

# Research of Image Segmentation Algorithm Based on Clustering

Wan Pu, Wang Lisha

Physics Institute, Zhaotong University, Zhaotong, Yunnan, PR China

**Keywords:** image segmentation; clustering; FCM; edge detection; adaptive threshold.

**Abstract.** Through the study of existing image segmentation algorithms, this paper improves the standard FCM-based image segmentation algorithm. And according to the principles of effective treatment on neighborhood image noise by mean filtering and median filtering, proposes a new similarity distance calculation method, which gives full consideration to the gray information and neighborhood information of pixels. Experimental results show that this improved algorithm has higher segmentation efficiency, while having a high segmentation accuracy and strong noise immunity.

## INTRODUCTION

Image as the visual foundation of human perception of the world, is an important means of human to obtain, express and transmit information, image segmentation is important image analysis techniques. Image segmentation in image processing has a very important position, is one of the important steps in image analysis, and the first step of many subsequent image processing and analysis tasks. Image segmentation is widely used, almost all areas of image processing associated with image segmentation applications. In a variety of imaging applications, image segmentation is so inseparable for image target extraction, measure, such as: industrial automation, process control, document image processing, image coding, biomedical image analysis, as well as military, sports, agriculture engineering<sup>[1]</sup>.

Among the image segmentation methods, the method based on clustering is one of the widely used. Image segmentation algorithm based on clustering, compared with other algorithms, the biggest advantage is that it is an unsupervised learning algorithm, does not require prior knowledge of the classification criteria, further the training sample, so it gets a high importance in the image field. Currently, clustering-based problem though has good development prospects, but it is also a very challenging research work. Because its influence in academia growing, many scholar have in-depth research on it, made a lot of solutions and classic algorithms<sup>[2]</sup>. However, although the existing clustering algorithms are able to solve division problems of some images, but the image complexity causes them difficult to solve the problem of all the images, that is, there is no universally adaptable clustering algorithm, so this academic field has yet to be more studied.

## Image segmentation and clustering

**Image segmentation methods.** Image segmentation means the technology and process to divide the image into regions with different characteristics and extract the interested object. In which the characteristics may be the image texture, spectrum, grayscale and color, the interested object may corresponds to a plurality of regions or single area.

There are many image segmentation methods, and some segmentation operations may be directly applied to any image, while others can only be applied to special classes of images<sup>[3]</sup>. Some algorithms need rough segmentation first, because they need the information extracted from the image. Some widely used image segmentation algorithms as shown in figure 1.

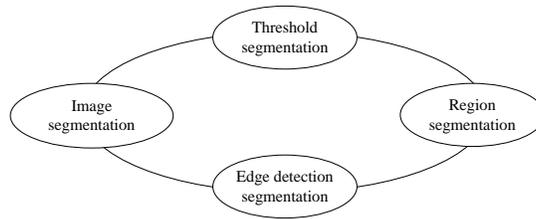


Figure 1. Classification of commonly used image segmentation algorithms

### Clustering segmentation and algorithm.

1. Basic framework of clustering. Clustering is a process to part a set of data objects into several subsets based on a criterion, making the data objects within same subset similar each other, while the data objects within different subsets dissimilar. However, clustering and classification has a different nature. For classification, has known classification rules before dividing, as long as the application of classification rules the data objects are divided, and the result is accurate. But for the clustering, has not known classification rules, needs to part the data objects into several sub-clusters based on a certain guideline, and the result is not accurate. Division criteria of clustering is generally to make the data objects within the same subset have a similarity, and among different subsets have the opposite feature<sup>[4]</sup>. Among them, the similarities and dissimilarities can be evaluated based on the attribute value of the data object. Clustering segmentation is generally divided into five steps: data preparation, feature extraction, clustering segmentation, division assessment and division transformation, the specific process as shown in figure 2. Between each step can be fed back to each other, in order to improve the clustering result.

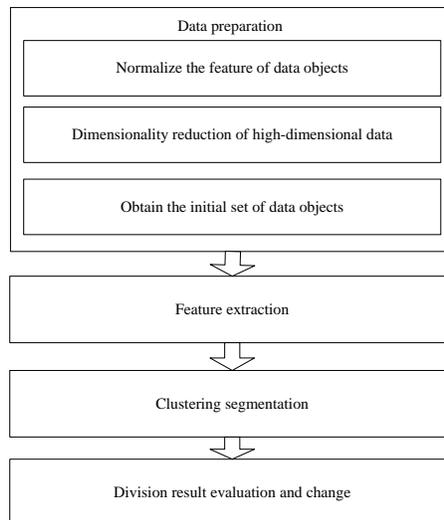


Figure 2. General flow of clustering segmentation

Feature extraction: feature is the basic properties extracted from the original object, constitutes a clustering feature space, a good feature space is conducive to resist the effects of noise samples and generate the correct clustering. Clustering segmentation: by using a certain similarity measure and classification evaluation criteria to classify the samples in feature space. Different similarity measures will produce different clustering results. Division assessment: for a given sample set, regardless of whether there is some kind of distribution, each clustering algorithm will have a division. Usually different clustering algorithms will obtain different clustering results, and for the same algorithm, sometimes different parameters or data input order will also change the results of clustering<sup>[5]</sup>. Division

transformation: based on the need for further processing, different applications will transform the clustering results into a different form, and make different judgments.

Clustering result namely the generated several sub-clusters have two main characteristics: the data objects within the same sub-cluster have similarity, and those in different sub-clusters have dissimilarity with each other. The ideal sub-clusters obtained from accurate clustering can be defined that a set of data objects is compact and isolated. If a sub-cluster is compact, then it meets the homogeneity criteria of clustering, and if a sub-cluster is isolated from other sub-clusters, then it meets the separability criterion of clustering.

## 2. Commonly used clustering algorithms

### (1) K-means clustering

K-means clustering algorithm is the C-means clustering, its main idea is: divide  $n$  samples  $x_i (i=1,2,\dots,n)$  into  $m$  groups  $G_j (j=1,2,\dots,m)$ , making the target function minimized, while access to the cluster center  $m_j$  for each group. Each class is expressed by the average value of objects in it. This algorithm is a classical method to solve the clustering problem, and its main advantage is simple and fast. However, it may obtain different clustering results for different values of  $K$ , and converge to local minima.

### (2) Fuzzy C-means clustering (FCM)

FCM algorithm is to divide  $n$  samples  $x_i (i=1,2,\dots,n)$  into  $m$  fuzzy groups  $G_j (j=1,2,\dots,m)$ , making the target function minimized, while access to the clustering center  $m_j$ . The difference of it between means division is that it indicates the degree of each sample belonging to a group by the interval value within  $[0,1]$ , but the belonging degree of each sample must sum to 1. The general processing of FCM algorithm is shown in figure 3.

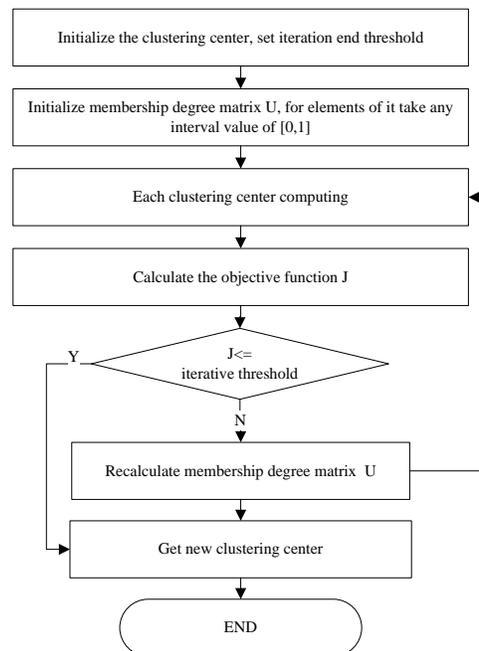


Figure 3. General processing of FCM clustering algorithm

The algorithm output is  $m$  clustering center vector and a fuzzy partition matrix  $m \times n$ , and this membership degree matrix shows the extent of each sample belonging to each class. According to the principle of maximum membership in fuzzy set and based on this matrix, each sample in which class

can be determined. Clustering center is the deputy of this class, showing the average characteristics of each class.

### Improved FCM image segmentation algorithm

**Performance analysis of FCM image segmentation algorithm.** Although the standard FCM is able to obtain better effect when segmenting the image without noise pollution, but for the image with noise pollution and fuzzy edges the segmentation effect is not ideal. This sensitivity to noise is substantially to only use Euclidean distance which considers gray information as a standard to measure the distance between the pixel and the clustering center, but no use of the spatial position information of the divided pixel, ignoring the neighborhood information between pixels, so for low SNR image the segmentation effect is far from ideal. In the segmentation process, both to reduce all kinds of noise and accurately classify original image information is a direction of the mage segmentation algorithm based on standard FCM to be improved. Scholars have developed many solutions to improve this algorithm to improve the segmentation performance. Such algorithms use the diagonal elements in a two-dimensional histogram which is composed by the image pixel gray and neighborhood gray, to reflect the relatively stable image information, and the operation is only related to the number of image gray levels, thus achieving noise image segmentation<sup>[6]</sup>. But these programs have their own limitations, and segmentation accuracy is still not very satisfactory.

**Improvement of FCM image segmentation algorithm.** Neighborhood averaging method is an algorithm for local spatial domain processing. For an image  $f(x, y)$  with  $M \times N$ , Neighborhood averaging treatment to replace the gray value of each pixel by the arithmetic mean of pixels around it, and these pixels surrounding are neighborhood, set the processed image is  $f'(x, y)$ , then the entire process can be expressed as  $f'(x, y) = \frac{1}{T} \sum_{(i,j)}^{(M,N)} f(i, j)$ , in which  $T$  is the number of pixels for the neighborhood. The size of the neighborhood  $m \times n$  determines the neighborhood averaging treatment effect, if set too large, the image will be excessively smoothed, become fuzzy. Neighborhood averaging method is simple, but may make the image details and object edges blurred.

Combining with Euclidean distance to improve the similarity distance function in traditional FCM algorithm, set  $f'(x, y)$  as the gray value after replacement by the mean,  $c_k$  is the clustering center, then the new clustering similarity distance as  $D_k(i, j) = \sqrt{\|f'(x, y) - c_k\|^2}$ . Why median filter can effectively suppress noise is due to replacing the gray value of the image itself with the median, considering the spatial neighborhood information of pixels, so this new similarity distance calculation method based on spatial neighborhood information is introduced in FCM algorithm, to make better use of the spatial information of pixels, thereby overcoming the inaccuracy caused by only using the gray information for clustering. The new distance formula is  $D_k(i, j) = \sqrt{\|f''(x, y) - c_k\|^2}$ , in which  $f''(x, y)$  as the median of all pixel gray values in the window with a center as  $f(x, y)$ .

Standard FCM-based image segmentation method is inhomogeneous sensitive to image noise and gray, this limitation is mainly due to only using the intrinsic characteristics rather than spatial neighborhood information in the clustering process<sup>[7]</sup>. To overcome it, we propose a similarity distance combining with the spatial characteristics to improve the objective function of FCM algorithm. From the combination of the mean and median similarity distances we can get a new similarity distance formula, objective formula:

$$D_k(i, j) = \sqrt{\delta_1 \|f'(x, y) - c_k\|^2 + \delta_2 \|f''(x, y) - c_k\|^2}$$

$$J' = \sum_{i=1}^M \sum_{j=1}^N \lambda \sqrt{\delta_1 \|f'(x, y) - c_k\|^2 + \delta_2 \|f''(x, y) - c_k\|^2}$$

In which  $\lambda \in [0,1]$ , and  $\delta_1, \delta_2$  are the control parameters of neighborhood similarity distance.

### **Performance analysis of improved FCM algorithm**

To verify the effect of the new similarity distance applied in FCM image segmentation, we respectively use standard FCM segmentation algorithm and improved algorithm to segment the noisy image, and then compare to the experimental results. First, construct a binary image with resolution  $200 \times 200$ , while the gray values including two categories of pixel are 51 and 153, and add Gaussian noise mean 0, variance 0.02. The traditional FCM algorithm in dealing with noisy image, the segmentation effect is difficult to satisfactory. The fundamental reason for this situation is that only considered the current pixel gray value while ignoring other valuable information. The improved method due to a combination of the mean and median information of the currently divided pixel neighborhood can largely reduce the influence of noise pixels, showing good noise immunity. For different images, we can change the control parameters of neighborhood similarity distance, after performing the segmentation process for many times obtain the best segmentation results.

### **Conclusion**

Image segmentation is both hot issue and difficult problem in the field of digital image processing, and as the basis for image analysis, image understanding and other more advanced stages of processing, image segmentation has been taken the most researcher attention, there have been thousands of segmentation algorithms. In recent years, some experts and researchers attempted image segmentation by using fuzzy clustering, and because that the fuzzy clustering theory can better describe the image fuzziness and uncertainty, so image segmentation algorithms based on it have made many deep and effective results. This paper analyzes the classic FCM algorithm. The traditional FCM in similarity distance calculation uses Euclidean distance, only considering the gray information, leading to sensitive for noise and uneven distribution of gray, so cannot make effective segmentation on noisy images. In response to this limitation, this paper proposes a new similarity distance formula, combined with the mean and median information of neighborhood pixels. This new similarity distance effectively combines the neighborhood information of the divided pixel, also considers the gray information, so can improve the noise immunity.

### **References**

- [1] C.C. Hung, S. Kulkarni, B.C. Kuo, A New Weighted Fuzzy C-Means Clustering Algorithm for Remotely Sensed Image Classification, IEEE Transactions on Signal Processing, 2011 vol.5, pp.543-553.
- [2] L. Szilagyii, Z. Benyo, S.M. Szilagyii, H.S. Adam, MR Brain Image Segmentation Using an Enhanced Fuzzy C-Means Algorithm, in 25th Annual International Conference of IEEE EMBS, 2003, pp.17-21.
- [3] Mohamad Forouzanfar, et al. Parameter Optimization of Improved Fuzzy C-Means Clustering Algorithm for Brain MR Image Segmentation[J]. Engineering Applications of Artificial Intelligence, 2010, 10 (2): 160-168.
- [4] Xia Y, Feng D.G., Wang T.J., Image Segmentation by Clustering of Spatial Patterns [J]. Pattern Recognition Letters, 2007, 28: 1548-1555.

- [5] Max Mignotte. Segmentation by Fusion of Histogram-Based K-Means Clusters in Different Color Space [J], IEEE Transaction Image Processing, 2008, 17(5):780-787.
- [6] Li Wei-wei, Huan Hui-xian, Zhang Dong-bo, et al. A Color Image Segmentation Method Based on Automatic Seeded Region Growing[C], 2007 IEEE International Conference on Automation and Logistics, Washington DC: IEEE Computer Society, 2007: 1925-1929.
- [7] N.Isa, S.Salamah, U.Ngah. Adaptive Fuzzy Moving K-Means Clustering Algorithm for Image Segmentation [J]. IEEE Trans. on Consumer Electronics, 2009, 55(4):2145-2146.