TELKOMNIKA, Vol.14, No.3, September 2016, pp. 981~986
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v14i3.2757

Fuzzy C-Means Clustering Based on Improved Marked Watershed Transformation

Cuijie Zhao*[1,2], Hongdong Zhao[1], Wei Yao[3]
[1]School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, Tianjin, China
[2]Tianjin University of Finance and Economics Pearl River College, Tianjin 301811, Tianjin, China
[3] School of Economics and Management, Tianjin University of Science and Technology, Tianjin 300222, Tianjin, China
*Corresponding author, e-mail: 36872763@qq.com

Abstract

Currently, the fuzzy c-means algorithm plays a certain role in remote sensing image classification. However, it is easy to fall into local optimal solution, which leads to poor classification. In order to improve the accuracy of classification, this paper, based on the improved marked watershed segmentation, puts forward a fuzzy c-means clustering optimization algorithm. Because the watershed segmentation and fuzzy c-means clustering are sensitive to the noise of the image, this paper uses the adaptive median filtering algorithm to eliminate the noise information. During this process, the classification numbers and initial cluster centers of fuzzy c-means are determined by the result of the fuzzy similar relation clustering. Through a series of comparative simulation experiments, the results show that the method proposed in this paper is more accurate than the ISODATA method, and it is a feasible training method.

Keywords: *adaptive median filtering, marked watershed segmentation, fuzzy similarity relation, fuzzy C-Means clustering*

1. Introduction

At present, remote sensing technology has been gradually developed, the detected image resolution can achieve decimeter counting. In remote sensing monitoring and geographic information acquisition, remote sensing technology can quickly and accurately obtain the relevant data. People's daily work and life cannot be separated from the vast amount of remote sensing data. However, how to process more scientificly and accurately needs further discussion. With the increasing resolution of remote sensing data, the focus of the research is the kind of image processing techniques to extract more accurate quantitative information and multi-scale information. In the extraction of remote sensing image data, the basic is image classification [1]. Because the remote sensing image contains rich spectral information, and the data is very large, which leads to poor accuracy of classification of remote sensing image.

There are two main ways to classify remote sensing image, which are supervised classification and non-supervised classification. Non-supervised classification has become one of the main methods in the field of remote sensing image classification [2]. Fuzzy clustering analysis is one of the main techniques for unsupervised machine learning. It uses the fuzzy theory to analyze the important data, and establishes uncertain descriptions for each sample. It can objectively reflect the real world, and it has important theoretical and practical value. With the further development of the application, the research of the fuzzy clustering algorithm is constantly enriched. Among the numerous fuzzy clustering algorithms, the fuzzy c- means clustering is the most widely used and successful algorithm. It can obtain the sample membership by optimizing the objective function, and determine a series of sample data, finally achieve the goal of automatic classification of the sample data.

Fuzzy clustering optimization gain the attention of many international scholars in related fields rapidly since its advent. In 2010 Yandi Zarnegarnia and Humid Alavi Majd put forward the clustering algorithm bases on similar matrix [3]. It could identify the included proteins in esophagus, stomach and colon cancers based on similarity of Gene Ontology annotation, the

Received May 2, 2016; Revised June 21, 2016; Accepted July 7, 2016

similar degree between the pixels could be calculate. But the algorithm cannot be applied to the remote sensing images, because of the large amount of calculation. Then, in 2011 Beliakov G, Jams S introduced the fuzzy c-means clustering algorithm, it is one of dynamic clustering algorithm that minimize the error sum of squares of samples and the clustering center of a dynamic clustering algorithm [4]. It decides the samples belonging to which category type [5], the clustering algorithm cannot fully analyze the gray scale characteristics of the samples and the connection degree of adjacent pixels [6, 7], but the fuzzy c-means clustering algorithm was effective to noise information.

This paper introduces the theory of fuzzy c-mean clustering algorithm, and describes the marked watershed segmentation, then elaborates the improved algorithm. It needs to construct the fuzzy similarity matrix. The remote sensing image contains many kinds of objective information. it needs to find the pixel of each object, and then determines the initial cluster center of the algorithm. Through comparative experiments, this paper shows the feasibility and superiority of the fuzzy c-mean clustering optimization algorithm. It can be seen that the improved fuzzy c-means clustering algorithm cannot only improve the classification accuracy, but also improve the ability to avoid local extremum.

2. Improved Marked Watershed Segmentation And Fuzzy Similar Relation

2.1. Improved Maked Watered Segmentation Algorithm

All local minimum value of image corresponding to the segmentation region is given by the improved marked watered segmentation algorithm, the local minimum value not only corresponds to the minimum of real images but also the pseudo local minimum value caused by the texture details and background noise. The image is split to tens of thousands of small area for the pseudo local minimum value, the divided areas seriously influence on the extraction of image target. Image pretreatment methods are applied to image. Morphology minimum calibration technology that based on morphology and fuzzy distance transformation is proposed to reduce the number of false local minimum point and restrain over-segmentation. The extraction algorithm based on marker-based watershed is proposed, the algorithm can extract the low frequency part of the gradient image, and tag the local minimum value, and it forces the local minimum value of the original image with morphology minimum calibration technology to shield the original minima in the original gradient image [8]. The algorithm although effectively solve the over-segmentation problem, but it make the edge profile or the details of the images positioning not accurately.

In order to avoid image contour fuzzy orientation deviation and solve the over-segmentation problem, firstly, the color image must be process by adaptive median filter. The process eliminates the impulse noise and preserves the detail of the image [9]. The adaptive median filter includes two layers.

$$\begin{cases} A_1 = Z_{med} - Z_{\min} \\ A_2 = Z_{med} - Z_{\max} \end{cases} \tag{1}$$

$$\begin{cases} B_1 = Z_{xy} - Z_{\min} \\ B_2 = Z_{xy} - Z_{\min} \end{cases} \tag{2}$$

Among which, ”Z_{med}” is the median of image, ”Z_{min}” is the minimum of image, ”Z_{max}” is the maximum of image, ”Z_{xy}” is the value of pixel. If $A_1>0, A_2>0$, the second layer enters the active state, and the size of filter window increases. If $B_1>0, B_2>0$, ”Z_{xy}” and ”Z_{ed}” are output values. Secondly, the gradient image is dealt with the open and close operations, and it is reconstructed by morphological. Morphological open and close operations based on the expansion of the morphology and corrosion. The definition of morphological expansion operation is shown in the equation below:

$$D_{nb}(f,r) = \min(f \oplus nb, r) \tag{3}$$

The definition of the morphological expansion corrosion is shown in the equation below:

$$E_{nb}(f,r) = \max(f \Theta nb, r) \quad (4)$$

The definition of morphological close and open operations for processing gradient image in this paper are shown in the equation below:

$$C_{nb}(f,r) = E_{nb}(f \bullet nb, r)$$
$$O_{nb}(C_{nb},r) = D_{nb}(C_{nb} \circ nb, r) \quad (5)$$

"nb" is the disc structure element, "r" is the radius of the structure element. The latest gradient image is expressed by "I_c", then the regional minima are extracted from the modified gradient image, the extracted minima constitute the binary marker image I_C^{mark}. The markers are the minima of the original image by H-minima technology.

H-minima eliminates the marker regional minima which is lower than threshold value "h", the binary marker image was represented by I_C^{mark}.

$$I_C^{mark} = IMMIN(I_C \mid I_C^{mark}) \quad (6)$$

IMMIN() is the calibration operation of morphology minima. The following figures are results of watershed segmentation.

Through the following comparative simulation experiments, it can be seen the improved marker based watershed algorithm is more accurate than the previous algorithm. For complex images, the marked watershed algorithm segmentation results are not ideal. these complex images refer to the internal noise disturbance and meticulous close irregular complex image, and images with little difference between target and background. The reason is that the low pass filtering algorithm can filter the image edge information with small gradient amplitude in the image.The improved algorithm not only solve above problems, but also effectively preserve the edge details of the image. The improved marker based watershed algorithm segmentation results of is more accurate, and it is a feasible training method

Figure 1. The original image

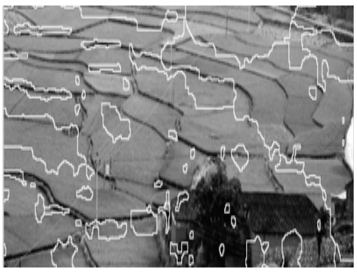

Figure 2. Watershed segmentation

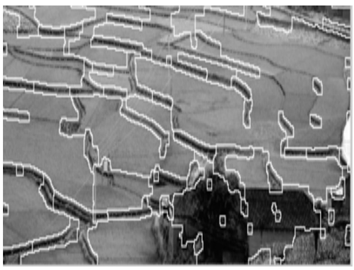

Figure 3. Improved maker watered segmentation

2.2. Fuzzy C-Means Based on Fuzzy Similar Relation

The fuzzy c-means algorithm is high sensitivity to the initial clustering center. In order to avoid the incorrect clustering effect caused by wrong selecting initial cluster centers, the initial clustering center is determined by fuzzy similarity relation in this paper.

Fuzzy similar matrix "R" is established by introducing similar relationship "r_{ij}", "r_{ij}" represents the similarity degree between "x_i"and "x_j", the form of fuzzy similar matrix is shown with the equation below:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad (7)$$

"r_{ij}" is the normalized result of Euclidean distance for median value.

$$\begin{cases} r_{ij} = 1 - \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_i - x_j)^2} \\ R_{ij} = r_{ij} / \max(r) \end{cases} \quad (8)$$

The classification numbers and the samples of each category are determined by establishing fuzzy similar matrix. Fuzzy c-means algorithm can make the sample points of vector space according to a certain distance measurement, but it cannot fully research the pixels' gray scale characteristics and the correlation between adjacent pixels, it makes the fuzzy c-means algorithm fairly sensitive to noise. In addition, the fuzzy c-means algorithm is fairly sensitive to initial clustering center and the input classification number, so adaptive median filtering is proposed to eliminate the noise, the fuzzy similar relation algorithm clustering result is used to determine the initial clustering center and the number of categories of the fuzzy c-means algorithm.

Fuzzy c-means make the data set $X = \{x_1, x_2, ..., x_n\}$, it consists by the samples, and the dimension is "p" to "c" classes. The initial clustering centers matrix "U" is matrix that consisted by the pixels median of each category.

If the "u_{ik}" is the membership degree matrix element between "x_i" and "x_k", the result of fuzzy c-means clustering is represented by fuzzy membership degree matrix. The fuzzy membership degree matrix is shown in the equation below:

$$u_{ij}^{(t)} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}^{(t)}}{u_{kj}^{(t)}}\right)^{\frac{2}{l-1}}} \quad (9)$$

"d_{ij}" is the difference between the clustering center of number "i" and the sample of number "j", "l"is the weighted index. The clustering center needs to be constantly iterated. Clustering centers iteration formula in this paper is shown in the equation below:

$$P^{(t)} = \sum_{i=1:}^{m}\sum_{j=1}^{n}(u^{(t)} \cdot U^{(t)} \cdot f) / I \cdot \sum_{i=1}^{m} u^{(t)} \quad (10)$$

"f" is the matrix of original image , "$U^{(t)}$"is the image partition matrix iterated "t" times. "I" is the matrix with "m" lines and one column that each element is 1. The following figures are clustering results.

Figure 4. The original image

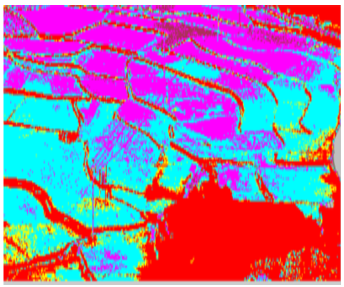

Figure 5. Fuzzy c-means based Improved maker watered segmentation

3. Results and Analysis

ISODATA algorithm is one of the clustering algorithm, the clustering process is realized by constantly iteration. Samples can be transferred from a polymerization class to another.

ISODATA clustering algorithm needs to constantly adjust the clustering results through the continuous iterative clustering center. ISODATA algorithm is very sensitive to the initial clustering center, but there is much noise information, and many parameters need to be set up in this algorithm. Image classification based on ISODATA algorithm is shown in the following figure.

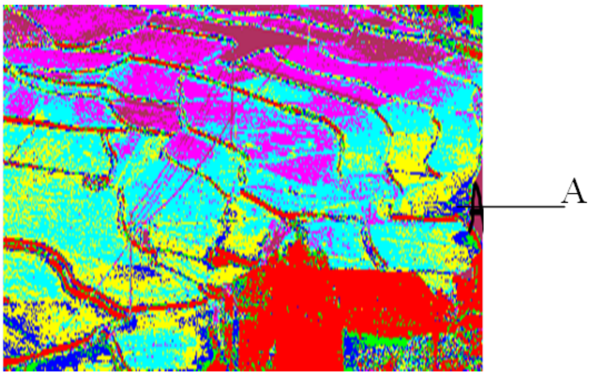

Figure 6. ISODATA clustering result

The road area in the image that marked A is divided into farm in Figure 6, but it is divided into road in Figure 5. Another example is the processing of Figure 7, figure 8 and figure 9 are the results of the classification, the contrast is very obvious. through comparative experiments, it can be seen the improved algorithm is more accurate than the ISODATA algorithm. Optimized algorithm not only eliminates the impulse noise effectively, but also avoids over-segmentation problem and reduce the noise on the clustering.

Figure 7. Original image

Figure 8. ISODATA clustering

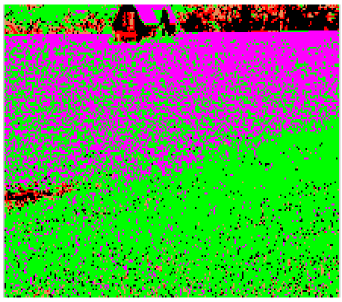

Figure 9. Fuzzy c-means clustering

4. Conclusion

This paper mainly introduces the fuzzy C- means clustering optimization algorithm. It starts from the introduction of basic theories, and gradually explores the particle marker-based watershed segmentation. And then elaborates the whole algorithm by analyzing the pros and cons of the algorithm and introducing the algorithm procedure. In order to verify the superiority of the algorithm proposed here, this paper adopts the fuzzy c -means clustering optimization algorithm and the ISODATA algorithm to respectively classify samples. Through the comparative simulation experiments, it can be seen from the experiment data analysis that the fuzzy c-means clustering optimization algorithm cannot only improve classification accuracy, but also accelerate the convergence speed and enhance the ability of avoiding local extremes. The results show that the method proposed in this paper is more accurate than the ISODATA method, and it is a feasible training method.

Acknowledgements

This work was supported by the research projects of the social science and humanity on Young Fund of the ministry of Education 15YJCZH208.

References

[1] Liqian Cheng. Remote sensing image classification based on optimized support vector machine. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(2): 1037-1045.
[2] B Sai Chandana. Clustering Algorithm Combined with Hill Climbing for Classification of Remote Sensing Image. *International Journal of Electrical and Computer Engineering*. 2014; 4(6): 923-930.
[3] Y Zarnegarnia, HA Majd, MR Tavirani, N Khaier, AAK Maboodi. Application of fuzzy clustering in analysis of included proteins in esophagus, stomach and colon cancers based on similarity of Gene Ontology annotation. *Koomesh*. 2010; 12(01): 14-22.
[4] Beliakov G, James S, Li G. Learing choquet-integral-based matrices for semi supervised clustering. *IEEE Transactions on Fuzzy Systems*. 2011; 19(3): 562-574.
[5] Julian Fan, Jing Li. A Fixed Suppressed Rate Selection Method for Suppressed Fuzzy C-Means Clustering Algorithm. *Applied Mathematics.* 2014; 5(8): 1275-1283.
[6] Ji Ze-xuan, Chen Qiang, Sun Quan-sen, Xia De-shen. Image Segmentation with Anisotropic Weighted Fuzzy C-Means Clustering. *Journal of Computer Aided Design & Computer Graphics*. 2009; 21(10): 1451-1459.
[7] Wang Xiaopeng, Hao Chongyang, Fan Yangyu. Watershed Segmentation Based on Morphological Scale-Space and Gradient Modification. *Journal of Electronics & Information Technology*. 2006; 28(3): 485-489.
[8] Gao Li, Yang Shuyan, Xia Jie, et al. A New Marker-Based Watershed Algorithm. *Acta Electronica Sinica*. 2006; 34(11): 2018-2023.
[9] Basran Parminder S, Robertson Andrew, Wells Derek. CT image artifacts from brachytherapy seed implants: a post processing 3D adaptive median filter. *Medical Physics*. 2011; 38(2): 712-718.