



# Object Detection in The Image Using the Method of Selecting Significant Structures

Vladimir Mokshin<sup>1</sup>, Ildar Sayfudinov<sup>2</sup>, Svetlana Yudina<sup>2</sup>, Leonid Sharnin<sup>1</sup>

<sup>1</sup>Kazan National Research Technical University named after AN Tupolev-KAI, Karl Marx Str., 10, Kazan, 420111, Russia

<sup>2</sup>Almetyevsky branch of the Kazan National Research Technical University named after AN Tupolev-KAI, Builders Ave., 9, Bld, B,

\*Corresponding author E-mail: [Almetyevsk, 423457, Russia](mailto:Almetyevsk, 423457, Russia)

## Abstract

The approach to image segmentation is reviewed in the article. The method of highlighting significant contours in the image is reviewed. Some structures in the image attract attention more than others due to certain distinctive properties. The article reviews the approach of highlighting significant structures in the image representing the areas of candidates identifying the object in the video frame for mobile platforms. For example, such shapes can be smoother, longer and closed. Such structures are called significant. It would be expedient to use only these significant structures to increase the speed of image recognition by computer vision methods focused on the contour selection. This approach allocates the computing resources only to significant structures, thus reducing the total computation time. Since the image consists of many pixels and links between them, which are called edges, significant structures can be measured. The article presents an approach to measuring the structure significance that largely coincides with human perception. Some image structures attract our attention without the need for a systematic scan of the entire image. In most cases, this significance represents the structure properties as a whole, i.e. parts of the structure cannot be isolated. This article presents a measure of significance based on the measurement of length and curvature. The measure highlights structures characteristic of human perception, and they often correspond to objects of interest in the image. A method is presented for calculating significance using an iterative scheme combined into a single local network for processing elements. The optimization approach to represent a processed image highlighting significant locations is used in the network.

**Keywords:** pattern recognition, structural significance, image map, organization of perception, visual attention, segmentation.

## 1. Introduction

Automated video surveillance systems (video analytics) are widely used in many areas of human activity today, especially in state bodies and manufacturing enterprises. Using these systems allows to automate the alert functions and improve the performance of employees by providing direct control of the performance discipline. Systems described in paper [1] can be used in reviewing such systems as simulation models. Another significant issue is security in such systems. Security approaches are considered in paper [2]. At the same time, due to a large number of problems solved in video analytics, a need arises for an efficient method of selecting significant structures in the image that identify an object in a video frame for mobile platforms [3-7]. Many methods have been proposed for detecting the edge of an object's area on an image, which apply local operations to each point of the input image in order to extract short edge elements. Some of these operations are very simple, like subtracting background [8], but they do not allow to be applied to dynamic images without a priori representation of the image, while other operations are more complex and allow to eliminate most of the local noise [9-11], though they are more demanding in computing. When small edge elements are obtained, they need to be joined together to form edges; edge algorithms are used for this operation. To detect significant edges in noisy images, global information about the image edges and the structure of objects they represent are required for the following procedures, and this heuristic information is usually embedded in the program [12-14]. However, this approach

has a drawback – minor changes in the peculiarities of the desired edges may require reprogramming large parts of the program.

Since mobile platforms have limited computational resources, an approach is needed for object detection and recognition that would reduce the computational burden on segmentation methods. This is possible through the input of a part of the image where the object is supposed, rather than the whole image, to the segmentation algorithm.

## 2. Analysis of literature and formulation of the problem

Analysis of the data has revealed that the possibility of the object detection by a human is overestimated [15, 16]. The human observing system can pay attention only to a certain part of the scene at each specific moment in time. This would mean that the observer is "blind" to other parts of the scene. The traditional object detection can be divided into upward methods based on low-level functions [17-19] and downward methods involving high-level knowledge [20]. Areas of interest that attract human attention are called significant areas/locations in the upward attention. There must be a sufficient difference in important characteristics with respect to this reaction in relation to the surrounding objects. Attention, which is caused by the cognitive factor – for example, knowledge, current goals and expectations, is called the downward attention. There are many examples of this kind of intuitive attention: car drivers see gas stations on some streets or bicyclists are looking for cycle lanes.

Optimization methods play an important role in a computer vision. Optimization is a powerful paradigm for expressing and solving problems in a wide range of areas, and has been successfully applied to many vision problems. Methods of discrete optimization are of a particular interest, as they often provide nontrivial quality assurance solutions carefully using the problem structure.

Problems with a computer vision are inherently ambiguous, and one can only expect "enlightened guesses" about the content of images in general. A natural approach to solving this problem is to develop an objective function that measures the quality of the potential solution and later uses the optimization method to find the best solution. It is common knowledge that the basics of optimization are well suited to eliminate noise and other sources of uncertainty, such as uncertainties in image data. Moreover, when formulating problems from the standpoint of optimization, several sources of information can be easily taken into account. Another advantage of the optimization approach is that it actually ensures a clear separation between how the problem is formulated (objective function) and how problems are solved (optimization algorithm). Unfortunately, the optimization problems that arise in the computer vision are often very difficult to solve.

Perceptual organization of Lowe [21] is more closely related to the problem reviewed in this article. The processes proposed by Lowe reveal cases of collinearity, coterminal and parallelism between straight lines, and will not be efficient in cases where these conditions do not play a major role [22-24]. Most of the previous approaches to segmentation do not meet the requirements of time and accuracy constraints, and they largely depend on the complexity of the background curves [25-27].

The optimization task includes a set of possible solutions  $S$  and an objective function  $E : S \rightarrow \mathbb{R}$  that measures the solution quality. In general, the search space  $S$  is implicitly defined and consists of a very large number of possible solutions. The objective function  $E$  can measure adequacy or inadequacy of the solution; when  $E$  measures inadequacy, the problem of optimization is often called the problem of minimizing energy, and  $E$  is called the energy function.

Since many papers in the field of a computer vision are related to optimization as energy minimization, this agreement is followed, and it is considered that the optimal solution  $x \in S$  is the one that minimizes the energy of the function  $E$ . An ideal solution to the optimization problem would be the candidate's solution, which is the global minimum of the energy function,  $x^* = \arg \min_{x \in S} E(x)$ .

It is tempting to review the energy function and optimization methods as completely independent. This involves the development of the energy function that completely captures the limitations of the vision problem and then applies the general energy minimization method. However, this approach does not take the computational problems into account, which are of great practical importance. No algorithm can find the global minimum of an arbitrary energy function without exhausting the enumeration of the search space because any candidate that has not been evaluated may appear the best [28, 29]. As a result, any completely general minimization method, such as genetic algorithms [30] or MCMC [31], is equivalent to an exhaustive search. However, sometimes an effective algorithm can be developed to solve problems in a particular class through using the structure of search space  $S$  and the energy function  $E$ .

A new emphasis has been placed on discrete optimization techniques recently, such as dynamic programming or graph algorithms for solving problems of a computer vision. There are two key differences between discrete optimization methods and the more classical continuous optimization methods commonly used in

computer vision. Firstly, these methods work with discrete solutions. Secondly, discrete methods often provide nontrivial quality assurance solutions that they find. These theoretical guarantees are often accompanied by high efficiency in practice.

### 3. Goal and objectives of research

The goal of the article is to increase the speed of the object detection and recognition in the image for computational platforms with limited computing resources through the decrease in the information content of the data supplied to the segmentation algorithm.

The following tasks have been set in order to achieve the goal:

1. To develop a measure that takes the weights of the curves based on smoothness and curvature into account.
2. To build a significance network for all orientation elements. The significance network will be a set of significant orientation elements. The orientation elements are edges corresponding to the linear segment of the image.
3. To evaluate the significance network for false positive detection.

### 4. Developing a measure of significance and building a significance network

Orientation elements are the main computational elements in the

network. Each element  $P_i$  is associated with a processor that can perform some calculations based on its state and the state of its  $k$  neighboring processor. This defines a single network containing  $kn^2$  processing blocks with local links.  $k$  is 48 in the current implementation, which ensures a reasonable angular resolution. Let

us turn to a related sequence of orientation elements  $P_i, \dots, P_{i+N}$ , each element being a linear segment or an empty gap as a curve of length  $N$  (curves can be continuous or with any number of gaps). The optimization problem is formulated as maximizing the measure of significance of the curve  $\Phi_N$  over all curves of length  $N$ , starting with  $P_i$ .

$$\max_{(P_{i+1}, \dots, P_{i+N}) \in \delta^N(P_i)} \Phi_N(P_i, \dots, P_{i+N}), \tag{1}$$

where  $\delta^N(P_i)$  is the set of all possible curves of length  $N$ , starting with  $P_i$ .

For a certain class of measures  $\Phi(\square)$ ,  $\Phi_N$  can be found through repeating simple local computations. Let us consider first curves just three elements long to illustrate this. The problem in this case is the following:

$$\max_{(P_{i+1}, P_{i+2}) \in \delta^2(P_i)} \Phi_2(P_i, P_{i+1}, P_{i+2}), \tag{2}$$

i.e. for a given element  $P_i$ ,  $P_{i+1}$  is found (one of  $P_i$ 's  $k$  neighbors), as well as  $P_{i+2}$  (neighbor  $P_{i+1}$ ) such that  $\Phi_2(P_i, P_{i+1}, P_{i+2})$  is maximum. A naive approach will again require the study  $k^2$  of various curves. Suppose, however, that  $\Phi_2$  meets the following condition:

$$\max_{\delta^2(p_i)} \Phi_2(p_i, p_{i+1}, p_{i+2}) = \max_{p_{i+1}} \Phi_1(p_i, \max_{p_{i+2}} \Phi_1(p_{i+1}, p_{i+2})) \quad (3)$$

In this case, the  $\Phi_2$  maximization can be achieved through repeating the  $\Phi_1$  application over the shorter curves. The general approach is formulated in a similar way:

$$\max_{\delta^N(p_i)} \Phi_N(p_i, \dots, p_{i+N}) = \max_{p_{i+1} \in \delta(p_i)} \Phi_1(p_i, \max_{\delta^{N-1}(p_{i+1})} \Phi_{N-1}(p_{i+1}, \dots, p_{i+N})), \quad (4)$$

where  $\delta(p_i)$  means  $\delta^1(p_i)$ . As such, the search area needed for each curve of length  $N$ , is reduced from  $p_i$  to size  $kN$  instead of  $k^N$ , which is required for a naive approach. The principle in (4) is related to the principle of optimality underlying all multistage decision-making processes, and is a special case of dynamic programming, in particular. This refers to a family of functions that obey principle (4) as extensible functions.

Two factors play a role in the measurement of significance. The first factor is related to the length of the curve, and the second factor is related to its shape. The length of the curve is measured as the number of elements on the curve that have the actual curve (rather than a gap) passing through them. These elements are called active elements, while the elements associated with gaps are referred to as virtual elements. Local significance  $\sigma_i$  is associated with each element  $p_i$ . If it is an active element,  $\sigma_i$  is set to a positive value, which is set to 1, while  $\sigma_i$  is set to 0 for the virtual element. The measure is associated with the length of the curve  $p_i, \dots, p_{i+N}$ :

$$\sum_{j=i}^{i+N} \sigma_j \quad (5)$$

The measure above is the sum of the local values of the significance of active elements along the curve. The attenuation function associated with the curve  $p_i, \dots, p_j$  is then found as follows:

$$\rho_{i,j} = \prod_{k=i+1}^j \rho_k, \quad (6)$$

where  $\rho_{i,i} = 1$ . The measure in (5) is modified using the attenuation coefficients:

$$\sum_{j=i}^{i+N} \rho_{i,j} \sigma_j. \quad (7)$$

The measure in (7) is the weighted contribution of local significance values  $\sigma_i$  along a curve where weights are inversely proportional to the number of virtual elements along  $p_i, \dots, p_j$ .

In order to measure the shape of the curve, a measure is used which is in inverse relationship to the total curvature of the curve. The total

curvature of the curve  $\gamma$  is defined as  $\int_{\gamma} \left(\frac{d\theta}{ds}\right)^2 ds$ , where  $\theta(s)$  is the slope along the curve, and  $\frac{d\theta}{ds}$  at point P is known as the local curvature at this point (the inverse of R, the radius of curvature). It is required to use the total curvature to obtain a limited measure that

is inversely proportional to the total curvature. The following measure meets these requirements:

$$\exp^{-\int_{\gamma} \left(\frac{d\theta}{ds}\right)^2 ds} \quad (8)$$

Let us denote the difference of orientation between the  $k$ -th element and its successor as  $a_k$ , and the length of the orientation element as  $\square s$  to obtain a discrete approximation to the measure in (8). A discrete approximation to the full curvature of a measure along  $p_i, \dots, p_j$  is thus as follows:

$$C_{i,j} = \prod_{k=i}^{j-1} f_{k,k+1}, \quad (9)$$

where

$$f_{k,k+1} = \exp^{-\frac{2a_k \tan \frac{a_k}{2}}{\square s}} \quad (10)$$

$C_{i,j}$  plays the role of the weight assigned to each value of local significance  $\sigma_i$  along the curve.

A measure that ensures a high score on long curves with low overall curvature is now defined as follows:

$$\sum_{j=i}^{i+N} C_{i,j} \rho_{i,j} \sigma_j. \quad (11)$$

Measure (11) is the weighted contribution of local significance values  $\sigma_i$  along the curve. The curves that will receive a high measure on (11) are long curves, as straight as possible, and have the least number of gaps.

## 5. Significance network

Each element  $p_i$  is associated with a state parameter denoted as  $E_{p_i}$  and a set of three attributes, which includes local significance  $\sigma_i$ , orientation  $\theta_i$ , and attenuation coefficient  $\rho_i$ . Each element  $p_i$  updates its state of variable  $E_{p_i}$  iteratively through local calculations. Notation  $E_{p_i}$  is used in this case, explicitly indicating that the variable is associated with the  $p_i$  element in the network. At the end of the iteration  $N$ ,  $E_{p_i}$  contains the measure of significance obtained in (11), which is maximum through all possible curves of length  $N$ , starting with  $p_i$ , where these curves are continuous or with any number of gaps.

$E_{p_i}$  is updated through the following calculation:

$$E_{p_i}^{(0)} = \sigma_i$$

$$E_{p_i}^{(n+1)} = \sigma_i + \rho_i \max_{p_j \in \delta(p_i)} E_{p_j}^{(n)} f_{i,j}, \quad (12)$$

where  $P_j$  is one of  $k$  possible neighbors  $P_i$ , and  $f_{i,j}$  are the "coupling constants" defined in (10). To expand the recurrence formula above, the given curve  $\gamma$  is selected being represented as  $\theta_i, \dots, \theta_{i+N}$ , where each element along the curve has only one neighboring element to link to. The following statement relates the value of the state of variable  $P_i$  to a measure in (11).

Statement 1:

$$E_{P_i}^{(N)} = \sum_{j=i}^{i+N} C_{i,j} \rho_{i,j} \sigma_j$$

The proof can be obtained through the induction on the length of the curve. Along with the fact that this measure is extensible, the

statement above implies that among all possible curves  $\gamma_i$  of length  $N$ , starting with  $P_i$ , continuous or with any number of gaps,  $E_{P_i}$  will be found along the curve that is maximal with respect to the measure in (11), i.e.:

$$E_{P_i}^{(N)} = \max_{\gamma_i} \sum_j C_{i,j} \rho_{i,j} \sigma_j,$$

taken by all  $\gamma_i$ . It must be noted that if the measure  $\Phi$  is extensible, it does not mean that the optimal contour through  $P$  simply expands as iterations continue. In fact, the optimal curve on the stage  $N+1$  may differ from the optimal curve on the stage  $N$ .

The state values of elements on the network form new representation of the image, which is a "biased" representation of the visual environment, highlighting interesting or prominent places.

## 6. Comparison of false positive detection with other methods

A comparison between the percentage of false positive detection was made to analyze the significance network. Test samples were used consisting of short oriented segments arranged uniformly along the perimeter of the circle in the background of segments with random positions and orientations (Figure 1). The significance of forms and noise segments was found using each of the four methods: thresholding [32], watershed [33, 34], morphology [35], and the proposed approach. Segments were sorted in ascending order by their significance: significance of the most significant

segment  $\phi_1$  and significance of a less significant segment  $\phi_n$ . Let us define false positive segments as noise assigned a value greater than  $\phi_{m+1}$  for given  $m$  segments of the form. A false positive estimate for each method was found for samples consisting of different number of shape edges and noise edges. A false positive estimate for each combination (for example, 20 shape segments and 70 noise segments) was estimated through averaging a false positive estimate over ten tests using different noise samples. The right half of Figure 2 is a graph of the percentage of false positives from the number of constituent noise segments for a circle with 20 segments.

All methods are applied well (less than 10% of false positive detections) at a low noise level (40 noise segments or less). The

results of the methods begin to diverge at higher noise levels. It must be noted that thresholding exceeds the morphological processing, it is also noticeable that the watershed is applied comparably with the proposed approach, although the watershed method is more expensive to calculate. Finally, watershed and the proposed approach have significantly lower false positive estimates at lower signal-to-noise ratios.

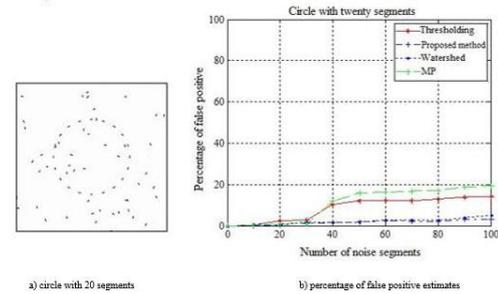


Figure 1: Circle with 20 segments: a – circle with 20 segments, b – percentage of false positive estimates.

The next comparison was identical to the first one, except that the shape segments were formed by an unlimited sinusoid (Figure 2). There is a graph of the percentage of false positive triggers relative to the number of noise segments for a 10-segment sinusoidal curve in the right half of Figure 2. Comparatively low indicators of the proposed method compared with other methods can be attributed to its apparent dependence on closure. However, it still outperforms thresholding for higher signal-to-noise ratios and has an error rate comparable to thresholding (i.e. within 5%) at lower signal-to-noise ratios.

A background consisting of correlated noises (i.e. a dipole) was used in the third comparison (Figure 3). The dipole consists of two collinear segments separated by a gap equal to the distance between neighboring segments of the circle. Since the two segments forming the dipole are collinear, the degree of closeness between the segments forming the dipole is greater than between adjacent segments of the circle. Therefore, it is impossible to distinguish noise segments from the shape segments using only a local measurement. It can be seen from the graph that thresholding and morphological processing methods have almost the 100% false positive level, while watershed and the proposed method are much better at solving this problem.

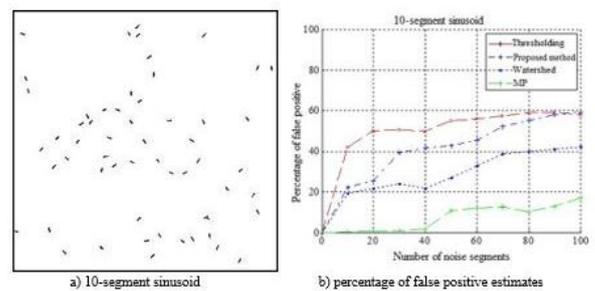


Figure 2: 10-segment sinusoid: a – 10-segment sinusoid, b – percentage of false positive estimates.

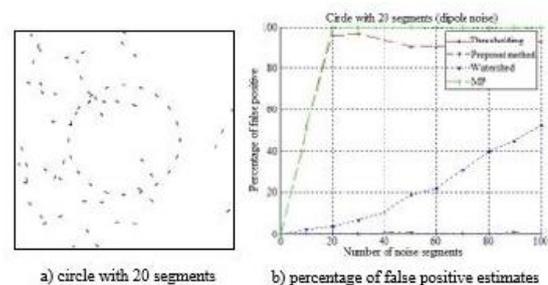


Figure 3: Circle with 20 segments (dipole noise): a – circle with 20 segments (dipole), b – percentage of false positive estimates.

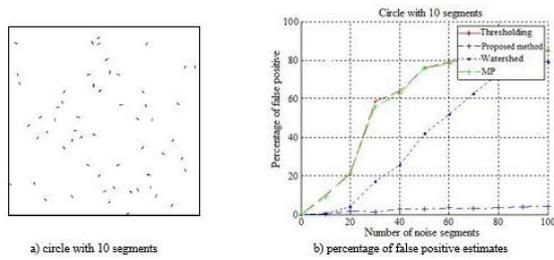


Figure 4: Circle with 10 segments: a – circle with 10 segments, b – percentage of false positive estimates.

A 10-segment circle is used in the fourth comparison (Figure 4). This is a complex picture because the sampling frequency is so small that there is only one segment for 36 degrees of the circle. Most methods do not work well even at relatively high signal-to-noise ratios. For the noise level 80, thresholding and morphological processing methods are applied at a false positive level of 90%. The watershed method is applied a little better, with a false positive level of 70%. In contrast, the false positive level for the proposed method is less than 5%.

Let us consider the examples of practical use of the above methods in solving the problem of object segmentation on the image in the video analytical system focused on mobile platforms for registering the dump truck runs. The image segmentation results are shown below for thresholding (Figure 5), watershed (Figure 6), morphological processing (Figure 7), and the proposed method of selecting significant structures (Figure 8).

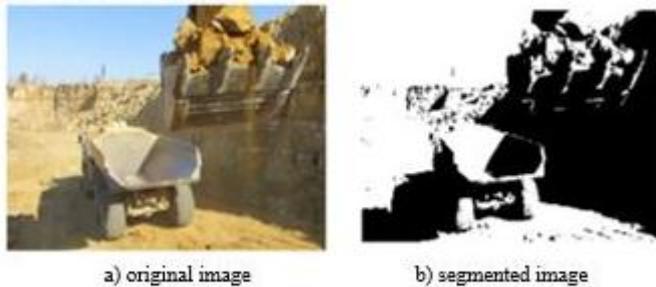


Figure 5: Segmentation. Thresholding: a – original image, b – segmented image

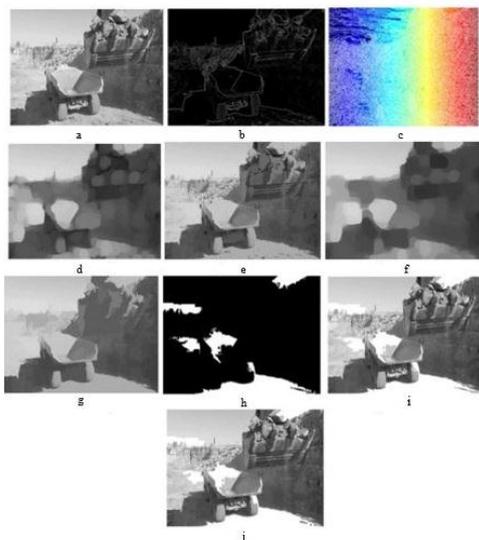


Figure 6: Stages of segmentation using watershed: a – original image in shades of gray, b – calculation of the gradient, c – transformation of the gradient value watershed, d – operation "opening", e – operation "closing", f – operation "opening-closing", g – "opening-closing" by reconstruction, h – local maxima of "opening-closing", i – local maxima superimposed on the original image, and j – markers and boundaries of the object superimposed on the original image

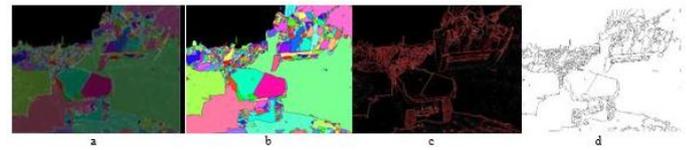


Figure 7: Stages of segmentation in morphological processing: a – superimposed basins, b – catch basins, c – overlaid dams, and d – watershed lines

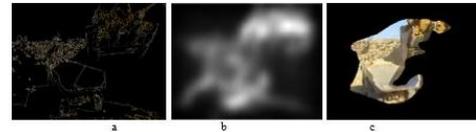


Figure 8: Stages of segmentation of the proposed method: a – image preprocessing using Sobel operator, b – image significance map, c – selected area of the image

The above segmentation examples illustrate that the proposed segmentation method requires fewer stages than the watershed and morphological processing. It is also shown that the proposed method segments the object in accordance with perceptual expectations, in contrast to thresholding.

## 7. Discussion of the results of researching the significance measure

A significance network is a mechanism used to identify significant curves in images based on length and straightness. This method has several distinctive properties. Firstly, the significance measure normally prefers long and smooth curves to short or wavy curves. Besides, it is guaranteed that the network will find the most significant curve in accordance with the measure. In this case, the network gaps are filled, and noise is allowed. The network itself is locally interlinked, and its size is proportional to the image size. The distinctive features of the place are further emphasized, since the contribution of distant elements to the result of this element is weakened with the curvature and the length of the gap separating distant elements from this element. The significance network differs not only in emphasizing significant locations, but also in the presence of curves that make these locations significant. This requires grouping a curve fragment that divides gaps and intersections, which requires choosing from an exponential set of possible image curves.

Based on the method properties, it is applicable for image preprocessing in order to reduce the data to be supplied into any segmentation classifier. Further research will help find out whether variations in the current computations can be determined that allow to find the significance map and the most significant curves, for more efficient calculation.

## 8. Conclusion

1. A method for selecting significant structures in an image based on the significance measure for elements representing the object contours has been developed. It is assumed that direct perception includes processes for detecting significant image structures that the subsequent processes can focus on, such as segmentation and recognition. The object significance is divided into two sources – local and structural. Of them, the structural significance is more problematic from a computational point of view, since it requires the efficient calculation of certain global properties.
2. A method that connects the elements into a single local element processing network has been developed. The network has the following properties:
  - 1) calculations are local and simple,
  - 2) number of calculations amounts to dozens or up to hundreds of iterations,
  - 3) small dependence on the image complexity,

- 4) gaps in the curves are filled during the calculation,
  - 5) contours are smoothed out when creating a significance map,
  - 6) the network creates coupling information to allow tracing the curve along the transitions and gaps, and
  - 7) the network is reliable in terms of some processing units' failure having little impact on the network performance.
3. Aside from developing a method for selecting significant structures, false positive detection was performed with other segmentation methods: thresholding, watershed, and morphological processing. The research has revealed that the proposed method for isolating significant structures showed the fastest and most accurate results. For instance, in case of a 10-segment circle with a frequency of 1 segment for 36 degrees, most methods for noise level 80 are true at 70-90% false positive level, while the proposed method remains at a false positive level of less than 5%

## References

- [1] Yakimov, I., Kirpichnikov, A., Mokshin, V., Yakhina, Z., Gainullin, R. 2017. The comparison of structured modeling and simulation of queuing systems. *Communications in Computer and Information Science (CCIS)*, 800, 256-267.
- [2] Tutubalin, P.I., Mokshin, V.V. 2017. The Evaluation of the cryptographic strength of asymmetric encryption algorithms. 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC), IEEE, 180—183. doi:10.1109/RPC.2017.8168094
- [3] Mokshin, V.V., Saifudinov, I.R., Kirpichnikov, A.P., Sharnin, L.M. 2016. Recognition of vehicles on the basis of heuristic data and machine learning. *Bulletin of Kazan Technological University*, 19(5), 130-137.
- [4] Emmanouil, T.A., Treisman, A. 2008. *Perception & Psychophysics*. Springer, 70(6), 946-954.
- [5] Gurevich, I.B., Yashina, V.V. 2008. Descriptive approach to image analysis: image models. *Pattern Recognition and Image Analysis*. Springer, 18(4), 518-541.
- [6] Vapnik, V. Vashist, A. 2009. A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5-6), 544-557.
- [7] Abramov, S.K., Zelensky, A.A., Lukin, V.V., Ponomarenko, N.N. 2009. The use of the TID2008 database in the development of visual quality metrics and image processing methods. *Radiotekhnika i elektronika*, 4, 99-109.
- [8] Mandellos, N.A., Keramitsoglou, I., Kiranoudis, Ch.T. 2011. A background subtraction algorithm for detecting and tracking vehicles. *Expert Systems with Applications*, 38, 1619-1631.
- [9] Cheng, X., Zhang, Y., Chen, Y., Wu, Y., Yue, Yi. 2017. Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, 141, 351-356.
- [10] Jiaying, L., Dexiang, Z., Jingjing, Z., Jun, Z., Lina, X. 2017. Facial Expression Recognition with Faster R-CNN. *Procedia Computer Science*, 107, 135-140.
- [11] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. 2016. You only look once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 779—788.
- [12] Long, J. Shelhamer, E., Darrell, T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640—651.
- [13] Chen, L.C. Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. 2017. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, V, 99.
- [14] Goferman, S. Zelnik-Manor, L., Tal, A. 2012. Context-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10), 1915—1926.
- [15] Wagemans, J., Kubovy, M., Peterson, M.A. Elder, J.H., Palmer, S.E., Singh, M. A Century of Gestalt Psychology in Visual Perception: I. Perceptual Grouping and Figure-Ground Organization. *Psychological Bulletin*, 138(6), 1172-1217.
- [16] Saifudinov, I.R., Mokshin, V.V., Kirpichnikov, A.P. 2017. Grouping of contours of objects of structural images on the basis of network of element visibility. *Bulletin of Kazan Technological University*, 20(9), 120-123.
- [17] Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y. 2011. Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2), 353—367.
- [18] Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A. 2012. Saliency filters: Contrast based filtering for salient region detection. *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, 733—740.
- [19] Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.M. 2015. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3), 569—582.
- [20] Smirnov, E.A., Timoshenko, D.M., Andrianov, S.N. 2014. Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks. *AASRI Procedia*, 6, 89-94.
- [21] Lowe, D.G. 1984. *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, Boston, Mass.
- [22] Li, G., Yu, Y. 2015. Visual saliency based on multiscale deep features. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 5455—5463.
- [23] Li, G., Yu, Y. 2016. Deep contrast learning for salient object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 478—487.
- [24] Liu, N., Han, J. 2016. Dhsnet: Deep hierarchical saliency network for salient object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 678—686.
- [25] Zitnick, C.L., Dollar, P. 2014. Edge boxes: Locating object proposals from edges. *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science, Springer, Cham, 8693, 391-405.
- [26] Arbelaez, P., Pont-Tuset, J., Barron, J.T., Marques, F., Malik, J. 2014. Multiscale combinatorial grouping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(1), 128—140.
- [27] Feng, J., Wei, Y., Tao, L., Zhang, C., Sun, J. 2011. Salient object detection by composition. *International Conference on Computer Vision*, Barcelona, 1028—1035.
- [28] Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., Mech, R. 2016. Unconstrained salient object detection via proposal subset optimization. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 5733—5742.
- [29] Goferman, S., Zelnik-Manor, L., Tal, A. 2012. Context-aware saliency detection. *IEEE Trans Pattern Anal Mach Intell*, 34(10), 1915—1926.
- [30] Poli, R., Langdon, W.B., McPhee, N.F. 2008. *A Field Guide to Genetic Programming*. <https://www.lulu.com/>
- [31] Amit, Y., Kong, A. 1996. Graphical templates for model registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):225-236.
- [32] Luessi, M., Eichmann, M., Schuster, G.M., Katsaggelos, A.K. 2009. Framework for efficient optimal multilevel image thresholding. *Journal of Electronic Imaging*, 18(1), 013004.
- [33] Vincent, L., Soille, P. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), 583—598.
- [34] Shakawat, Md., Sarker, Z., Tan, W.H., Logeswaran, R. 2007. Morphological based technique for image segmentation. *Int. J. Inf. Tech.*, 14(1), 55—80.
- [35] Arbelaez, P., Maire, M., Fowlkes, C. 2008. Contour detection and hierarchical image segmentation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 33(5).