration and involves the solving of a certain differential equation by iterations which amounts to finding the restored image. At the second step, using the restored image, the boundaries are calculated. This can be done by using some simple methods without solving differential equations. In other words, the basic feature of the methods is restoration. It is important to note that the restoration could be implemented by using a simpler functional than the initial one, used for simultaneous restoration and boundary extraction. We call such functionals *modified* MS, GR and PL functionals.

In [5] we have tested the quality of boundary extraction by the above mentioned methods. Now we test their restoration capabilities. Namely, we find quantitative estimates for the capabilities a) to reduce noise on the image b) to preserve contrast discontinuities of the image. The results obtained are presented as diagrams and 3D graphs.

3.1 Artificial Test Images

The methods are studied using the following test images from Picasso's database: (Fig 1): degenerating ridge a), degenerating step d), snail e) and junction f). These images involve domains of slowly varying contrast as well as domains of contrast discontinuities. The value of discontinuity also varies. At the upper row of Fig.1 we additionally show the boundaries (b) of the image "degenerating ridge" and its noisy version (c). Before processing an image, the procedure of edge detection and adding some noise is applied to each test image.

Certainly, one can instead consider natural images in order to evaluate noise reduction capabilities. Such images could be taken as "ground truth". After adding a noise, the images could be restored by energy methods. However, in order to evaluate edge preserving capabilities, the explicit edges are needed. The PICASSO's image database is very convenient for that purpose since it already contains the edges of any image.

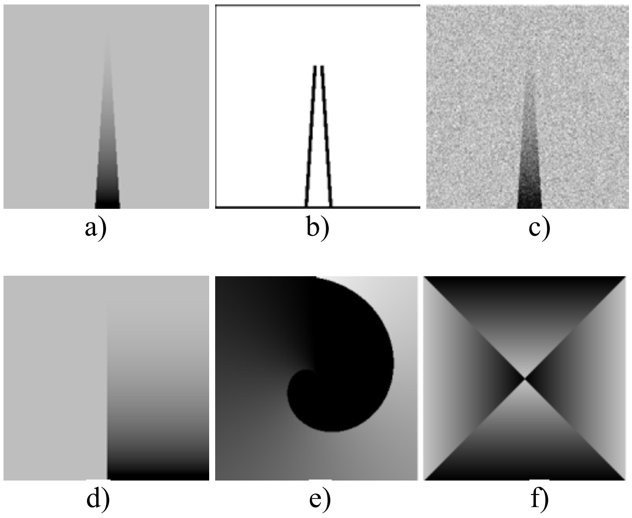


Fig.1. a) degenerating ridge, b) boundaries of the image a), c) Gaussian noise added to the image a); d) degenerating step, e) snail, f) junction

3.2 Performance Metrics

Noise reduction property is evaluated as follows. Let I be an image presented as a set of pixels I_{pq}, where p,q is the position of a pixel. Consider the image norm $\| I \|^A = \sqrt{\sum_{(p,q) \in A} I_{pq}^2}$, where A is a domain of the image (not necessarily of the entire image). Next, define the metric

$$d^A(I_1,I_2) = \| I_1 - I_2\|^A . \quad (1)$$

Let I_{init} be initial image without noise, and $I_{noise,\sigma}$ be a sequence of noisy images obtained from I_{init} by adding to it Gaussian additive noise with variance σ. Here $0 \le \sigma \le 30$. The values $d^A_{noise,\sigma} = d^A(I_{noise,\sigma}, I_{init})$ show the distances between initial and noisy versions of an image. After processing of $I_{noise,\sigma}$ by an energy method, we get a set of restored images $I_{proc.\ \sigma}$. Similarly the values $d^A_{proc,\sigma} = d^A(I_{proc,\sigma}, I_{init})$ can be calculated. These values show the distance between processed images and initial ones.

To evaluate the restoration capability, we use two similar metrics of the above type. Namely, denote by B a small neighborhood of the boundary in the initial image. For example, such a neighborhood is presented as two black lines on Fig.1b). In our calculations, we take 4-pixel width neighborhood. Denote by d^B the metric defined by (1) for $A=B$. This metric is applied to contrast discontinuities domains aiming to evaluate edge preserving capability of a method. Usually d^B is large if the boundaries are blurred after image processing.

Similarly we define the second metric: $d^{\Omega\backslash B}$, where Ω is the entire domain of the image. It is white on Fig.1b) and occupies the largest part of the image. This metric is applied to slowly varying contrast domains in order to evaluate noise reduction capability of a method.

3.3 Formulas of Functionals

The formulas for the restoration versions of modified GR and MS functionals have been obtained in [4-5]. The energy E for modified GR and PL functionals is calculated as

$$E(u) = (1-\lambda)\int (u-I)^2 + \lambda\int c^2 \cdot \varphi\left(\left(\frac{\|\nabla u\|}{c}\right)^2\right), \quad (2)$$

where $c^2 = \propto^2 \max_{(p,q)} \|\nabla I(p,q)\|^2 \, ; 0 < \lambda, \mu \le 1$.

Here I is the initial image, u - the solution of the minimization task, Ω - image's domain, λ, μ are parameters of both methods. For GR method, $\varphi(r) = r/(1+r)$ and for PL method, φ is piecewise linear: $\varphi(r)=r$, if $0 \le r \le 1$, and $\varphi(r)=1$, if $r > 1$.

The parameter λ regulates the relative contributions of two terms in the right hand side of (2) aiming to minimize $E(u)$. Small values of λ restrict a solution to be close to the initial image. Large λ force the solution u to be smooth. The parameter μ controls the properties of the restoration term. The smaller the value of μ is, the better is edge preserving qualities and the worse is noise reduction.

The modified MS functional takes the form

$$E(u) = \int \frac{r}{1+r}, \quad r = (1-\lambda)(u - I/c)^2 + \lambda \cdot \|\nabla u\|^2,$$

where λ, c are parameters of this method. The smaller the value of c is, the better is edge preserving quality and the worse is noise reduction.

3.4 Results

In order to reach a better comparison, the performance results for all the methods are presented within common graphs (Fig. 2-3). Here we set λ=0.8 for all methods, μ=0.5 for GR and PL and c=50 for MS. For each test image on Fig.1, the 4-pixel width boundary neighborhood was found (domain B) and to each image a Gaussian noise with zero mean and variance σ was added. All calculation results were averaged over four test images a), d), e), f) (Fig1).

Fig.2 shows the distance between the initial ground truth image and a) noisy images before processing and b) noisy images after processing using different methods. The distance is calculated in domains of slowly varying contrast using $d^{\Omega\backslash B}$ - metric. This graph shows mainly noise reduction properties of GR, PL and MS methods.

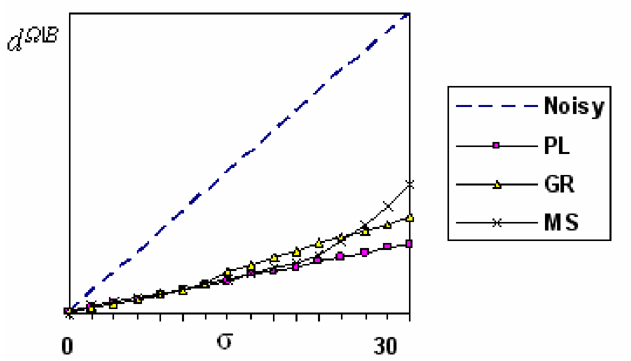


Fig.2. Distances between domains of slowly varying contrast in initial image and in a) noisy images (dashed) b) processed noisy images using GR, PL, MS methods. Here σ is the variance of Gaussian additive noise.

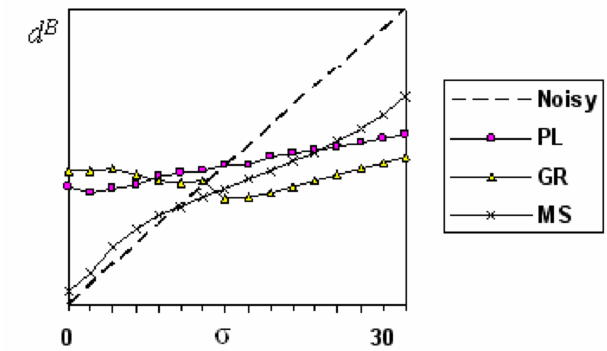


Fig.3. Same as Fig. 2, for boundary domains

Fig.3 shows the distance d^B between the boundary area of the initial image and boundary areas of a) noisy images before processing and b) noisy images after processing using the above methods.

For PL and GR methods, d^B is considerable even if the noise level is small. It means that these methods blur boundaries, especially if contrast discontinuities are small. PL and GR methods are good to restore highly noisy images.

Conversely, for MS method, d^B is small if the noise level is small. It means that MS method preserves well small contrast discontinuities. This method is good to refine images with small level of noise, but it fails when the noise is high enough.

For Laplacian and uniform noise distributions we obtained similar results.

In order to reveal more properties of a method being tested, it is useful to vary not only σ, but also the parameters of the method. As we have found out, the parameter λ is less informative than μ, so it is possible to fix λ=0.8±0.1. This guarantees good convergence of the iteration process.

Figs.4 and 5 represent the distances between the initial image degenerating ridge and restored images after adding Gaussian noise with zero mean. The upper graphs of Fig.4 (GR method) and Fig.5 (MS method) are calculated in areas of slowly varying contrast; down graphs – in the neighborhood of boundaries. For PL method the corresponding graphs are similar to Fig 5.

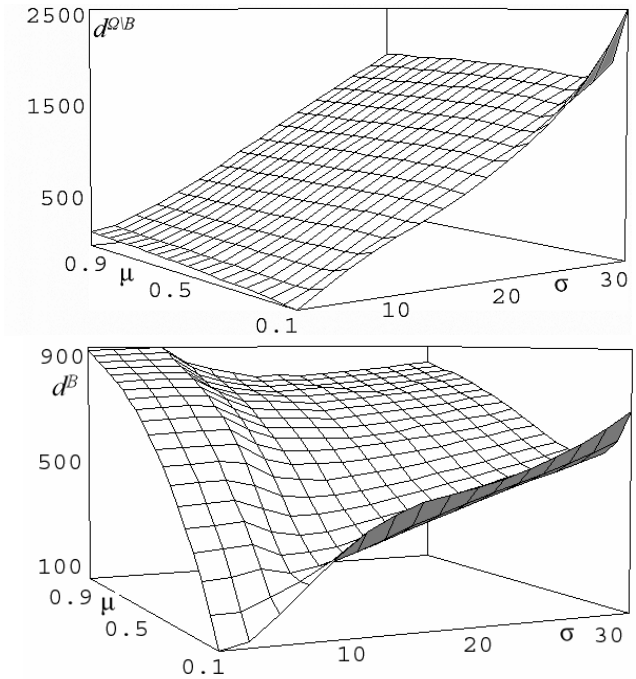


Fig.4. Distances between "ground truth" image and noisy images processed using GR method. Upper graph: distances for domains of slowly varying contrast; lower graph – for boundary domains. Here λ=0.8, Gaussian noise variance σ = 0..30 and parameter μ =0.15..0.95.

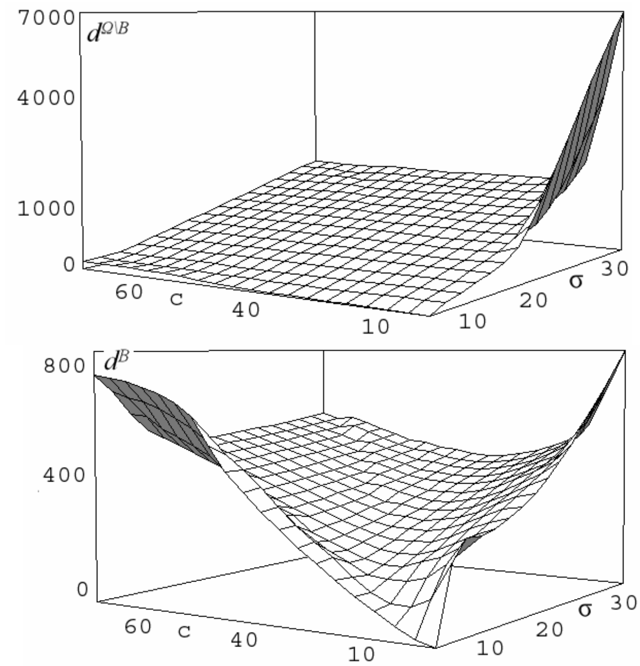


Fig.5. Same as Fig.4, for MS method. Here λ=0.8, Gaussian noise variance σ = 0..30 and parameter c =5..70.

Note that $d^{\Omega\backslash B}$ grows quickly at large values of on σ, and small values of μ (Fig.4) or c (Fig.5). This can be explained by the fact that at small values of parameters, the methods take noisy pixels for edges and try to preserve them. Also, the minimal value of d^B on Fig.5 equals 100. It means that GR method blurs boundaries even at zero noise level. Conversely, minimal value of d^B on Fig.6 equals to zero. This means good edge preserving capabilities of MS method at small noise levels.

The graphs on Fig. 2-6 help a user choose appropriate parameters for image restoration methods in order to achieve optimal performance.

4 Evaluation of Algorithms for Segmenting an Image into Textured and Nontextured Regions

4.1 Algorithms

Applying of image processing algorithms to "unsuitable" images often provides meaningless results (for example, using edge detection algorithms for processing a textured image, or extracting texture features from nontextured image). To avoid this drawback the input image may be preliminary divided into textured and nontextured regions. In what follows, several appropriate algorithms will be applied to suitable regions of the image. We evaluate here three algorithms of preliminary segmentation.

In [10] the algorithm, based on computing the density of local extrema of the image function was studied. A pixel at (r_0,c_0) is a local row maximum if:

$$I(r_0,c_0-1) < I(r_0,c_0) > I(r_0,c_0+1),$$

where $I(r,c)$ – the image function with row coordinate r and column coordinate c. If an interval $\{(r_0,c) \mid a \leq c \leq b\}$ has a constant gray value $I(r_0,c)=g$, such that $g > I(r_0,a-1), I(r_0,b+1)$, then the center pixel $(r_0,(a+b)/2)$ is also a local row maximum. Local row minimum, column minimum and column maximum are defined in a similar way. A pixel is a local extremum if it is both a local row extremum and a local column extremum.

The density of local extrema is calculated in a small window around the pixel. It characterizes texture coarseness. A pixel is considered as textured if the local extrema density in its neighborhood falls in the specified range of values.

The operator MAXDIF was proposed in [11]:

$$MD(F,K)[i,j] = max\{W(F,K,i,j)\} - min\{W(F,K,i,j)\},$$

where image $MD(F,K)$ is the result of processing input image F using $K \times K$ window, $W(F,K,i,j)$ is the subset of input image pixels belonging to $K \times K$ window which center is placed at (i,j) pixel.

This operator sets the center pixel of a window to be the difference between the maximum and the minimum gray level values within that window. This difference is high for textured image regions and low for nontextured ones. A pixel is labeled as textured if its gray value is greater than a specified threshold.

The third algorithm, used in our comparative study, is based on counting the edge number in pixel neighborhood. Some edge detection algorithm is initially applied to an input image (we have used the Canny detector). Then an edge density (the edge number divided by the area of the window) is calculated for each pixel. High value of this density corresponds to textured image regions.

4.2 Evaluation Technique

In order to evaluate the above mentioned algorithms we have used a set of test images and corresponding ground truths from the PICASSO's image database. This set contains both real and synthetic grey scale images (two test images are depicted in Fig. 6 as an example). The real image subset is represented by five aerial images. Four synthetic images were specially designed for assessing the ability of the algorithms to discriminate textured regions characterized by different coarseness and grey level variations from homogeneous and noisy regions.

For quantitative evaluation of segmentation quality we use the modified Pratt's figure of merit (FOM) [12]:

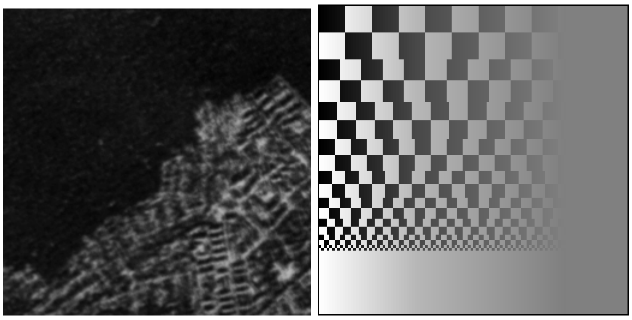
$$FOM = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+.d_i^2},$$

where N – number of pixels in the image,

d_i – Euclidean distance between the ith pixel of segmented image and its correct class,

γ – scaling constant which can be used to change a contribution of errors to the FOM.

FOM =1 if the segmentation is absolutely correct.

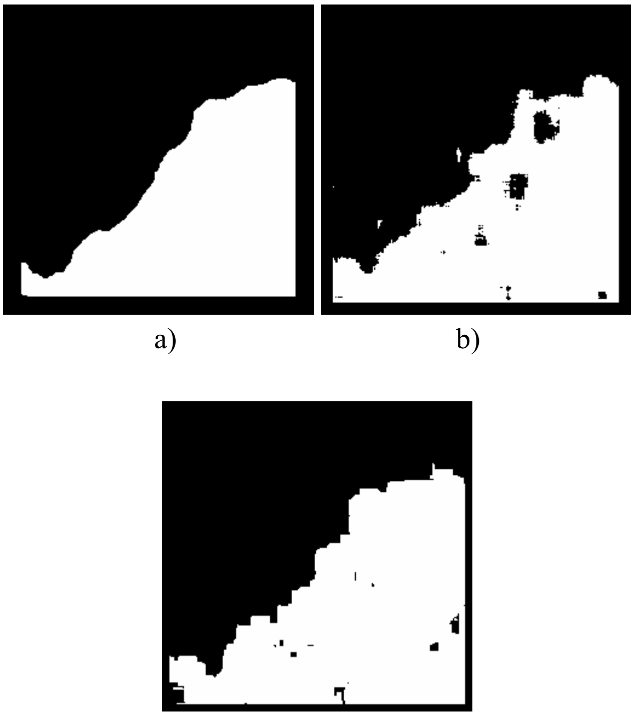


a) b)

Fig.6. Examples of test images: a) aerial image b) synthetic image.

4.3 Results

Each test image is processed by the algorithms using different values of window size. For the aerial image from Fig.6a) the processing results are shown in Fig.7. We evaluate them by comparing the segmented images with the corresponding ground truths. Values of performance measure FOM are calculated separately for real and synthetic image subsets. A summary diagram is presented in Fig.8.



a) b) c)

Fig.7. Results of processing the image Fig.6a) by different algorithms: a) Edge Density b) Local Extrema Density c) MAXDIF.

Our results show that the edge density algorithm demonstrated the best performance in comparative study of the algorithms being tested. The MAXDIF's performance is only slightly lower.

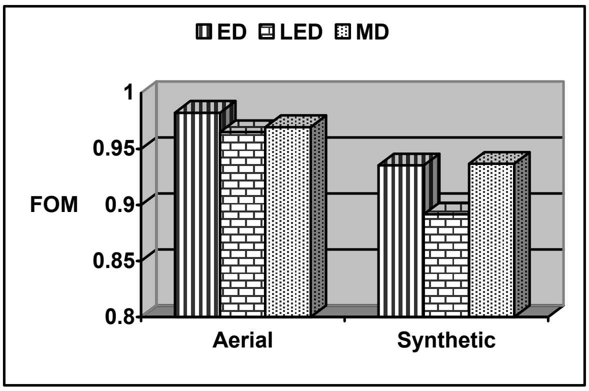


Fig.8. Summary of algorithmsэperformance.

A runtime is another important characteristic of preliminary segmentation algorithms. The runtime of MAXDIF is smallest in the test, but as can be seen in Fig.9 MAXDIF's segmentation quality drops substantially when unsuitable window size is used. So it may be necessary to repeat the process some times using various window sizes. It cancels out the speed advantage of this algorithm over others.

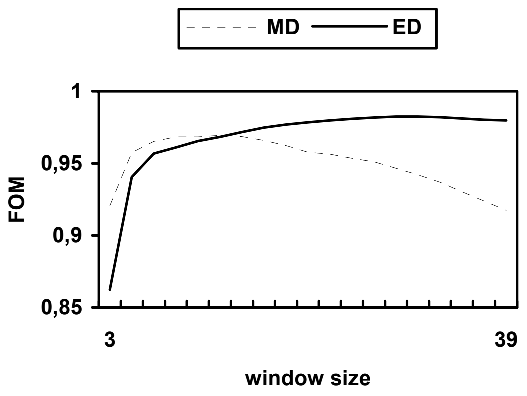


Fig.9. FOM as a function of window size.

5 Conclusions

In view of the above, we can conclude that the current version of PICASSO system represents an effective tool for a comparative study not only of edge detectors, but also of some algorithms intended for image restoration and texture analysis. It makes possible to quantitatively evaluate the performance of the algorithms being tested, find their advantages and disadvantages and compare the results in the automatic mode.

Next steps on the way towards the creation of an adaptive system of real image segmentation will be devoted to boundary improvement. To meet that task we are developing within PICASSO 2 a bank of test images for performance evaluation of snakes and balloon methods and region growing methods. We are planning to devote a separate paper to a thorough study of that issue.

References:

[1] T. Pavlidis, Why Progress in Machine Vision is So Slow, *Pattern Recognition Letters*, Vol.13, No.4, 1992, pp. 221-225.

[2] P. Courtney and N.A. Thacker, Performance Characterisation in Computer Vision: The Role of Statistics in Testing and Design, In *Imaging and Vision Systems: Theory, Assessment and Applications*, J. Blanc-Talon and Popescu (Eds.), NOVA Science Books, 2001.

[3] R.M. Haralick, Performance Characterization in Computer Vision, *CVGIP: Image Understanding*, Vol. 60, No 2, 1994, pp. 245-249.

[4] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Koutsaev, V.K. Nikolaev, A.V. Zakharov, PICASSO – Edge Detectors Evaluating System Based on the Comprehensive Set of Artificial Images, *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics*, Vol. 9, 2002, pp 88-93.

[5] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Koutsaev, V.K. Nikolaev, A.V. Zakharov, Testing of Energy Minimizing Methods in Image Preprocessing Using the PICASSO System, *Proceedings of the 8th World Multiconference on Systemics, Cybernetics and Informatics*, Vol. 6, 2004, pp 233-238.

[6] D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, *Commun. Pure Appl. Math.* Vol. XLII, 1989, pp. 577-685.

[7] D. Geman and G. Reynolds, Constrained restoration and the recovery of discontinuities, *IEEE TPAMI*, Vol. 14, No.3, 1992, pp. 376-383.

[8] G.A. Hewer, C. Kenney, and B.S. Manjunath, Variational Image Segmentation Using Boundary Functions, *IEEE Trans. Image Processing*, v.7, No. 9, 1998, pp.1269-1282

[9] M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, Edge Detector Comparison: Initial Study and Methodology, *Computer Vision and Image Understanding*, Vol.69, No.1, 1998, pp. 38-54.

[10] K. Karu, A.K. Jain and R.M. Bolle, Is There Any Texture in the Image? *Pattern Recognition*, Vol.29, No.9, 1996, pp. 1437-1446.

[11] I. Dinstein, A.C. Fong, L.M. Ni and K.Y. Wong, Fast Discrimination Between Homogeneous and Textured Regions, *Proc. of the Seventh International Conference on Pattern Recognition*, Vol.1, 1984, pp. 361-363.

[12] K.C. Strasters, J.J. Gerbrands, Three-Dimensional Image Segmentation Using a Split, Merge and Group Approach, *Pattern Recognition Letters*, Vol.12, No.5, 1991, pp. 307-325.