Investigation of optimal heuristical parameters for mixed ACO-k-means segmentation algorithm for MRI images

Samer El-Khatib*, Sergey Rodzin**
Department of software engineering
ICTIS, Southern Federal University
Taganrog, Russian Federation
samer_elkhatib@mail.ru*, srodzin@yandex.ru**

Yuri Skobtsov
Telematics department
IAMM, Peter the Gr. St. Petersburg Polytechnic University
Saint-Petersburg, Russian Federation
ya_skobtsov@list.ru

***Abstract* — The parameters of the modified mixed Ant Colony Optimization (ACO) - *k*-means image segmentation algorithm are considered. There have been investigated such parameters as *n* – the number of ants; heuristic coefficients of ACO algorithm and their dependence on the image scale and number of iterations before and after parameters correction. The proposed algorithm and sub-system for the study of coefficients, as part of the medical image segmentation system, have been implemented. Operation of the algorithm with and without the use of optimal parameters was applied. Optimal parameters were studied for 6 groups of MRI images: brain, heart, lungs, liver, bone structures, and others. The results are displayed in the final table. Images from Ossirix image dataset and real patients' images were used for testing.**

Keywords— MRI image segmentation, ant colony optimization, k-means, swarm intelligence

I. INTRODUCTION

Modern medical diagnosis is not possible without image processing. Medical images are complex and variable. Image segmentation is one of the most difficult tasks in image processing. All the subsequent image-processing steps such as classification, identification and feature extraction are directly dependent on the segmentation results. This problem has been studied in many papers and a large number of image segmentation algorithms has been proposed [1, 2].

Segmentation is the process of the image fragmentation into disjoint parts characterized by a definite feature. This procedure is used to solve a wide range of tasks: finding objects in satellite images (mountains, rivers, etc.), face recognition, medical image processing – (MRI, Computer Tomography, X-Ray), etc.

The main difficulty in the process of segmentation is the availability of additional factors inherent to the pictures: the variability of background, the presence of noise in the images, the difference between the parts of images. There are two main classes of segmentation methods:

- Automatic methods (participation of the user is not required).
- Interactive methods (in the process of clarification the additional data are required from the user).

Edge detection operators (filter of Roberts, Sobel, Prewitt, Canny), histogram methods and graph-based algorithms are among the most known and widely used automatic segmentation methods. Since the beginning of 2000-s the most attention of researchers has been focused on the interactive methods.

Artificial intelligence methods have been used more often for image segmentation lately. In particular, Genetic Algorithms have shown some interesting results. One of the reasons for this approach is the ability of genetic algorithm to operate with a huge, complex search space with minimum information on the fitness function. For example, the majority of the existing color image segmentation algorithms have many parameters which should be adjusted for a specific type of images. The corresponding search space has a high dimension and there are complex interactions between the parameters.

Artificial neural networks are also applied for image segmentation and, in case of optimal parameter settings, good results can be achieved. Region-growing method is used for solving this problem.

K-means algorithm, based on quadratic error minimization, can be used for image segmentation. This algorithm can provide good clasterization for pixels, but it cannot eliminate some extra information, i.e. a noise. *K*-means algorithm builds *k* different clusters. As a rule, the selection of a number of clusters *k* is based on the results of previous experiments or on human evaluation. The general idea of the algorithm is, that for a given number of clusters k it is necessary to find such partition, so that average values in different clusters differ from each other as much as possible.

Recent research results have shown the prospects of application of nature-inspired techniques for image segmentation task such as ant colony optimization, particle swarm optimization and bee colony optimization.

The given article is an expansion of the previous articles [3, 4], where there was proposed a modification of algorithm based on the ant-colony optimization image segmentation technique. Investigations of the proposed algorithm have

shown that it can be improved through the use of optimum initial meta-heuristic parameters. These parameters include number of ants and heuristic coefficients α and β of ACO algorithm.

II. Ant Colony Optimization

Ants are social insects, formed in teams. The collective system can meet the challenges of implementation of the general work, the fulfillment of which by each element of the system would be impossible without the joint co-ordination and control. In this case, we use such term as swarm intelligence.

The basis of the behavior of ant colony is its self-organization, which ensures the success of the common objectives by means of low-level interaction. Ant colony is a structure without any centralized control; its main peculiarity is local information transfer only between certain agents. Direct exchange is in the food, chemical and visual characteristics. There is also an indirect exchange, which is used in ant colony optimization algorithms. Therefore, in general, we consider ants who cannot perceive how close the food is.

Ant's behavior is subjected to simple rules, despite the fact that the behavior of an ant colony is a complex process. Ant is seen as an agent who is affected by something and generating reaction to this affection (in this case it goes about pheromones and the choice of action based on this). Ant is a computational agent. The behavior of an artificial ant and a real ant is similar (on aspects of interest to us).

Indirect exchange or stigmergy is a kind of cooperation, in which one member of the team makes a change in a certain field of the environment, while others may use it only when they get to this area. It is known from biology, that such ants' interaction takes place thanks to a special chemical substance - pheromone. Pheromone is a special enzyme that is secreted when ant moves. Depending on the concentration of the pheromone, we can determine the weight movement along a certain path.

Adaptability of behavior is implemented by the pheromone evaporation. In nature, the pheromone is able to be perceived by ants for several days.

The general idea of ant colony application for solving optimization problems was introduced by Marco Dorigo [5] and is described in Fig. 1. Introdution to metaheuristical optimizations, such as ant colony optimization can be found in [6].

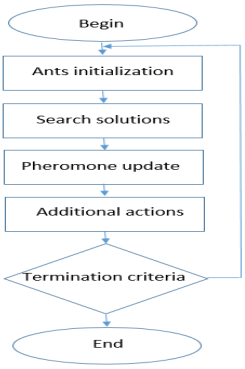


Fig. 1. Common scheme of ant colony optimization algorithm

III. Image segmentation using ACO-k-means algorithm

To obtain an efficient algorithm for image segmentation we have developed a method in which all the advantages of k-means and ACO algorithms are used.

The first step is to set the number of clusters and initialize their centers. Then, according to the clustering algorithm k-means, we need to determine the belonging of each image pixel to a particular cluster. At this stage, the most important role is played by the ACO algorithm. It defines the relationship of each pixel with clusters of the image. This is done according to the probability, which is inversely proportional to the distance between the pixel, cluster center and the variable τ, which represents the pheromone level. The pheromone level is determined in proportion to the minimal distance between each pair of cluster centers and inversely proportional to the distance between each pixel and its center. Thus, the pheromone level value increases with increase of a distance between the centers of clusters, as well as with increase in compactness of pixels in the cluster. Under the same conditions, the probability of the pixel attachment to the cluster increases.

The evaporation of the pheromone is calculated in order to reduce the impact of the previously made choices which are of lower priority. Similarly to k-means algorithm, the distributed cluster centers are updated by recalculation of the average value of pixels in each cluster. It lasts as long as the change in value of the cluster center does not vary substantially. In contrast to the k-means algorithm, the developed method does not stop at this stage. The process continues to perform clustering of m ants, each of which eventually finds a solution. The criteria for finding the best solutions and the updated pheromone level are prior for the next group of m ants respectively. When the stopping criterion is reached, the clustering is completed and the best solution is found.

In the previous articles [3, 4] there was examined a modified ant colony optimization algorithm for image segmentation. It consists of the following steps:

Algorithm 1. ACO-k-means segmentation algorithm

```
Begin
    Initialize([number clusters], [number ants]);
    Repeat
    For each ant do
      M: For each pixel do
        Calc(probability belonging pixel to cluster) (1);
      End
      Update (cluster center);
      If (NewCenter <> OldCenter) then
        goto M;
      Else
        Save(current solution);
    End
    Select Best Solution From All Ants (5);
    Update(for each pixel) (6, 7);
    Correct(common solution);
    Until criteria not reached
End
```

Software implementation of the algorithm starts with determination of the level of pheromone τ and assignment of

heuristic information η for each pixel. Then, each ant determines pixel's belonging to the cluster with probability P, which can be calculated in the following way (1):

$$P_i(X_n) = \frac{[\tau_i(X_n)]^{\alpha}[\eta_i(X_n)]^{\beta}}{\sum_{j=0}^{K}[\tau_j(X_n)]^{\alpha}[\eta_j(X_n)]^{\beta}} \quad (1)$$

where $\tau_i(X_n)$ and $\eta_i(X_n)$ - information about pheromone and heuristic variable of belonging of a pixel to the cluster i respectively, α and β - heuristic coefficients of ACO algorithm, k – number of clusters. Heuristic information is obtained as follows:

$$\eta_i(X_n) = \frac{k}{CDist(X_n, CC_i) * PDist(X_n, PC_i)} \quad (2)$$

where X_n n[th] pixel, CC_i - i[th] spectral cluster center, PC_i - i[th] spatial cluster center, $CDist(X_n, CC_i)$ - distance between (X_n, CC_i) according to color characteristics – (3), $PDist(X_n, PC_i)$ - the geometrical distance between (X_n, PC_i) - (4), k – constant value.

$$CDist(X_n, CC_i) = |Int(X_n) - Int(CC_i)| \quad (3)$$

where $Int(X_n)$ is the intensity of the pixel X_n.

$$PDist(X_n, PC_i) = \sqrt{(X_n.x - PC_i.x)^2 + (X_n.y - PC_i.y)^2} \quad (4)$$

where $X_n.x$ and $X_n.y$ are x and y coordinates respectively for the pixel X_n.

In the given modification we suggest to use the following set of simple rules as the target function (optimality criterion of a better solution):

- $\max_{i=1..m} \left(\sum_{k=1..K-1} \sum_{j=k+1..K} CDist(C_k, C_j) \right)$ - the maximum sum value of the color distance between the cluster centers for all ants (the distance between the clusters, in terms of color characteristics, should be at a maximum, then clusters will be different from each other), where CDist – the color distance between two pixels, C_k - center of the cluster k.
- $\min_{i=1..m} \left(\sum_{k=1..K} \sum_{p=1..S_k} PDist(C_k, X_p) \right)$ - the minimum value of the sums of geometric distances between the centers of the clusters and the pixels belonging to the cluster (the sum of the Euclidean distances between the center of the cluster and each of its pixels should be minimal, , according to the spatial characteristics, respectively, the cluster will be more homogeneous), where S_k - the number of pixels in the cluster k, PDist – Euclidian distance between two pixels, C_k - center of the cluster k.
- $\min_{i=1..m} \left(\sum_{k=1..K} \sum_{p=1..S_k} CDist(C_k, X_p) \right)$ - the minimum sum value of the color distances between the centers of the clusters and the pixels belonging to the cluster (sum of the Euclidean distances between the center of the cluster and its each pixel, according to color characteristics, should be minimal, then, the cluster will be more compact), where S_k - the number of pixels in the cluster k, CDist – the color distance between two pixels, C_k - center of the cluster k.

The fitness function of the ant colony is 3-criterial and can be obtained as follows:

$$f(m_i) = \begin{cases} f_1 = \sum_{k=1..K-1} \sum_{j=k+1..K} CDist(C_{m_i k}, C_{m_i j}) \\ f_2 = \sum_{k=1..K} \sum_{p=1..S_{m_i k}} PDist(C_{m_i k}, X_{m_i p}) \\ f_3 = \sum_{k=1..K} \sum_{p=1..S_{m_i k}} CDist(C_{m_i k}, X_{m_i p}) \end{cases} \quad (5)$$

Choice of the best solution can be presented as:

$$f(best) = \sum_{i=1..m} \begin{cases} \max(f_1) \\ \min(f_2) \\ \min(f_3) \end{cases} \quad (6)$$

Pheromone updating is calculated as in standard ACO algorithm according to:

$$\tau_i(X_n) \leftarrow (1-\rho)\tau_i(X_n) + \sum_i \Delta\tau_i(X_n) \quad (7)$$

and pheromone evaporation is calculated as:

$$\Delta\tau_i(X_n) = \begin{cases} \frac{Q * Min(k')}{AvgCDist(k',i) * AvgPDist(k',i)}, & if\ X_n \subset cluster\ i \\ 0, & otherwise \end{cases} \quad (8)$$

where Q is a constant value, $Min(k')$ - the minimum color outer cluster distance found by the most successful ant, $AvgCDist(k',i)$ and $AvgPDist(k',i)$ - the average value of the spatial Euclidean and color distances between each pixel and cluster centers for the most successful ant.

IV. Empirical Studies of the Algorithm Parameters

The developed algorithm is supposed to conduct a pilot study of such parameters as n_k - number of ants, α and β - heuristic coefficients of ACO algorithm. For the tests there were used MRI images from Ossirix image dataset [7] as well as images of real patients. Preliminary, all the pictures were divided into 6 groups: brain, heart, lungs, liver, bone structures and others. The word "others" refers to the images that do not have a sufficient number of examples and impossible to classify. The study provides a simulation of the behavior of the algorithm with various settings in order to determine the influence of the initial parameters on the total convergence of the algorithm, the number of iterations, the proximity to the optimal solution found, the degree of identity to the reference image processing result. Some additional tools have been developed to obtain the results shown in Figure 2.

For each test case, the processing results were stored in Microsoft SQL Server database. Based on the stored information, it was possible to calculate the final experimental coefficients by using SQL queries and obtain dependencies to sum up.

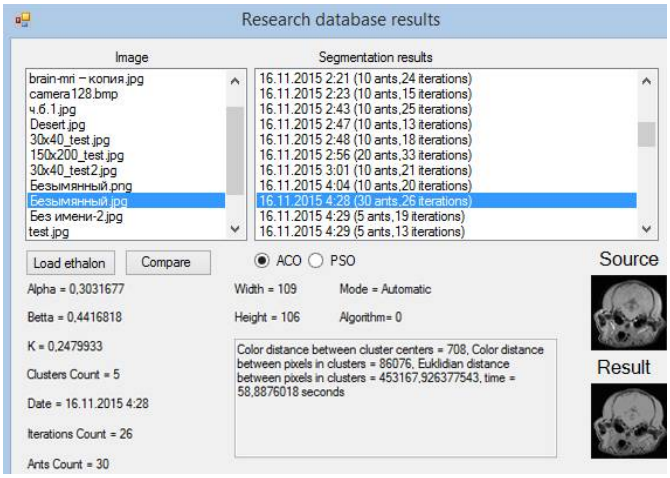


Fig. 2. Study results for ACO-k-means combined algorithm

An important part of the research is to assess the quality of the output result of processing accuracy. Jaccard Similarity Index was used in order to assess the quality of the algorithm [8]. It is calculated in the following way:

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|}=\frac{|A\cap B|}{|A|+|B|-|A\cap B|} \tag{9}$$

A hundred of images from Ossirix dataset were used to obtain the optimal parameters. All images were presented with different initial conditions – with presence of noise, in contrast mode, blurred, in different sizes as well as images of good quality.

Diagrams for finally obtained results are presented further. There we present dependency of image dimensions on the number of ants (figure 3), averaged diagram of algorithm iterations before and after use of optimal heuristic parameters (figure 4) and dependency of heuristics on the image scaling (figure 5).

Figure 3 shows, that for images lower than 50x50 pixels it is rational to use no more than five computational agents. For large values - up to 800x800 pixels, it is appropriate to take no more than 60 ants. These values were obtained experimentally, however, due to probabilistic algorithms and the various options for the initial arrangements of ants combinations, optimal or near-optimal solutions are also achieved with less and more computational agents, but in a much smaller number of cases. Therefore, the data in Figure 3 are the results of the generalized automatic segmentation method of ant colonies and k-means.

Figure 4 shows a histogram of dependence on the number of steps before and after heuristical coefficients correction. As we can observe, use of optimal parameters allows to reduce the number of iterations and to accelerate algorithm.

In the investigation of the algorithm it is also of interest how the coefficients depend on the scaling. Figure 5 shows a histogram of heuristic coefficients dependency on image scaling. As we can observe, the coefficients are resistant to image scaling.

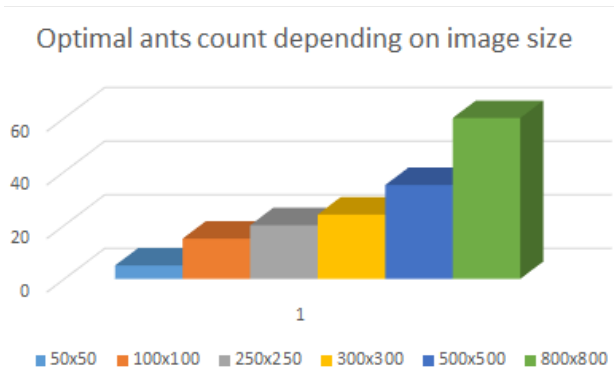


Fig. 3. Optimal ants count depending on image size

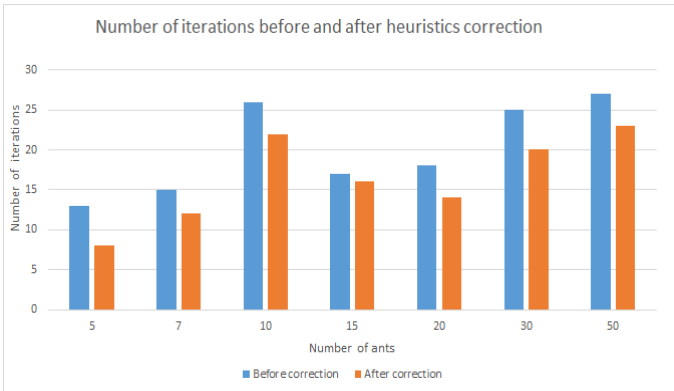


Fig. 4. Histogram of iterations number before and after coefficients correction

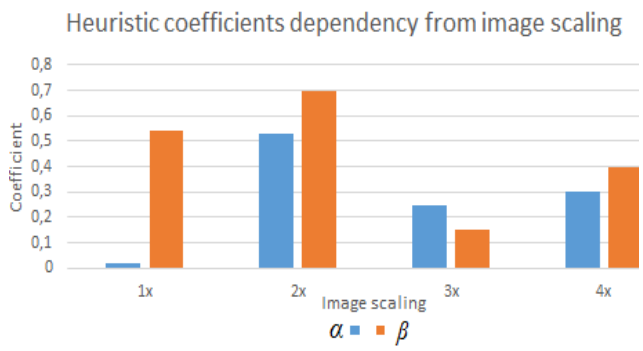


Fig. 5. Heuristic coefficients dependency on image scaling

Table 1 presents the average values of heuristics after the research of various image groups.

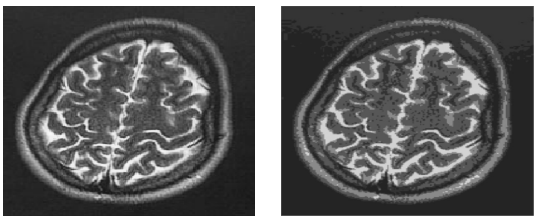
TABLE I. EXPERIMENTAL HEURISTIC COEFFICIENTS FOR DIFFERENT MRI IMAGE GROUPS

Image group	Image condition			
	Noise	***Contrast***	***Blur***	***Good quality***
Brain	$\alpha = 0.12$; $\beta = 0.52$	$\alpha = 0.22$; $\beta = 0.7$	$\alpha = 0.32$; $\beta = 0.6$	$\alpha = 0.14$; $\beta = 0.44$
Heart	$\alpha = 0.20$; $\beta = 0.43$	$\alpha = 0.17$; $\beta = 0.33$	$\alpha = 0.37$; $\beta = 0.23$	$\alpha = 0.12$; $\beta = 0.5$
Lungs	$\alpha = 0.21$; $\beta = 0.46$	$\alpha = 0.17$; $\beta = 0.36$	$\alpha = 0.27$; $\beta = 0.26$	$\alpha = 0.31$; $\beta = 0.44$
Liver	$\alpha = 0.31$; $\beta = 0.36$	$\alpha = 0.37$; $\beta = 0.46$	$\alpha = 0.47$; $\beta = 0.56$	$\alpha = 0.21$; $\beta = 0.26$
Bone structures	$\alpha = 0.30$; $\beta = 0.20$	$\alpha = 0.20$; $\beta = 0.10$	$\alpha = 0.30$; $\beta = 0.20$	$\alpha = 0.31$; $\beta = 0.30$
Others	$\alpha = 0.15$; $\beta = 0.35$	$\alpha = 0.25$; $\beta = 0.34$	$\alpha = 0.35$; $\beta = 0.24$	$\alpha = 0.35$; $\beta = 0.65$

V. RESULTS

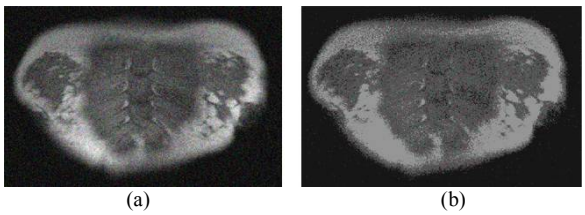
In order to demonstrate how the algorithm operates, we suggest to consider the results of the segmentation of some images from Ossirixx image dataset.

Figure 6 shows MRI of brain: 6a – source image, 6b – segmentation result. Figure 7 shows MRI of the heart: 7a – source image and 7b – segmentation result.



(a) (b)

Fig. 6. Brain MRI. (a) - Source image. (b) – Segmented image



(a) (b)

Fig. 7. Heart MRI. (a) - Source image. (b) – Segmented image.

On the presented images, we can observe the localization of the most complex and difficult elements and their separation from the background.

In order to test the segmentation method, its results are usually compared to other segmentation techniques. We have carried out a comparison of the proposed method with the optimal settings to well-known existing methods, such as C-means [9] and manual method – Magic Wand.

To measure the degree of accuracy for used algorithms we have applied an error matrix methodology [10].

$$Accuracy\ percent = \frac{N-D}{N} * 100 \qquad (10)$$

where N – the total number of pixels, D - number of pixels other than the standard.

Figures 8-10 present segmentation accuracy of different image quality for different segmentation algorithms.

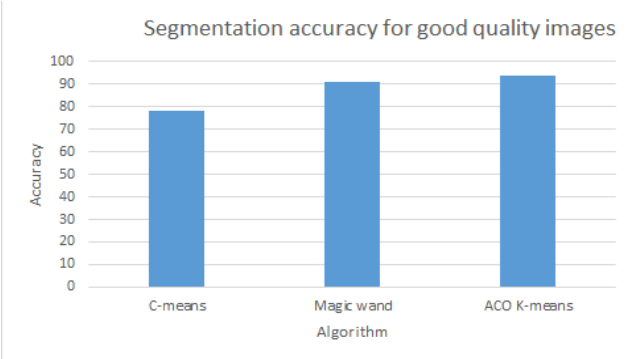


Fig. 8. Segmentation accuracy for good quality images

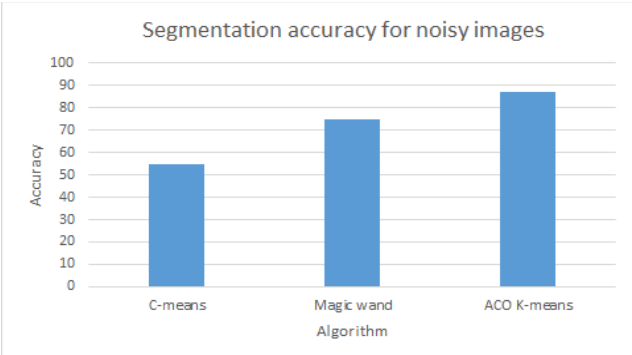


Fig. 9. Segmentation accuracy for noisy images

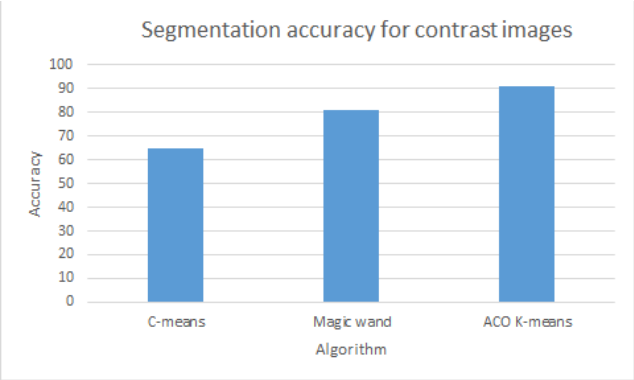


Fig. 10. Segmentation accuracy for contrast images

As we can observe from the diagrams, in all cases the proposed algorithm outperformed other used algorithms. Mainly enhancing was obtained for noisy images – 15%.

VI. Conclusion

In the presented paper, there has been introduced the combined ACO-k-means algorithm and proposed investigation of its parameters. The presented algorithm based on the ants population. Each ant makes its own decision, using the information on the pheromones secreted by other ants of the colony. There have been proposed the experimental values of heuristics and the optimal number of ants, depending on the size of the images, for obtaining solutions. In general, the use of optimal parameters allowed us to obtain acceleration of the algorithm by 15% in comparison with common pseudorandom values.

A comparison of the algorithms' results with C-means and Magic Wand methods has been presented. In all cases, the algorithm makes a better final value of the result accuracy, comparing to the studied techniques. The greatest improvement in the accuracy of the result has been obtained for noisy images - by 15% compared to the existing investigated methods.

The proposed algorithm can be enhanced by means of the additional research, especially of pseudorandom heuristic coefficients, as well as their impact on convergence and the final result of the processing.

These experiments could reduce the computational complexity and, as result, accelerate the algorithm, without using the parallelization approaches, such as MPI or CUDA.

References

[1] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd edition, Prentice-Hall, 2008.

[2] W.K. Pratt, Introduction to Digital Image Processing, CRC Press, 2013.

[3] S.A. El-Khatib, Yu.A. Skobtcov, System of Medical Image Segmentation Using Ant Colony Optimization. St. Petersburg state polytechnic university journal "Computer Science. Telecommunications and control systems", 2(217) – 3(222), Saint Petersburg, 2015, pp. 9-19. DOI: 10.5862/JCSTCS/1

[4] S.A.El-Khatib S.A., Yu.A. Skobtsov, Computer system of medical image segmentation using ant colony optimization method. Scientific journal "Radio Electronics, Computer Science, Control", #3(34), Zaporizhzhya, 2015, pp.49-57. DOI: 10.15588/1607-3274-2015-3-6

[5] M. Dorigo, Optimization, Learning and Natural Algorithms. Phd Thesis. Politecnico de Mi-lano, 1992.

[6] Y.A. Skobtcov, E.E.Fedorov, Metaheuristics. Donetsk National Technical University, pub-lisher "Nowledg", 2013, 426p.

[7] Ossirix image dataset. http://www.osirix-viewer.com/

[8] P. Jaccard, Distribution de la flore alpine dans le Bassin des Dranses et dans quelques re-gions voisines. Bulletin de la Societe Vaudoise des Sciences Naturelles, V. 37(140), pp. 241—272

[9] Zhou H. Fuzzy C-Means Techniques for Medical Image Segmentation. / H. Zhou, G. Schaefer, C. Shi // Fuzzy Systems in Bioinformatics and Computational Biology Volume 242 of the series Studies in Fuzziness and Soft Computing. – 2009. – p. 257-271.

[10] Confusion matrix [Electronic resource]. – Access mode : https://en.wikipedia.org/wiki/Confusion_matrix