Research on Volume Segmentation Algorithm for Medical Image Based on Clustering

Li Xinwu
Electronic Business Department,
Jiangxi University of Finance and Economics,Nanchang,Jiangxi,330013, China
Liyue7511@163.com

Abstract

Direct 3D volume segmentation is one of the difficult and hot research fields in 3D medical data field processing. Using K-means clustering techniques, a new clustering segmentation algorithm is presented. Firstly, According to the physical means of the medical data, the data field is preprocessed to speed up succeed processing. Secondly, the paper deduces and analyzes the clustering and segmentation algorithm and presents some methods to increase the process speed,including improving cluster seed selection, improving calculation flow, and amending pixel processing and operational principle of algorithm. Finally, the experimental results show that the algorithm has high accuracy when used to segment 3D medical tissue and can improve process speed greatly.

1. Introduction

The segmentation of 3D medical data field has always been an extremely challenging subject due to imaging principle, fuzzy tissue and other factors. In the past more than 20 years, people had addressed a large number of segmentation algorithms. However, the complexity of human body structure, irregularity of tissue organs as well as difference among different individuals makes the segmentation of medical data field no common theory so far. Although the segmentation of 3D medical data field is very difficult, it is one of the key technologies for data field processing and system analysis and understanding, and an extremely important step of data field visualization. Only can accurate segmentation of the data field obtain reasonable models for subsequent renderings. It can be said that the realization of 3D visualization of the medical data field is to carry on correct and reasonable segmentation[1] of the image data at first.

Haralick and Shapira regard image segmentation as a clustering process. The clustering means mathematically that a large number of d -dimensional data samples (n units) are clustered into k classes ($k << n$) so as to maximize the similarity of the samples in same class and minimize the similarity of the samples in different class. The clustering process is to make continuous classification of the data objects containing several attributes by the clustering algorithm automatically, and the data is cut into several classes by the identification of the data characteristics. Therefore, the algorithms can be explored by the clustering rules and the clustering basis of various targets found[2,3,3,4,5,6,7], and then based on which, the image is identified and segmented.

K-means algorithm is a basic division method in the clustering methods and has better scalability, so it is widely used. In addition, while this algorithm that takes error square and criterion function as the clustering criterion function involves the clustering result into local solution easily and make it dependent on the initial value, a large amount of 3D medical image data causes bad algorithm timeliness. In view of these K-means algorithm shortcomings, a newly improved K-means algorithm of medical image volume segmentation is provided on the basis of the thinking of seeking the optimal initial value and improving algorithm's procedure.

2. Data field pretreatment

Each voxel's gray value (or color value) is given according to the people's habits or the users' requirements, and not owned by substances. Therefore, the difference of adjacent data has a certain meaning while the absolute value of each data is of no importance in the data field. This algorithm suggests that the function values of original 3D data field is within the range of 0-255 in the normalization integration, and processed data replaces original data to give the gray level value so as to provide a gray level field for the feature extraction, decrease the post-treatment memory demand and improve the post-treatment speed.

Although it is not simple for such problems as the time consumption that the data field turns 16-bit and 12-bit gray level images into 8-bit under the premise of keeping up the key information of the images in the process of the normalization pretreatment, the process is over in data format conversion of the pretreatment and any data field will be treated only once so as not to affect the efficiency of the whole algorithm.

978-0-7695-3488-6/08 $25.00 © 2008 IEEE
DOI 10.1109/KAM.2008.34

3. K-means algorithm principle

Steps for K-means clustering algorithm are[4,5] (see Fig.1):

(1) Select n objects as the initial cluster seeds on principle;

(2) Repeat (3) and (4) until no change in each cluster;

(3) Reassign each object to the most similar cluster in terms of the value of the cluster seeds;

(4) Update the cluster seeds, i.e., recompute the mean value of the object in each cluster, and take the mean value points of the objects as new cluster seeds.

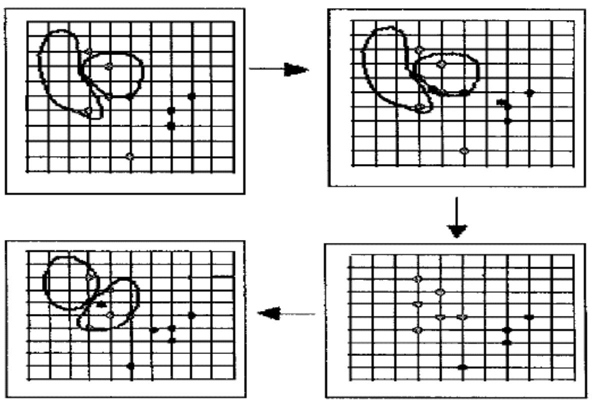

Figure 1 K-means algorithm procedures

3.1. Improving cluster seed selection

When calculating the K turn of clustering seeds with the improved algorithm, those data in the cluster having a great similarity to the K-1 category seeds should be adopted to calculate their mean points (geometrical center) as the clustering seed of the K turn and the specific calculation method is below:

(1) For the cluster $c_{i(k-1)}$ obtained through the K-1 turn of clustering, the minimum similarity $sim\_\min_{i(k-1)}$ of the data in the cluster to the clustering seed $s_{i(k-1)}$ of the cluster is calculated;

(2) The data in the cluster $c_{i(k-1)}$ is calculated that has a similarity of more than $1-\beta*(1-sim\_\min_{i(k-1)})$ to the clustering seed $s_{i(k-1)}$ (among, β is a constant between 0-1), and the data set is recorded as $cn_{i(k-1)}$;

(3) The mean points of the data in $cn_{i(k-1)}$ are calculated as the clustering seed of the K turn.

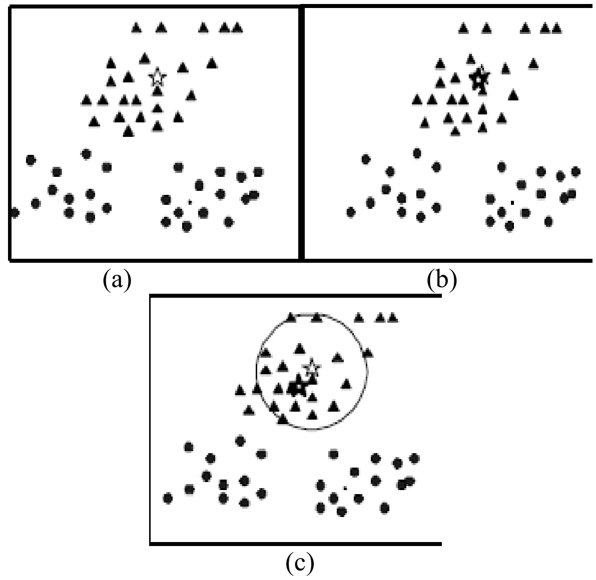

Figure 2 Comparing Pictures of k-means Algorithm and Its Improved Algorithm

In Fig.2, (a) shows cluster i of the K-1 turn and its seeds, (b) shows cluster i of the K turn and the new seeds (initial algorithm), and cluster i of the K turn and the new seeds (improved algorithm).Indication of the symbols in Fig 1 can be listed as follows: ▲ means data point i in the cluster, □ means seed of the cluster i in the K-1 turn, □ means new seed the cluster i in the K turn, ● means other data points, ○ means the points within its range are used to calculate the new seeds.

As seen in Fig. 2, the new clustering seeds are obviously moving toward the data intensive zone. The improved algorithm could achieve a good clustering effect on the cluster sets containing isolated points. For the processing of big sets, this improved algorithm, as same as k-means algorithm, is relatively flexible and high-effective. Its time complexity is $O(nkt)$, of which, n is the number of all objects, k is the number of the clusters, while t is the iteration number of the algorithm, and generally, $k \langle\langle n$ and $t \langle\langle n$.

3.2.Improvement of algorithm flow

The general K-means algorithm is a gradient ascent iteration algorithm, each time of iteration could cause the corresponding increase of the target function values, and the iteration might be ended in the limited steps. However, such an algorithm also has some disadvantages, for example, the algorithm is easily trapped in the local maximum solution and such a solution depends on the selection of initial partition. Therefore, the means algorithm is used as local searching process to be inlaid in the local search structure of the iteration in order to obtain better clustering results through the relationship between balancing the reinforcement of the local search and extending the searching range.

In the image clustering problems, D neighborhood of a partition refers to the partition obtained through randomly selecting D different clusters in a certain partition and redistributing them into other clusters. In other words, the neighborhood of the current partition means the partition obtained through randomly selecting one cluster and redistributing it into other cluster. The calculation flow of the K-means-based iteration clustering algorithm of local searching clusters is displayed in the following.

Input: the number k of the results' clusters, containing the data set of N clusters.

Output: k clusters, ensuring that the clusters in all clusters are similar or correlated.

Step 1 Randomly select an initial partition $P_k=\{C_1,C_2,..C_k\}$ and calculate the corresponding concept vector $c(C_i), i=1,2,..k$, then initialize the current maximum target function value $fopti$ and determine the ending conditions of the algorithm, the parameter value $\varepsilon(\varepsilon \succ 0)$ receiving the conditions and the maximum iterating times n that the target function value is not improved any more.

Step 2 Repeat.

Step 3 Perform the local search on P_k with the means cluster clustering algorithm to obtain a local maximum target function value $fopt$ and its corresponding partition P_k^*.

Step 4 If $fopt \succ fopti$, the current best partition is $P_k^{'} = P_k^*$, $fopti = fopt$, the current partition is not improved any more, and the iteration times $t := 0$.

Step 5 Repeat

Step 6 Randomly generate a cluster $x_i (i = 1,2,...N)$ and repeat the following processes:

(1) If x_i is beyond the tabu list, it will be redistributed into other cluster to calculate the increase Δf of the target function value and the times of iteration without improvement is $t : t = t + 1$; while if x_i is in the tabu list, Step 6 will be repeated.

(2) If $\Delta f \succ \varepsilon$, P_k is the partition of redistribution, the target function value is $f_{opt} = f_{opt} + \Delta f$, x_i is added into the tabu list, and the tabu length of other tabu objects is deducted 1.

(3) If $fopt \succ fopti$, $fopti = fopt$, $P_k^{'} = P_k$, and the times of iteration is $t := 0$.

(4) If x_i is tested throughout all clusters and the times of iteration without improvement is $t \prec n$, Step 6 will be repeated.

Step 7 "Until $t = n$" means there is no improved partition generated in the successive n times of iteration.

Step 8 Randomly select several clusters from P_k and redistribute them into other clusters to obtain the new partition P_k.

Step 9 Until the ending conditions are met.

3.3. Pixel processing and operational principle of algorithm

In order to quicken the efficiency and the ability of the algorithm to process the large-scale data field, the pixel operation and processing have to comply with the several following principles:

(1)Pre-segment the data field in the phase of the data field segmentation pretreatment, i.e. use the methods of manual interaction and model guidance.

(2)According to prior knowledge of the structure shape and the position that medical data field dissected the tissue, give the interactive definition to several seed points and take these seed points as the initial samples.

(3)According to the probability distribution of an established characteristic, directly classify the pixel points that the selected obvious characteristic belonged to a seed point, namely the points that have the obvious characteristics and definitely belong to a class will be marked as a class directly, and not be calculated.

(4)Calculate the points that have no obvious characteristics and are classified strictly through mathematical algorithm, namely the points that are possible to belong to different classes only for the edge region or the edge transitional region carry on the algorithm operation and segmentation.

(5)As for these points to be calculated, use the dot interlaced sampling to carry on the sampling calculation in the space. Namely make the sampling to calculate whether a point belongs to A class from the surrounded seed point A. When the calculation of the n point finishes, next point to be calculated will be selected as n+2 but not n+1 according to the space order if the n point belongs to A class; if the calculation result is that the n+2 point belongs to A class, the n+1 point will be fallen into A class directly; if the calculation result is that the n+2 point does not belong to A class, the attribute that the n+1 point directs towards A class will be calculated repeatedly.

Such this reduces the blindness of the defined initial sample points greatly so as to enhance the accuracy of the segmentation, and also reduces the data quantity calculated by the algorithm greatly to enhance the algorithm efficiency.

4. Experimental verification

This paper's algorithm performs the experiment on PC. PC configurations are P4 1.8G CPU and 512M memory. In this paper, the algorithm is simulated in the utilization of the data sets with different sizes and various distributions. Due to the limited space, the following only introduces the segmentation results of actual data fields measurement for a group of complex medical organization MRI data field. The date field size is $128 \times 128 \times 128$. The experiment result shows the comparison with general K-means algorithm[5]. (1) for segmentation effect (see Fig. 3). Fig3. (a) is original image, Fig3. (b) clustered and segmented by general K-means algorithm[4] lacks many details obviously, and Fig3. (c) segmented by this paper's algorithm has high segmentation precision and good visual effects; (2) this paper's algorithm is also improved significantly from the algorithm efficiency (see Tab. 1).

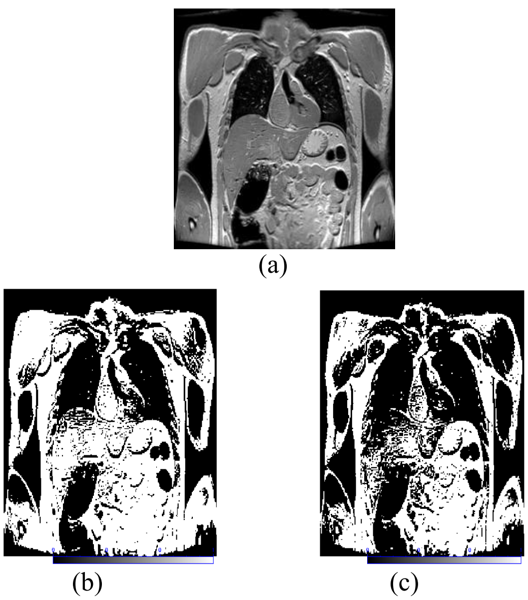

Figure 3 Original image and segmentation effect image with different algorithms

Table 1 Segmentation efficiency and accuracy of different algorithm

Algorithm	This paper's algorithm	General K-means algorithm
Time consumption (s)	19	62
Segmentation accuracy	95%	78%

5. Conclusion

General K-means clustering algorithm is vulnerable to the local solution in practical application, so this paper fully utilizes prior knowledge of the segmentation object to perform several pretreatments in the course of detailed computation by the thinking of seeking the optimal initial value in several samplings and one clustering as well as the improved K-means clustering algorithm procedure. As a result, large reduction of the processing units and great improvement of the algorithm anti-interference make the algorithm improve not merely convergence speed but also segmentation accuracy. Besides, the practical application of K-means clustering segmentation algorithm is greatly improved in 3D medical data field segmentation.

Acknowledgement

The paper and the research are supported by the education department of Jiangxi Province(2007259).

References

[1] Y.Cheng and F. Huang (2007). "A Coupled Minimization Problem for Medical Image Segmentation with Priors". *International Journal of Computer Vision*. Springer Netherlands, pp.259-272.

[2] Haralick R.M.,Shapiro L. G (1985)."Survey:Image segmentation techniques". *Computer Graphics and Image Processing*. Vol. 29, No. 4, pp.100-132.

[3] Bradley P. S., Managasarian L. (2000). "K-plane clustering". *Journal of Global Optimization*. Vol. 16, No. 1, pp.23-32.

[4] Tang Y.,Rong Q. S (2004). "An implementation of clustering algorithm based on K-means". *Journal of Hubei Institute for Nationalities*. Vol. 22, No. 1, pp.69-71.

[5] Jiao C.L., Gao M.T., Shi Y. K (2007). "Image clustering and segmentation based on Improved clustering neural network". *Computer Engineering and Application*. Vol. 43, No. 20, pp.93-96.

[6] Yi S (2006). "Global optimization for NN training". *IEEE Computers*. Vol. 19, No. 3, pp.45-54.

[7] Zhang Y.F., Mao J. L. (2003). "An improved K-means algorithm". *Computer Application*. Vol. 23, No. 8, pp.21-23.