



## FUZZY CLUSTERING METHODS IN MULTISPECTRAL SATELLITE IMAGE SEGMENTATION

Rauf Kh. Sadykhov <sup>1)</sup>, Valentin V. Ganchenko <sup>1)</sup>, Leonid P. Podenok <sup>2)</sup>

<sup>1)</sup> Computer Systems Department,  
Belarusian State University of Informatics and Radioelectronics,  
6 P. Brovka st, Minsk, Belarus  
rsadykhov@bsuir.by, ganchenko@lsi-bas-net.by, http://www.bsuir.by

<sup>2)</sup> Laboratory of System Identification,  
United Institute of Informatics Problems,  
National Academy of Sciences of Belarus,  
6 Surganov st, Minsk, Belarus  
podenok@lsi-bas-net.by, http://uiip.bas-net.by

**Abstract:** Segmentation method for subject processing the multi-spectral satellite images based on fuzzy clustering and preliminary non-linear filtering is represented. Three fuzzy clustering algorithms, namely Fuzzy C-means, Gustafson-Kessel, and Gath-Geva have been utilized. The experimental results obtained using these algorithms with and without preliminary nonlinear filtering to segment multi-spectral Landsat images have approved that segmentation based on fuzzy clustering provides good-looking discrimination of different land cover types. Implementations of Fuzzy C-means, Gustafson-Kessel, and Gath-Geva algorithms have got linear computational complexity depending on initial cluster amount and image size for single iteration step. They assume internal parallel implementation. The preliminary processing of source channels with nonlinear filter provides more clear cluster discrimination and has as a consequence more clear segment outlining...

**Keywords:** *Fuzzy Systems, Clustering, Image Segmentation, Multi-spectral Images, Satellite Images.*

### 1. INTRODUCTION

It is known that forests and wetland are the main factors preventing the decline in biodiversity on the Earth in aggressive conditions of human activity. The main problem is agricultural expansion and deforestation. Deforestation is the consequence of two main reasons – agricultural expansion and accidental events. But forestry and agriculture are inseparable and condemned to work hand in hand. Significant part of forest is damaged by fire, pests, irrational agricultural politics leading to change of ground water level and as result leads to sickness and wreck. At now it is possible to discriminate forest areas on early stage of damaging using multi-spectral images of high spatial resolution received from satellites. That technology started about 40 years ago to monitor Earth surface at now is the effective instrument of ecological and agricultural monitoring such the regions as forests and wetland and preventing any accidents. Multi-spectral satellite images are able to bring us information in both visible and invisible spectral bands about vegetation,

water temperature and land cover.

Multi-dimensional cluster analysis and segmentation are base procedures in thematic processing the multi-spectral images received from remote sensing satellites. There are lot of clustering and segmentation methods which have different benefits and imperfections. The special class of such the methods is represented by fuzzy clustering ones.

### 2. METHOD DESCRIPTION

Three clustering algorithms based on fuzzy methods where developed and utilized as part of segmentation software. There are fuzzy c-means [1] (FCM) and its variants – Gustafson-Kessel clustering algorithm [2, 3] and Gath-Geva one [4].

FCM is a method of data clustering which allows one data objects to be a member of two or more clusters. This method developed by [5] and improved by [6] is based on minimization of the following objective function:

$$J = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - s_j\|, \quad (1)$$

where  $m$  is real number  $\geq 0$ ,  $\mu_{ij}$  – degree of membership of  $x_i$  in cluster  $s_j$ ,  $x_i$  is the multi-dimensional data object,  $s_j$  is the multi-dimensional center of the cluster, and  $\|\cdot\|$  is any norm expressing the similarity between any measured data and the cluster center. Fuzzy clustering is carried out through an iterative optimization of the  $J_m$  with the update of membership matrix  $\mu_{ij}$  and the cluster centers  $s_j$  using following algorithm:

1. The data set  $X = (x_j) = (x_{j1}, x_{j2}, \dots, x_{jp})^T$  given. We choose the number of clusters, denoted by “ $c$ ”,  $1 < c < n$ , where  $p$  – dimension of data set.
2. Initialize the partition matrix  $\mu_{ij}$  using random number generator in range  $[0, 1]$ :

$$\sum_{i=1}^c \mu_{ij} = 1 \quad (2)$$

3. Calculate the cluster centers  $s_i = (s_{i1}, s_{i2}, \dots, s_{ip})^T$  ( $i = 1, 2, \dots, c$ ):

$$s_i = \frac{\sum_{j=1}^n \mu_{ij}^2 x_j}{\sum_{j=1}^n \mu_{ij}^2} \quad (3)$$

4. Calculate the error:

$$E = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^2 (s_i - x_j)^T (s_i - x_j) \quad (4)$$

If error is small enough, then stop iterations, else go to item 5.

5. Calculate the “new” partition matrix  $\mu_{ij}$ , where  $d_{mj}^2$  is similarity measure:

$$\mu_{ij} = \left( \frac{\sum_{m=1}^n d_{ij}^2 / d_{mj}^2}{\sum_{m=1}^n d_{ij}^2 / d_{mj}^2} \right)^{-\frac{1}{2}} \quad (5)$$

and then go to step 3.

Fuzzy c-means, Gustafson-Kessel, and Gath-Geva clustering algorithms are distinguished in the definition of distance function between the objects to be classified: Fuzzy C-means:

$$d_{ij}^2 = (s_i - x_j)^T (s_i - x_j) \quad (6)$$

Simple Euclidean distance provides hyper-spherical form of clusters. Gustafson-Kessel (relation (7)) and Gath-Geva (relation (8)) provide ellipsoidal form of clusters and more comprehensive partitioning the multi-dimensional data.

$$d_{ij}^2 = (s_i - x_j)^T (\det(F_i)^{\frac{1}{2}} F_i^{-1}) (s_i - x_j) \quad (7)$$

$$d_{ij}^2 = \frac{\det(F_i)^{\frac{1}{2}}}{\alpha_i} \exp\left(\frac{1}{2}(s_i - x_j)^T (F_i^{-1})(s_i - x_j)\right) \quad (8)$$

$F_i$  is covariance matrix and  $\alpha_i$  is calculated in following manner:

$$\alpha_i = \frac{1}{n} \sum_{j=1}^n \mu_{ij} \quad (9)$$

Covariance matrix is used to form non-spherical clusters which are more suitable for multi-dimensional data partitioning. That also leads to visible difference in algorithm convergence, effectivity and performance of processing the multi-spectral images.

Distance function for Gustafson-Kessel and Gath-Geva algorithms uses covariance matrix to take into account different metrics (scales) in different dimensions. The large images to be processed might be the reason of weak covariance matrix conditionality that results in heavy losses of aptitude for discrimination for Gustafson-Kessel and overflows for Gath-Geva when fixed digit capacity arithmetic based on atomic numerical types is used. To keep of those problems standard normalization methods and arbitrary (multi-precision) arithmetic are the sufficient efforts. When  $d_{ij}^2$  becomes small the overflow may occur. To prevent that case the distant function is confined from below by some smallest value  $d_{\min}$ .

Due to source signal noise the some spatial segment granularity occurs. To reduce that granularity the weak nonlinear filtering using algorithm [7] was applied to source channels. That algorithm may be used both for edges extraction and for nonlinear filtering. It does not lower the sharpness of transitions of channel brightness the boundaries of cover types remain clear. As a result the boundaries of spatial segments after clustering process also remain legible. Algorithm works with values of brightness and pixels coordinates simultaneously. To form the homogeneous areas or search the edges on the gray-scale image a round

mask is used. Usually the mask's radius is 3.4 pixels which gives mask of 37 pixels size. The mask is placed at each point of the image and the brightness of each pixel of the mask is compared with the brightness of the mask central pixel (relation (10)), where  $I(r_0)$  – the brightness value of the mask's center,  $I(r)$  – the brightness value of the mask pixel,  $t$  specific threshold.

$$c(\mathbf{r}, \mathbf{r}_0) = \begin{cases} 1, & I(\mathbf{r}) - I(\mathbf{r}_0) \leq t \\ 0, & I(\mathbf{r}) - I(\mathbf{r}_0) > t \end{cases} \quad (20)$$

$$n(\mathbf{r}_0) = \sum_{\mathbf{r}} c(\mathbf{r}, \mathbf{r}_0) \quad (31)$$

The result of the comparisons is conform to the relation (11), where  $n$  is the quantity of pixels in the USAN (Univalve Segment Assimilating Nucleus). Then that sum should be minimized, so the algorithm is called SUSAN (Smallest USAN).

Parameter  $t$  means the maximum of ignored noise. Then  $n$  is compared with it's thresholding value  $g$ , which is  $3n_{\max}/4$ , where  $n_{\max}$  – maximum value which could be assigned to  $n$ .

### 3. EXPERIMENTAL RESULTS

To test the fuzzy clustering algorithms the multi-spectral Landsat images have been used. Fig. 1 shows channels 2-5 which were chosen for processing. Landsat channels shown on fig.1 have different dispersion and were equalized using `netpbm` utilities to make them visible more clearly. The channel data to be fed into clustering software remain untouched because of clustering algorithm takes in to account real distribution of channel signal level on one's own. Raw output data of developed segmentation software is 2-d vector field over  $R^N$ , where  $N$  is a number of clusters. Each component of that field is strictly increasing function of probability for pixel to be member of one of  $N$  clusters.

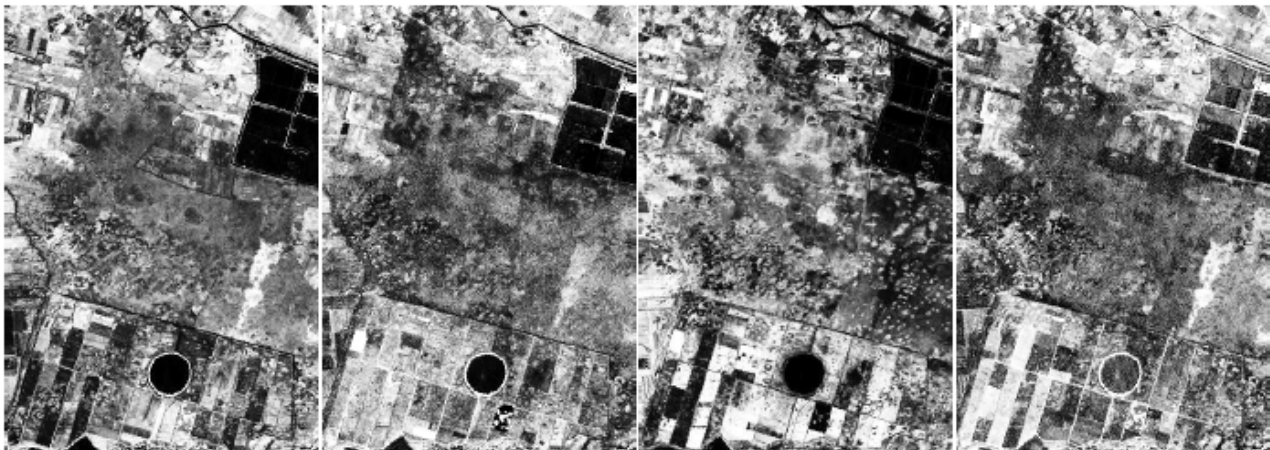


Fig. 1 – Channels 2-5 of multi-spectral Landsat image to be processed with fuzzy clustering algorithm.



Fig. 2 – Nonlinear filtering samples: (a) original image, (b) brightness threshold = 7, (c) brightness threshold = 15, (d) brightness threshold = 20.

In order to obtain exploitable data for a classification scheme, we first needed to extract relevant information of raw Nomarski's microscopy issued images. We proposed to proceed in two steps [2]: first a detected items' images extraction phase and then an appropriated coding of the extracted images.

On presented images one can see some well discriminated land cover types. There are open water, forests, wetlands, bushes and agriculture areas. Results of segmentation on 15 clusters using fuzzy C-means without any preliminary filtering show that there also are fuzzy boundaries of segments on the cluster map going on to sand in

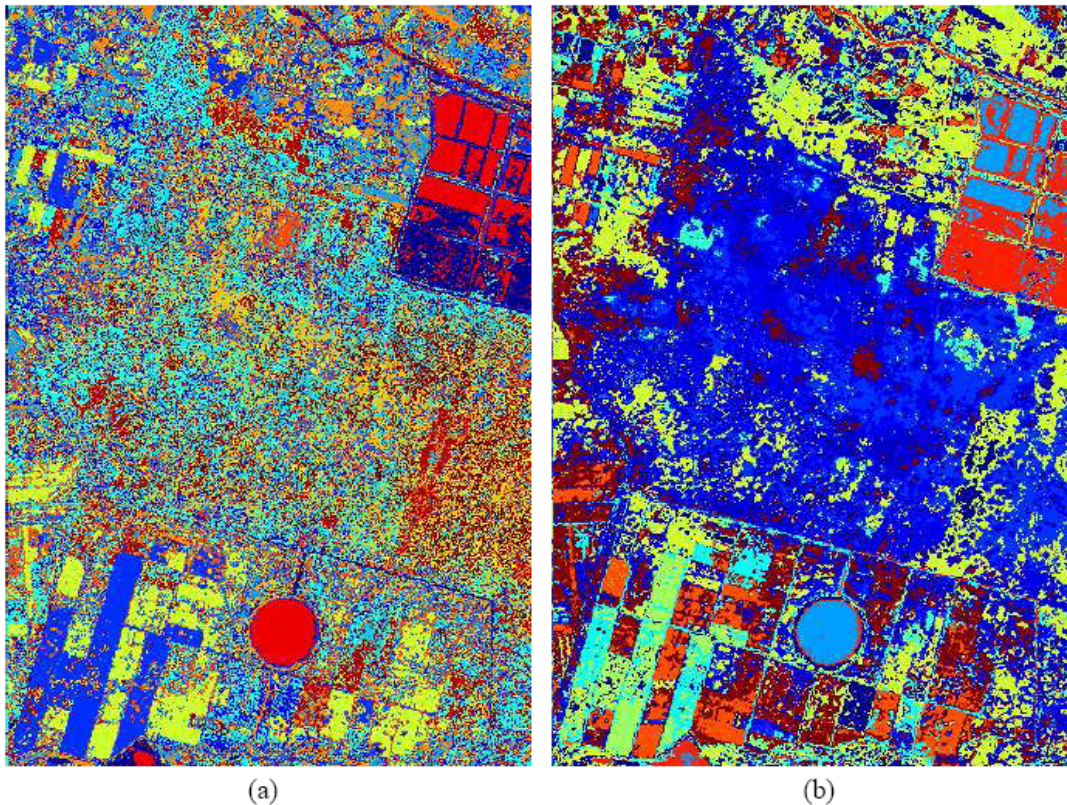


some places. To diminish such the phenomena one can use filter which will smooth slightly changing areas rather blur clearly discriminated land cover type edges. Most appropriate filter having got such the behaviour is the non-linear one described in [7]. The fig. 2 demonstrates smoothing properties of that filter when brightness threshold was chosen be equal to 7, 15, and 20. Besides the brightness threshold parameter the filter has spatial one – distance threshold which was set into default value corresponding to radius equal to about 3.4 and was not changed across experiments. As one can see, smoothing property of that filter increases with brightness threshold grows. At the same time some part of edges remains sharp. Amount of sharp edges and consequently cluster granularity depend on brightness threshold. Thus we get additional degree of freedom in segmentation process control that could dramatically improve the segmentation results.

Nonlinear filtering lowers intra-class covariance but practically doesn't touch interclass ones, so

cluster discrimination doesn't suffer. Smoothing the channels before segmentation allows to eliminate fine granularity of output segment map and to merge the small cluster with large ones. In reality this means disappearance of some real small objects, for example, bushes in wetland areas. Nevertheless, sometimes it is necessary to get generalized segment map. The SUSAN filter allows fulfilling nonlinear processing varying the threshold in wide region achieving the smoothing results from negligible small up to very strong.

Fig. 3-6 demonstrate examples of segmentation using non-linear filtering with different brightness threshold for Gustafson-Kessel and Gath-Geva clustering algorithms. The cluster number was chosen taking into account real diversity of land covers for territory being investigated. Actually a lot of experiments with different cluster number were fulfilled. As soon as segmentation become to look stable cluster number was stated.



**Fig. 3 – Segmentation results using Gustafson-Kessel (a) and Gath-Geva (b) algorithms at 20 clusters without any filtering.**

Obtained results have transformed into scalar field using simple maximal probability solver to get the possibility of visual evaluation and have presented as colour map of spatial partitions. As a result every pixel became a member of one cluster. All pixels of same colour are members of same cluster. No any intelligent algorithm was used to colourize segment map so the same areas on

different pictures have different colours. As one can see the growth of the threshold leads to raising the segment area when target number of clusters is fixed. However it should be pointed that general structure of segment arrangement and its structure in relation with land cover types is not significantly changed.



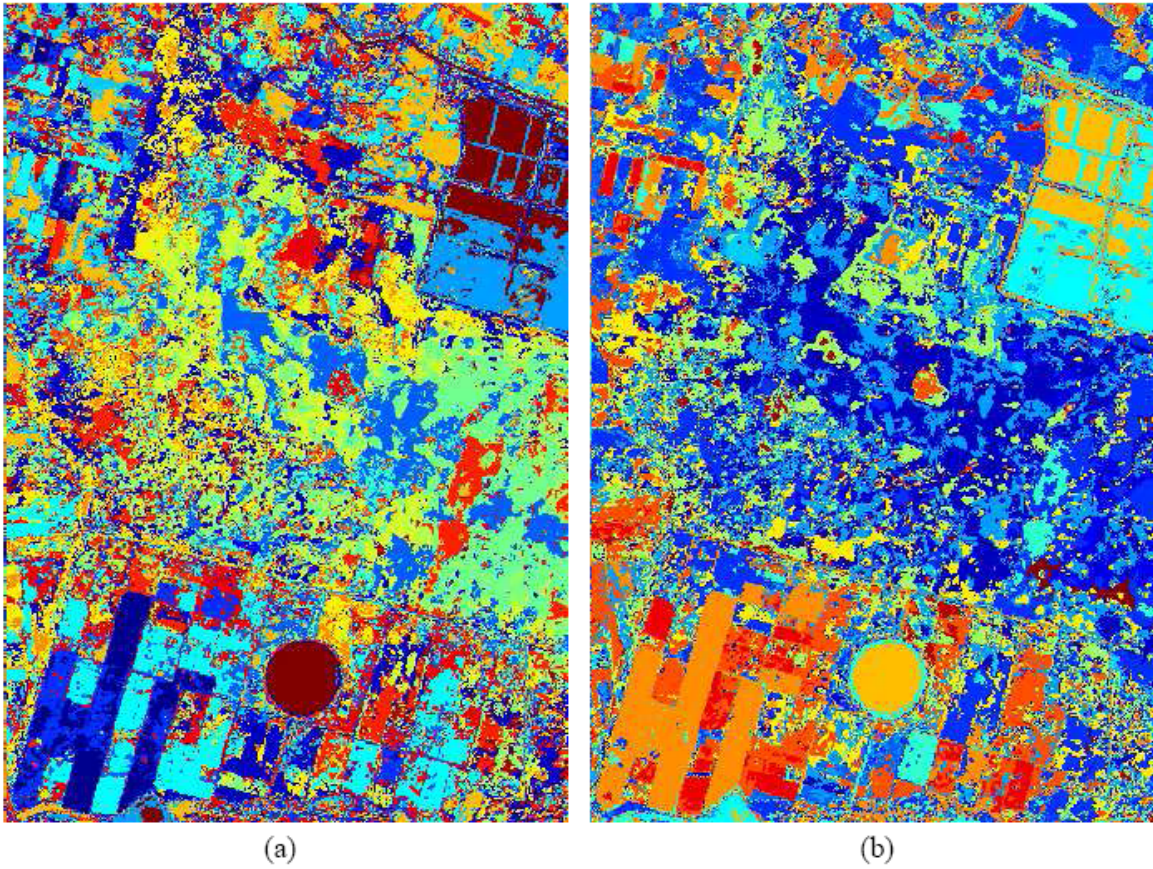


Fig. 4 – Segmentation results using Gustafson-Kessel (a) and Gath-Geva (b) algorithms at 20 clusters using filtering with threshold = 7.

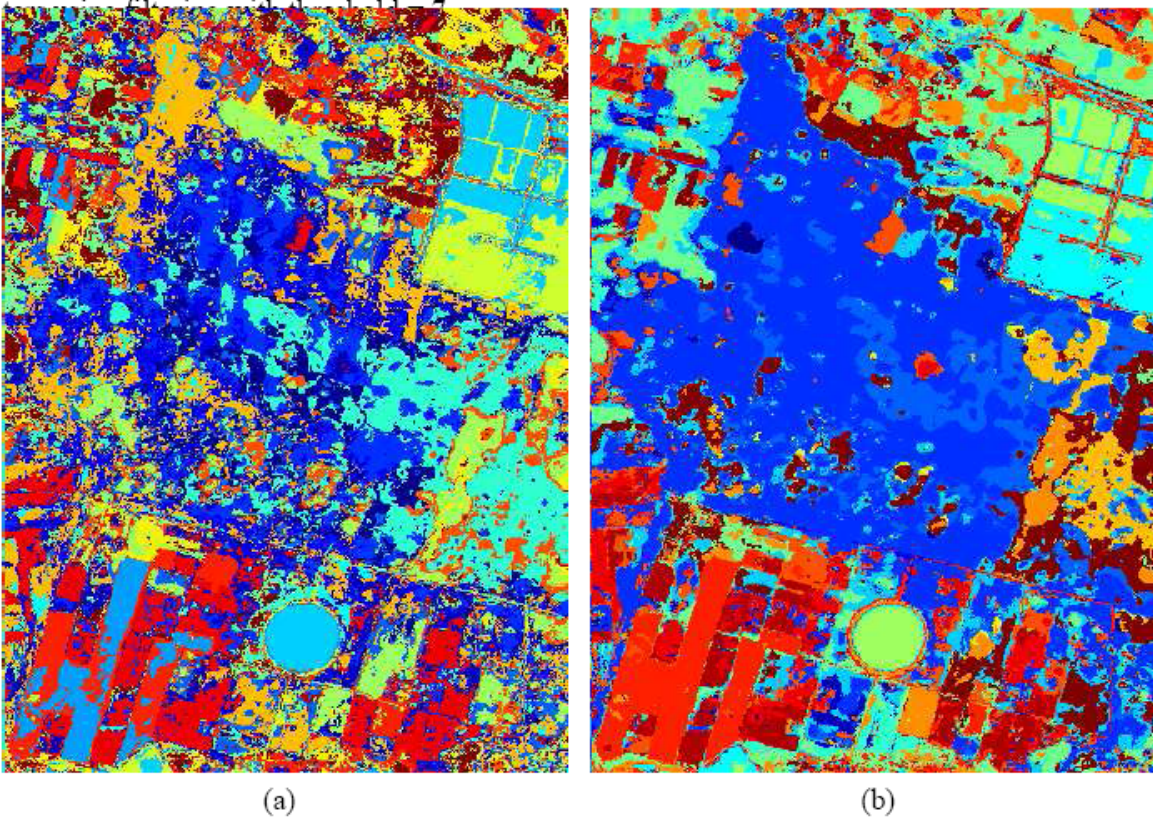


Fig. 5 – Segmentation results using Gustafson-Kessel (a) and Gath-Geva (b) algorithms at 20 clusters using filtering with threshold = 15.



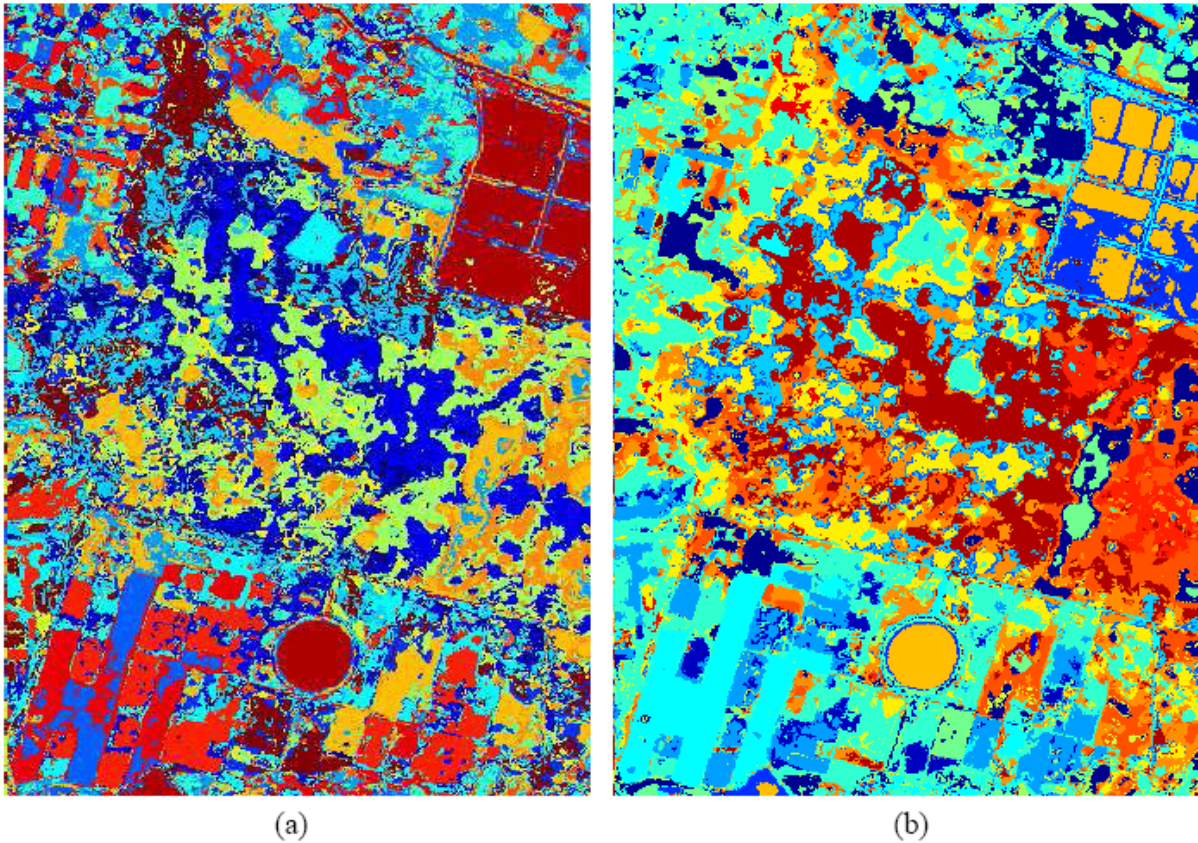


Fig. 6 – Segmentation results using Gustafson-Kessel (a) and Gath-Geva (b) algorithms at 20 clusters using filtering with threshold = 20.

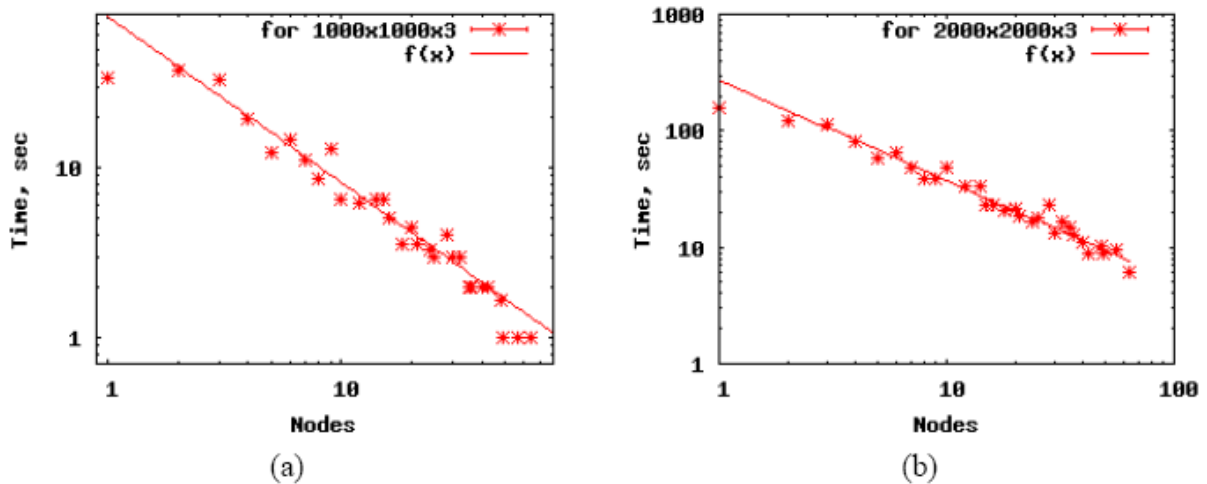


Fig. 7 – Dependence of processing time from node count for colour images with size of 2000x2000 (a) and 1000x1000 (b) pixels.

#### 4. ACCELERATION OF CALCULATIONS

For increase in usability and calculation efficiency can be made with use in systems of mass parallelism, and also have connection with GIS. As a basis for parallel processing interface MPI (The Message Passing Interface) is chosen. Tests of system of processing of images were made on a supercomputer "SKIF K-1000". At carrying out of

experiments it was involved from 1 up to 64 computing units for colour images in the size 2000x2000 and 1000x1000 pixels. Dependence of a system operating time of on amount of the involved computing units is resulted on the schedules resulted on fig. 7.

## 5. CONCLUSION

Segmentation results of multispectral Landsat images obtained using fuzzy clustering methods such as Fuzzy C-means, Gustafson-Kessel, and Gath-Geva with and without preliminary nonlinear filtering testify that segmentation using fuzzy clustering methods provides good-looking discrimination of land cover types that occurs in such the complex cases as wetland, water-meadow, and bush areas. The discrimination quality of segmented images was tested and approved by land-improvement specialists using data of land-based expedition. The non-linear filtering with guided smoothing is the convenient instrument of preliminary processing for semi-automatic segmentation of complex land covers.

Implementations of Fuzzy C-means and Gustafson-Kessel algorithms have got linear computational complexity depending on initial cluster amount and image size for single iteration step. All the algorithms assume internal parallel implementation for MPP computer. The preliminary processing of source channels with nonlinear filter provides clearer cluster discrimination and has as a consequence more clear segment outlining and provides operated generalization of output segment maps.

Really, segmentation software returns the results as 2-d vector field  $v_i$ , which is the strictly increasing function of probability  $p_i$  that pixel is a member of  $i$ -th cluster. When using more complex tool than simple maximal probability solver, for example, maximal likelihood one, it is possible to significantly improve the results of segmentation and to avoid large part of manual processing the multi-spectral data.

## 6. REFERENCES

- [1] Hoppner, F., Klawonn F., Kruse, R., Runkler, T.: Fuzzy Cluster Analysis. Wiley, Chichester (1999).
- [2] Gustafson, D. E., Kessel, W. C.: Fuzzy clustering with fuzzy covariance matrix. In Proceedings of the IEEE CDC, INSTICC Press, San Diego (1979) 761–766.
- [3] Babuska, R., van der Veen, P. J., Kaymak, U.: Improved covariance estimation for Gustafson-Kessel clustering. IEEE International Conference on Fuzzy Systems (2002) 1081–1085”.
- [4] Gath, I., Geva, A. B.: Unsupervised optimal fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence (1989) 7:773–781.
- [5] Dunn, J. C.: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting

Compact Well-Separated Clusters. Journal of Cybernetics (1973) 3: 32-57.

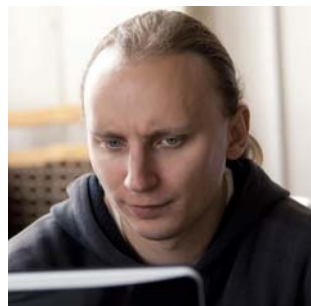
- [6] Bezdek, J. C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981).
- [7] Smith, S. M., Brady, J. M.: SUSAN – a new approach to low level image processing. International Journal of Computer Vision May (1997) 23(1):45-78.



**Rauf Kh. Sadykhov**, in 1967 graduated from Azerbaijan Polytechnic Institute (Baku) on the specialty "Mathematical and Computing the Instruments and Devices". After graduation from the Institute he attended the postgraduate course at the Institute of Engineering Cybernetics in Minsk. In 1991 he defended his thesis for a scientific degree of a doctor of engineering science in the field of computing science and in 1992 has obtained a professor's scientific rank.

Since 1995 R.Kh. Sadykhov is a head of Computer System Department in Belarusian State University of Informatics and Radioelectronics and simultaneously he is a head of System Identification laboratory, Institute of Engineering cybernetics of the Belarusian Academy of Sciences. R.Kh. Sadykhov has published more than 350 scientific works, including books, patents, papers and reports at the International Conferences, Symposiums and Workshop.

The area of the scientific investigations includes: digital signal and image processing, recognition of handwritten symbol and signature identification, remote-sensing object recognition, computer vision system for the control and recognition, intellectual neural systems, multi-agent systems, parallel architectures for digital signal and image processing. Prof. R. Sadykhov is the vice-chairman of Belarusian Association of Pattern Recognition (IAPR) and Belarus SIG of International Neural Network Society (INNS), the member of IEE (United Kingdom).



**Valentin V. Ganchenko** in 2006 graduated from Belarusian State University of Informatics and Radioelectronics (Minsk) on the specialty "Computers, Systems and Networks". After graduation from the University he attended the postgraduate course at the "United Institute of Informatics Problems of the National Academy of Sciences of Belarus" (UIIP NAS of Belarus) in Minsk.

Since 2006 Valentin V. Ganchenko is a collaborator of the Laboratory of System Identification in United Institute of Informatics Problems of the National Academy of Sciences of Belarus.

The area of the scientific investigations includes: digital image processing, remote-sensing object recognition, parallel architectures for digital image processing.



**Leonid P. Podenok** in 1982 graduated from Belarusian State University (Minsk) on the specialty Physics.

Leonid P. Podenok is a collaborator of the Laboratory of System Identification in United Institute of Informatics Problems of the National Academy of Sciences of Belarus.

The area of the scientific investigations includes: multispectral satellite image processing, parallel processing, space telemetry processing, public key cryptography and fast arithmetics.