http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (23): 7489-7494, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7489.7494

Medical Image Segmentation Based on Genetic Clustering Algorithm

[1,2]Peng Taile and [1]Zhang Wenjun
[1]School of Film and TV Arts and Technology,Shanghai University, 200072, Shanghai, China
[2]School of Computer Science and Technology,Huaibei Normal University, 235000, Huaibei Anhui, China

Abstract: In view of medical image segmentation greatly influenced by noise and low segmentation accuracy, a image segmentation algorithm based on genetic clustering is put forward. Firstly, according to sliding window we divide image into several sub-domains and using genetic algorithm searches the optimal population in each sub-domain. The following thing is the most optimal population is treated as clustering initial value of fuzzy c-means clustering algorithm for image segmentation. Finally, we obtain a number of better result of image segmented. Experimental results show that the algorithm has significant improvement to image segmentation relative to Otsu algorithm.

Key words: Sliding window, fuzzy clustering, initial clustering centers, image segmentation

INTRODUCTION

Image segmentation is one of the basic computer vision problems. Aim of image segmentation is to segment image into discrete barriers, different details and significance of parts on the basis of the physical characteristics of image.

In past few years some experts applied neural network, mathematical morphology, fuzzy mathematics and wavelet analysis to the image segmentation and obtained many meaningful research results. Nowadays, image segmentation has a wide range of applications in medical image processing. For example, it can be used to determine the focal lesions and to detect the size of focal lesions, etc.

At present, medical images from medical equipment (CT, MRI, PET, etc.) have more or less noise. The noise makes medical image segmentation more difficulty. Some image segmentation models based on region information, such as C-V model and LBF model, can reduce the effect of noise, which are easy to fall into local optimum (Chan and Vese, 2001; Li *et al*., 2008; Wang *et al*., 2012).

Based on the point of view of statistics, image segmentation can be understood in terms of pixels for the physical characteristics of the process of clustering and clustering algorithm can achieve good segmentation effect.

Usual clustering algorithm have K-means algorithm, hierarchical clustering algorithm, fuzzy clustering algorithm, spectral clustering algorithm (Xu *et al*., 2012), there are else many based on clustering algorithm, such as neighbors propagation clustering (Frey and Dueck, 2007; Mezard, 2007). As for the traditional clustering algorithm, such as k-means algorithm, EM algorithm (Khan *et al*., 2009), etc., when the sample space is a convex these algorithms are easy fall into local optimum.

Genetic algorithm is adaptive global optimization search algorithm based on biological evolutionary. In view of the global optimization of genetic algorithm characteristics, in this paper a genetic clustering image segmentation algorithm is put forward based on sliding window to solve the problem that fuzzy c-means clustering is easy to converge to a local minimum and even the clustering process is not in convergence.

First of all, according to sliding window we segment image into dynamic block and the pixel of window is optimized by the genetic algorithm to produce the initial clustering center, then it make fuzzy c-means clustering algorithm produce segmentation results.

Due to the size of the patch of image is smaller, the number of pixels is less, which is easy to make the genetic algorithm achieve the global optimal, even short of global optimal, it makes this effect is reduced to a small area in a piece of image.

RELATED ALGORITHMS

A Genetic fuzzy c-means clustering algorithm: Genetic Fuzzy c-means clustering algorithm is named as GFCM algorithm, which is easy to apply the characteristic of global optimization to Fuzzy c-means clustering and solve the Fuzzy c-means clustering algorithm that is easy to fall into local optimum problem.

Giving a data set D:

Corresponding Author: Peng Taile , School of Film and TV Arts and Technology,Shanghai University, 200072, Shanghai, China

$$D = \{D_1, D_2, \bullet \ldots D_r, \ldots, D_n\} \tag{1}$$

The data set D is divided into c classes. Every class contains a sample at least and degree of each sample depending on some class can be expressed by fuzzy membership. D_j is a random pixel in D, u_{ij} is membership of D_j corresponding to the I class membership D_j,a arbitrary pixel in D, to I kind of membership for u_{ij}, classification results can be expressed by using a fuzzy membership matrix U and U is $\{u_{ij}\}$,which is in [0,1]. Define the objective function:

$$J(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} \tag{2}$$

$$V = \{V_1, V_2, \bullet \ldots, V_c\} \tag{3}$$

Where:
C: No. of clusters;
U_{ij}: Degree of membership of D_j to the clustering center V_i:

$$\sum_{i=1}^{c} U_{ij} = 1, v_i = \sum_{j=1}^{n} u_{ij} D_j / \sum_{j=1}^{n} u_{ij},$$
$$d_{ij} = \left|D_j - v_i\right|$$

m: Weighted index, selecting the U and V make J (U, V) minimum in GFCM

Steps of algorithm are as follows:

Step 1: According to the genetic algorithms, the initial clustering center V is determined
Step 2: Compute matrix of membership U:

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}\right]^{-1} \quad j=1, 2 \text{ and } n \tag{4}$$

Step 3: Update the clustering center:

$$V_i = \frac{\sum_{j=1}^{n} (u_{ij})^m D_j}{\sum_{j=1}^{n} (u_{ij})^m} \quad i=1, 2 \text{ and } c \tag{5}$$

Step 4: Repeat step 2 and step 3, until making the Eq. 2 convergence

B Genetic algorithm: Genetic algorithms (Goldberg,1989) have become increasing important for researchers in solving clustering problems since they can provide feasible solutions in a limited amount of time. They were first proposed by Holland in 1975 and have been successfully applied to the fields of optimization machine learning, fuzzy logic controllers. When genetic algorithms are used to solve a problem, a representation that describes the problem states must first be defined. The most common form applied is the bit string. Next, an initial population is defined and genetic operations are performed to generate the next generation. The same procedure is then repeated until the termination criterion is satisfied. The simple genetic algorithm is described as follows:

Step 1: Define a suitable code system of the problem to be solved
Step 2: Define an initial population of P individuals for evolution
Step 3: Estimate fitness of the individuals
Step 4: Perform crossover and mutation operations to generate possible offspring
Step 5: Evaluate the fitness value of each individual.
Step 6: Select superior N individuals according to their fitness values
Step 7: If find the optimal solution, the algorithm Stops, otherwise, goes to Step 4

ALGORITHMS OPERATE

A Choice of initial population: According to the classification method of pixel gray value, the algorithm selects the initial population. Because gray-level histogram is a good way to describe the statistical characteristics of image pixel description, this paper uses the trough of the histogram to determine number of image clustering. To a image P, P_,$R^{I\times J}$, the pixels space D is for D = $\{D_1, D_2, \ldots, D_{I\times J}\}$.We use median filter to reduce the noise. then, image is converted to grayscale. According to sliding window image block can be selected from image, the pixel space of block is D_i, where: D_i is $\{D_1, D_2, \ldots, D_{M\times N}\}$,M<I,N<J. then, we Obtains gray-level histogram of each block, As shown in Fig. 1.

To image block P_r, its image gray value is in [G_{r1}, G_{rm}]. According to the probability density function of image grey distribution, pixels are divided into k classes. According to the gray value of pixels, there is a kind of class set C, that is $\{c_1, c_2, \ldots, c_k\}$,K value is in 2 to:

$$\sqrt{M \times N}$$

B Coding method: In this study the clustering center is coded in real number coding way and there are N

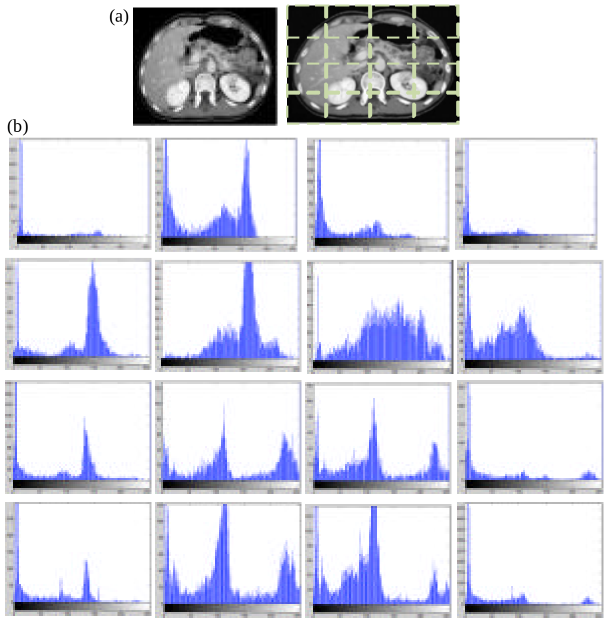

Fig. 1(a-b): a CT image is divided into several blocks, then we Obtains gray-level histogram of each block

individuals in a population. To each individual, it is a group pixels which is randomly selected from k classes of set C and it is a group of clustering center.

The number of clustering center k_i is changing and the largest number of clustering center is k_{max}, coding length of individual D_i is fixed and coding length is k_{max}. In medical ultrasound image segmentation, k_{max} tends to be 2.

C Determinate of fitness function: Genetic algorithm treats fitness function as the basis to direct the evolutionary search process and uses the fitness value of each individual guides search. So selection of the fitness function is very important, the function directly influence to the genetic algorithm convergence speed and whether we can find the optimal solution. Optimized clustering centers have the following characteristics: Between the cluster center and points in the same class has high degree of polymerization, Between the cluster center and points in another class has high degree of separation. According to the characteristics we design the following optimization target function Q (C).

Q (C) is a clustering validation index on measure of class in the polymerization degree and separation degree. Internal degree of polymerization with classes is expressed by square of the distance between two data.

$$\text{scatter}(x, y) = \text{Max}\Big(\sum_{D_x, D_y \in C_i} \|D_x - D_y\|^2\Big) \tag{6}$$

where, Scatter (x,y) is the sum of squares between any two data in class, $|D_x\text{-}D_y|$ is the Euclidean distance between point D_x and D_y.

Separating degree between classes is:

$$\text{Sep}(x, j) = \text{Min}\Big(\sum_{j=1}^{k} \sum_{D_x, D_j \in C_i} \|D_x - D_j\|^2\Big) \tag{7}$$

$$Q(C) = \text{Scat}(C) / \text{Sep}(C) \tag{8}$$

The optimal clustering result corresponds to balance point of in degree of polymerization of the class and the degree of separation between class (Sun *et al*., 2004; Fan and Wu, 2002), in numerical value which show that index function Q (C) should get minimum value.

D Selection strategy: The choosing operation of genetic algorithm may used to guide the search direction, which is very important to the clustering result. In a general way, genetic algorithm is easy to fall into local optimum .To local optimum of the *K* value, the genetic algorithm may lead to mistake. In this paper a kind of selection operator based on proportional selection and optimal preservation strategy is used. Using fitness function calculate the probability of an individual being selected and then using roulette wheel chooses the individual.

When k is unknown, the first is to acquire the optimal clustering number and then on the basis of the optimal clustering number k_b, the best clustering center will be found in $\{c_1^*, c_2^*, \ldots, c_{kb}^*\}$. Literature (He and Tan, 2012) discuss how seek for to the best clustering center in detail discusses populations consistency function. In the base of the Literature Literature (He and Tan, 2012), we define a convergence function H_{con}:

$$H_{con} = \frac{\sum n_i}{N} \tag{9}$$

where, n_i is individual of clustering number of no repeat. Assuming the fitness of the individual D_i is $f(D_i)$, the D_i choice probability function is:

$$p(i) = \frac{f(D_i)}{\sum_{j=1}^{N} f(D_j)} \tag{10}$$

In Eq. 10, f (D_i) is fitness function on D_i. Assuming the clustering number of individual D_i is $k_{i\ \text{and}}$ probability of D_i being chosen equals probability of k_i being chosen. If individuals have different clustering number, at the same time H_{con} Approaches to a given decimal ε, So under each k, having individuals with different class center go parallel evolution, which avoids bad results on account of choosing the fixed initial clustering center.

When H_{con} approaches to a given decimal ε, each population individual clustering number is k_b but the clustering center Z_i of these clustering is different. Select option for individual X_i is repeatedly chosen, finally the best clustering center Z_b is determined.

Preferred strategy is to save the best solution of the population to the next generation. In the evolutionary process, using the optimum individuals of preceding generations replace the worst individuals of new group, thus evolve into a new population. All of these can prevent genetic operation from destroying current the best individual in population fitness.

E Cross regulation: Before Crossover operation, N individuals in mating pool randomly match and form N / 2 pairs of individual. Crossover operation is done between two individuals of these matching individual groups.

GENETIC CLUSTERING ALGORITHM BASED ON SLIDING WINDOW

According to the genetic algorithm as mentioned above, firstly the global clustering center of image can be got, the following thing is that image clustering is completed by the fuzzy clustering method to achieve the purpose of image segmentation. So the genetic fuzzy clustering algorithm steps can be summarized as follows:

Step 1: Input image P (x, y), filter the noise and then transform image into grayscale. the following thing is initialization setting, including to define sliding window and input iterations G, cross rate P and population number N. The initial clustering center row vector K is $\{k_1,k_2,\ldots,k_N\}$.we preset normalized uniformity measure (NU) (Zhang, 1996),which is defined as δ_0

Step 2: Calculate the gray histogram of the image block. According to the gray value, these pixel points are divided into k class, there comes a kind of set C, which make up of $\{c_1,c_2,\ldots,c_k\}$. To set C, N group real number is randomly come into being and each group have k real numbers (k real numbers are k clustering centers), code string made up of the k real numbers is regarded as individual *D*, thus the initial group of scale N is formed

Step 3: According to the Eq. 7 ,fitness value of each individual is calculated

Step 4: Choose the best individuals to reserved to the next generation, these individuals participate evolution of population

Step 5: New generation individual are produced by the genetic operation. Calculate individual fitness value of the new group and check whether they meet the stop termination criterion. if they meet, operation will stop and turn to step 6, otherwise, turn to step 4

Step 6: Treat The individual with the smallest fitness value in last generation of group as the result. the k value and Z value of individual are decoded as further fuzzy clustering initial clustering center number and cluster center

Step 7: According to Eq. 3 and 4 fuzzy clustering will be going on. On the basis of membership function value of each point of the image, we can finish image segmentation and calculate NU value δ. If δ is greater than δ_0,the algorithm will be finished. Otherwise, the algorithm turns to step 2

EXPERIMENTAL RESULTS AND ANALYSIS

In order to prove the effect of image segmentation, we select several CT images of the liver, lungs and kidney as inputting image. The computer configuration for this experiment is as follows: Inter core2 Quad CPU Q8300, 4 GB memory and Matlab2010b. Our algorithms in this paper compare with frequently-used Otsu algorithms based on medical image segmentation (Otsu,1979).

On subjective visual evaluation:

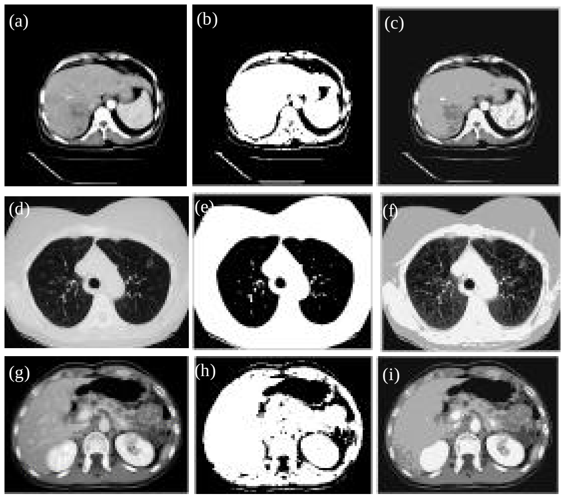

Fig.2(a-i): Result from Otsu algorithm and our algorithm, (a) original liver picture (b) Otsu algorithm (c) our algorithm, (d) original lung CT picture (e) Otsu algorithm (f) our algorithm, (g) original kidney CT picture (h) Otsu algorithm and (i) our algorithm

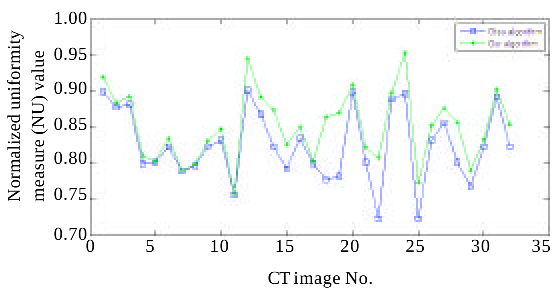

Fig. 3: This algorithm compare the Otsu algorithm in NU

From Fig. 2 and 3 we can see, to the liver, lung CT image segmentation, two kinds of algorithm has achieved good effect. On the reserving details of image, our algorithm works better than Otsu algorithm, texture of organ organization is reserved perfectly, The result of image segmentation have better level feels, which is very important to sure lesions and also easy to use objective evaluation parameters to evaluate the quality of objectivity after segmentation.

The effect of image segmentation is measured by the regional consistency (NU), To the liver CT image, lung CT images and kidney CT image, In Fig3 NU value from two different segmentation method can be drawn a conclusion that our algorithm is better than the Otsu algorithm. This is mainly because speckle noise and texture feature of medical ultrasonic image is very stronger. To these images, image segmentation algorithm based on the fuzzy clustering method is better than traditional algorithm based on probability statistical method. finally, subjective evaluation and objective evaluation made a basic consistent result.

CONCLUSION

Traditional clustering image segmentation algorithm based on fuzzy c-means clustering on the initial clustering point is sensitive, the initial clustering point selection directly influence the result of image segmentation. In the algorithm what we told, image segmentation is transformed into local image segmentation by using sliding window, which reduce the initial clustering point to select the whole image of the segmentation results of influence. In addition, the use of genetic algorithm increases the algorithm converges to the global optimal solution of the probability, further gains the fuzzy c-means clustering the initial clustering point. Results show that the algorithm has good image segmentation effect and higher efficiency, more image details are retained , in order to further determine the focal provides reference.

ACKNOWLEDGMENT

This work was financially supported by the University Science Research Key Project of Anhui China (Grant No.KJ2010A304).

REFERENCES

Chan, T.F. and L.A. Vese, 2001. Active contours without edges. IEEE Trans. Image Process., 10: 266-277.

Fan, J. and C. Wu, 2002. Clustering validity function based on possibilistic partition coefficient combined with fuzzy variation. J. Electron. Inform. Technol., 24: 1017-1021.

Frey, B.J. and D. Dueck, 2007. Clustering by passing messages between data points. Science, 315: 972-976.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Pub. Co., New York, USA., ISBN: 9780201157673, Pages: 412.

He, H. and Y.H. Tan, 2012. A novel clustering method based on dynamic genetic algorithm. Acta Electronica Sinica, 40: 254-259.

Khan, J.F., R.R. Adhami and S.M.A. Bhuiyan, 2009. A customized Gabor filter for unsupervised color image segmentation. Image Vision Comput., 27: 489-501.

Li, C., C.Y. Kao, J.C. Gore and Z. Ding, 2008. Minimization of region-scalable fitting energy for image segmentation. IEEE Trans. Image Process, 17: 1940-1949.

Mezard, M., 2007. Where are the exemplars? Science, 315: 949-951.

Otsu, N., 1979. A threshold selection method from gray-level histogram. IEEE Trans. Syst. Man Cybern., 9: 62-66.

Sun, H., S. Wang and Q. Jiang, 2004. FCM-Based model selection algorithms for determining the number of clusters. Pattern Recognit., 37: 2027-2037.

Wang, S.F., X.X. Ji, J.W. Zhang, Y.J. Chen, L. Fang and T.M. Zhan, 2012. Improved coupled model for MR images segmentation and bias restoration. J. Image Graph., 17: 1175-1180.

Xu, X.L., Z.M. Lu, G.S. Zhang, C. Li and Q. Zhang, 2012. Color image segmentation based on improved affinity propagation clustering. J. Comput.-Aided Des. Comput. Graph., 24: 514-519.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation Pattern Recognit., 29: 1335-1346.