Research on Fuzzy Clustering Image Segmentation Algorithm based on GA and Gray Histogram

Baoyi Wang[1, a], Long Kang[1, b], Shaomin Zhang[1, c]

[1] School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

[a]email: wangbaoyiqj@126.com, [b]email: 846738303@qq.com, [c]email: zhangshaomin@126.com

Keywords: Image Segmentation; Genetic Algorithm; FCM; Gray Histogram

Abstract. When the traditional FCM algorithm is applied to image segmentation, it is sensitive to the selection of clustering center and can not automatically determine the number of clustering center. According to these problems, this paper proposes a fuzzy clustering image segmentation algorithm based on genetic algorithm and gray histogram. This algorithm uses the global search characters of genetic algorithm to avoid falling into local optimal solution, and the objective function is optimized and the number of the cluster center is determined by gray histogram. Experimental results show that the proposed method has strong robustness and good segmentation effect.

Introduction

Fuzzy c-means clustering algorithm is widely used in the field of image segmentation [1], but it is sensitive to the initial clustering center [4] and can not predict the number of clustering. Genetic algorithm searches the optimal solution by simulating the natural evolution process of the population, and can avoid the local optimal solution. When the gray value is used to calculate the objective function, the repeated calculation is large. Statistical characteristics of the gray histogram can be used to improve the convergence rate, and because the clustering center usually appears in the maximum value of the histogram, the number of clustering center can be determined by the number of maximum values [2]. In this paper, a fuzzy clustering image segmentation algorithm based on GA and gray histogram is proposed. The algorithm uses gray histogram to optimize the objective function and determine the number of clustering center. The global optimal approximate solution of clustering center is obtained by the genetic algorithm [3], and it is used as the initial clustering center of FCM to get the global optimal solution.

Image Segmentation based on Traditional FCM

The general expression of the objective function is:

$$J_m\left(U,V\right)=\sum_{k=1}^{c}\sum_{i=1}^{n}u_{ki}^{m}d_{ik}^{2} \tag{1}$$

$$\sum_{k=1}^{c}u_{ki}=1 \tag{2}$$

$$0\le u_{ki}\le 1 \tag{3}$$

U is a fuzzy membership matrix; *V* is the clustering center matrix, and each clustering center is the gray value of the image; *m* is the weighted index, and $m\in[1,+\infty)$; *c* is the number of cluster centers; *n* is the number of all the pixels; u_{ik} is fuzzy membership of the i-th sample with respect to the k-th clustering center; d_{ik} is the distance between the i-th sample and the k-th cluster center.

When $J_m(U,V)$ gets the minimum value $min\{J_m(U,V)\}$, the algorithm gets the best clustering center. Lagrange number multiplication is used to solve the problem.

$$u_{ki}=\frac{\left(d_{ik}\right)^{\frac{2}{1-m}}}{\sum_{r=1}^{c}\left(d_{ir}\right)^{\frac{2}{1-m}}} \quad (4)$$

$$v_k=\frac{\sum_{i=1}^{n}\left(u_{ki}\right)^m x_i}{\sum_{i=1}^{n}\left(u_{ki}\right)^m} \quad (5)$$

The general process of FCM algorithm for image segmentation：

Initialization：initialize the cluster number c, the weighted index m, the maximum number of iterations max and the iteration threshold ε; initialize iteration counter $t=0$; initialize cluster center $V^{(0)}=\{v_1,v_2\ldots v_c\}$ randomly，where $1\leqslant i\leqslant c$ and v_i represents the gray value of the image；

Step 1: Calculate the fuzzy membership matrix $U^{(t)}$;

$$u_{ki}^{(t)}=\frac{\left(d_{ik}^{(t)}\right)^{\frac{2}{1-m}}}{\sum_{r=1}^{c}\left(d_{ik}^{(t)}\right)^{\frac{2}{1-m}}} \quad (6)$$

Step 2: Calculate cluster center $V^{(t+1)}$;

$$v_k^{(t+1)}=\frac{\sum_{i=1}^{n}\left(u_{ki}^{(t)}\right)^m x_i}{\sum_{i=1}^{n}\left(u_{ki}^{(t)}\right)^m} \quad (7)$$

Step 3: Calculate the distance between $V^{(t+1)}$ and $V^{(t)}$. If $||V^{(t)}-V^{(t+1)}||\leqslant\varepsilon$, then the iteration is over and jump to Step 5;

Step 4: $t=t+1$，if $t<max$ then jump to Step 1；

Step 5: Output the clustering center V and the membership matrix U.

Fuzzy Clustering Image Segmentation based on Genetic Algorithm and Gray Histogram

1) Determine encoding scheme of GA

Real-coded genetic algorithm reduces the complexity of the algorithm and can improve the efficiency, so it is used in this paper. Encode clustering center $V=\{v_1,v_2\ldots v_c\}$, where c is the number of the clustering center and v_i $(0\leqslant v_i\leqslant 255)$ is the gray value of the image.

2) Determine the number of clustering center

Gray level histogram of the image reflects the distribution of gray level, and there are experiments show that the clustering center is usually in the maximum value of the histogram [3]. Therefore, the number of the maximum of gray histogram is used as the number of clusters.

3) Determine the fitness function of GA and the objective function of FCM

When the objective function of FCM obtains the optimal solution, genetic algorithm fitness function reaches the maximum. So the fitness function is defined as follows：

$$f=\frac{1}{J_m\left(U,V\right)+\delta} \quad (8)$$

In order to avoid the emergence denominator is zero, a small constant $\delta(0<\delta)$ can be added in the denominator, and δ can take the values of 1.

When using the clustering algorithm to segment the gray image, the gray values of the image are taken as samples. Because the gray value is in the range of L, the gray histogram can be used to count the number of pixels of each gray value, and when the statistic value of a gray value is not zero, it is used a sample. This method can reduce the amount of computation by avoiding the repeated calculation of the distance between gray value and clustering center.

Define the sample set $X=\{l_0,l_1,l_2\ldots l_i\ldots l_n\}$, and the statistics value of the sample $G=\{g_0,g_1,g_2\ldots g_i\ldots g_n\}$, where $0\leqslant n\leqslant 255$ and l_i is gray value, and g_i is the number of pixels corresponding to gray value l_i. Then,

$$J'_m(\mathrm{U},\mathrm{V}) = \sum_{k=1}^{c}\sum_{i=1}^{n} u_{ki}^m d_{ik}^2 g_i \quad (9)$$

$$u'_{ki} = \frac{(d_{ik})^{\frac{2}{1-m}}}{\sum_{r=1}^{c}(d_{ir})^{\frac{2}{1-m}}} \quad (10)$$

$$v'_k = \frac{\sum_{i=1}^{n} g_i (u'_{ki})^m x_i}{\sum_{i=1}^{n} g_i (u'_{ki})^m} \quad (11)$$

Finally, the fitness function is changed to

$$f' = \frac{1}{J'_m(U,V)+\delta} \quad (12)$$

4) Initialize population in GA

Randomly generated n individuals as the initial population, and each individual is represented as $\{v_1,v_2\ldots v_c\}$. And the gray value corresponding to the maximum values of the gray histogram compose an individual which replace an individual randomly in population.

5) Determine the probability of crossover and mutation

This paper uses the adaptive algorithm proposed by Ren Ziwu [6] et al to solve the phenomenon of random roaming. The crossover probability and mutation probability are determined as follow:

$$P_c = \begin{cases} P_{c1} - \dfrac{(P_{c1}-P_{c2})(f'-f'_{avg})}{f'_{max}-f'_{avg}} & f' \ge f'_{avg} \\ P_{c1} & f' < f'_{avg} \end{cases} \quad (13)$$

$$P_m = \begin{cases} P_{m1} - \dfrac{(P_{m1}-P_{m2})(f'-f'_{avg})}{f'_{max}-f'_{avg}} & f' \ge f'_{avg} \\ P_{m1} & f' < f'_{avg} \end{cases} \quad (14)$$

f'_{max} is the maximum fitness; f'_{avg} is the average fitness ; f' is the fitness of the individuals in the mutation or the large value in the fitness of the two individuals in the crossover.

6) Determine the genetic operator

Roulette wheel selection [7] is used as selection operator. Single crossover is used as crossover operator. And the random mutation operator is used to carry on the mutation operation.

7) Elite reservation strategy

To ensure the best individual of each generation is not destroyed, the elite reservation strategy is used in this paper. Through the elite reservation strategy, it can prevent the loss of the best individual of current population [5].

The Algorithmic process

Step 1: Initialize samples. The infrared image is processed by gray scale, and the Gauss noise of the image is eliminated by using the neighborhood averaging method with 3×3 window. Finally, the sample set $X=\{l_0,l_1,l_2\ldots l_i\ldots l_n\}$, and the statistical value set $G=\{g_0,g_1,g_2\ldots g_i\ldots g_n\}$ are extracted as the input of the algorithm from the gray histogram of the image, and $1 \leqslant n \leqslant 255$;

Step 2: Initialize parameters. Initialize crossover and mutation probability P_{c1}, P_{c2}, P_{m1}, P_{m2}; Initialize generation gap g, population size n, weighted index m, and the number of clusters c; set the number of iterations $t=0$ and Initialize the maximum number of iterations max;

Step 3: Initialize population $P=\{V_1,V_2\ldots V_i\ldots V_n\}$, and $V_i=\{v^1{}_i,v^2{}_i\ldots v^3{}_i\}$;

Step 4: Calculate the fitness of population according to formula (12);

Step 5: Do the selection operation based on Roulette wheel and the generation gap is g;

Step 6: Do the single crossover operation, and cross probability is P_c;

Step 7: Do the random mutation operation, and mutation probability is P_m;

Step 8: Calculate the fitness of the next generation;

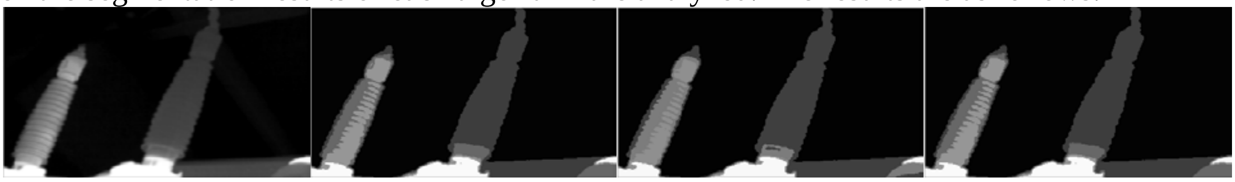
Step 9: Execute the elite retention strategy, and randomly generate individuals to fill sub population to ensure that the population size is *n*;

Step 10: To make $t=t+1$, if $t<max$, then jump to Step 4 and continue the iteration;

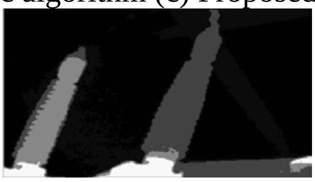
Step 11: The optimal cluster centers, which are obtained by genetic algorithm, is used as the initial clustering center of FCM and FCM is used to segment the image.

Experiment and Result Analysis

According to the above model and algorithm, this paper uses Matlab R2015a as the programming tool, and selects the 340×240 sized image the 110kV transformer bushing and the line joint as the experimental object. Initialize the parameters as $P_{c1}=0.9$, $P_{c2}=0.6$, $P_{m1}=0.1$, $P_{m2}=0.01$, $m=2$, $g=0.9$, $n=20$. Experiments were conducted using the proposed algorithm, genetic algorithm and random initial clustering center FCM algorithm for infrared image segmentation, then the segmentation results of each algorithm are analyzed. The results are as follows:

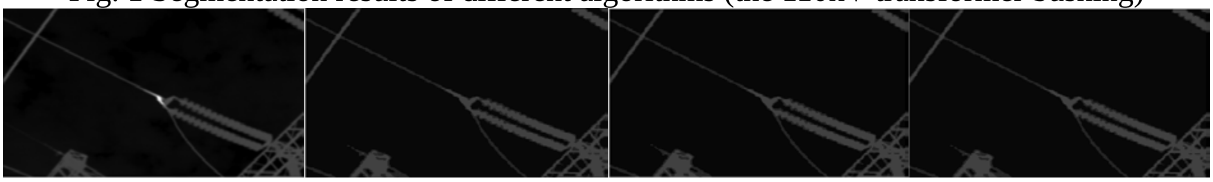


(a)The original image (b) Genetic algorithm (c) Proposed algorithm (d) Random FCM (1)



(e) Traditional FCM (2)

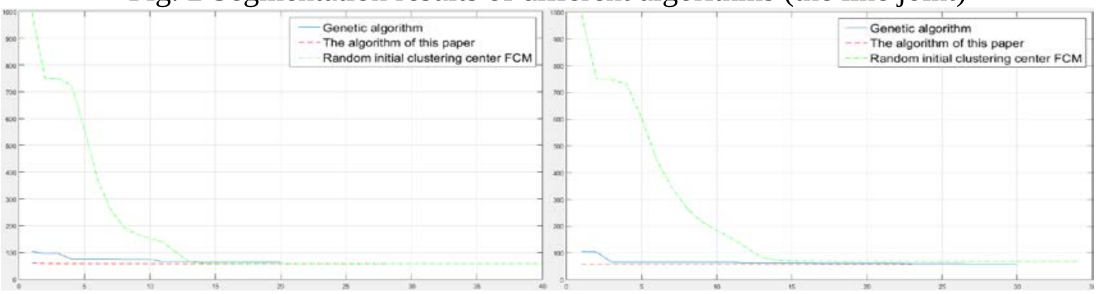
Fig. 1 Segmentation results of different algorithms (the 110kV transformer bushing)



(a)The original image (b) Genetic algorithm (c) Proposed algorithm (d) Random FCM (1)

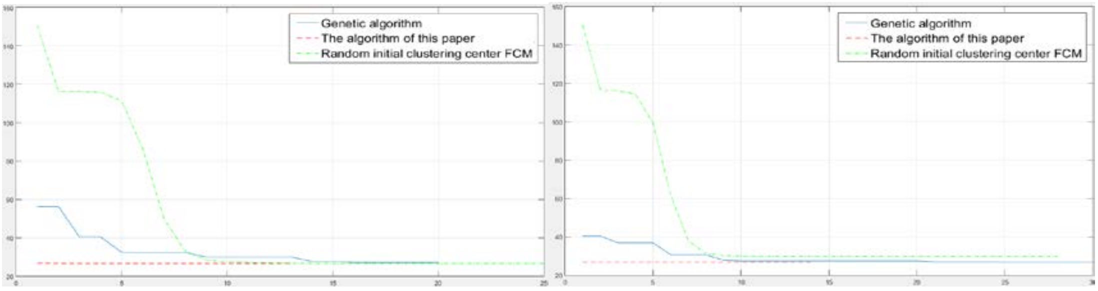
(e) Traditional FCM (2)

Fig. 2 Segmentation results of different algorithms (the line joint)



(a) Experiment 1 (b) Experiment 2

Fig. 3 Optimal solution curve (the 110kV transformer bushing)



(a) Experiment 1 (b) Experiment 2

Fig. 4 Optimal solution curve (the line joint)

Table 1 Iteration times, time consuming and clustering centers of different algorithms (The 110kv transformer bushing and c=5)

		Genetic algorithm	The proposed algorithm	Random FCM (1)
Experiment 1	iteration times	20	28	40
	time consuming (s)	32.37	2.40	3.12
	clustering centers	0.9548 0.6114 0.4632 0.2330 0.0141	0.9782 0.5869 0.4030 0.2409 0.0138	0.9782 0.5867 0.4028 0.2408 0.0138
Experiment 2		Genetic algorithm	The proposed algorithm	Random FCM (2)
	iteration times	30	19	37
	time consuming (s)	48.57	3.06	11.59
	clustering centers	0.9876 0.6008 0.3894 0.2308 0.0096	0.9782 0.5867 0.4028 0.2408 0.0138	0.9714 0.5275 0.2610 0.0432 0.0050

Table 2 Iteration times, time consuming and clustering centers of different algorithms (The line joint and $c=3$)

		Genetic algorithm	The proposed algorithm	Random FCM (1)
Experiment 1	iteration times	20	13	25
	time consuming (s)	19.05	0.67	1.13
	clustering centers	0.2630 0.1551 0.0331	0.2559 0.1454 0.0313	0.2559 0.1453 0.0313
Experiment 2		Genetic algorithm	The proposed algorithm	Random FCM (2)
	iteration times	30	14	28
	time consuming (s)	28.44	0.74	1.32
	clustering centers	0.2490 0.1332 0.0312	0.2559 0.1453 0.0313	0.2383 0.0428 0.0207

The following characteristics can be obtained by analysis:

1) The segmentation results of figure 1 (a)~(c) and figure 2 (a)~(c) show that, all of the proposed algorithm, genetic algorithm and random initial clustering center FCM algorithm can effectively segment the image; but as shown in figure 1 (d) ~ (e) and figure 2 (d) ~ (e), as well as table 1 and table 2, because of the sensitivity of the random initial clustering center FCM algorithm to the initial clustering center, the final clustering center of FCM is unstable, which affects the clustering results and the robustness is poor.

2) According to figure 3 and figure 4, as well as table 1 and table 2, when compared to the genetic algorithm, the optimal solution and clustering center of the algorithm proposed by this paper have smaller volatility and stronger robustness.

Conclusion

In this paper, a fuzzy clustering image segmentation algorithm based on GA and gray histogram is proposed to deal with the problem that the traditional FCM is easy to fall into local optimal solution and can not predict the number of clustering center. The number of cluster centers is determined by the number of peaks of gray histogram, and the statistical values of the gray level are used to reduce the amount of computation in the process of clustering. And the clustering center of the FCM is initialized by the global optimal solution of genetic algorithm. Experiments show that the proposed algorithm is more robust than the traditional FCM, and has a better segmentation effect.

Acknowledgement

In this paper, the research was sponsored by Scientific research project of Hebei Province (Project No. Z2012077).

References

[1] Li Wei. Research on image segmentation algorithm based on fuzzy C means clustering [D]. Harbin Engineering University. 2013.

[2] Zhao Yan. Research on image segmentation technology based on FCM algorithm [D]. Harbin Institute of Technology. 2012.

[3] Lou Yinxia, Cheng Ming, Wen Gaojin, et al. Image fuzzy clustering analysis based on FCM and genetic algorithm [J]. Computer engineering and Application, 2010 46(35) 173-176, 195.

[4] Bai Suqin, Hui Changkun, Wu Xiaojun, Wang Shitong. A fuzzy clustering algorithm based on genetic algorithm and its combination with FCM algorithm [J]. Journal of East China Shipbuilding Institute (NATURAL SCIENCE EDITION), 2001 15(6) 40-43.

[5] Liu Yu. Research and application of fuzzy clustering based on genetic algorithm [D]. Henan University. 2008.

[6] Ren Ziwu, San zhi. Research on the improvement of adaptive genetic algorithm and its application in system identification [J]. Journal of system simulation, 2006 18(1) 41-66.

[7] Tian Xiaomei, Gong Jing. A review of real encoding genetic algorithm [J]. Journal of Hunan Vocational College of Environmental Biology, 2005 11(1) 25-31.