UDC 004.932.72

T. Mandziy

Nonoptimizational Approach to Invariant Object Detection

Предложен новый подход к инвариантному обнаружению объектов. Рассмотрен общий случай инвариантности к необходимому набору преобразований. Модельные экспериментальные результаты, подтверждающие эффективность предложенного подхода, представлены задачей обнаружения объектов изображений.

A new approach to invariant object detection is proposed. A general case solution for invariance to required set of transformation is considered. Basic experimental results that justify proposed approach are presented on the task of image object detection.

Запропоновано новий підхід до інваріантного виявлення об'єктів. Розглянуто загальний випадок інваріантності до необхідного набору перетворень. Модельні експериментальні результати, які підтверджують ефективність запропонованого підходу, представлені задачею виявлення об'єктів зображень.

1. Introduction. Object detection and recognition is one of the most difficult tasks in computer science. Up to date there exist many approaches to object detection and recognition. Majority of those approaches aims to solve the problem for a certain narrow subset of object classes (faces, cars, textures, etc). The problem of such variety of approaches consists mainly in absence of general solutions to the object detection and recognition tasks. For instance, majority of methods and algorithms used for face recognition are inapplicable to texture recognition tasks and vise versa.

Let us consider general object detection on the example of image object detection task. There are many approaches for image object detection [1–3, 9]. Basically those approaches can be divided on two groups. First group encapsulates approaches for template based object detection [1–3]. This group of methods aims to detect objects similar to the given template in an input image. Template can be represented as a set of features [4], a contour representation [2, 9], a learned pattern [5], an actual image of the target object [3], etc. This group of methods is able to locate the object on entire input image. Problems arise when some sort of transformation or distortion is introduced to the target object in an input image. In a presence of such transformations as scale, rotation, projection, shape variations etc. the problem rises to a new level of algorithmical and computational complexity. Presence of such transformations makes conventional approaches more than useless. Existing adaptations of those methods to such transformations are usually restricted to some small subset of them, and do not solve the problem in general.

To the second group of object detection methods we refer to as object segmentation methods. Those are methods, in general, incapable to locate an object position in entire input image but given such initial position they can adjust model parameters to fit a model image to an input image. This second group includes mostly generative models that are able to efficiently model object appearances under some transformations on input image given the approximal initial position of that object in input image [6–8].

Those two groups of approaches to object detection and object modeling exist and develop separately. They both solve only a part of a more general object detection task.

Absence of some kind of "holistic" approach to object detection is a concern of this paper. It presents an attempt to combine advantages of existing object detection and object modeling techniques to produce a new and to a degree general approach to invariant object detection.

2. Invariant image object detection

Under invariance in object detection it is often understood insensitivity of object detection algorithm to a certain set of transformations. Problem of invariance in general case is very complicated and unsolved task. Many papers were dedicated to the solution of invariance problem [1–5, 9, 10, 13]. Most of them attempt to solve this problem for a certain limited set of transformations and those solutions usually do not allow for further extension of transformation set.

Probably the most natural requirement is the invariance to scale and rotation. They appear naturally as result of image acquisition procedure. Most of image object detection tasks have to cope with these two transformations. Methods for invariant, with respect to scale and rotation, object detection and representation are given significant amount of attention [1–5, 9, 10]. Shape changes and variations are another transformations that some object detection and recognition algorithms where designed to be invariant to [6–9, 11, 12].

Many of existing approaches to invariant object detection suffer from lack of fundamental generality of their solutions. Mostly offered solutions designed specifically for certain type of invariance that cannot be applied to transformations of different nature. As the result, absence of some kind of general approach leads to continuous growth of proposed solutions applicable only for narrow set of transformations and object classes.

2.1. Direct MAX computation

The goal of invariant image object detection is to detect a target object on an input two-dimensional image $I(x, y)$ invariant to certain set of possible transformations of target object on $I(x, y)$. Let $M(\xi)$ be a mathematical model of target object image with some parameter vector ξ. Every component of ξ is responsible for certain type of transformation. For instance, first component of parameter vector ξ_1 can be rotation angle φ or scale s of target object, second component ξ_2 can correspond to appearance changes of target object (for instance shape, texture, illumination etc), etc. Let $C(I, M(\xi), x, y)$ be some similarity measure that measures similarity of an input image $I(x, y)$ with target object model $M(\xi)$ at (x, y).

So the objective is to detect object of interest $M(\xi)$ with arbitrary allowed parameter vector ξ on arbitrary input image $I(x, y)$.

Basically described general task of invariant object detection can be represented as follows:

$$C^{inv}(x, y) = \max_{\xi}\{C(I, M(\xi), x, y)\}. \qquad (1)$$

Any existing object detection algorithm can be represented in form of (1). The difference is in the way a particular algorithm solves $\max\{C(I, M(\xi), x, y)\}$ task. In practice this task falls into optimization theory where $\max\{C(I, M(\xi), x, y)\}$ is formulated in terms of some conventional optimization technique (lest squares, dynamic programming, gradient based methods etc) or it is solved inexplicitly by some technique applicable only to a very narrow class of objects. In general $C(I, M(\xi), x, y)$ is a complex function of many variables and local minimums. Thus optimization of (1) is difficult and generally unsolvable with conventional methods task.

The only known general solution for (1) is brute force approach. For object detection task, it requires explicit computation of all possible outcomes of similarity measure C for all possible values of ξ. In discrete case it implies dividing the domain D of allowed values for ξ by $\Delta\xi$. Even though such approach guaranties the solution of (1) it is too computationally expensive for majority of real life object detection tasks.

Computational complexity of brute force solution for (1) consists of two components. The first is computation of values of $C(I, M(\xi), x, y)$ for all possible $\xi \in D$ and the second is computation of max function for obtained values. In discrete case, computational complexity of the second component is neglectable in comparison to first component. But there is a case when computational complexity of those two components can be reversed. For the sake of simplicity and without loss of generality of final solution, let us consider one-dimensional parameter vector ξ. As was mentioned, to solve (1) numerically in terms of brute-force paradigm, $C(I, M(\xi), x, y)$ should be computed for all possible values of $\xi \in D$. In this case (1) takes the next form:

$$C^{inv}(x, y) = \max\{C(I, M(\xi_1), x, y), \ldots, C(I, M(\xi_1 + k\Delta\xi), x, y), \ldots, C(I, M(\xi_2), x, y)\}, \qquad (2)$$

where $\xi \in [\xi_1, \xi_2]$.

For "straightforward" solution of (2) it is proposed to use analytical representation of max function. In discrete case analytical representation of max function for N variables is the following:

$$\max(x_1, x_2, \cdots, x_m) = \lim_{n\to\infty} \sqrt[n]{x_1^n + x_2^n + \cdots + x_m^n}. \qquad (3)$$

Rewriting of (2) in terms of (3) gives the next expression:

$$C^{inv}(x, y) = \lim_{n\to\infty} \sqrt[n]{\sum_{k=0}^{N} C(I, M(\xi_1 + k\Delta\xi), x, y)^n}. \qquad (4)$$

At this point expression (4) only overcomplicates the solution by increasing computational complexity of max function. But the transition from discrete values of ξ to continuous case in (4) dramatically decreases computational complexity of abovementioned firs component – computation of values of $C(I, M(\xi), x, y)$ for all possible $\xi \in D$. By heading $\Delta\xi$ to zero, sum in (3) transforms in the definite integral:

$$C^{inv}(x, y) = \lim_{n \to \infty} \sqrt[n]{\int_{\xi_1}^{\xi_2} C(I, M(\xi), x, y)^n d\xi} . \quad (5)$$

Derived expression (5) is the definition of so-called maximum norm for analytical functions:

$$\| f \|_n = \lim_{n \to \infty} \left(\int_D | f |^n d\mu \right)^{\frac{1}{n}}, \quad (6)$$

the only difference is in the condition for $C(I, M(\xi), x, y)$ in (5) to be nonnegative.

So as one can see, given the analytical solution to definite integral in (5) brings the complexity of computation of $C(I, M(\xi), x, y)$ for all possible values of $\xi \in D$ practically to zero. On the other hand complexity of max component computation heavily increased.

Image object detection. For the set of transformations target object shape changes, scale s and rotation φ were chosen. Let $M(b, s, \varphi)$ be a mathematical model of target object image with some parameter vector b responsible for shape changes, scale s and rotation φ. Let $C(I, M(b, s, \varphi), x, y)$ be some similarity measure that measures similarity of an input image $I(x, y)$ with target object image model $M(b, s, \varphi)$ at (x, y).

The objective is to detect object of interest $M(b, s, \varphi)$ with arbitrary allowable parameter vector b on arbitrary input image $I(x, y)$ regardless to affine transformations (in this particular case scale s and rotation φ) of target object on input image.

So described task of invariant image object detection according to (1) can be represented as the following:

$$C^{inv}(x, y) = \max_{b,s,\varphi} \{C(I, M(b, s, \varphi), x, y)\}, \quad (7)$$

where $C^{inv}(x, y)$ is some invariant similarity measure, $C(I, M(b, s, \varphi), x, y)$ is a similarity measure sensitive to affine transformations (s, φ) and appearance changes b.

Reformulating of (7) in terms of (5) gives the following:

$$C^{inv} = \lim_{n \to \infty} \left(\iiint_D C(I, M(b, s, \varphi), x, y)^n \, db\, ds\, d\varphi \right)^{\frac{1}{n}}, \quad (8)$$

So basically such problem formulation brings image object detection task down to “simple” integration of n th-power of similarity measure $C(I, M(b, s, \varphi), x, y)$ over a set of model parameters b and affine transform parameters s and φ.

2.2. Practical difficulties

Analytical representation. Even though theoretically (3) can be used for general object detection and recognition tasks, it is crucial for practical reasons to build proper analytical model $M(\xi)$ and similarity measure $C(I, M(\xi), x, y)$. The main purpose of that is practical integrability of (5) and simplicity of final result. To fully exploit all advantages of proposed approach it is required for $C(I, M(\xi), x, y)^n$ to be analytically integrable function with respect to parameter vector ξ. In practice it can be very difficult to build such function depending on object representation and a set of transformations. So in some practical cases integration over part of components of ξ may have to be done numerically.

Computational complexity. Computational complexity of max part in (1) given values of $C(I, M(\xi), x, y)$ for all possible values of $\xi \in D$ is neglectable in comparison to computational complexity of similarity measure $C(I, M(\xi), x, y)$ values for all possible values of $\xi \in D$. In contrast to classical brute force approach, proposed solution inverts the computational load for those two stages. All computational complexity of similarity measure $C(I, M(\xi), x, y)$ for all possible values of $\xi \in D$ collapses practicly to zero once analytical solution of integral in (5) is found. So now all computational complexity switched to computation of max component and depends on the value of parameter n.

Computational complexity is one of the biggest drawbacks of that approach. The main reason for that is representation of input image and model of target object. Importance of that is explained on example where for similarity measure cross-correlation is chosen. Generally speaking practical representation of input image I and target object model image $M(\xi)$ would always be in a form of superposition of their parts: $I = \sum_i^N I_i$ and $M = \sum_j^K M_j$ respectively. Given above, the n th-power of similarity measure can be represented as follows:

$$(C)^n = (I * M)^n = \left(\left(\sum_i^N I_i \right) * \left(\sum_j^K M_j \right) \right)^n = \left(\sum_i^N \sum_j^K (I_i * M_j) \right)^n . \quad (9)$$

Thus computation complexity grows polynomially with the growth of parameter n.

In practice b does not go to infinity but is chosen depending on type of object of interest and input image $I(x, y)$ content, to be sufficiently large enough to separate useful correlational peaks from noise ones.

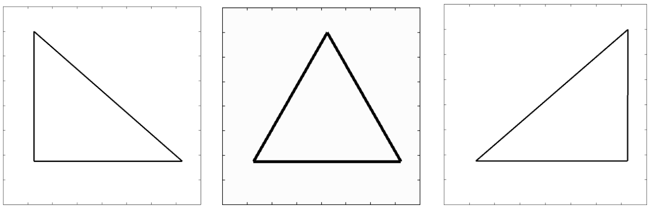
3. Basic experimental results

In this section basic experimental results obtained for described above approach for image object detection are presented. To make computation as simple as possible triangle was chosen as target object. Triangle was represented as a superposition of three line segments. For similarity measure cross-correlation measure was chosen.

Computational results sown on fig. 2 (in columns 2 and 3) depict shape invariant detection of triangle on the input images (first column of fig. 2). Contour image of triangle was modeled by means of ASM [2, 5] with one-dimensional parameter vector b. Fig. 1 shows the possible shape changes of target object depending on value of shape parameter b. The difference between second (for $n = 1$) and third (for $n = 3$) columns of fig. 2 shows that proposed approach allows to significantly amplify useful correlational signal by simply increasing values of parameter n.

Fig. 3 shows results of affine invariant detection of target object. Each input image (first column on fig. 3) contains two triangles. The rightmost triangle corresponds to model image of target object and the leftmost triangle is a target object triangle under some scale s and rotation φ transformations. Second column depicts values of C^{inv} computed for $n = 2$. As fig. 3 shows, picks of C^{inv} correctly indicate the location of the target object, subjected to affine transformations.

Experimental results shown on fig. 4 demonstrate affine invariant detection of target object in noisy input image. The noise takes a form of a randomly placed line segments on input image. The strongest pick of C^{inv} corresponds to a true location of target object on input image. Due to presence of noise C^{inv} contains numbers of additional picks with smaller amplitudes.



a b c

Fig. 1. ASM generated triangle shape samples: *a*) $b = -1$; *b*) $b = 0$; *c*) $b = 1$

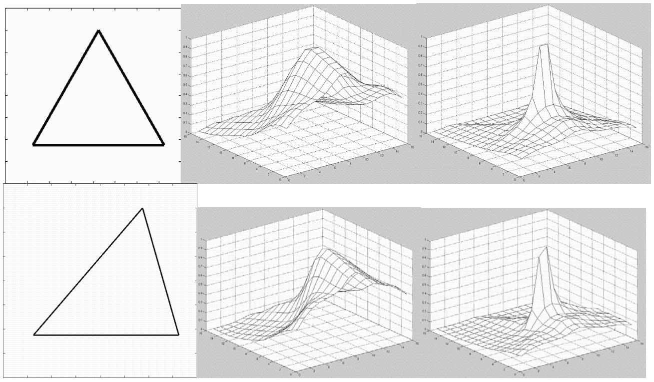
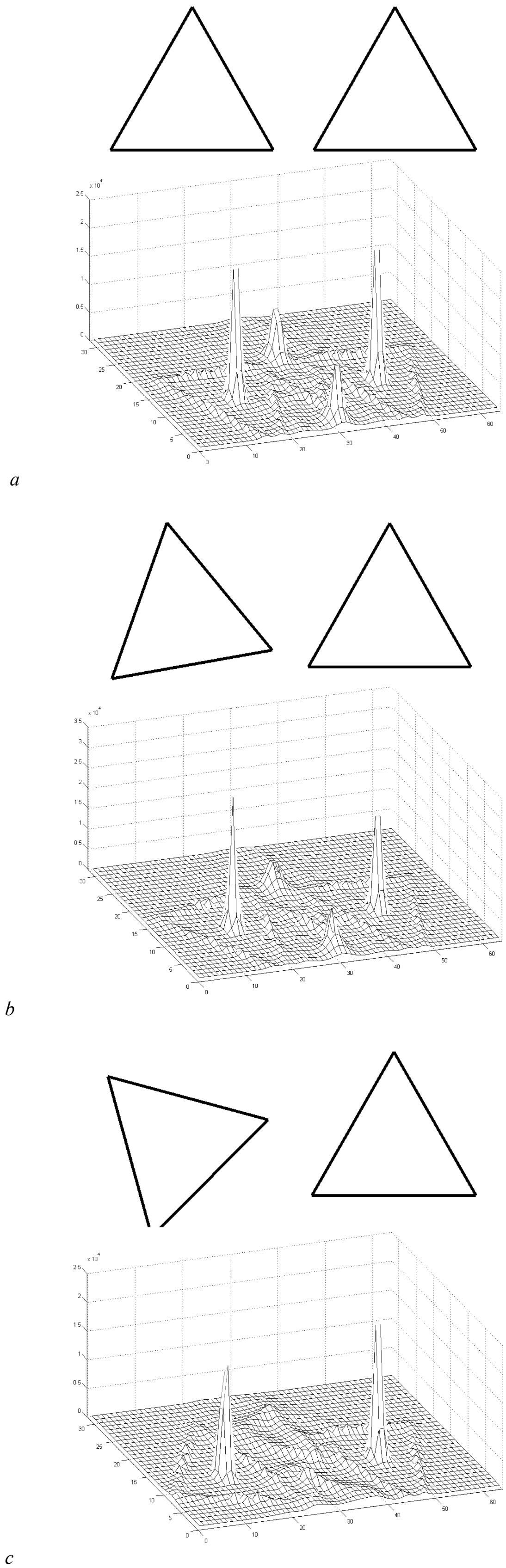


Fig. 2. Shape invariant triangle detection results: first column shows input images (top corresponds to object with $b = 0$, bottom-$b = 0{,}5$); second and third columns show shape invariant C^{inv} for $n = 1$ and $n = 3$ respectively.

Even though experiments were conducted on a such simple object as triangle, it should be noted that there is no algorithmical restrictions on the complexity and topology of a target object shape. It is only a computational complexity (or to be more precise, the number of parts the target object is represented with) that can put restrictions on chosen target object.



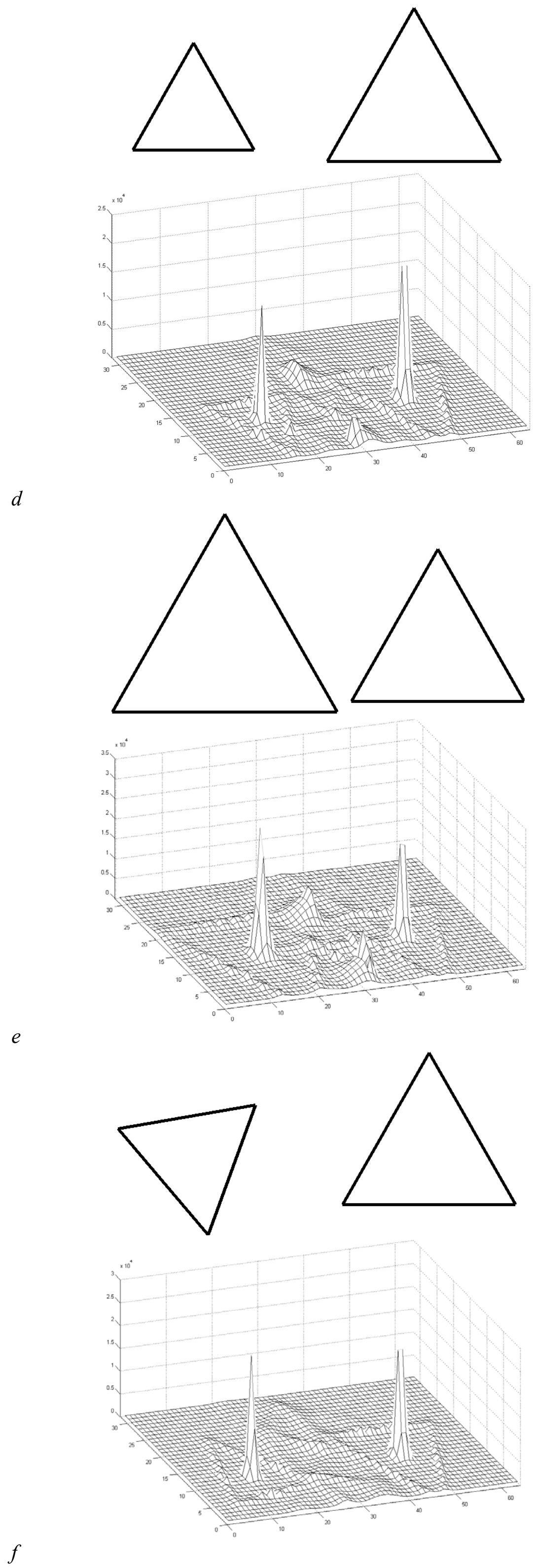


Fig. 3. Affine invariant triangle detection results: the first column shows input images that contains triangles with different scales and rotations: *a* – s = 1, φ = 0°; *b* – s = 1, φ = 10°; *c* – s = 1, φ = 45°; *d* – s = 0,7, φ = 0°; *e* – s = 1.3, φ = 0°; *f* – s = 0,8, φ = 70°); the second column show values of C^{inv} for $n = 2$

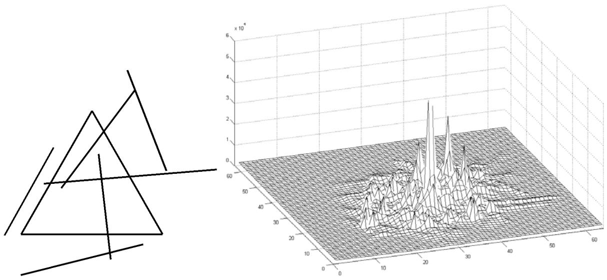


Fig. 4. Affine invariant triangle detection results: left image: input image with random noise; right: values of C^{inv} for input image

4. Conclusions

Proposed in this paper approach to invariant object detection has two main advantages. The first advantage of formulated in (5) approach is its generality. Expression (5) neither put any constrictions on the type and nature of objects to be detected, nor it puts any constrictions on set of transformations we want to achieve invariance to. The second advantage is absence of optimization procedure. The detection process is straightforward computation of expression (5). It is a strong statement of this approach because majority of existing fundamental object detection techniques strongly rely on the search of the optimum solution to a certain optimization task. And as a result they suffer from all the existing problems in the domain of multivariable nonlinear functions optimization.

To be able fully exploit advantages of the proposed approach one would have to cope with few of its drawbacks. The major disadvantage is the requirement of analytical integrability of $C(I, M(\xi), x, y)^n$ over ξ. Another disadvantage is high computational complexity of (5) for large values of n.

Presented experimental results for image object detection justify the validity of the proposed approach. It allows one to build an object detection system invariant to certain set of transformation as long as they can be properly represented in $C(I, M(\xi), x, y)$.

1. *Brunelli R*. Template Matching Techniques in Computer Vision: Theory and Practice. – Wiley, 2009.
2. *Template*-Based Object Detection through Partial Shape Matching and Boundary Verification / Ge Feng., Liu Tiecheng, Song Wang et al. // Int. J. of Information and Communication Engineering, 2008. – **4**, N 2. – P. 148–157.
3. *Morgan McGuire*. An image registration technique for recovering rotation, scale and translation parameters // NEC Tech Report, Feb. 1998. – 29 p.
4. *Tuytelaars T.*, *Mikolajczyk K.* Local invariant feature detectors: a survey, Found. Trends. Comput. Graph. – 2008. – **3**, N 3. – P. 177–280.
5. *Viola P.*, *Jones M.J.* Robust Real-Time Face Detection // Int. J. of Computer Vision. – 2004. – 57(2). – P. 137–154.
6. *Active* shape models – their training and application / T.F. Cootes, C.J. Taylor, D.H. Cooper et al. // Computer Vision and Image Understanding. – Jan. 1995. – 61(1). – P. 38–59.
7. *Cootes T.F.*, *Edwards G.J.*, *Taylor C.J.* Active Appearance Models // Proc. Fifth Europ. Conf. Comp. Vision / Ed. by H. Burkhardt, B. Neumann. – 1998. – **2**. – P. 484–498.
8. *Blanz V.*, *Vetter T.* A Morphable Model for the Synthesis of 3D Faces // SIGGRAPH'99 Conf. Proc.– 1999. – P. 187–194.
9. *Schindler K.*, *Suter D.* Object Detection by Global Contour Shape // Pat. Recognition. –2008. – **41**, N 12. – P. 3736–3748.
10. *Kountchev R.*, *Todorov V.*, *Kountcheva R.* Invariant Object Representation with Modified Mellin-Fourier Transform // 14th WSEAS Int. Conf.on Computers. – 2010. – **I**. – P. 232–236.
11. *Belongie S.*, *Malik J.*, *Puzicha J.* Shape Context: A new descriptor for shape matching and object recognition // NIPS. – 2000. – P. 831–837.
12. *Ferrari V.*, *Jurie F.*, *Schmid C.* Accurate Object Detection with Deformable Shape Models Learnt from Images // Proc. CVPR. – 2007. – P. 1–8.
13. *Flusser J.*, *Suk T.* Pattern recognition by affine moment invariants // Pat. Recog. – 1993. – **26**, N 1. – P. 167–174.

E-mail:teodor_mandziy@ipm.lviv.ua

●