ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

Improved Research on C-means Clustering Algorithm based on Relieft Feature Weighting

Jiangtao Wang

College of Network Communication, Zhejiang Yuexiu University, Shaoxing, China

*E-mail: 3068525606@qq.com

Abstract

Image segmentation is the most critical step in image processing and image analysis. I have studied many articles on image segmentation techniques. First, the C-means clustering algorithm is clarified and related concepts are explained, and the basic principles and clustering criteria of the clustering method are analyzed. Then, in view of the advantages and disadvantages of the algorithm and the existing shortcomings, the C-means clustering algorithm is improved, and Relieft technology is introduced. Because many features are involved in image segmentation, the core of the improved algorithm is the weighting process during feature extraction, and the final design An effective and robust color image segmentation process is developed. Through a large number of experiments, the segmentation results of the C-means image segmentation algorithm before and after the improvement are compared. The improved clustering algorithm can indeed achieve better image segmentation results; through the experimental data placed in different environments Statistics can verify that the improved segmentation algorithm is more robust. Finally, the author analyzes the shortcomings of the algorithm model, and also points out the direction of future research.

Keywords

Clustering; C-means; Feature Weighting; Relieft Technique; Fuzzy Set.

1. Introduction

Image segmentation [1] is one of the core technologies of computer vision. The early image segmentation technology used gray information to extract target information from the background, which had relatively large limitations. Since most of the targets in reality have rich color information, color image segmentation processing is obviously more able to meet the development requirements of pattern recognition. Color image segmentation refers to separating a specific target area from a complex background. Experts, scholars, and research institutes at home and abroad have done a lot of research in this field. The author of this article has conducted research on color clustering technology, especially fuzzy technology because it can express and deal with uncertainty problems well, so it will be used in the field of color image segmentation. There will be broader application prospects.

2. Fuzzy C-means Clustering Algorithm

2.1 Definition of Clustering

The definition of clustering [2] : The process of grouping or classifying objects in a specific data set. The classification is based on similarity. The basis of similarity is the measurement function, specifically the Euclidean distance between data objects. The purpose of clustering algorithm is to find the optimal metric value between data objects in a given data set.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

2.2 Basic Knowledge of Fuzzy Sets

Before discussing fuzzy sets [3], first discuss the membership function [4]: "The function of the degree to which a data object belongs to the set A", usually denoted as $\mu_A(x)$, with a value range [0,1]. A membership function in space is a fuzzy subset defined in the universe of discourse $X = \{x\}$. One-to-one correspondence between them:

$$\underset{\sim}{A} = \{\mu_{\mathrm{A}}(x_i) : x_i \in \mathrm{X}\} \tag{1}$$

In the process of clustering, we can regard the grouping generated after clustering as a fuzzy set. Therefore, the membership interval of each sample point to be clustered belongs to the grouping is also [0,1].

2.3 Fuzzy C-means Clustering

Pattern recognition has many theoretical foundations, and cluster analysis is an important method of unsupervised pattern classification [5]. Traditional clustering algorithms have clear requirements for the clustering of the identified objects, but in practical applications they are not always true, because the sample objects may have intersection areas between multiple classes. If the fuzzy clustering method is used for division, better results can be obtained. Basic idea: Assuming that the sample set $X = \{x_0, x_1, \cdots, x_{n-1}\}$ is divided into C sub-categories, the fuzzy matrix U can be used to represent the classification results. The basic properties of the matrix:

$$\mu_{ik} \in [0,1], \quad \sum_{i=0}^{c-1} \mu_{ik} = 1, \forall k, \quad 0 \le \sum_{k=0}^{n-1} \mu_{ik} \le n, \forall i \tag{2}$$

Among the many existing fuzzy clustering analysis methods, the fuzzy C-means clustering [6] method does a good job in feature extraction and automatic clustering. In 1973, Bedeck proposed the FCM clustering algorithm to improve C-means clustering and fuzzy objective function:

$$J(U,V) = \sum_{k=0}^{n-1} \sum_{i=0}^{c-1} (\mu_{ik})^m (d_{ik})^2 \tag{3}$$

Among them, $\mu_{ik} \in [0,1], m \in [1,\infty)$, is the weighted index; d_{ik} is the Euclidean distance from the DD sample object to the class, which is defined as:

$$(d_{ik})^2 = \|x_k - v_i\|^2 = (x_k - v_i)^T (x_k - v_i) \tag{4}$$

$\|x_k - v_i\|^2$ represents the optimal metric value between data objects. In formula (3), the calculation formula of the fuzzy objective function J(U,V) is given, which represents the weighted sum of squares of the optimal metric value of the data sample, and the weight is the m power of the sample x_k to the i-th class membership μ_{ik}, and finally The clustering criterion of the fuzzy C-means clustering algorithm is converted to the minimum value of the objective function min$\{J(U,V)\}$.

As follows, we construct a new objective function $J'(U,V)$, and obtain the minimum value of formula (3) from the new objective function.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

$$J'(U,V) = J(U,V) + \sum_{k=1}^{n} \lambda_k (\sum_{i=1}^{c} \mu_{ik} - 1) \tag{5}$$

Taking the derivation of all μ_{ik}, the necessary condition for the minimum of formula (3) is:

$$c_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m x_k}{\sum_{k=1}^{n} \mu_{ik}^m} \tag{6}$$

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)} \tag{7}$$

According to formula (6) and formula (7), the best fuzzy C mean value can be obtained. The clustering algorithm solves the extreme value of the objective function $J(U,V)$ through a series of iterative processes:

Step 1: Initialize the number of clusters $c(2 \le c \le n)$ and the weighted index $m(m \in [2,\infty))$;

Step 2: Preset the initial value of the matrix U is $U^{(l)} = [\mu_{ik}{}^{(l)}]$ and set l=0;

Step 3: Calculate the cluster centers of each cluster according to formula (5):

$$v_i = \sum_{k=0}^{n-1} (\mu_{ik})^m x_k / \sum_{k=0}^{n-1} (\mu_{ik})^m \tag{8}$$

Step 4: Update the fuzzy clustering matrix $U^{(l)}(l = l+1)$, calculate I_k and $\bar{I}_k$: $I_k = \{i \mid 0 \le i < c; d_k = \|x_k - v_i\| = 0\}$, $\bar{I}_k = \{0,1,\cdots,c-1\} - I_k$. If $I_k = \phi$,then:

$$u_{ik} = 1 / \sum_{j=0}^{c-1} (d_{ik} / d_{jk})^{2/(m-1)} \tag{9}$$

otherwise, for all $i \in \bar{I}_k$, set $\mu_{ik} = 0$ and take $\sum_{i \in I_k} \mu_{ik} = 1$;

Step 5: Calculate the calculation result of $\|U^{(l-1)} - U^{(l)}\|$ and compare the size of the initial threshold. If it is less than the initial threshold, the iteration stops; otherwise, go to step 3.

Until the iterative algorithm converges and the threshold is set to α, the cluster segmentation can be expressed as: $\mu_{ik} = \max_i \{\mu_{0k}, \mu_{1k}, \cdots, \mu_{c-1k}\} \ge \alpha$, then x_k belongs to the i-th category.

In image segmentation, C-means clustering algorithm has been widely used. The algorithm requires a preset number of initial clusters, and the individual characteristics of the classification object can

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

have a certain degree of influence on the segmentation effect. How to overcome the above problems? This requires improvements to the C-means clustering algorithm.

2.4 Modified Fuzzy C-means Clustering Algorithm

In order to solve the problems of the fuzzy C-means clustering algorithm and improve the segmentation effect of classification, this paper revised the fuzzy C-means clustering algorithm. The algorithm uses adaptive selection of the initial number of clusters and combines the Relieft technology to perform feature weighting. Decompose the image.

2.4.1 Adaptive Selection of the Number of Initial Clusters

The adaptive selection of the number of initial clusters [7] is an extremely important preliminary work of the fuzzy C-means clustering algorithm. Unreasonable initialization may cause the final clusters to be inconsistent with the real structure of the given data set. As a result, the final clustering fails. Determining the optimal number of initial clusters is the key. This article will introduce the cluster validity index integrated by Xie and Beni-Xie-Beni index, whose mathematical definition is as follows:

$$V_{XB}(U,V;X)=\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m\left\|x_j-v_i\right\|^2/(n\min_{i,k=1,i\neq k}^{c}\left\|v_i-v_k\right\|^2) \tag{10}$$

$V_{XB}(U,V;X)$ represents the Xie-Beni effectiveness index, also known as the compactness measure, and n represents the number of samples. The smaller the value of V_{XB}, the better the clustering effect.

The adaptive selection algorithm for the number of initial clusters is as follows:

Step 1: preset the number of clusters $c=c_{\min}$ and the optimal number of clusters $c^*=1$, where $c\in[c_{\min},c_{\max}]$, $c_{\min}=2$, $c_{\max}=7$;

Step 2: Given initial cluster center $V^{(0)}$;

Step 3: Calculate $V^{(b+1)}$ and $U^{(b)}$ according to formula (8) and formula (9);

Step 4: Calculate the objective function difference $\Delta J=J_m-J_{m-1}$. If $\Delta J<0$, go to step 3;

Step 5: Calculate the Xie-Beni effectiveness index V_{XB} according to formula (10);

Step 6: If $V_{XB}<V_{XB}^*$, then $V_{XB}=V_{XB}^*, c^*=c$;

Step 7: Let $c=c+1$. If $c=c_{\max}$, stop; otherwise, go to step 2;

When determining the number of initial clusters, an extreme value strategy of the number of clusters is adopted. The determination of the extreme value is determined based on prior knowledge. In the sense of color, it is not fully adaptive. The following research will solve this problem.

2.4.2 Relieft Technology for Feature Weighting

In 1992, Kira and Rendell proposed [9] the basic Relieft evaluation technology. Due to the limitations of the research conditions at the time, the evaluation technology could not handle incomplete data. Later Kononenk was extended to solve regression problems and multi-class division problems. Basic principle: "Combine weights with clustering features to evaluate the ability of features to distinguish short-distance samples, so that similar samples are closer, and heterogeneous samples are farther away." The features of Relieft evaluation technology are as follows: First, it has higher efficiency; second, there are no strict restrictions on the classified data objects; third, compared to other evaluation technologies, Relieft evaluation technology is more robust; fourth, can handle incomplete data and noisy data well [10], and can better remove irrelevant features.:

Assuming that the entire set of data objects to be clustered is $X=\{x_1,x_2,\cdots,x_n\}$, the N eigenvalues of any sample: $x_i=[x_{i1},x_{i2},\cdots,x_{iN}]^T$, the difference between h_j and x_i in the N×1 matrix:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

$$diff\_hit = \sum_{j=1}^{R} \frac{|x_i - h_j|}{\max(X) - \min(X)} \tag{11}$$

Use diff_miss to be an N×1 matrix, which represents the difference in features between m_{lj} and x_i.

$$diff\_miss = \sum_{l \neq class(x_i)} \frac{P(l)}{1 - P(class(x_i))} \sum_{j=1}^{R} \frac{|x_j - m_{lj}|}{\max(X) - \min(X)} \tag{12}$$

In the above formula, $P(l)$ represents the probability of the first category, which is the total number of samples in the data set divided by the number of samples in the first category. Repeat this several times to get the weight value of each feature in the feature set [11].

2.5 Image Segmentation Process

Improved fuzzy clustering algorithm implementation process: first convert the image to YIQ color space [12], and extract the brightness component Y, initialize the number of clusters through the effectiveness index, introduce Relieft technology for feature weighting, and extract image features to achieve the purpose of image segmentation.

The specific flow chart is as follows:

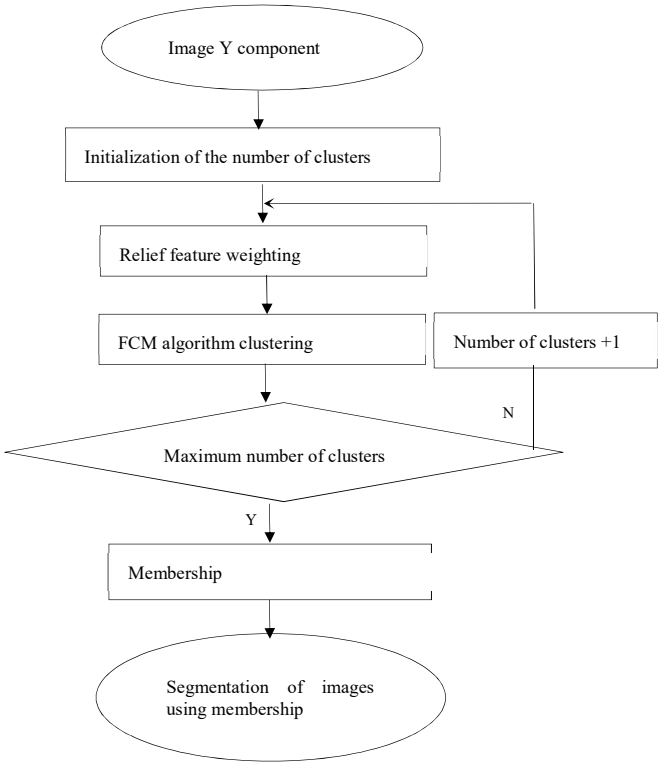

Fig. 1 The improved process of C means clustering algorithm

3. Experiment

3.1 Experimental Platform

To verify the effectiveness of the C-means clustering algorithm, under the Microsoft Windows10 Ultimate version of the operating system, the CPU is an Intel Core i5-4200 and a 4.00GB notebook with memory, and Matlab7.0 is used for simulation experiments.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

3.2 Verification and Comparison of Segmentation Algorithms

Due to space problems, take the segmentation results of 3 images in the collected traffic sign image library as an example:

(1) The original image is as follows:

Fig. 2 The original image

(2) The grayscale image is as follows:

Fig. 3 The original gray level image

(3) The segmentation effect of traffic sign images using fuzzy C-means clustering:

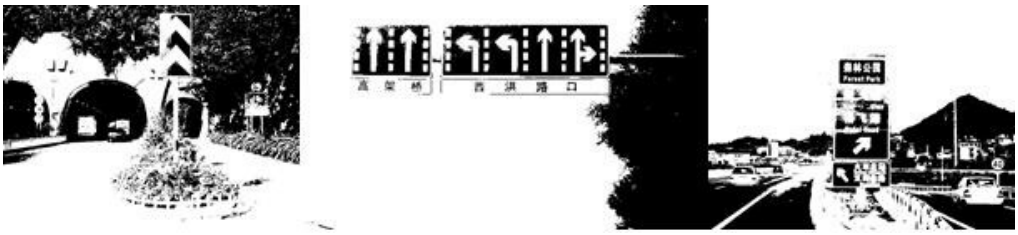

Fig. 4 By using the fuzzy C mean clustering segmentation effect

(4) The Relieft technology is introduced for feature weighting, and the segmentation algorithm of fuzzy C-means clustering is improved. The following figure 5 is the experimental result:

Fig. 5 C means clustering algorithm to introduce Relieft technology after the segmentation effect

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

(5) In order to further demonstrate the effect of the improved C-means clustering segmentation algorithm, a set of random signs before and after the introduction of Relift technology is tested. The comparison results are shown in Figure 6 and Figure 7:

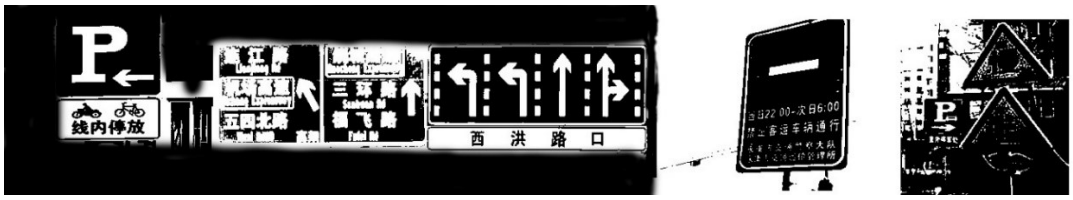

Fig. 6 The introduction of Relieft technology before the segmentation effect

Fig. 7 The introduction of Relieft technology after the segmentation effect

3.3 Comparison of Experimental Data in Different Environments

The author collected 1000 sets of experimental data under different weather conditions in sunny, cloudy and rainy days, as shown in Table 1:

Table 1. Comparison of the experimental data under different climatic conditions

	Sunny day success rate	Cloudy success rate	Rainy day success rate
Fuzzy C-means clustering	90.54%	81.28%	77.25%
Relieft feature weighted C-means clustering	93.77%	90.76%	89.52%

4. Experimental Results and Analysis

In order to verify the effectiveness of the segmentation algorithm proposed in this paper, a large number of experiments have been carried out on the images of traffic signs and road signs.

(1) Comparing the experimental results in Fig. 4 with Fig. 5, it can be found that:

Introducing the Relieft technology, the segmentation results of the feature-weighted C-means clustering algorithm are more accurate, and it can get a clearer edge under the premise of removing the noise more obviously.

(2) Comparison of experimental data in Table 1:

In sunny weather conditions, the detection success rate of the two algorithms is higher, and the improved C-means clustering algorithm is 3.23 percentage points higher, and the advantage is not obvious; but under the conditions of cloudy and rainy days, the improved C-means clustering algorithm The detection success rate of the algorithm is 12.27% higher. Through the comparison of experimental data, it is obvious that the color image segmentation algorithm that introduces Relieft feature weighting is more robust.

5. Conclusion

The development of computer vision and pattern recognition requires the theoretical support of color image segmentation algorithms. In this paper, based on the color image segmentation algorithm of fuzzy C-means clustering, Relieft technology is introduced, and the clustering featuresare weighted.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0016

A large amount of experimental data shows that the research and improvement of the C-means clustering algorithm based on Relieft feature weighting can effectively achieve color image segmentation while removing the effect of noise, and greatly improve the adaptability of the segmentation algorithm, that is to say, the robustness of the segmentation algorithm is good. It lays a good foundation for the practical application of color image segmentation, provides theoretical guarantee for the design of road traffic sign detection system, and has certain theoretical guiding significance in the application of security monitoring, intelligent transportation system and so on.

Color image segmentation algorithms have a lot of research results, segmentation algorithms are not the same, each has its own adaptation background and advantages. I and my team members have limited scientific research capabilities. The combination of this segmentation algorithm and other segmentation algorithms, such as: comprehensive watershed and regional merge clustering, or how to combine color space to expand the scope of application of the segmentation algorithm and improve the efficiency of the image segmentation algorithm is the direction that I and the research team will work hard in the future.

References

[1] L R, L Y. Suppressed Non-local Spatial Intuitionistic Fuzzy C-means Image Segmentation Algorithm[J]. Journal of Electronics & Information Technology, 2019,41(06):1472-1479.

[2] W Z, ZHANG C S, QIAN J B. Fuzzy C-mean image segmentation algorithm with dynamic parameters and edge subdivision, 2020,44(3):288-295.

[3] WEI G X, ZHOU X Z, PU X B. Gray image segmentation based on fuzzy C-means and artificial bee colony optimization[J]. Journal of Lanzhou University: Natural Sciences, 2019,55(2):250-254.

[4] Cady S T, Domínguez-García A D, Hadjicostis C N. A distributed generation control architecture for islanded AC microgrids [J]. IEEE Transactions on Control Systems Technology, 2015, 23(5) : 1717-1735.P. Kundur, Power System Stability and Control, New York: McGraw-Hill, 1994.

[5] Jia Long, Hu Zechun, Song Yonghua al. Joint planning of distribution networks with distributed energy storage systems and electric vehicle charging stations [J]. Proceedings of the CSEE, 2017, 37(1) : 73-83.

[6] ADHIKARI S K,SING J K,BASU D K,et al.Conditional Spatial Fuzzy C-means Clustering Algorithm with Application in MRI Image Segmentation//MANDAL J,SATAPATHY S,KUMASM,et a.,eds. Information Systems Design and Intelligent Applications.Berlin, Germany: Springer,2015: 539-547.

[7] WU X Y, CHEN S C. Online Local Adaptive Fuzzy C-means Clustering Algorithm. Pattern Recognition and Artificial Intelligence. 2013, 26(11): 1026-1032.

[8] LI Y L,SHEN Y.Fuzzy C-means Algorithm Based on the Spatial Information for Image Segmentation. Journal of Huazhong University of Science and Technology (Natural Science Edition) , 2009, 37(6): 57-59.

[9] WU S H, WU Y Q, ZHOU J J. Multi-level thresholding for remote sensing image of urban area based on line intercept histogram[J]. CAAI Transactions on Intelligent Systems, 2018, 13(2):227-235.

[10] Mahdavi N, Braslavsky J H, Seron M M, et al.Model predictive control of distributed air-conditioning loads to compensate fluctuations in solar power[J]. IEEE Transactions on Smart Grid, 2017, 8(6) : 3055-3065.

[11] de Melo H Nytorv J P Peregrina P G,et al.A controllable bidirectional battery charger for electric vehicles with vehicle-to-grid capability[J].IEEE Transactions on Vehicular Technology,2018,67 (1):114-123.

[12] YU J P, ZHENG J, MEI H B. K-means clustering algorithm based on improved artificial bee colony algorithm[J]. Journal of Computer Applications, 2014, 34(4):1065-1069,1088.

[13] S. Raju, G. Pillai, “Design and implementation of type-2 fuzzy logic controller for DFIG-based wind energy systems in distribution networks,” IEEE Transactions on Sustainable Energy, vol. 7, no. 1, pp. 345–353, 2016.

[14] MA Y R,PENG Y J.Fast Image Segmentation Model Combined with Fuzzy C-means Method and Curve Evolution[J]. Journal of Electronics & Information Technology, 2017,39(6):1379-1386.

[15] HEMANTH D J, ANITHA J, BALAS V E. Fast and Accurate Fuzzy C-means Algorithm for MR Brain Image Segmentation. International Journal of Imaging Systems and Technology, 2016, 26(3): 188-195.