

# Applications Research on the Cluster Analysis in Image Segmentation Algorithm

Jiangyan Sun, Huiming Su, Yuan Zhou

School of Engineering, Xi'an International University, Xi'an, 710077, China

**Keywords:** Image segmentation algorithm, Cluster analysis, Fuzzy cluster.

**Abstract.** Clustering analysis is an unsupervised classification method, which classifies the samples by classifying the similarity. In the absence of prior knowledge, the image segmentation can be done by cluster analysis. This paper gives the concept of clustering analysis and two common cluster algorithms, which are FCM and K-Means clustering. The applications of the two clustering algorithms in image segmentation are explored in the paper to provide some references for the relevant researchers.

## Introduction

It is a structured analysis process for people to understand the image. People should decompose the image into organized regions with specific properties constitutes before the next step. To identify and analyze the target, it is necessary to separate them based on the process. It is possible to further deal with the target. Image segmentation is the process of segmenting an image into distinct regions and extracting regions of interest. The segmented regions have consistent attributes, such as grayscale, color and texture. Image segmentation plays an important role in image engineering, and it is the key step from image processing to image analysis. It relates to the target feature measurement and determines the target location; at the same time, the object hierarchy and extracting parameters of the target expression, feature measurement helps establish the image based on the structure of knowledge is an important basis for image understanding. Image segmentation has been widely used in real life, such as image retrieval, image coding and satellite image monitoring. If the data space of the clustering is the edge feature of the pixel, then the image segmentation method based on the clustering analysis and the edge based image segmentation method are combined. If the data space of the clustering is the regional feature of pixels, it can be considered that the image segmentation method based on the clustering analysis and the region based image segmentation method are combined. Therefore, clustering algorithm can not only be applied to image segmentation according to specific features, but also can be used in combination with other image segmentation methods. This shows that the clustering method applied to image segmentation has a very wide range of research and application fields. Compared with other segmentation methods, the image segmentation method based on clustering analysis has three main features. The image segmentation algorithms are supervised, efficient and adaptive.

## Concept of Cluster Analysis

The definition of the cluster is to divide the sample collection based on a similarity between samples to produce a collection. Each sample in a collection must belong to a class and belong to only one class. The division is based on the similarity of samples within the class, while the samples between classes are not similar. By introducing similarity measures between samples, it is shown that the similarity between the two sample points within a class is greater than that of any other sample within the class and in another class not within the class. Clustering can also be described in pairs of high-dimensional spaces, where the density distribution of uneven sample points is segmented, forming regions of high density points, continuous regions, and low-density points. The low-density

region divides the regions of the high-density points. Finally, each subset partition generated by collection should be non-empty, because a sample may not exist in air concentration. In clustering analysis, we usually classify some things of similar nature into one kind, and classify the things not of the same nature into different classes, and then classify them according to the similar degree of the nature. Two the better the sample is, the more similar they are. To realize the objective and practical classification, we must classify the samples according to the measurement of the relevant attribute variables. So, we need to describe the statistics of the degree of closeness of each variable or sample. Commonly used statistics are the distance between samples, the class of feature vectors, the distance between classes. An important step of data mining is data preparation, including the standardization, integration and pretreatment of the selected data and so on, this is the premise of data mining, but also is a prerequisite for the normal implementation of the clustering algorithm. To cluster data objects, the most important premise is to calculate the distance between the data objects based on statistical methods. Clustering algorithm is an unsupervised learning method, it does not need training samples, so many clustering methods have great advantages in solving image segmentation, and have been widely used.

## Two Common Algorithms of Cluster Analysis

**Fuzzy C-Means Algorithm.** Fuzzy C-Means clustering algorithm fuzzy c-means algorithm on the number of fuzzy clustering algorithm, Fuzzy C- Means (FCM) algorithm is the most widely used and successful. It obtains each sample point for all the class center membership, which determines the sample point is to achieve the automatic classification of sample data through the optimization of the objective function for. Due to the classification, there are categories of uncertainty, to describe the nature of Dunn, the fuzzy theory is introduced into K clustering algorithm. The criterion function for each sample is each type of distance with the membership of the weighted square. For a given sample set  $s$ ,  $K$  initial class mean elements are selected. The membership degree of each sample was calculated. Calculate the mean of each kind respectively. According to the current membership degree, each sample is classified into the class with the highest degree of membership. Fuzzy C-Means algorithm is compared with the value of each pixel space of each center point clustering, finally to each pixel is assigned a value that is closer to the center point of the pixel where clustering. Fuzzy clustering analysis is one of the main techniques of unsupervised machine learning method, is important to data analysis and modeling using fuzzy theory, established the sample category uncertainty description, can objectively reflect the real world, it has been effectively used in large-scale data analysis, data mining, vector quantization, image segmentation, pattern recognition, has important theoretical and practical value, with the further development of the application of fuzzy clustering algorithm, and constantly enrich.

**K-Means Algorithm.** The K-means method is proposed by Mac Queen in 1967, where each class is represented by the average value of the object in that class. It is a classical method to solve clustering problem. The main advantage is that the algorithm is simple and fast. However, this method may produce different clustering results for different  $K$  values and converge to local minima. Second, the standard function of the method is not suitable for classes with large differences, and is very sensitive to noise samples because a small amount of noisy data can have an enormous impact on average values. For a given sample set  $S$ , select the  $L$  of the initial class mean; according to the mean of all kinds of current, each sample is assigned to the nearest class mean class; according to the current class, computing the mean; if the criterion function of  $J$  relative to the previous iteration does not change, return the current otherwise go to step second division. K means clustering algorithm is first randomly selected  $K$  objects as the initial clustering center. Then calculate the distance between each object and each seed cluster center and assign each object to its nearest cluster center. The clustering centers and the objects assigned to them represent a cluster. Once all the objects have been assigned, the clustering centers of each cluster are recalculated according to the existing objects in the cluster. The process will be repeated until a termination condition is met. The termination condition can be that no object is redistributed to different clusters, and no clustering center changes again, with the sum of squared errors being locally minimum.

## Image Segmentation Algorithm Based on Cluster Analysis

**Image Segmentation Algorithm Based on FCM.** FCM algorithm is used for image segmentation, which is to blur the pixels of the image and calibrate each class of pixels, to realize the image segmentation. We take the pixels of image as the sample points of data sets. Then, the feature of pixels is regarded as the feature of sample points. Therefore, the segmentation problem of images is transformed into the following optimization problems:

$$\text{Minimize } J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^m (u_{ik})^m (d_{ik})^2$$

It meets the stated constraints:

$$\sum_{i=1}^c u_{ik} = 1, 1 \leq k \leq n;$$

$$u_{ik} \geq 0, 1 \leq k \leq n, 1 \leq i \leq n$$

The goal of image segmentation is to separate the target from the background. To make the fuzzy threshold description suitable for the target recognition task, the fuzzy threshold is described to blur. We use the maximum membership function degree method to blur. Image segmentation based on fuzzy C means clustering avoids the problem of setting region min value, and solves the problem of multi class segmentation which is difficult to solve by closed value segmentation. At the same time, FCM clustering does not need manual intervention in the iterative segmentation. Clustering iterative optimization is a mountain climbing method, which belongs to local search, and is easy to fall into local extremum. Simulated annealing has been used to achieve global optimization, although in some cases it can improve the optimization effect, but its computational capacity is enormous. The extreme value of the final convergence of the FCM algorithm is related to the selection of the initial values of the class center. Therefore, many improvement methods avoid local maxima by estimating good initial values. The potential function method is an efficient initialization method, but its computational complexity is an exponential function of the sample dimension.

In addition, density function method, morphological method and fuzzy measure method are used to estimate good initial values to avoid local extremum. For a monotonically convergent sequence, if the initial value of the sequence is close to the convergence, then the number of iterations decreases. At the same time, for non-monotonic sequences, the farther the initial value of the sequence is from the optimal convergence value, the greater the possibility of convergence to the local extremum; otherwise, the more easily it converges to the optimal clustering. The FCM algorithm needs to preset the initial class center, so the clustering effect of FCM algorithm is very sensitive to the initial value. Before clustering, you must set the number of categories, so that clustering can produce a division. However, in practical applications, especially in automation applications, the number of actual classes is not clear. The most difficult part of image segmentation is the determination of the appropriate number of classes. To avoid the default number of clusters, a heuristic method of stepwise optimization is used to determine the number of classes to achieve automatic segmentation. This method avoids the determination of class number to a certain extent, but it is not suitable to calculate one by one. Many class number determination methods can be regarded as exhaustive methods, but the exhaustive method is too large to be practical. In practical applications, the number of categories is artificially set according to prior knowledge and practical experience in specific fields.

**Image Segmentation Algorithm Based on K-Means Algorithm.** As the K-Means clustering method is easy to be affected by noise samples and does not make good use of the region features of the image, the particle noise in the image results in many isolated fragments. Median filtering can reduce noise and retain the edge information of the image. On the other hand, it allows each sample to represent a neighborhood after filtering. The median filter is defined as:

$$f(x, y) = \text{middle}(\sigma)$$

$$\sigma = \{im(i, j) \mid i \in [x - w, x + w]; j \in [y - w, y + w]\}$$

Among the above formulas, middle(.) is the median value of a sample in a collection.

To reduce the segmentation error caused by image noise, smoothing filtering can be used to reduce noise, but smoothing also loses detail information, such as edge or edge of image. This paper considers the use of median filtering to remove noise and uneven image brightness. The median filtering method can well preserve the edge information of the image while removing the noise of clustering algorithm is a prerequisite for iterative optimization with the distribution of variance in the samples of all types, and the optimization is based on having a set of correct initial classes on the average. If the above condition is not satisfied, the final segmentation results K-Means clustering will fall into local optimal and incorrect segmentation is proposed in this paper a screening method of the initial mean. This method is a split clustering method can filter accomplished in time to facilitate the establishment of a K-Means clustering segmentation conditions at the same time, with the probability distance as the basis for the re-division.

The flow of the image segmentation algorithm based on K-Means is as follows: image space transformation, color images using Lab space representation. The median filter is used to remove the independent noise in the image, and then Canny algorithm is used to mark the edges of the image. The edge pixels of sampled images are taken as samples, and the initial class mean of the sample set is estimated by the split clustering method. From the estimated initial class mean, the K-Means clustering is used to partition the sample set, and then the whole image is divided according to the convergent mean value. Each pixel is selected from the nearest class average. The edge of the segmented image is compared to the Canny algorithm, which is close to the edge.

## Conclusion

Image segmentation is a basic problem in computer vision, and it is the basis of image understanding. In the absence of prior knowledge, image segmentation is an unsupervised classification. Clustering analysis is an effective unsupervised classification method, so it is widely used in image segmentation. But the image samples are usually noisy, which makes the segmentation of the clustering algorithm slow and the segmentation unprecise. It is still a hot research topic to achieve fast and accurate image segmentation based on cluster analysis.

## Acknowledgements

This research was financially supported by the Key Research and Development Program (General Program) of Shaanxi Provincial science and technology Department(Program No. 2017GY-094), the Key Research and Development Program (General Program) of Shaanxi Provincial science and technology Department (Program No. 2017GY-062) and the Special Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 17JK1102).

## References

- [1] Cui Xixi, Wu Chengmao. Neutrosophic C- means clustering in kernel space and its application in image segmentation [J]. Journal of Image and Graphics, 2016, 21(10): 1316-1327.
- [2] Zhou Tingting, Zhang Hongbo, Pan Lijing. Application of Three Kinds of Clustering Algorithm in Building Image Segmentation [J]. Modern Computer, 2017(1): 76-80.
- [3] Li Xuelian. The research and application of image segmentation based on clustering analysis [J]. Journal of Baoji University of Arts and Sciences (Natural Science), 2014, 34(2): 66-69.
- [4] Zhang Hongqi, Liu Yu, Li Haijun. The Research on Strawberry Image Segmentation based on Fuzzy Clustering Analysis [J]. Agricultural Science & Technology and Equipment, 2014(7): 21-23.