American Journal of Remote Sensing
2022; 10(2): 30-38
http://www.sciencepublishinggroup.com/j/ajrs
doi: 10.11648/j.ajrs.20221002.11
ISSN: 2328-5788 (Print); ISSN: 2328-580X (Online)

About One Approach to the Construction of Clustering and Classification Grid-Type Algorithms

Anatolii Kuzmin[1], Leonid Grekov[2], Nataliia Kuzmina[3], Oleksii Petrov[2]

[1]Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

[2]Scientific Production Enterprise "Agroresurssystems", Kyiv, Ukraine

[3]Faculty of Mathematics, Informatics and Physics, National Pedagogical Dragomanov University, Kyiv, Ukraine

Email address:

kuzmin_a_b@ukr.net (Anatolii Kuzmin), n.m.kuzmina@npu.edu.ua (Nataliia Kuzmina), petroff@ukr.net (Oleksii Petrov)

To cite this article:

Anatolii Kuzmin, Leonid Grekov, Nataliia Kuzmina, Oleksii Petrov. About One Approach to the Construction of Clustering and Classification Grid-Type Algorithms. A*merican Journal of Remote Sensing*. Vol. 10, No. 2, 2022, pp. 30-38.
doi: 10.11648/j.ajrs.20221002.11

Received: August 10, 2022; **Accepted**: August 29, 2022; **Published**: September 5, 2022

Abstract: Applied problems of studying the earth's surface using satellite images of remote sensing of the Earth are considered for the study of forest, agricultural, water and other natural resources, where clustering and classification algorithms are instrumental research methods. It is noted that the most well-known procedures for classifying and segmenting multispectral space images in GIS systems, such as ArcGIS, ERDAS, ENVI, are built-in. The need to expand the toolkit for a more efficient solution of applied problems of this class is noted. New universal clustering and classification algorithms based on a unified approach are proposed. Both methods belong to grid-type algorithms, and at the first stage of their work they group points of a set of n - dimensional vectors into grid cells, each cell saves only the numbers of points belonging to it and is characterized by a unique code. The vector grid spacing is a parameter of the method and is set by the user using a single integer value. At the next stage, the clustering algorithm combines the cells and the points belonging to them into clusters using the cell neighborhood principle. In this case, the algorithm does not attach the next cell to the cluster in the case when its density is less than the specified value. The classification algorithm refers the points of the cell of the main set to the class to which the cell with the same code of the training set belongs. The algorithms can be used to process large data sets of large spatial dimensions, including satellite images. Clustering and classification algorithms do not require a preliminary specification of the number of clusters and information about the nature of the distribution of points in the input set.

Keywords: Remote Sensing Methods, Images Segmentation, Clustering Algorithms, Classification Algorithms, Grid Methods, Neighborhood Relation, Cell Density

1. The Use of Clustering Algorithms for Solving the Applied Problems of Analyzing Satellite Images

The range of applications of machine learning algorithms is very wide: these algorithms are used in some measure in archeology, medicine, psychology, chemistry, biology, public administration, philology, anthropology, marketing, sociology, geology and other disciplines. We will not claim to be universal, but will dwell on the application of these algorithms in the field of space imagery analysis in the spectral range of the average spatial resolution from 10 to 30 m per pixel for the analysis of various natural resources, primarily agricultural, forestry, water and others. Typical applied problems of analyzing multispectral remote sensing satellite images using cluster analysis algorithms are image segmentation into conditionally homogeneous zones (clusters), for example, arable land, forests, meadows, residential buildings, a surface occupied by water and other types of surface. The questions of studying agricultural resources using remote sensing data are considered in the researches of A. Kuzmin, L. Grekov etc. [1, 2].

As a rule, satellite images with a spatial resolution of more than 20 m per pixel are used to solve such problems, and

extensive territories are analyzed. Data from several spectral ranges (R, G, B, NIR) can be used as initial data, and various vegetation indexes, such as NDVI or others. To reduce the dimension of the feature space and eliminate the correlation between them, when using classification and clustering methods, the Principal Components method is used [3]. Freely distributed satellite images obtained from the NASA Landsat 8, 9 satellite and the European Space Agency's Sentinel l, 2 satellites are very popular. The tasks of segmentation of a smaller spatial scale include clustering forest areas in order to identify different species composition of wood, identifying arrays of different levels of productivity, dividing sown areas into clusters of winter and spring crops, identifying soil differences within one agronomic field or an array of neighboring fields. For tasks of this class, satellite images with a higher spatial resolution of 3–10 m per pixel may turn out to be more promising.

A large set of applied problems of classifying satellite images arises in the analysis of agricultural resources in the interests of large agricultural producers or state land control bodies, designed to monitor compliance with agricultural technologies, in particular crop rotations. Which implies the allocation on a given array of sown areas of fields sown with major crops (wheat, barley, corn, sunflower, rapeseed, sugar beets, soybeans, peas, and others). To solve a problem of this class, a priori information about crops grown in individual fields is used, and a part of the satellite image in selected spectral ranges from such fields is used as a training set. The rest of the satellite image is then analyzed using classification algorithms.

To solve the applied problems of this class, it is convenient to use GIS systems for processing satellite images such as ArcGIS [4], ERDAS [5], ENVI [6] and others, which allow you to effectively perform a number of auxiliary procedures for preliminary preparation of a satellite image for solving clustering or classification problems, and also display the results of these procedures in a convenient form [7]. Note that these GIS systems themselves have built-in procedures for unsupervised and supervised classification of multichannel images. Such as Isodata, K-Means, Mean shift method, support vector machine, minimum distance method, maximum likelihood method, nearest neighbor method, random tree method, DBScan method [8-11].

The purpose of this work is to propose efficient clustering and classification algorithms designed to process large arrays [12] – graphic information obtained from satellite images of the Earth remote sensing in the visible spectrum. The amount of data in such tasks contains $10^5 - 10^7$ pixels in 2 - 5 spectral ranges. The task was also set to minimize the number of customizable parameters of the proposed methods and make them as understandable as possible for applied specialists in the field of remote sensing. Satellite image segmentation methods are considered in a number of works [12-15].

Both algorithms described below are based on general principles and can be classified as grid-type algorithms. At the first stage of the algorithm, objects are grouped by cells of a regular grid, which are assigned a unique vector integer code. At the next stage, the algorithms work only with the codes of the corresponding cells. Cells are combined into clusters using the neighborhood principle or a cell belongs to a certain class when the cell codes of the training and main sets match.

At the same time, the clustering algorithm does not require specifying the number of desired clusters; for its work, it uses two parameters. *Nmean* - which determines the size of the grid steps in the space of the data set. This parameter can be interpreted as the average number of points that fall into a grid cell when they are uniformly distributed in a multidimensional parallelepiped that bounds the original set. *Nmin* – the value of the minimum density of the number of points in a cell, the addition of which to clusters is no longer performed by the algorithm.

In the proposed classification algorithm, in addition to the main and training sets, two parameters are set. $Nmean$, which determines the grid steps built on the training set parallelepiped and $\gamma \le 1$ - the threshold of dominance in the cells of the training set of points of the same class. A grid with the same steps is also used to process the main set.

The grid approach to the construction of classification and clustering algorithms was considered in the works [12, 14, 16 -18], among which such well-known methods as CLIQUE and MAFIA can be distinguished.

2. Description of the Grid Clustering Algorithm

Working title of the method: Method of Uniting Neighboring Cells – UNC Method.

2.1. Initial Data

The set of points of the d-dimensional Euclidean space $P = \{p_1, p_2, ...p_N\}$ [1], where $p_i \in R^d$.

Method parameters: *Nmean* - determines the grid step on the set of points P.

Nmin - sets the minimum cell density, determines the degree of separability of clusters.

2.2. Defining Grid Parameters

$$B\min = \min_{1 \le i \le N} p_i,\ B\max = \max_{1 \le i \le N} p_i,\ l = B\max - B\min,$$

$$NN = \left[\left(\frac{N}{Nmean} \right)^{1/d} \right],\ h = \frac{l}{NN}. \qquad (1)$$

All operations in (1, 2) are performed on vector quantities.

2.3. The Procedure for Grouping Points of the Initial Set P by Grid Cells

For each point of the set P, d - dimensional code of the cell, to which it belongs, is calculated

1 Expressions in { } are sets and allow set-theoretic operations

$$Rc_i = \left[\frac{p_i - B\min}{h} \right] + \mathbf{1},\ i = 1..N \quad (2)$$

We form a table (dictionary) Tpoint.

$$\mathrm{Tpoint}[Rc_j] = j_1, j_2, ...j_k \quad (3)$$

The key of the table records is the cell code, the content of the record are the numbers of points, belonging to the cell with the corresponding code.

We convert the table into a two-dimensional list

$$\mathrm{Lpoint} = \left[...\left[Np_k, Rc_k, j_{k,1}, j_{k,2}, ..j_{k,s} \right] ... \right] \quad k = 1..Nc \quad (4)$$

The list Lpoint is sorted in descending order by the key Np - the number of points (density) in the cell, Nc - the number of cells with non-zero density.

2.4. Grouping Cells into Clusters

The contents of the two-dimensional list Lpoint are sequentially viewed. The cell's code of the first element of the list Rc_1 is written to the first cluster $F_1 = \{Rc_1\}$. For the cell's code of the next element of the list Rc_2 we determine the set of neighboring cells $S(Rc_2)$, the maximum number of neighbors of each cell is 3^d. Compute the intersection $F_1 \cap S(Rc_2)$, if $F_1 \cap S(Rc_2) = \varnothing$ then we form a new cluster $F_2 = \{Rc_2\}$, otherwise $F_1 = F_1 \cup \{Rc_2\}$.

Let clusters $F_1, F_2, ...F_r$ be formed in m steps of list Lpoint processing, for the cell's code of the next element of the list Lpoint Rc_{m+1} we calculate $l_j = \left| S(Rc_{m+1}) \cap F_j \right|, j = 1..r$ - the number of intersections of the set of neighbors with the formed clusters. If $l_j = 0,\ i = 1..r$, then we form a new cluster $F_{r+1} = \{Rc_{m+1}\}$.

Otherwise, if

$$l_j \neq 0,\ j \in \{j_1 < j_2 < ... < j_s\}, s \leq r, \quad (5)$$

Then all clusters with which there is an intersection $S(Rc_{m+1})$ are combined into one $F_{j_1} = F_{j_1} \cup F_{j_2} \cup ...F_{j_s} \cup \{Rc_{m+1}\}$. All other clusters are numbered while maintaining their order.

The process of joining cells ends either with the exhaustion of the list Lpoint, or if the density of the next cell is $Np_{m+1} < N\min$.

2.5. Grouping Points into Clusters

The result of the previous stage is clusters containing many cell codes $F_i = \{Rc_{i1}, Rc_{i2}, ...Rc_{ik}\}$. For each cluster F_i, we form a cluster containing a set of point numbers Fpoint_i. $\mathrm{Fpoint}_i = \mathrm{Fpoint}_i \cup \{\mathrm{Tpoint}[Rc_j]\}, j \in \{i1, i2, ..ik\}$.

2.6. Computational Complexity of the Algorithm

Note that the clustering procedure itself, points 4 and 5 are performed using natural numbers: cell codes and point numbers of the original set. Floating-point calculations are performed only in the point 3 of the algorithm to determine the cell code and require $N \cdot d$ division operations and $N \cdot d$ addition operations, as well N as operations for adding numbers of points of the initial set to the table Tpoint. Note that this part of the algorithm has the ability to implement parallel computing.

At stage 3, the intersection of the neighbors of the cell to be joined with already existing clusters is searched. Computational complexity of the operation of intersection of sets A and B - $O(\min(|A|,|B|))$. At the m-step, the computational costs are $O(\min(m, 3^d)) \leq O(3^d)$. Thus, the total complexity of the operations of intersection of sets is estimated by the value $O(Nc \cdot 3^d)$

When condition (5) is satisfied, it is required to perform a union of sets, the complexity of the operation of combining two sets A and B is $O(|A| + |B|)$. Taking into account that the total power of the merged sets at each step increases by one and does not exceed Nc, the total computational cost estimate for the stage of merging sets is defined as $O(Nc^2)$. Thus, the total cost estimate is estimated as $O((Nc)^2 + Nd)$.

3. Description of the Grid Classification Algorithm

Working name of the method: Grid Pattern Classification Method GPC Method.

3.1. Initial Data

The labeled set of points in a d - dimensional Euclidean space $Pl = \{(pl_1, k_1), (pl_2, k_2), ...(pl_{N_1}, k_{N_1})\}$, where $pl_i \in R^d$, $k_i \in \{1, 2, ...K\}$. The set Pl will be called training.

The initial set $P = \{p_1, p_{2,....} p_N\}$, $p_i \in R^d$. $N >> N_1$.

$Nmean$ - determines the grid step on the set of points P and Pl, γ - the threshold value for assigning a cell to a specific cluster.

3.2. Defining Grid Parameters

The grid parameters of the classification method are determined by the formulas (1) with the only difference in the calculation of $B\min = \min_{1 \leq i \leq N_1} pl_i$, $B\max = \max_{1 \leq i \leq N_1} pl_i$, $l = B\max - B\min$, i.e. the grid parameters are determined by the points of the training set of points.

3.3. The Procedure for Grouping Points of the Training and Initial Sets by Grid Cells

For each cell of the sets P and Pl by the formula (2) we

calculate d - dimensional code of the cell, to which it belongs $Rc_i, i=1..N$ and $\mathrm{Rcl}_j, j=1..N_1$.

We form tables Tpoint, $\mathrm{Tpoint}[\mathrm{Rc}_s]=s_1,s_2..s_m$, where the key of the table is the cell's code, and the contents of the record of the numbers of all points, that belonged to the cell with the corresponding code and TLpoint, $\mathrm{TLpoint}[\mathrm{Rcl}_i]=l_{i,1},l_{i,2},...l_{i,K}$, where $l_{i,j}, j=1..K$ is the number of points in the training set with the label j, that belong to the cell with the code Rlc_i.

It should be noted that the set of codes of non-empty cells $\{\boldsymbol{Rcl}\}$ constructed for the training set of points Pl and the main set $\{\boldsymbol{Rc}\}$ of points P, as a rule, does not match. The best option will be when $\{\boldsymbol{Rc}\}\subseteq\{\boldsymbol{Rcl}\}$ or $\{\boldsymbol{Rc}\}/\{\boldsymbol{Rcl}\}$ contains the minimum number of non-empty cells, which will provide the most correct classification process.

3.4. Set $\{\boldsymbol{Rc}\}$ Cell Classification Procedure

We sequentially look through the keys of the table Tpoint. For each key Rc_s, we determine the presence of a record with a similar key in the table $\mathrm{TLpoint}[\mathrm{Rc}_s]$.

If the record $\mathrm{TLpoint}[\mathrm{Rc}_s]=l_{s,1},l_{s,2},...l_{s,K}$ exists, then we check the condition for the dominance of points of the same class.

$$\frac{\max\limits_{1\le j\le K} l_{s,j}}{\sum\limits_{j=1}^{K} l_{s,j}}\ge\gamma \qquad (6)$$

When condition (6) is satisfied, we mark all points $\mathrm{Tpoint}[\mathrm{Rc}_s]$ as belonging to the class $s_0=\arg\max\limits_{1\le j\le K} l_{s,j}$. If condition (6) is not satisfied, then the label 0 is assigned to the points of the cell, i.e. the points of such cell are not assigned to any class.

Note, that the presence in the table TLpoint of a large number of cells, for which condition (6) is violated, indicates about insufficient quality of the training set, this may also be evidenced by the presence of a significant number of records in the main set table Tpoint, for which there is no record with the same key in the training set table TLpoint. The next paragraph deals with the procedure for assigning a class label to some cells of this type.

3.5. The Procedure for Joining to the Formed Classes of Cells of the Main Set Whose Neighbors Have Intersection with the Cells of the Training Set

We calculate $\{\boldsymbol{TTset}\}=\{\boldsymbol{Rc}\}\setminus\{\boldsymbol{Rcl}\}$. If this set is not empty, then for each cell of the set $Rc_i\in\{\boldsymbol{TTset}\}$ we calculate the intersection of the neighbors of the cell with the set $\{\boldsymbol{Rcl}\}$, $\{\boldsymbol{TTin}\}=S(Rc_i)\cap\{\boldsymbol{Rcl}\}$. $\{\boldsymbol{TTin}\}$ - contains cells of the training set that are adjacent to the current cell from $\{\boldsymbol{TTset}\}$ and thereby determine the class of points of the cell.

To determine the class label, we find the value $l_1,l_2,...l_K=\sum\limits_{\mathrm{Rcl}_j\in\{TTin\}}\mathrm{TLpoint}[\mathrm{Rcl}_j]$ and calculate the class label of all points in the cell according to criterion (6). If criterion (6) is not met for the current cell, then the label 0 is assigned to such cell and points, which corresponds to the impossibility of assigning the points of such cell to any classes. Thus, a set of points is generated that do not belong to any of the classes.

3.6. Computational Complexity of the Algorithm

At the stage 3, the cell codes for the training and main point sets are calculated with the total number of floating-point operations $2(N+N_1)d$. At the stage 4 of the algorithm to check the logical condition (6) it is necessary to perform $(K+2)Ncl$ of arithmetic operations, where $Ncl=len(\{Rcl\})$ is the number of non-empty cells of the training set Pl. Also for each cell of the set $\{\boldsymbol{Rcl}\}$, by its code, an object with the same code in the set $\{\boldsymbol{Rcl}\}$ is selected, which requires Nc operations where $Nc=len(\{\boldsymbol{Rc}\})$.

4. The Results of Computational Experiments

The article presents computational experiments on the operation of clustering and classification algorithms using two test sets of points in a space R^2 with the conditional name "Gauss" with a volume of 450,000 points and "Figures" with a volume of 500,000 points.

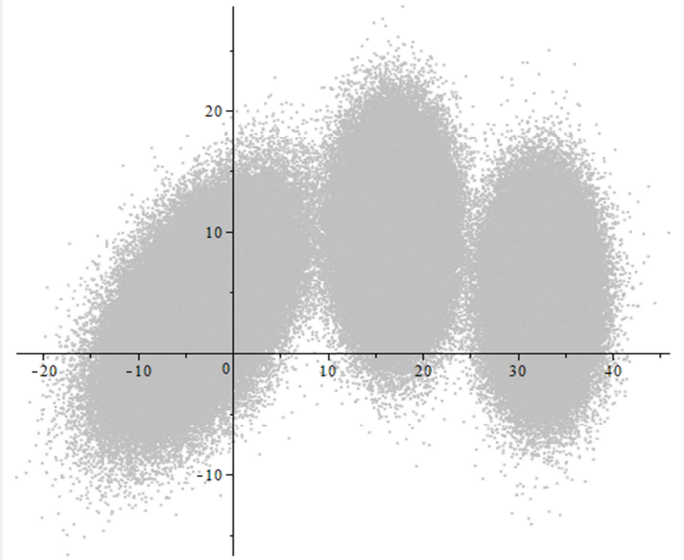
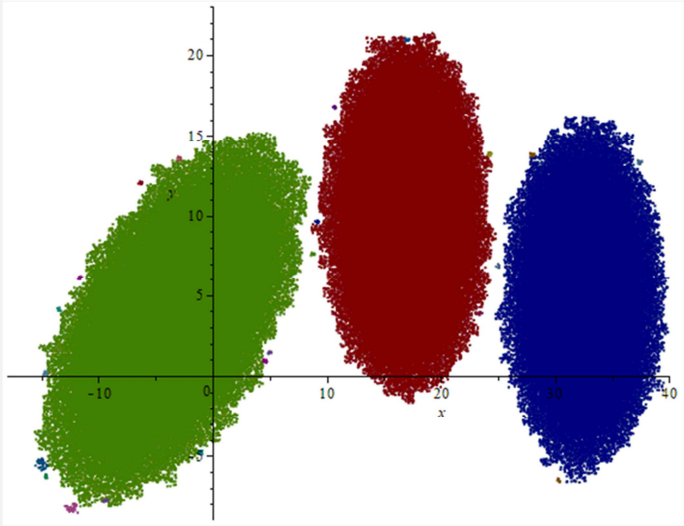
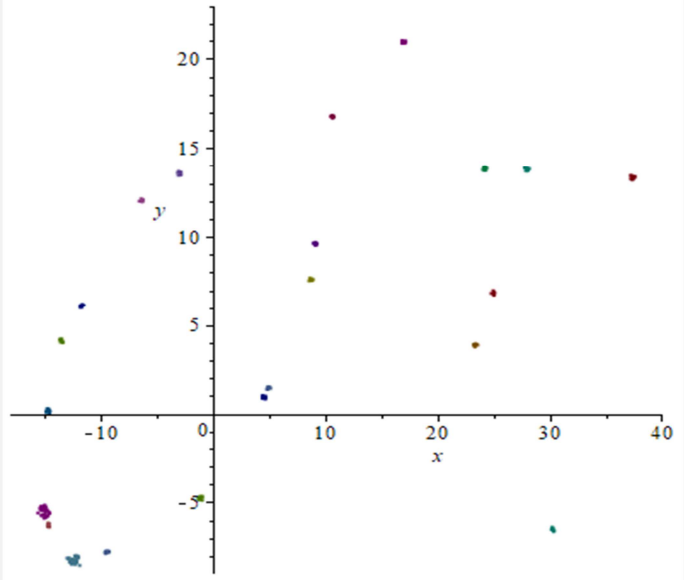
The "Gauss" set contains three clouds of points distributed according to the normal law with different parameters of mean values and dispersions. In this case, the point clouds are partially overlapped by their peripheral parts.

The set "Figures" contains 7 fragments of various shapes and densities, including those that are not linearly separable. Additionally, the "Shapes" set is "seeded" with a set of evenly distributed points that simulate spikes and noise in the amount of 2000 points.

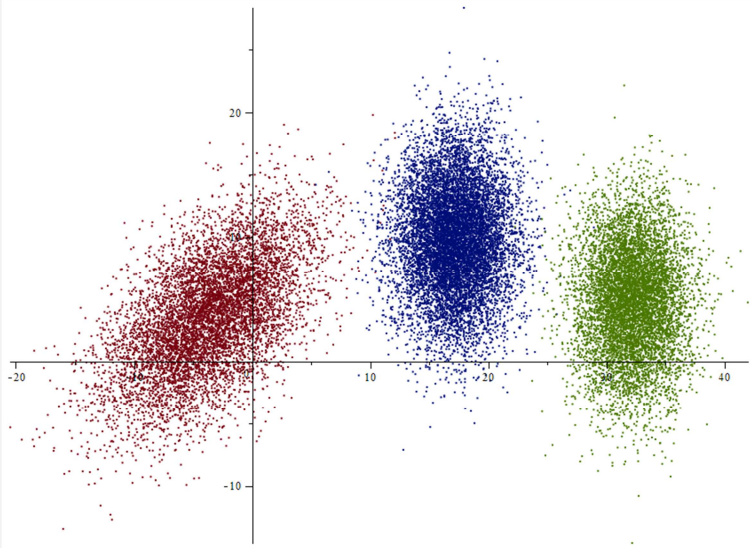
For the classification algorithm to work, training sets of points are randomly selected from the described sets, containing 5-7% of the volume of the original sets, to which local perturbations are introduced.

4.1. Test Results with the Set of Points "Gauss"

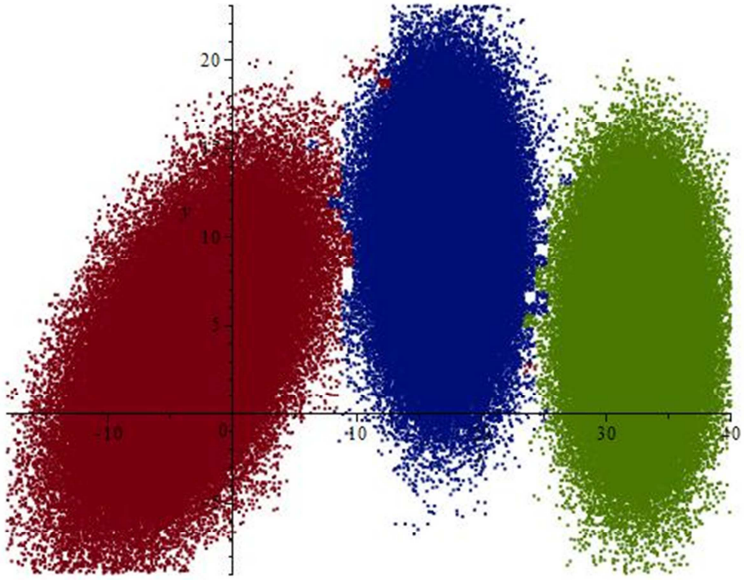
Table 1. *Test "Gauss".*

Initial set of points "Gauss"	
Clustering results. The algorithm assigned 489865 points to the three main clusters, which is 98% of the total. A grid of 182x182 cells was used. $Nmean$ =15. $Nmin$ =8 Clustering error is ≈ 2%	
False clusters of small volume on the boundaries of the main clusters	

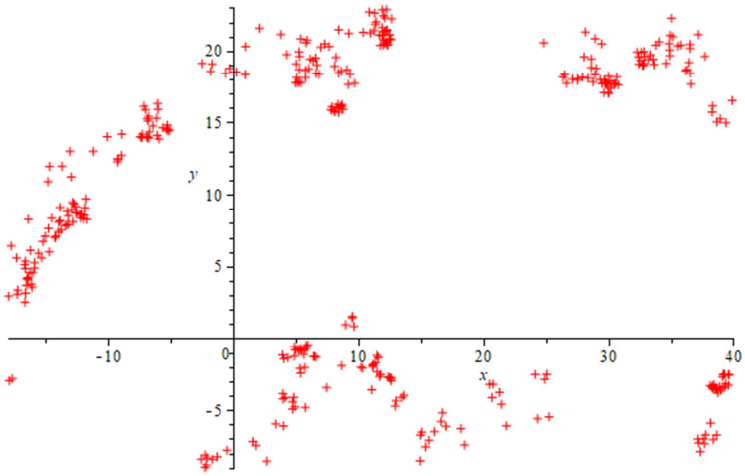
Training set 22,500 points.



Classification results

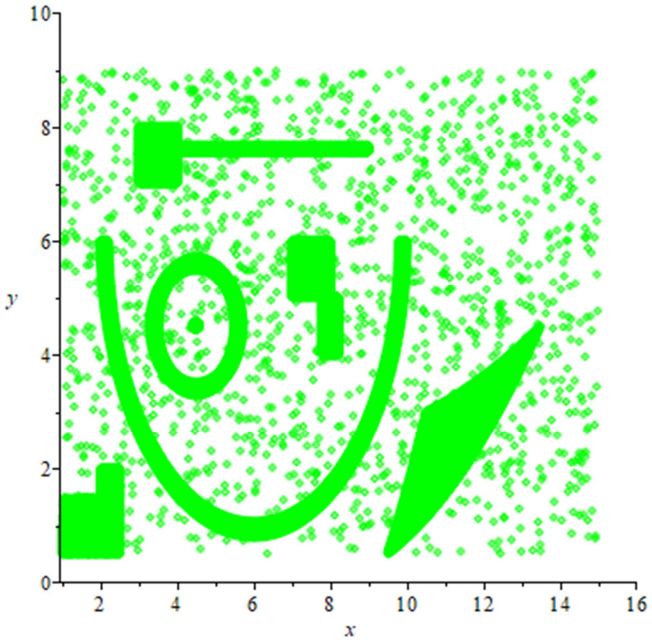
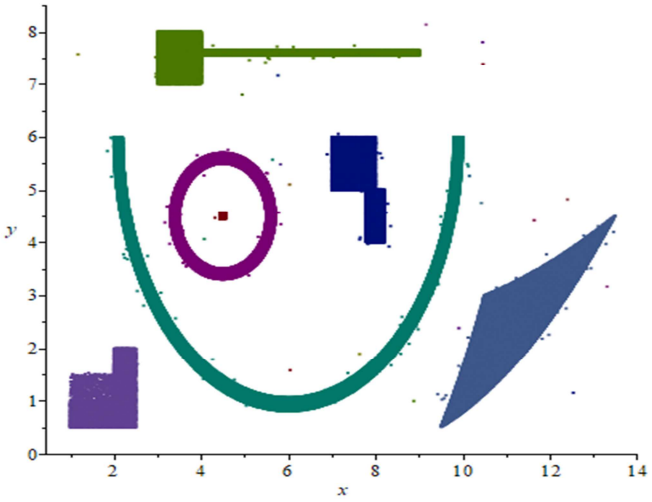
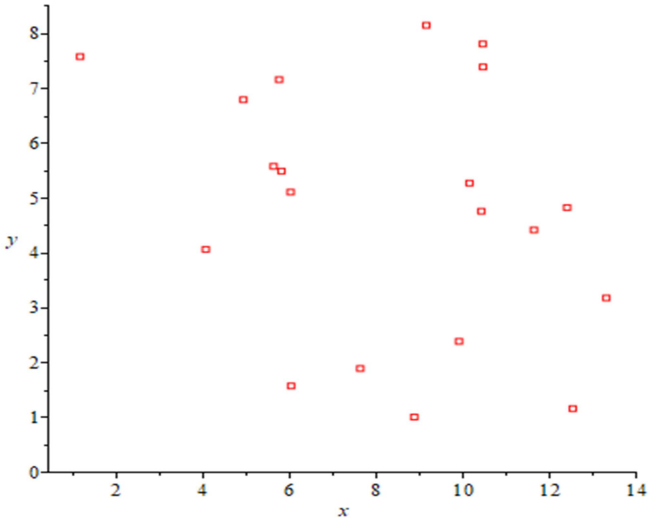


Points that left unclassified by the algorithm 881 points ≈ 0.18%.

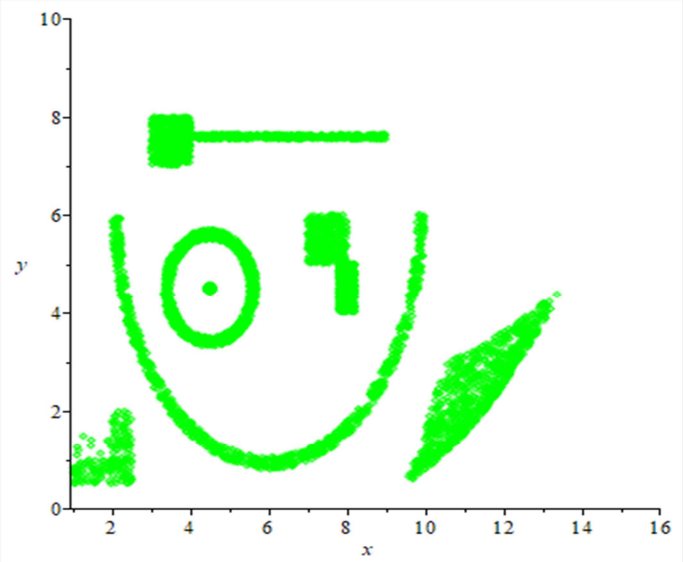


4.2. Test Results with Set of Points "Shapes"

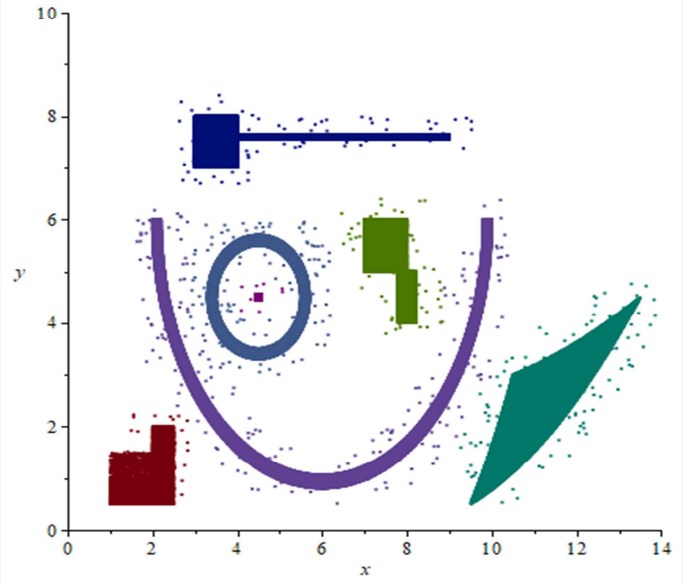
Table 2. *Test "Shapes".*

Initial set of points "Figures"	
Clustering results. The clustering algorithm identified seven clusters and attached the nearest points from the set of outlier points to individual clusters. In total, 448866 points were assigned to clusters. Clustering error was ≈ 0.25% The grid of 150x150 cells was used $Nmean$ =20. $Nmin$ =2	
False clusters of small volume consisting of points of a set of outliers.	

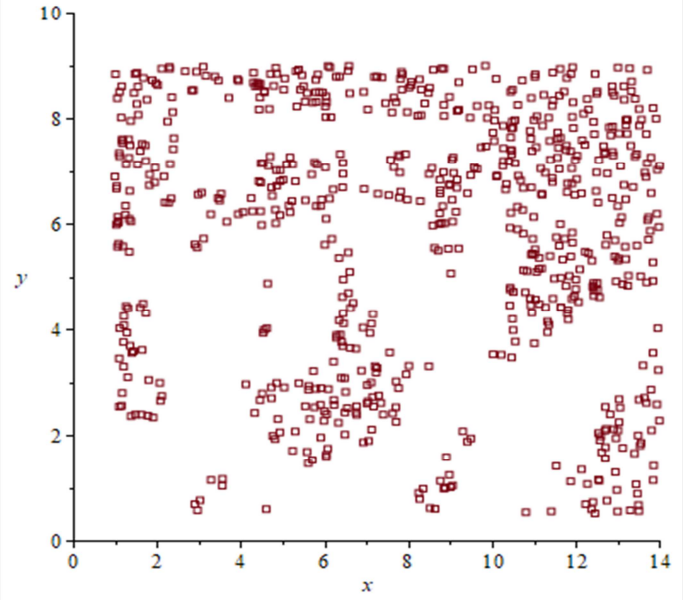
The training set consists of 4500 points



Classification results.
The algorithm correctly identified 7 classes by attaching the nearest points from the set of outliers to each cluster



The points that the algorithm left unclassified are 816 points out of 450,000.



5. Conclusions

The modern development of space technologies is characterized by a significant increase in the number and quality of remote sensing satellites, including those with high spatial resolution. To date, dozens of countries in the world already have national space facilities for remote sensing of the earth's surface and corresponding centers for processing satellite images. This allows researchers from different countries to study the Earth's resources in more detail in order to preserve the climate, water bodies, forests, develop cities, search for minerals, and increase the efficiency of agricultural production. Such studies are associated with the need for regular automated processing of very significant volumes of satellite information. Such processing requires the use of not only high-performance computing systems, but also efficient satellite image analysis algorithms, in particular, fast classification and clustering algorithms with a minimum set of adjustable parameters. Despite the existence of a large number of such algorithms, the personal practical experience of the authors of satellite image processing convinced us of the need to conduct research in the direction of developing cluster analysis algorithms that can efficiently process large amounts of information. The article proposes new approaches to the construction of clustering and classification algorithms applicable to the analysis of large data sets received from remote sensing satellites. The algorithms have been tested and demonstrated their effectiveness. We believe that they can be used as basic analysis tools in existing GIS systems.

Further research involves the use of the proposed algorithms for solving a number of applied problems, in particular, for the analysis of forest plantations by remote sensing on a national scale.

References

[1] A. Kuzmin, L. Grekov, O. Petrov, O. Medvedenko (2017) Computational procedures for thematic processing of satellite images in the interest of monitoring agricultural resources (part 1). Environmental safety and nature management, № 1-2 (23), 70-78.

[2] A. Kuzmin, L. Grekov, N. Kuzmina, O. Petrov, O. Medvedenko (2020) Computational procedures for thematic processing of satellite images in the interest of monitoring agricultural resources (part 2). Environmental safety and nature management, № 1 (33), 87–93. https://doi.org/10.32347/2411-4049.2020.1.87-94.

[3] Tou J. T, Gonzalez R. C. Pattern recognition principles. Boston, MA, USA: Addison-Wesley Publ. Company, 1974. 395 p.

[4] ArcGIS Desktop https://desktop.arcgis.com/ru/arcmap/10.3/main/get-started/arcgis-tutorials.htm.

[5] ERDAS ER Mapper. Hexagon Geospatial. https://www.hexagongeospatial.com/brochure-pages/erdas-ermapper-professional-benefit-brochure.

[6] ENVI— Environment for Visualizing Images. Harris Geospatial Solutions https://www.l3harrisgeospatial.com/docs/using_envi_home.html.

[7] N. Abramov, D. Makarov, A. Talalayev, V. Fralenko Modern methods of intellectual data processing of remote sensing data. Software systems: Theory and applications vol. 9, № 4 (39), p. 417-442.

[8] Scheinberg K. An efficient implementation of an active set method for svms // J. Mach. Learn. Res. — 2006. — Vol. 7. — Pp. 2237-2257.

[9] Yizong Cheng Mean Shift, Mode Seeking, and Clustering // IEEE Transactions on Pattern Analysis and Machine Intelligence. — IEEE, 1995. — August (vol. 17, rel. 8).— doi: 10.1109/34.400568.

[10] Ester, Martin; Kriegel, Hans-Peter; Sander, Jörg; Xu, Xiaowei (1996). У Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M. A density-based algorithm for discovering clusters in large spatial databases with noise Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. p. 226 –231. ISBN 1-57735-004-9.

[11] Ester, M. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise / M. Ester, H.-P. Kriegel, J. Sander, X. Xu // In Proc. ACM SIGMOD Int. Conf. on Management of Data, Portland, OR, 1996. –P. 226-231.

[12] Sarmah S., Bhattacharyya D. K. (2012) A grid-density based technique for finding clusters in satellite image. Pattern Recognition Letters, V. 33, 589-604.

[13] I. Pestunov, Y. Siniavsky (2012) Clustering Algorithm in Satellite Image Segmentation Problems. Bulletin of the Kemerovo State University №4 (52) vol. 2, 110-125.

[14] I. Pestunov, S. Rylov. «A Method for Constructing an Ensemble of Grid Hierarchical Clustering Algorithms for Satellite Image Segmentation», Regional problems of remote sensing of the Earth, Materials of the international scientific conference, Siberian Federal University, Krasnoyarsk, 2014, p. 215-223.

[15] Y. Kulikova, I. Pestunov, Y. Siniavsky. «Nonparametric Clustering Algorithm for Processing Large Data Arrays», Proceedings of the 14th scientific conference "Mathematical methods for pattern recognition", MAKS Press, M., 2009, p. 149-152.

[16] Agrawal, R. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications / R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan // In Proc. ACM SIGMOD Int. Conf. on Management of Data, Seattle, Washington, 1998. -P. 94-105.

[17] Nagesh, H. MAFIA: Efficient and Scalable Subspace Clustering for Very Large Data Sets / H. Nagesh, S. Goil, A. Choudhary // Technical Report Number CPDC-TR-9906-019, Center for Parallel and Distributed Computing, Northwestern University, 1999. 20 p.

[18] Rongjun Qin, Tao Liu A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability, Remote Sens. 2022, 14, 646. https://doi.org/10.3390/rs14030646