UDC 004.932
DOI: 10.15587/1729-4061.2023.282374

# DETERMINATION OF THE NUMBER OF CLUSTERS ON IMAGES FROM SPACE OPTIC-ELECTRONIC OBSERVATION SYSTEMS USING THE K-MEANS ALGORITHM

**Hennadii Khudov**
*Corresponding author*
Doctor of Technical Sciences, Professor, Head of Department*
E-mail: 2345kh_hg@ukr.net
**Oleksandr Makoveichuk**
Doctor of Technical Sciences, Associate Professor
Department of Computer Sciences and Software Engineering
Academician Yuriy Bugay International Scientific and Technical University
Mahnitohorskyi lane, 3, Kyiv, Ukraine, 02094
**Volodymyr Komarov**
Doctor of Military Sciences, Professor, Head of Research Department
Scientific-Research Institute of Military Intelligence
Yuriia Illienka str., 81, Kyiv, Ukraine, 04050
**Vladyslav Khudov**
PhD, Junior Researcher
Department of Information Technology Security
Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166
**Irina Khizhnyak**
PhD*
**Volodymyr Bashynskyi**
Doctor of Technical Sciences, Chief Researcher
The Central Research Institute of the Armed Forces of Ukraine
Povitroflotskyi ave., 28B, Kyiv, Ukraine, 03049
**Stanislav Stetsiv**
Assistant Professor
Department of Missile Forces
Hetman Petro Sahaidachnyi National Army Academy
Heroiv Maidanu str., 32, Lviv, Ukraine, 61023
**Yevhen Dudar**
Deputy Head of Department
Department of Troop Training**
**Andrii Rudiy**
Senior Lecturer
Department of Armoured Vehicles**
**Mykhailo Buhera**
PhD, Senior Researcher
Scientific and Organizational Department
Central Scientifically-Research Institute of Armaments and Military Equipment of the Armed Forces of Ukraine
Povitrofloski ave., 28, Kyiv, Ukraine, 03049
*Department of Radar Troops Tactic
Ivan Kozhedub Kharkiv National Air Force University
Sumska str., 77/79, Kharkiv, Ukraine, 61023
**Hetman Petro Sahaidachnyi National Army Academy
Heroiv Maidanu str., 32, Lviv, Ukraine, 61023

***The object of research is the process of clustering images from space optical-electronic surveillance systems. The main hypothesis of the study assumed that experimental studies would make it possible to determine the number of clusters on images from space optical-electronic surveillance systems when using the k-means algorithm.***

***The method of clustering images from space optical-electronic surveillance systems using the k-means algorithm, unlike the known ones, implies:***

***– splitting the source image into Red-Green-Blue brightness channels;***

***– determination of the Euclidean distance between pixels;***

***– distribution of the entire set of image pixels into clusters;***

***– recalculation of "centers" of each subset;***

***– reassignment of new "centers" of each cluster;***

***– minimization of the total intracluster variance.***

***Experimental studies were conducted on the clustering of the original image using the k-means method at different values of k. It was established that with an increase in the value of k, the visual quality of clustering improves, and it is possible to visually determine a larger number of clusters in the images.***

***To determine the number of clusters, the sum of clustering errors of type 1 and 2 at different values of k was evaluated. It was established that when the value of k increases, the sum of errors of the 1st and 2nd kind initially decreases exponentially. A further increase in the value of k does not lead to a significant decrease in errors of the 1st and 2nd kind. It was established that for a typical image from the space optical-electronic observation system, the value of k in the clustering method based on the k-means algorithm should be equal to 4. At the same time, the sum of errors of the 1st and 2nd kind is 31.3 %.***

***Further research is directed to the development of clustering methods that reduce the sum of errors of the 1st and 2nd kind***

***Keywords: image clustering, space observation system, k-means, errors of the 1st and 2nd kind, number of clusters***

*Received date 30.03.2023*
*Accepted date 16.06.2023*
*Published date 30.06.2023*

***How to Cite:*** *Khudov, H., Makoveichuk, O., Komarov, V., Khudov, V., Khizhnyak, I., Bashynskyi, V., Stetsiv, S., Dudar, Y., Rudiy, A., Buhera, M. (2023). Determination of the number of clusters on images from space optic-electronic observation systems using the k-means algorithm. Eastern-European Journal of Enterprise Technologies, 3 (9 (123)), 60–69. doi: https://doi.org/10.15587/1729-4061.2023.282374*

## 1. Introduction

The most widespread method of optical-electronic image clustering is the method based on the *k*-means algorithm. This statement also applies to images from space optical-electronic surveillance systems [1]. The advantages of the clustering method based on the *k*-means algorithm are simplicity, flexibility, fast convergence, and the ability to check the distinction between selected clusters. Usually, the number of clusters k is two [1]. This statement is justified when it comes to image clustering into background objects and objects of interest [2]. In this case, the task of clustering images from space optical-electronic surveillance systems is transformed into the task of segmentation into two segments [3].

When solving tasks in the interests of agriculture during clustering using the method based on the *k*-means algorithm, the number of clusters is an a priori known value. The value of *k* is chosen from 2 to 4 depending on the problem being solved.

When clustering images from space optical-electronic observation systems, the class of objects is much wider. Known methods of segmentation of such images assume that the value of k should be chosen equal to two. This is due to the need to solve the task of dividing the source image into two segments – background objects and objects of interest (for example, objects of military equipment).

The experience of modern armed conflicts shows the variety of objects of interest in images from space optical-electronic surveillance systems [4]. Such images contain a large number of distinguishing elements, objects of interest have a complex morphological structure and are heterogeneous [5]. This greatly complicates the clustering of images using the method based on the *k*-means algorithm. Therefore, determining the number of clusters in images from space optical-electronic surveillance systems using the *k*-means algorithm is an urgent task.

## 2. Literature review and problem statement

The general classical approach to clustering and examples of data clustering are discussed in [6]. It was established that the classical approach works well when performing unsupervised classification of patterns by groups. In the presence of differences, data clustering is already slow, that is, with greater time costs. The disadvantage of [6] is that it does not take into account the features of image clustering since such a task cannot be attributed to the uncontrolled classification of patterns by groups.

A hierarchical approach to semantic clustering of images is proposed in [7]. The essence of the hierarchical clustering algorithm in machine learning [7] is that a representative image is selected to denote a cluster at each intermediate stage. The disadvantage of [7] is the loss of information because the representative image denotes any other one that belongs to the cluster. In order to automatically obtain the total number of clusters, such loss of information is monitored, which is an advantage of [7].

The application of k-means and C-means clustering for image segmentation is considered in [8]. The advantages of these methods in their classical form are simplicity, comprehensibility, and flexibility in use, fast convergence during the operation of the algorithm, and the possibility of checking the statistical significance of the differences between the segmented clusters. The main disadvantages of [8] are the dependence of the clustering result on the initial selection of cluster centers and the slowing down of the algorithm when clustering images into a large number of clusters.

Image segmentation using subtractive clustering is proposed in [9]. The essence of the proposed data clustering method [9] is that each pixel in the image can be the center of a potential cluster, and the center of the cluster, in turn, is generated based on the potential value of the pixel. The advantage of [9] is the application after the k-means algorithm to the resulting image of a median filter in order to remove the "unwanted" area on this image. The disadvantage of [9] is the presence of a large number of "unwanted" regions after subtractive clustering.

Combining the K-means clustering method and the fuzzy K-means clustering algorithm was proposed in [10]. This complex method is called the K-C-means fuzzy clustering algorithm and has shown good results in the segmentation of medical images. The advantages of [10] are the speed of segmentation of medical magnetic resonance images and ease of use. Disadvantages of [10] are the need to determine the number of clusters before starting the algorithm and segmentation of only simple medical images presented in shades of gray.

Clustering of color images is proposed using an improved hybrid method of K-means and fuzzy C-means algorithms in [11]. The advantage of [11] is the ability to quickly cluster not only medical color images. The main disadvantage of [11] is the clustering of simple images into a limited number of clusters.

In [12], a method of unsupervised segmentation by incremental clustering was proposed to eliminate the shortcoming of the dependence of clustering results on the initial selection of cluster centers. The essence of [12] is an incremental approach and correlation clustering with a gradual increase in the number of clusters and constant refinement during segmentation. The advantages of [12] are the lack of determination of the initial number of clusters and the avoidance of repeated clustering of the entire image when the number of clusters increases. The main disadvantage of [12] is computational complexity.

In [13], a method of segmentation of radar images with a synthesized aperture is proposed, which is based on the k-means clustering method and the Otsu thresholding method. Additionally, in [13], at the final stage, morphological operations are performed in order to obtain more accurate segmentation results. The advantage of [13] is the good results of image segmentation with existing speckle noise. This is possible due to the use of the median filter in the method [13]. The main drawback of [13] is the segmentation of only the radar images of the radar.

For the segmentation of color images, work [14] proposed an adaptive method of k-means clustering, in which it is not necessary to enter the number of clusters beforehand. The essence of the method [14] is to transform the color models of the input image representation into the LAB color model. After that, k-means clustering with further morphological processing is performed. The advantages of [14] are the reduction of the influence of light on the segmentation results, the segmentation of color images of different color representation models, and the adaptive selection of the number of clusters. The disadvantage of [14] is the impossibility of universal clustering of both tone and color images and the need to represent color images in the LAB color model.

In [15], it is proposed to perform sequential k-median clustering without replacement in order to solve the issue of determining cluster centroids. The essence of [15] is the preliminary selection of a set of cluster centers with a good k-median value from the distribution that generated the sequence. The advantage of [15] is the preliminary determination of the centroids and the stability of the algorithm to anomalous values of the input data due to the use of the median. The disadvantage of [15] is the impossibility of replacing the values of the selected set of cluster centers during the operation of the algorithm.

In work [16] it is proposed to segment images using the k-means algorithm using the Mumford-Shah model. The essence of the approach [16] consists in the application of the k-means algorithm for the purpose of color quantization in color models of image representation. However, the pixels are grouped by clusters only in the color space without taking into account the connectivity between them, which, in turn, leads to the fragmentation of segments. It is to solve this problem that the Mumford-Shah model is proposed in [16]. The advantage of [16] is the optimization of both the shape of the segments and their content. The main drawback is the somewhat slow and complex mathematical optimization algorithm.

In [17], the concept of G-image was introduced to define such images that are represented in the domains of an irregular graph. In order to segment images of this type of representation, the use of C-means with a spatial information restriction in the wavelet space is proposed in [17]. The advantage of [17] is the preservation of similarity taking into account the spatial information of the image, which is possible due to performing actions in the wavelet space, not in the Euclidean space. The disadvantage of [17] is the good results of segmentation on synthetic G-images, the results of the operation of this approach on real images are not yet available.

In recent decades, artificial neural networks have been actively used both for solving the problem of classification and for the problem of image clustering. In [18], a general approach to image clustering based on an artificial neural network is considered. The advantages of this approach to image clustering are a parallel and distributed architecture of data processing, the possibility of adaptive study of own weights of relationships, presentation of the problem using only quantitative features, etc. The disadvantage of [18] is the impossibility of using simple one-level artificial neural networks for segmentation of complex images.

In [19], it is proposed to use an end-to-end unsupervised clustering neural network for image segmentation. The essence of [19] is the development of a separate clustering module consisting of a maximum unifying layer and a Gaussian block containing two sets of training parameters. The advantage of [19] is the possibility of simultaneous training of a separate clustering module with other layers of the neural network, adaptability to the situation when the number of clusters is not known in advance. The disadvantage of [19] is the need to use an autoencoder in order to obtain the hidden semantics of input features.

Paper [20] proposed the use of several pre-trained CNN convolutional neural networks to improve clustering results. A necessary condition of the approach [20] is training on the same data set. The advantage of [20] is good clustering results of natural images. The main drawback of [20] is the need to retrain all involved convolutional neural networks at once.

The use of an algorithm that allows performing adaptive clustering – the growing neural gas algorithm - was proposed in [21]. The advantage of using [21] for image clustering is an accurate description of the topological relationships between the clusters selected during the operation of the algorithm, and thus the problem of distributing all uncertain data between clusters is well solved. The disadvantage of [21] is the increase in computing costs with the growth of data for clustering.

In [22], the integration of a self-organized neural network of feature mapping and the spectral clustering algorithm is proposed for the clustering of face images. At the first stage of automatic feature selection, a convolutional neural network is used, and then the clustering stage is performed with the sequential application of a self-organized feature mapping neural network and the spectral clustering algorithm. The main drawback of [22] is the multi-step procedure of the method with sequential execution of them one after the other and, as a result, large time costs. The advantage is taking into account all the disadvantages and advantages of the approaches used in the integrated method [22].

For fast clustering of images, the use of a convolutional neural network and the binary k-means algorithm is proposed in [23]. The advantage of [23] is a quick and efficient search for the nearest initial cluster center. The disadvantage of [23] is the use of additional volumes of RAM for storing hash tables and their constant access and updating.

Paper [24] provides an overview of evolutionary algorithms, which have recently been popular and have shown good results in solving complex optimization problems; the possibility of their application to solving the problem of image clustering is considered. The advantage of using evolutionary algorithms for image clustering is the speed of finding the best values of the parameters of the algorithm, the possibility of integrating evolutionary algorithms with other clustering methods, etc.

A multi-objective evolutionary image clustering algorithm was proposed in [25] for the purpose of finding optimal cluster centers. The disadvantage of [25] is the dependence of its work efficiency on the choice of the objective function, namely, either maximization of the intercluster distance or minimization of intraclass compactness. The advantage of [25] is the generation during the operation of the algorithm of a set of "non-dominant" solutions, which are no longer considered.

In work [26], the use of a genetic algorithm is proposed for the clustering of space images. The advantage of [26] is the possibility of clustering masked objects, namely, when the color characteristics of the object are visually close to the characteristics of the background. The disadvantage of [26] is the need to carry out a clustering operation in each of the color channels of the color space of the image representation.

For the segmentation of remote sensing images of the earth in [27], a complex clustering approach using the k-means algorithm and the genetic algorithm is proposed. The advantage of [27] is the good results of clustering space images into a small number of clusters, however, this approach is not acceptable for clustering images from unmanned aerial vehicles.

Paper [28] proposed the use of the swarm intelligence algorithm, namely the particle swarm algorithm, for the segmentation of remote sensing images of the earth. The segmentation results made it possible to classify this method as a clustering method. The advantage of [28] is the reduction of segmentation errors of the 1$^{st}$ and 2$^{nd}$ kind in comparison with the classical method of k-means. The main drawback [28] is the need to perform clustering by the particle swarm algorithm in parallel in each brightness channel for a color image.

In [29], a particle swarm algorithm for clustering space images with complex structures is proposed. The advantages of [29] are the ability to effectively divide the image into a larger number of clusters compared to the k-means algorithm and to separate objects from the background even in the presence of masking. The disadvantage of [29] is increased resource costs in comparison with the k-means algorithm.

For the clustering of space images, a method based on the sine-cosine algorithm is proposed in [30]. The essence of the approach [30] is to determine the threshold level of segmentation using the particle swarm algorithm using the trigonometric functions of sine and cosine. The advantage of [30] is the possibility of highlighting uncertain areas and masked objects in the image. The disadvantage of [30] is the need to calculate the sine and cosine functions in each of the brightness channels for a color image.

In [31], the possibility of using another method of swarm intelligence - the method of an artificial bee colony - was considered. This method has shown good results in identifying the boundaries of objects of interest, but for performing clustering it is either unacceptable in the classical representation or requires complex application with other methods of image processing.

Thus, known methods of clustering based on the k-means algorithm use values of k from 2 to 4, depending on the task being solved. When segmenting images from space optical-electronic surveillance systems, the value k is chosen, as a rule, to be equal to two. This is due to the need to solve the task of dividing the source image into only two segments (background objects and objects of interest). Therefore, determining the number of clusters in images from space optical-electronic observation systems using the k-means algorithm is an urgent task.

## 3. The aim and objectives of the study

The purpose of this study is to experimentally determine the number of clusters on images from space optical-electronic surveillance systems using the k-means algorithm. This will make it possible to reduce the sum of type 1 and type 2 clustering errors.

To achieve the goal, it is necessary to solve the following tasks:

– to determine the main stages of the clustering method of images from space optical-electronic surveillance systems using the k-means algorithm;

– to perform clustering of the image from the space optical-electronic observation system at different values of $k$;

– to determine the sum of type 1 and type 2 clustering errors and determine the number of clusters in the image when using the k-means algorithm.

## 4. The study materials and methods

The object of our research is the process of clustering images from space optical-electronic surveillance systems.

The main hypothesis of the study assumed that experimental studies would make it possible to determine the number of clusters on images from space optical-electronic surveillance systems when using the k-means algorithm.

The following research methods were used during the research: mathematical apparatus of matrix theory, methods of digital image processing, methods of probability theory and mathematical statistics, clustering methods, iterative methods, methods of system analysis, methods of mathematical modeling.

The following limitations and assumptions were adopted during the research:

– only images from space surveillance systems are considered;

– the image is optical-electronic;

– the optical-electronic image is represented in the Red-Green-Blue (RGB) color space;

– only the method based on the k-means algorithm is used for clustering, other clustering methods are not considered;

– the sum of errors of the 1st and 2nd kind is selected as an indicator of the quality of clustering, other indicators of the quality of clustering are not considered;

– determination of the number of clusters is carried out by an experimental method;

– a typical image from the World View 2 space optical-electronic observation system was selected during the experimental research;

– there are no distorting factors.

## 5. Research results on determining the number of clusters when using the k-means algorithm

### 5. 1. The main stages of the image clustering method based on the k-means algorithm

The clustering of the original optical-electronic image $f(\mathbf{X})$ is the mapping (1):

$$f(\mathbf{X})\rightarrow f_c(\mathbf{X}), \tag{1}$$

where $f(\mathbf{X})$ is the starting optical-electronic image;

$f_c(\mathbf{X})$ is the resulting clustered image.

Clustering of the original optical-electronic image $f(\mathbf{X})$ in accordance with expression (1) involves dividing the entire set of objects (pixels) of the image into relatively homogeneous non-intersecting subsets (clusters). The distribution is carried out according to the feature of the object (pixel). The selection of objects (pixels) in the space of features according to the selected comparison criterion (metric) is carried out relative to the value of the specified object (pixel) (the so-called "center") of the subset (cluster).

The color intensity value is used as an object (pixel) feature. For a tonal image, this is the value of the color intensity in one-dimensional space; for RGB images – in three-dimensional space. The metric is the "distance" from each of the "centers" of the subset (cluster) to the object (pixel) of the image, followed by its assignment to the subset (cluster) whose distance to the "center" is minimal.

So, the result of clustering is function (2):

$$f(\mathbf{X})\rightarrow \mathbf{Y}, \tag{2}$$

where $\mathbf{Y}=\{1,2,...K\}$ is the set of "centers" (numbers) of subsets (clusters).

Function (2) matches any object $\boldsymbol{x_i}\in\mathbf{X}$ with only one "center" (number) of the cluster $y_i\in\mathbf{Y}$. At the same time, each object (pixel) $x_i\in\mathbf{X}$ is given the number of subsets (clusters) $y_i\in\mathbf{Y}$.

So, the mathematical formulation of the problem of optical-electronic image clustering in general takes the form (3):

$$\begin{cases} f(\mathbf{X})\rightarrow \mathbf{Y}, \text{ for } \mathbf{Y}=\{1,2,...,\mathrm{K}\}; \\ f(\mathbf{X})=\bigcup\limits_{k=1}^{K} fc_i(\mathbf{X}); \\ fc_i(\mathbf{X})\cap fc_{+1}(\mathbf{X})=\varnothing; \\ \forall x \exists y_k,\ k=\overline{1,K}; \\ x_i\in fc_K(\mathbf{X}),\ \text{if } d(x_i,y_k)\approx y_k,\ k=\overline{1,K}, \end{cases} \tag{3}$$

where $fc(\mathbf{X})=\{fc_1(\mathbf{X}), fc_2(\mathbf{X}),..., fc_K(\mathbf{X})\}$ are the clusters in the image $f_c(\mathbf{X})$;

$d(x_i,y_k)$ is a function of the distance between objects (pixels) $x_i$ and $y_k$;

$y_k$ – "centers" of subsets (clusters);

$I$ is the number of objects (pixels) on $f_c(\mathbf{X})$, $k=(1, 2, ..., K)$;

$K$ is the number of subsets (clusters) on $f_c(\mathbf{X})$, $k=(1, 2, ..., K)$.

The result of the clustering of optical-electronic images is the division of the image into subsets (clusters), the main requirement of which is proximity with respect to the selected metric within the subset and a significant difference with respect to the metric from different subsets. At the same time, the number of subsets (clusters) can be defined, and their "centers" (numbers) - undefined at the beginning of clustering. Any of the available options for the comparison criterion can be chosen as the metric and is determined before starting the method.

The main stages of the method of clustering images from space optical-electronic surveillance systems using the k-means algorithm are shown in Fig. 1.

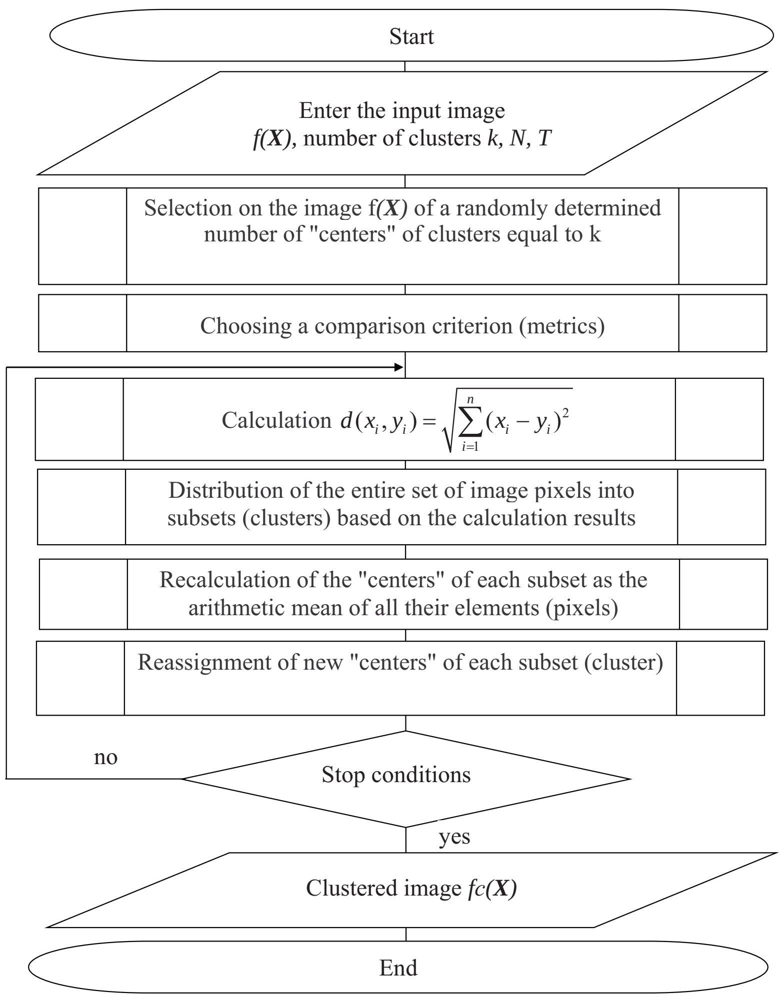


Fig. 1. The main stages of the method of clustering images from space optical-electronic surveillance systems using the k-means algorithm

The method of clustering images from space optical-electronic surveillance systems using the k-means algorithm involves the following:

1. Input of initial data:

– a digital image for clustering – $f(\mathbf{X})$, which is a matrix of elements (pixels), where each element is represented by a vector of characteristics (4):

$$f(\mathbf{X})=\{x_1, x_2, ..., x_i\}, \tag{4}$$

where $i$ is the dimension of the image characteristics space.

Each element (pixel) $x$ for clustering is represented by a vector of characteristics (5):

$$O=\{c_1, c_2, ..., c_n\}, \tag{5}$$

where $n$ is the dimensionally of the element (pixel) characteristics space.

Under the condition of tonal image clustering, the vector of characteristics (5) consists of the location coordinates of the elements (pixels) and the value of their color intensity in one-dimensional space. Under the condition of clustering of the color RGB image, the values of the elements of color intensity in three-dimensional space are added to the vector of characteristics (5);

– number of clusters – $k$;

– the maximum number of iterations of the algorithm – $T$;

– the number of iterations during which the values of the "centers" of the clusters are unchanged – $M$.

2. Selection of a randomly determined number of "centers" of clusters equal to $k$ on the input image $f(\mathbf{X})$. Such a choice is random only at the beginning of the k-means algorithm.

3. Selection of comparison criteria (metrics).

4. Calculation of the comparison criterion (metrics).

Beginning of the iterative process of clustering. Euclidean distance is chosen as a comparison criterion (metric) for image clustering, which is the distance between objects (pixels) in the $n$-dimensional space of characteristics.

This comparison criterion is calculated according to expression (6):

$$d(x_i,y_i)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}, \tag{6}$$

where $x_i$ is the pixel color intensity value;

$y_i$ is the color intensity value of the "center" of the cluster.

5. Division of the entire set of image pixels into subsets (clusters) based on the results of calculation of expression (6) and under the condition of the minimum value of such a distance to the "center" of the cluster.

6. Recalculation of the "centers" of each subset (cluster) as the arithmetic mean of all their elements (pixels).

7. Reassignment of new "centers" of each subset (cluster).

Points 4–7 of the k-means algorithm are repeated until one of the conditions for stopping the algorithm is met:

– execution of the maximum number of iterations of the algorithm;

– the invariance of the value of the "centers" of all subsets (clusters) during a certain number of iterations of the algorithm.

Therefore, the main goal of the k-means algorithm is to minimize the total intracluster variance $D$ (expression (7)):

$$D=\sum_{k=1}^{K}\sum_{O_j\in y_k}\left(O_j-y_k\right)^2. \tag{7}$$

Given the representation of the original image in the RGB color space, the main stages of the method (Fig. 1) are applied to the Red, Green, and Blue brightness channels.

So, the method of clustering images from space optical-electronic surveillance systems when using the k-means algorithm involves:

– splitting the source image into RGB brightness channels;

– determination of the Euclidean distance between pixels;

– division of the entire set of image pixels into subsets (clusters);

– recalculation of "centers" of each subset;

– reassignment of new "centers" of each subset (cluster);

– minimization of the total intracluster variance.

**5. 2. Image clustering from the space optical-electronic surveillance system**

For clustering, the original optical-electronic image from the space observation system WorldView-2 (USA) (Fig. 2) [32] was chosen.

Fig. 2. Original color optical-electronic image [32]

Experimental research was conducted on the clustering of the original image (Fig. 2) using the method based on k-means at different values of $k$, which we shall change from 2 to 10. The results of clustering are shown in Fig. 3–11.

Fig. 3. Clustering of the original image using the method based on the k-means algorithm ($k$=2)

Analysis of Fig. 3–11 indicates different visual quality of clustered images. As the value of $k$ increases, the visual quality of clustering improves, and it is visually possible to identify more clusters in the images of Fig. 3–11.

Fig. 4. Clustering of the original image using the method based on the k-means algorithm ($k$=3)

Fig. 5. Clustering of the original image using the method based on the k-means algorithm ($k$=4)

Fig. 6. Clustering of the original image using the method based on the k-means algorithm ($k$=5)

Fig. 7. Clustering of the original image using the method based on the k-means algorithm ($k$=6)

Fig. 8. Clustering of the original image using the method based on the k-means algorithm ($k$=7)

Fig. 9. Clustering of the original image using the method based on the k-means algorithm ($k$=8)

Fig. 10. Clustering of the original image using the method based on the k-means algorithm ($k$=9)

Fig. 11. Clustering of the original image using the method based on the k-means algorithm ($k$=10)

**5. 3. Determination of the sum of clustering errors of type 1 and 2 and the number of clusters**

To determine the number of clusters, we shall evaluate the sum of clustering errors of type 1 and 2 at different values of $k$ in Fig. 3–11. It is known that errors of the 1st and 2nd kind are determined by the maximum likelihood criterion [33, 34]. Clustering errors of type 1 and 2 were determined by known expressions (8), (9) [34, 35]:

$$\alpha_1 = \frac{S_1^i\left(fc(\mathbf{X})\right)}{S_2\left(f(\mathbf{X})\right)}, \tag{8}$$

$$\beta_2 = 1 - \frac{S_3\left(fs(\mathbf{X})\right)}{S_4\left(f(\mathbf{X})\right)}, \tag{9}$$

where $S_1^i\left(fc(\mathbf{X})\right)$ is the defined plane of objects of the $i$-th cluster, which is mistakenly assigned to other clusters on the clustered image $fs(\mathbf{X})$; $S_2(f(\mathbf{X}))$ is the defined plane of objects of other clusters, except for the $i$-th one in the original image $f(\mathbf{X})$; $S_3(fc(\mathbf{X}))$ is the defined plane of correctly clustered objects on the image $fc(\mathbf{X})$; $S_4(f(\mathbf{X}))$ is the defined plane of the objects of the defined cluster on the original image $f(\mathbf{X})$.

Table 1 gives the calculations of clustering errors of type 1 and 2 and the sum of these errors at different values of $k$. Clustering errors of type 1 and 2 for each value of $k$ in Table 1 were calculated using expressions (8) and (9). At the same time, images corresponding to the corresponding values of $k$ were chosen as clustered images (Fig. 3–11). The image (Fig. 2) was chosen as the initial image at different values of $k$. The sum of errors was defined as the arithmetic sum of errors of the 1st and 2nd kind.

Table 1

Results of calculation of clustering errors of the 1st and 2nd kind

| Clustering error | Value of $k$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 kind, % | 37.5 | 27.3 | 15.2 | 14.7 | 14.4 | 14.0 | 13.7 | 13.5 | 13.0 |
| 2 kind, % | 34.4 | 26.5 | 16.1 | 15.7 | 15.3 | 15.0 | 14.6 | 14.3 | 13.8 |
| The sum of errors of the 1st and 2nd kind, % | 71.9 | 53.8 | 31.3 | 30.4 | 29.7 | 29.0 | 28.3 | 27.8 | 26.8 |

Fig. 12 shows the dependence of the sum of errors of the 1st and 2nd kind on the value of $k$.

Analysis of Fig. 12 reveals that when the value of k increases, the sum of errors of type 1 and 2 initially decreases exponentially ($k$=2, 3, 4). After the value of the number of clusters $k$=4, the fall in the sum of errors of the 1st and 2nd kind slows down. A further increase in the value of $k$ does not lead to a significant decrease in errors of the 1st and 2nd kind. So, from the analysis of Fig. 12, it can be concluded that for a typical image from the space optical-electronic observation system, the value of $k$ in the clustering method based on the k-means algorithm should be equal to 4. At the same time, the sum of errors of the 1st and 2nd kind is 31.3 %.

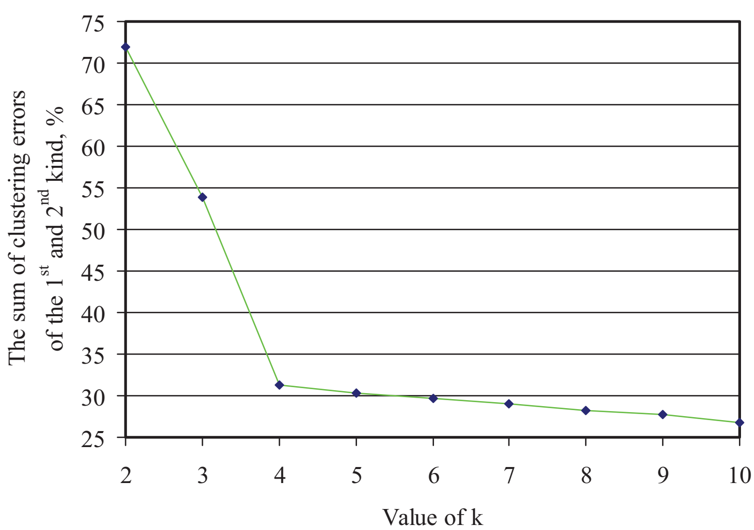


Fig. 12. The dependence of the sum of errors of the 1st and 2nd kind on the value of $k$

## 6. Discussion of results of determining the number of clusters using the k-means algorithm

A feature of the study is the determination of the number of clusters on images from space optical-electronic surveillance systems using the k-means algorithm. This determination of the number of clusters will allow, in contrast to, for example, [1, 9], clustering of the image with a reduced value of the sum of type 1 and type 2 clustering errors. Reducing the amount of clustering errors of type 1 and 2 becomes possible thanks to the selection of a certain value of the number of clusters when clustering an image from a space optical-electronic surveillance system when using the k-means algorithm.

The method of clustering images from space optical-electronic surveillance systems using the k-means algorithm (Fig. 1), unlike the known ones, involves:

– splitting the source image into RGB brightness channels;

– determination of the Euclidean distance between pixels;

– division of the entire set of image pixels into subsets (clusters);

– recalculation of "centers" of each subset;

– reassignment of new "centers" of each subset (cluster);

– minimization of the total intracluster variance.

Experimental research was conducted on the clustering of the original image (Fig. 2) by the method based on k-means at different values of $k$ ($k$=2–10). The results of clustering are shown in Fig. 3–11. Analysis of Fig. 3–11 shows the different visual quality of clustered images. As the value of $k$ increases, the visual quality of clustering improves, and a larger number of clusters can be visually determined in the images of Fig. 3–11.

To determine the number of clusters, the sum of clustering errors of type 1 and 2 at different values of $k$ was evaluated. It can be seen (Fig. 12) that when the value of $k$ increases, the sum of errors of the 1$^{st}$ and 2$^{nd}$ kind initially decreases exponentially ($k$=2, 3, 4). After the value of the number of clusters $k$=4, the fall in the sum of errors of the 1$^{st}$ and 2$^{nd}$ kind slows down. A further increase in the value of $k$ does not lead to a significant decrease in errors of the 1$^{st}$ and 2$^{nd}$ kind. So, from the analysis of Fig. 12, it can be concluded that for a typical image from the space optical-electronic observation system, the value of $k$ in the clustering method based on the k-means algorithm should be equal to 4. At the same time, the sum of errors of the 1$^{st}$ and 2$^{nd}$ kind is 31.3 %.

The uniqueness of the research is the determined number of clusters when clustering images from space optical-electronic surveillance systems using the method based on the k-means algorithm. This, in turn, led to a decrease in the sum of type 1 and type 2 clustering errors.

The following limitations and assumptions were adopted during the research:

– only images from space surveillance systems are considered;

– the image is optical-electronic;

– the optical-electronic image is represented in the RGB color space;

– only the method based on the k-means algorithm is used for clustering, other clustering methods are not considered;

– the sum of errors of the 1$^{st}$ and 2$^{nd}$ kind is selected as an indicator of the quality of clustering, other indicators of the quality of clustering are not considered;

determination of the number of clusters is carried out by an experimental method;

– a typical image from the World View 2 space optical-electronic observation system was selected during the experimental research;

there are no distorting factors.

The clustering method based on the k-means algorithm can be applied in the software and technical system of clustering images from space optical-electronic surveillance systems.

The disadvantage of the method is its use only for images from space optical-electronic surveillance systems.

Further research is aimed at the development of clustering methods that reduce the sum of errors of the 1$^{st}$ and 2$^{nd}$ kind.

## 7. Conclusions

1. The method of clustering images from space optical-electronic surveillance systems using the k-means algorithm involves:

– splitting the source image into RGB brightness channels;

– determination of the Euclidean distance between pixels;

– division of the entire set of image pixels into subsets (clusters);

– recalculation of "centers" of each subset;

– reassignment of new "centers" of each subset (cluster);

– minimization of the total intracluster variance.

2. Experimental studies on clustering of the original image by the method based on k-means at different values of $k$ were carried out. It was established that with an increase in the value of $k$, the visual quality of clustering improves, and it is possible to visually determine a larger number of clusters in the images.

3. To determine the number of clusters, the sum of clustering errors of type 1 and 2 at different values of $k$ was evaluated. It was established that when the value of $k$ increases, the sum of errors of type 1 and 2 initially decreases exponentially ($k$=2, 3, 4). A further increase in the value of $k$ does not lead to a significant decrease in errors of the 1$^{st}$ and 2$^{nd}$ kind. For a typical image from the space optical-electronic observation system, the value of $k$ in the clustering method based on the k-means algorithm should be equal to 4. At the same time, the sum of errors of the 1$^{st}$ and 2$^{nd}$ kind is 31.3 %.

## Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

## Funding

The study was conducted without financial support.

## Data availability

All data are available in the main text of the manuscript.

References


1. Green, M. (2020). K-Means Clustering for Surface Segmentation of Satellite Images. Available at: https://medium.com/@maxfieldeland/k-means-clustering-for-surface-segmentation-of-satellite-images-ad1902791ebf
2. Lafabregue, B., Gancarski, P., Weber, J., Forestier, G. (2022). Incremental constrained clustering with application to remote sensing images time series. 2022 IEEE International Conference on Data Mining Workshops (ICDMW). doi: https://doi.org/10.1109/icdmw58026.2022.00110
3. Lampert, T., Lafabregue, B., Dao, T.-B.-H., Serrette, N., Vrain, C., Gancarski, P. (2019). Constrained Distance-Based Clustering for Satellite Image Time-Series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12 (11), 4606–4621. doi: https://doi.org/10.1109/jstars.2019.2950406
4. Space, the unseen frontier in the war in Ukraine (2022). BBC News. Available at: https://www.bbc.com/news/technology-63109532
5. Khudov, H., Makoveichuk, O., Khizhnyak, I., Shamrai, B., Glukhov, S., Lunov, O. et al. (2022). The Method for Determining Informative Zones on Images from On-Board Surveillance Systems. International Journal of Emerging Technology and Advanced Engineering, 12 (8), 61–69. doi: https://doi.org/10.46338/ijetae0822_08
6. Samanta, S., Chatterjee, S. (2018). A Survey On Data Clustering Approaches. 1st International Business Research Conference (IBRC 2018), 34–42. Available at: https://www.researchgate.net/publication/341134327_A_Survey_On_Data_Clustering_Approaches
7. Pandey, S., Khanna, P. (2014). A hierarchical clustering approach for image datasets. 2014 9th International Conference on Industrial and Information Systems (ICIIS). doi: https://doi.org/10.1109/iciinfs.2014.7036504
8. Aktas, Y. C. (2021). Image Segmentation with Clustering. The Fundamentals of K-Means and Fuzzy-C Means Clustering and their usage for Image Segmentation. Towards Data Science. – 2021. Available at: https://towardsdatascience.com/image-segmentation-with-clustering-b4bbc98f2ee6
9. Dhanachandra, N., Manglem, K., Chanu, Y. J. (2015). Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. Procedia Computer Science, 54, 764–771. doi: https://doi.org/10.1016/j.procs.2015.06.090
10. Funmilola, A. A., Oke, O. A., Adedeji, T. O., Alade, O. M., Adewusi, E. A. (2012). Fuzzy k-c-means Clustering Algorithm for Medical Image Segmentation. Journal of Information Engineering and Applications, 2 (6), 21–33. Available at: https://core.ac.uk/download/pdf/234676965.pdf
11. Kishor Duggirala, R. (2020). Segmenting Images Using Hybridization of K-Means and Fuzzy C-Means Algorithms. Introduction to Data Science and Machine Learning. doi: https://doi.org/10.5772/intechopen.86374
12. NamAnh, D. (2015). Segmentation by Incremental Clustering. International Journal of Computer Applications, 111 (12), 23–30. doi: https://doi.org/10.5120/19591-1360
13. Niharika, E., Adeeba, H., Krishna, A. S. R., Yugander, P. (2017). K-means based noisy SAR image segmentation using median filtering and otsu method. 2017 International Conference on IoT and Application (ICIOT). doi: https://doi.org/10.1109/iciota.2017.8073630
14. Zheng, X., Lei, Q., Yao, R., Gong, Y., Yin, Q. (2018). Image segmentation based on adaptive K-means algorithm. EURASIP Journal on Image and Video Processing, 2018 (1). doi: https://doi.org/10.1186/s13640-018-0309-3
15. Hess, T., Sabato, S. (2020). Sequential no-Substitution k-Median-Clustering. 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), 108, 962–972. Available at: https://proceedings.mlr.press/v108/hess20a.html
16. Shah, N., Patel, D., Fränti, P. (2021). k-Means image segmentation using Mumford–Shah model. Journal of Electronic Imaging, 30 (06). doi: https://doi.org/10.1117/1.jei.30.6.063029
17. Wang, C., Pedrycz, W., Li, Z., Zhou, M., Ge, S. S. (2021). G-Image Segmentation: Similarity-Preserving Fuzzy C-Means With Spatial Information Constraint in Wavelet Space. IEEE Transactions on Fuzzy Systems, 29 (12), 3887–3898. doi: https://doi.org/10.1109/tfuzz.2020.3029285
18. Khosla, R. (2020). An Approach towards Neural Network based Image Clustering. Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2020/12/an-approach-towards-neural-network-based-image-clustering/
19. Li, H., Li, J., Zhu, M. (2023). End-to-end unsupervised clustering neural networks for image clustering. doi: https://doi.org/10.36227/techrxiv.22147559.v2
20. Guérin, J., Boots, B. (2018). Improving Image Clustering With Multiple Pretrained CNN Feature Extractors. Available at: https://homes.cs.washington.edu/~bboots/files/GuerinBMVC18.pdf
21. Benito-Picazo, J., Palomo, E. J., Dominguez, E., Ramos, A. D. (2020). Image Clustering Using a Growing Neural Gas with Forbidden Regions. 2020 International Joint Conference on Neural Networks (IJCNN). doi: https://doi.org/10.1109/ijcnn48605.2020.9207700
22. Zhang, L.-F., Li, C.-F., Wang, H.-R., Shi, M.-Y. (2018). Research On Face Image Clustering Based On Integrating Som And Spectral Clustering Algorithm. 2018 International Conference on Machine Learning and Cybernetics (ICMLC). doi: https://doi.org/10.1109/icmlc.2018.8526946
23. Ke, S., Zhao, Y., Li, B., Wu, Z., Liu, X. (2016). Fast image clustering based on convolutional neural network and binary K-means. Eighth International Conference on Digital Image Processing (ICDIP 2016). doi: https://doi.org/10.1117/12.2244263
24. Al-Qaisi, L., Hassonah, M. A., Al-Zoubi, M. M., Al-Zoubi, A. M. (2021). A Review of Evolutionary Data Clustering Algorithms for Image Segmentation. Algorithms for Intelligent Systems, 201–214. doi: https://doi.org/10.1007/978-981-33-4191-3_9

25. Abeysinghe, W., Wong, M., Hung, C.-C., Bechikh, S. (2019). Multi-Objective Evolutionary Algorithm for Image Segmentation. 2019 SoutheastCon. doi: https://doi.org/10.1109/southeastcon42311.2019.9020457
26. Khudov, H., Makoveichuk, O., Butko, I., Gyrenko, I., Stryhun, V., Bilous, O. et al. (2022). Devising a method for segmenting camouflaged military equipment on images from space surveillance systems using a genetic algorithm. Eastern-European Journal of Enterprise Technologies, 3 (9 (117)), 6–14. doi: https://doi.org/10.15587/1729-4061.2022.259759
27. Ruban, I., Khudov, H., Makoveichuk, O., Khudov, V., Kalimulin, T., Glukhov, S. et al. (2022). Methods of UAVs images segmentation based on k-means and a genetic algorithm. Eastern-European Journal of Enterprise Technologies, 4 (9 (118)), 30–40. doi: https://doi.org/10.15587/1729-4061.2022.263387
28. Ruban, I., Khudov, H., Makoveichuk, O., Butko, I., Glukhov, S., Khizhnyak, I. et al. (2022). Application of the Particle Swarm Algorithm to the Task of Image Segmentation for Remote Sensing of the Earth. Lecture Notes in Networks and Systems, 573–585. doi: https://doi.org/10.1007/978-981-19-5845-8_40
29. Khudov, H., Makoveichuk, O., Khizhnyak, I., Oleksenko, O., Khazhanets, Y., Solomonenko, Y. et al. (2022). Devising a method for segmenting complex structured images acquired from space observation systems based on the particle swarm algorithm. Eastern-European Journal of Enterprise Technologies, 2 (9 (116)), 6–13. doi: https://doi.org/10.15587/1729-4061.2022.255203
30. Khudov, H., Makoveichuk, O., Khudov, V., Maliuha, V., Andriienko, A., Tertyshnik, Y. et al. (2022). Devising a method for segmenting images acquired from space optical and electronic observation systems based on the Sine-Cosine algorithm. Eastern-European Journal of Enterprise Technologies, 5 (9 (119)), 17–24. doi: https://doi.org/10.15587/1729-4061.2022.265775
31. Ruban, I., Khudov, H., Makoveichuk, O., Khizhnyak, I., Khudov, V., Podlipaiev, V. et al. (2019). Segmentation of optical-electronic images from on-board systems of remote sensing of the earth by the artificial bee colony method. Eastern-European Journal of Enterprise Technologies, 2 (9 (98)), 37–45. doi: https://doi.org/10.15587/1729-4061.2019.161860
32. Satellite Imagery. Available at: https://www.maxar.com/products/satellite-imagery
33. Khudov, G. V. (2003). Features of optimization of two-alternative decisions by joint search and detection of objects. Problemy Upravleniya I Informatiki (Avtomatika), 5, 51–59. Available at: https://www.researchgate.net/publication/291431400_Features_of_optimization_of_two-alternative_decisions_by_joint_search_and_detection_of_objects
34. Khudov, H., Makoveichuk, O., Misiuk, D., Pievtsov, H., Khizhnyak, I., Solomonenko, Y. et al. (2022). Devising a method for processing the image of a vehicle's license plate when shooting with a smartphone camera. Eastern-European Journal of Enterprise Technologies, 1 (2 (115)), 6–21. doi: https://doi.org/10.15587/1729-4061.2022.252310
35. Khudov, H., Makoveichuk, O., Khizhnyak, I., Glukhov, S., Shamrai, N., Rudnichenko, S. et al. (2022). The Choice of Quality Indicator for the Image Segmentation Evaluation. International Journal of Emerging Technology and Advanced Engineering, 12 (10), 95–103. doi: https://doi.org/10.46338/ijetae1022_11