

Segmentation of MRI Volume Data Based on Clustering Method

Dongsheng Ji^{1,2,a}, Peng Chen, Chi Zhang³,

¹Gansu Radio & TV University, Lanzhou, 730000, China,

²School of information Science & Engineering, Lanzhou University, 730000, Lanzhou, China,

³State Grid Gansu Electric Power Company, Lanzhou, 730000, China.

Abstract. Here we analyze the difficulties of segmentation without tag line of left ventricle MR images, and propose an algorithm for automatic segmentation of left ventricle (LV) internal and external profiles. Herein, we propose an Incomplete K-means and Category Optimization (IKCO) method. Initially, using Hough transformation to automatically locate initial contour of the LV, the algorithm uses a simple approach to complete data subsampling and initial center determination. Next, according to the clustering rules, the proposed algorithm finishes MR image segmentation. Finally, the algorithm uses a category optimization method to improve segmentation results. Experiments show that the algorithm provides good segmentation results.

1 Introduction

Nuclear magnetic resonance (NMR) imaging technology can provide anatomical information of the heart including different fault plane image information. Compared with other medical imaging modalities, such as tomography x-ray radiography, echocardiography, the cardiac muscle tissue has a higher resolution in MR imaging. As it is a non-invasive technique, MR imaging poses no harm to the human body [1]. NMR technique can be used to effectively prevent and diagnose heart disease. As the left ventricle (LV) pumps oxygenated blood from the lungs to the rest of the body, LV plays an important role in cardiac function, and is the focus of our research program. Left ventricular surface reconstruction is based on inside and outside LV MR image contour segmentation, and can provide border restrictions on myocardial motion analysis. Therefore, quick and accurate segmentation of LV profiles is of great significance [2, 3]. Mainstream methods of left ventricular myocardial segmentation are based on edge [4, 5]. For example, Snake model, Level Set model, and Geometric active contour model are all improved models, but the mastoid muscle is often present and interferes with the MR image, due to local gradient maximum regions, weak edges, artefacts and other phenomena, resulting in a lot of difficulties in image segmentation methods based on edge. To solve these problems, an image segmentation model based on image priori knowledge was introduced to improve the robustness of the model. Thus it appears that newer methods are based on prior knowledge constraints and regional approaches. Recent studies [5,6] introduced the physiological cardiac structure constraint model to control the relative position of the LV internal and external contours, and introduced regional information to provide a category of information for image segmentation.

Other studies such as Paragios' paper[7] have developed algorithms on the basis of geometric active contour model, which combined the edge of the image and the regional information and physiological structure constraints. Split left ventricular internal and external contour.

Paragios method uses Gaussian mixture model fitting the histogram to obtain regional information; however, EM algorithm for Gaussian mixture model has some dependence on the initial parameter estimates; therefore it may be stuck in local maxima. Paragios' method also has some limitations, for example, the use of GVF (Gradient Vector Flow) [8,9] instead of gradient field guide evolution to the edge of the curve, which can expand the model capture range. [10] The performance of GVF will be affected for uneven greyscale images or noisy images[11, 12]. At this point, the evolution of the curve from weak border into another area causes border leaks or local maxima in the gradient within the region and stops at the edge of the isolated region, and it cannot move to the real border. In addition, the curve evolution formula in Paragios method is complex, and, in order to assure its stability, requires special numerical methods, so, the numerical computational complexity is higher. Through the analysis of the various methods stated above, in this paper, from the view of simplicity and practicality, we put forward an interior and exterior contour automatic segmentation method of the LV.

In this paper, proposed automatic segmentation method of left ventricular internal and external contours. Short axis image of left ventricle myocardial inner and outer profiles approximate rounded, firstly, using Hough transform to automatically locate initial contour of the left ventricle; Then, based on geometric active contour model, K-means clustering algorithm to be applied to a coarse segmentation of objects in an image, provides regional information and cardiac Physiology structure

^a jids8012@163.com

constraint, establish coupled curve evolution equations of myocardium interior and exterior contour, to realize the automatic segmentation of left ventricle.

2 Hough transform algorithm to describe MR left ventricular image edge

The extraction of features from shape of MR image is one of the principal responsibilities of an effective features descriptor. In this work, we proposed a contour descriptor, which characterizes local and global features extraction of left ventricular internal and external contours. Local features are extracted using Hough transform for describing the contour of the MR left ventricular. The descriptors are explained in the subsequent subsections. Hough transform is an effective technique for detecting lines from an MR image, which is widely used in computer vision and pattern recognition. Hough transform maps a line in the spatial domain to a point in the Hough parametric domain, The effectiveness of the Hough transform based local features that provide small variations in histograms for intra-class and large variations for interclass is demonstrated.[13, 14,15].

3 Incomplete K-means clustering algorithm

3.1 K-means clustering algorithm

Clustering is a certain similarity criteria to divide the data into groups having similar properties of subclasses, and internal similarities greater than the similarity between such. Measurement of distance or similarity is the basis of clustering algorithms.

K-means clustering algorithm is one of the most common iterative operator with adjustment K-means center of mass[7]. K-means is a clustering method based on squared error, is also very famous hard clustering algorithm. The algorithm is simple and fast, and as with other clustering algorithms, K-means is an iterative optimization process.

Assuming that the sample set $X = \{x_1, x_2, \dots, x_N\}$ are divided into K classes, first, choose an initial separation for K classes, and calculate the mean vector μ of these classes, and then the remaining each sample by Euclidean distance is assigned to the closest one.

$$d(x_i, u) = |x_i - u_i|. \quad (1)$$

Recalculated mean vector is assigned to each class as the new class of centers. Repeat this process until the mean vector μ converges.

Central idea of K-means clustering algorithm is to minimize the total distance in class. All methods are re-adjusted the cluster centers and the allocation of the sample to the nearest centroid in the same category:

$$E = \sum_{j=1}^k \sum_{x_i \in W_j} \|x_i - \mu_j\| \quad (2)$$

Algorithm steps are as follows:

Step1: Set the maximum number of iterations and the minimum allowable deviation threshold, initialize all class center $\mu_j^{(0)}$;

Step2: Distribute x_i to the nearest centroid $\mu_j^{(n)}$ represented clustering;

Step3: Calculate the new center of mass $\mu_j^{(n+1)}$ and $E^{(n+1)}$;

Step4: Repeat Step 2 and Step 3, until K is less than the threshold, or when they reach the maximum number of iterations.

K-means belongs to unsupervised learning methods. In the application of image segmentation, generally the target image is divided into different classes according to the gray value or combining histogram for image clustering segmentation. Image pixel clustering in the feature space is found in the spatial characteristics of similar pixel point group. According to the different methods of selecting initial cluster centers, sample allocation rules, centroid calculation method, and stopping criterion is different, as there are several different versions of the algorithm. When the K-means algorithm for clustering, the initial cluster centers are generally randomly generated, the number of clusters are also manually set. Several methods of choosing clustering center are possible: (1) select special samples, (2) select K samples; (3) arrange all samples according to the distance, choose distance equal sample; (4) select maximum distance between samples. Many new K-means algorithms are presented to automatically determine the number of clusters and the initial cluster centers.

In this paper, the traditional K-means clustering algorithm is put forward with a few improvements: First, use the sample space streamlined ways to reduce a single iteration time-consumption; secondly, use a simple method to determine the initial clustering center; an incomplete clustering method is proposed to further improve the efficiency of clustering; and finally, the classification method to optimize the clustering results to get rid of some small area in cluster results.

3.1 Streamline the sample space

At present, there are some main sample cutting strategies [8]: One is based on image pixels, image blocking, with a mean or median of block as the new samples. Another strategy is based on image feature, which have the same characteristics of samples (pixels) into one class. The sample cutting strategy of this paper is to improve in the second method. We initially block feature space, divide pixels by their characteristics into a block of cells, and, the cell's value to reflect such characteristics of the number of pixels. Compared with traditional K-means clustering in the process of the results of all samples for processing, this algorithm can significantly improve the efficiency of clustering.

3.2 Determine clustering center

There are many optimization methods in the choice of K-means cluster centers which deal with image segmentation issues and achieve better results. These methods include: LINDE-BUZO-GRAY[9], (LBG), max-min distance-based (MMX) [10], Density-based methods (DEN) [11], and Subset Farthest First (SFF) [12]. In this paper, we give the method to determine the initial center in the previous sample based on streamline, the maximum minimum distance method is introduced into density as a measure, the initial clustering center K choose sample K features of highest density sample, then, iteration to determine the initial clustering center.

Problem: Data set $X = (x_1, x_2, \dots, x_n)$ is divided into classes, making the error sum E of squares of the clustering is minimal.

Define 1: The data object $x = (x^1, x^2, \dots, x^p)$ and $y = (y^1, y^2, \dots, y^p)$, the distance between them is:

$$d(x, y) = \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)^2 + \dots + (x^p - y^p)^2} \quad (3)$$

3.3 Incomplete clustering

The traditional K-means clustering method in the process of cluster is to compute the distance between sample and each cluster center. As most of the sample is very obvious, this can be a way to quickly get it belongs to class, avoid to calculate the distance to each cluster center, and to shrink the clustering of time. Proposed algorithms will cluster near the center of the sample quickly to this clustering. The main idea is to proceed from the cluster center; cluster center R neighborhood of samples to the belonged class, it did not appear in any sample clustering center in the neighborhood, iteration to the clustering center no longer changes it due to recent class.

Define 2: Assume a point q in m-dimensional space $Z = (z_1, z_2, \dots, z_m)$, S is am-Korekata, V is a multi-dimensional region, if the following conditions are met:

- (1) q is the center of S .
- (2) R is a positive integer.
- (3) Value area for $[q[i] - R, q[i] + R]$,
- (4) V is a non-empty cell of S does not contain a q .

So, V is for q field (q, R) , expressed in $T(q, R)$.

For the sample x , if it is located within the R field of its cluster center, it is attributed to m-centric class. The size of the field will become the important factors influencing the clustering effect. Determining ideal field radius is very complicated; as a result, this paper puts forward three kinds of approximate methods to determine the domain size R. When ith cluster, clustering center

for $\phi^i = \{\phi_1^i, \phi_2^i, \dots, \phi_k^i\}$, jth clustering radius from the center for $R^{i,j}$, Three methods of computation formula is as follows:

1)Adjacency minimum distance:

$$R^{i,j} = \frac{1}{2} \cdot \min(\|\phi_m^i - \phi_j^i\|); 1 \leq m, j \leq k, m \neq j \quad (4)$$

2) Adjacent average:

$$R^{i,j} = \frac{1}{2} \cdot \min(\|\phi_m^i - \phi_j^i\|); 1 \leq m, j \leq k, m \neq j \quad (5)$$

3) Global average:

$$R^{i,j} = \frac{1}{2C_k^2} \sum_{1 \leq m, j \leq k, m \neq j} C_k^2 \|\phi_m^i - \phi_j^i\|^2 \quad (6)$$

The above are three ways to determine the domain radius R , it shows experiments on more than one image to get R .

3.4 Clustering optimization

Using “incomplete” K-means clustering method for image segmentation [13], clustering is based on features, and completely ignores the space information of image, therefore the clustering results have many small areas. This paper presents an optimization method of the classification result, this approach combines regional information, small areas on the clustering segmentation has been removed, and obtains good effect. The main idea is to use the method of hash $R \times R$ neighborhood of the statistical sample x that belongs to the range of sample number, again will belong to class x belongs to sample number compared with a threshold, if greater than threshold do not perform any operation; otherwise the sample x to neighborhood sample belongs to class most of the class.

Due to traverse the entire optimization process is based on the classification results, only in a single step to statistics class predetermined belongs to the neighborhood, so the entire complexity of clustering algorithm approach to $O(n)$. Compared to other more complex clustering algorithms, this algorithm can greatly improve the efficiency of clustering optimization.

4 Left ventricle initial positioning of the inside and outside profile by Hough transform

In order to overcome the initial contour effects on geometric active contour segmentation results, Hough transformation [15] is used to estimate the initial internal and external contours. Hough transformation is a method of detecting the resolution defined shape of the image [14]. Hough transformation can be based on the prior knowledge of the approximate model to search the exact shape of the objects not known.

Using left ventricular internal and external contours prior shape similar to the circle, we adopted Hough transformation to estimate LV internal and external contours. In the process of LV systolic, when the LV in contour approximates a circle because of the influence of the mastoid muscle, there may be undetectable left heart interior contour phenomenon. In the course of left ventricular systolic, when the inner contour of the left ventricle does not approximate a circle, due to the effect of mastoid muscle, the image may appear less than left ventricular contour detection. We used Hough transformation in detecting left ventricular

contour and then center of the outline of MR imaging parameters to estimate the radius of inner contour of the initial LV.

Left ventricular contour edges and greyscale consistency is better than the outer contour, so, left heart inner contour segmentation dependence on the initial contour is much smaller and does not need to be as accurate as outer contour. The initial outline to more accurately locate the real edge is close and the ease curve evolution converges to the correct contour.

Hough transformation can identify partial deformation or partially occluded objects, but, it has a feature of the large amount of calculation [16,17]. To reduce computational complexity, improve the accuracy of applying prior knowledge, the long axis and short axis plane intersection point must be within the ventricle, node in the left ventricular internal and external contours around the center point. Centered on the intersection a range of window (the window size by MR imaging parameters, can be obtained from MR DICOM image file header), calculate the Hough transformation in this window.

Hough transformation makes use of round transformation formulas for $(x - a)^2 + (y - b)^2 = \rho^2$, parameter space is (ρ, a, b) .

This initial contour algorithm:

Step 1: According to MR imaging parameters to estimate the range

$[R_{min}, R_{max}]$ of outdoor profile radius of left heart;

Step2: Hough transformation for circle radius in the

$[R_{min}, R_{max}]$, record the number of Hough transformation to detect under different radius of circle;

Step3: Take minimum radius as left ventricular outer contour radius (If $n_r > 1$ the center take close to the longest axis and minor axis plane intersection center averages);

Step 4: Taking outer contour center combined by MR imaging parameters estimation within the outline of the radius, initialize the left ventricle contour.

Manually set $[R_{min}, R_{max}]$, using the loop method or gradient descent algorithm to get Hough transformation position left ventricular initial contour.

The effectiveness of the Hough transformation is based on local features that provide small variations for intra-class and large variations for interclass is demonstrated in Fig.1. Sobel operator is applied to extract the left ventricular contour as reference, with the Hough transformation algorithm initialization profiles

Hough transformation strong anti-blocking methods in this section to MR short-axis images are tagged as effective, as they showed the effectiveness of the Hough transformation based on local features that provide small variations in histograms for intra-class and large variations for interclass as demonstrated in Fig.2. Sobel operator is applied to extract the left ventricular contour as reference, with the Hough transformation algorithm initialization profiles.



(A) MR short-axis images. (B) Sobel operator to extract the edge. (C) Initializes the internal and external contours.

Figure 1. Hough transformation position initial contour.



(A) Tagged MR short-axis images. (B) Sobel operator to extract the edge. (C) Initializes the internal and external contours.

Figure 2. Hough transformation position Tagged MR initial contour.

5 Analysis of the left ventricle MR images

Regional information provides target category information for segmentation model that is mainly clustering method based on statistical models (Gaussian mixture model) or other clustering methods based on distance to perform target classification.

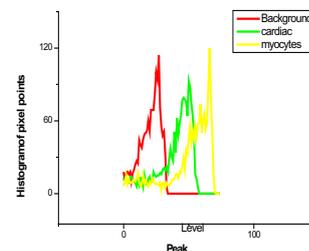
Paragios obtained regional information by Gaussian mixture model fitting histogram [18] and the K-means algorithms were used to classify image target, and obtain regional information. According to left ventricular intimal area, myocardial region and background region, Paragios' method put MR image into three peak images. With three Gaussian distribution fitted histogram, then, follow the Gaussian function obtained by fitting the calculated probability of category area.

Different MRI also varies and the majority of hospital MR images are as shown in Fig.3. Inconsistent greyscale of the image and the histogram are not an ideal distribution of the three peaks, and EM algorithms for Gaussian mixture model estimation of initial parameters have certain dependencies.

Fig.4 shows the comparison of two methods of classification and three Gaussian distribution (green, red and blank).The classification results show that, although the K-means algorithm to sort results; however, due to myocardial grey scale similar to the background, left ventricular myocardium and the background are divided into different categories. This causes the outer contour of the left ventricle segmentation errors as external contour segmentation without using region information is used in this method, instead of using cardiac physiology constraints as prior information. If there are any deviation estimates of initial value and true value, EM algorithm might stall at some local maximum, so incomplete K-means algorithm are better than the algorithm of classification.



(A) MR histogram of the three peaks.



(B) The histogram corresponds to three peaks.

Figure 3. Ideal MR histogram of the three peaks.

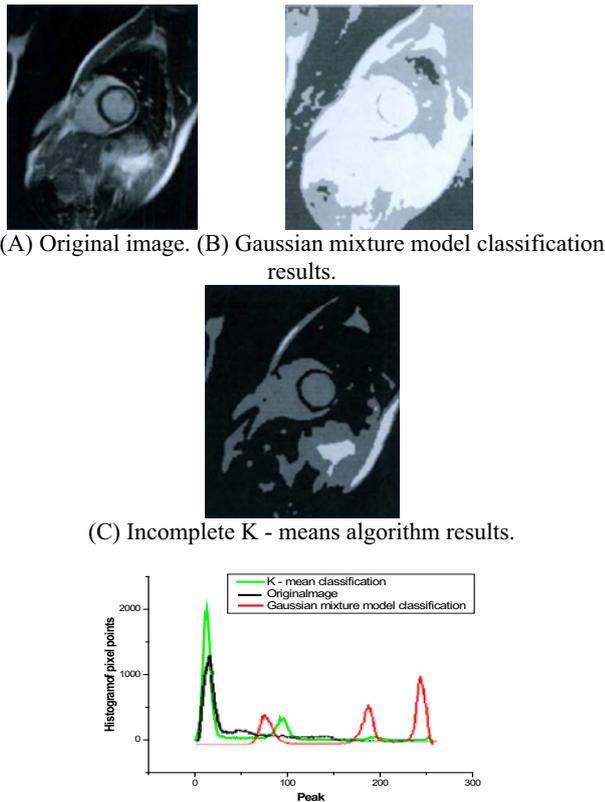


Figure 4. MR images of left ventricle classification results.

In order to compare the validity and robustness of the algorithm, the methods in this chapter are compared with geometric active contour model and its improved model Fig.5 shows that, in the case of the same iterative number, ($\alpha=0.4, \beta=0.3, \gamma=0.5, m=3, M=4, \delta=1$, time step $t=0.1$, number of iterations $n=350$). The method in this paper effectively partitioned the left ventricular internal and external contours, while other methods of segmentation results deviate from true profile (Fig.5E). It can be seen by comparing the role of various additions, contour area information; curve will swell when the curve evolves to left heart interior contour internal region; when the curve evolves beyond the inner contour, the curve will shrink. This acts as an adaptive balloon force. Left ventricular myocardial physiological structure constraint item control the relative position of the inner and outer contour evolution curve, to play the role of shape constraints.

Segmentation results compared with professional doctors hand-split, which is the gold standard for medical image segmentation. Comparing the results with manual segmentation results, using mean absolute distance (MAD) to measure the differences between the two, P is the algorithm to get the left ventricular contour lines, h for the hand-split result. MAD expressions are as follows:

$$e(P, H) = \frac{1}{2} \left\{ \frac{1}{n} \sum_{i=1}^n d(p_i, H) + \frac{1}{m} \sum_{j=1}^m d(h_j, P) \right\} \quad (7)$$

$P = (p_1, p_2, \dots, p_n)$ and $H = (h_1, h_2, \dots, h_m)$ are the coordinates point on the contour, $d(p_i, H)$ is the nearest distance of point p_i to the H coordinates:

$$d(p_i, H) = \min_j \|h_j - p_i\| \quad (8)$$

The methods in this paper and ASM model are presented in Table 1, in Fig.6 shows compare each method parameters average, variance and maximum, and the MAD measurement results of 30 images, wherein the algorithm can split the left ventricle inner

and outer contour and have better effect. The time complexity is close to that in geometric active contour model, and average of 100-200 iterations to converge.

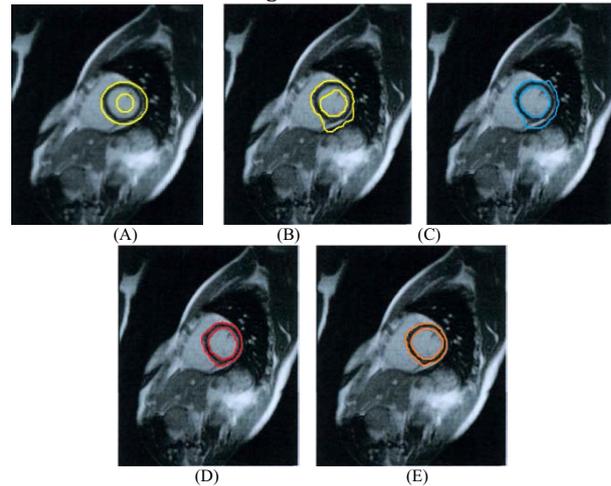


Figure 5. Comparisons of left ventricle segmentation result.

Table.1. MAD evaluation results with methods (pixels)

Method	Average	Variance	Maximum
Hybrid AAM	1.06	0.5	1.77
Shape particle filtering	1.10	0.3	1.90
Multi-view AAM	1.40	0.7	1.93
ASM	1.70	1.1	3.51
Extended AAM	1.06	0.48	1.34
Left ventricular external profile(IKCO)	0.80	0.11	1.05
Left ventricular internal profile(IKCO)	0.84	0.10	1.03

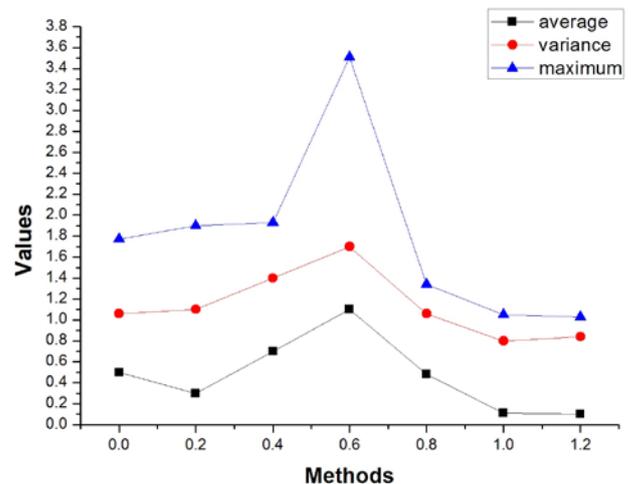


Figure 6. Compare each method parameters average, variance and maximum.

Results from evaluation: the methods in this paper achieve high levels of accuracy, the left ventricular internal and external contours with an average error of less than 1 pixel, the literature of between the segmentation accuracy is 1-2 pixels and good stability than other methods in this chapter, and the segmentation error is the standard deviation of less than 0.2.

The comparison of left ventricular internal and external contours MAD value shows that the methods in this paper achieve

high levels of accuracy, the left ventricular internal and external contours with an average error of less than 1 pixel, and the segmentation error is the standard deviation of less than 0.2.

6 Conclusion

MR image left ventricle inside and outside contour segmentation is the basis of left ventricular surface reconstruction, it can also provide boundary restrictions on myocardial motion analysis, at the same time provide carrier for left ventricular. Mainstream methods of left ventricular myocardial segmentation are image segmentation methods which are edge-driven, such as Snake model, Level Set model, Geometric active contour model and improved models. However, the mastoid muscle is often presents interference in the MR image resulting in local gradient maximum regions, weak edges and artifacts and other phenomena, thus results in many difficulties in image segmentation methods based on edge.

Hough transformation algorithm is robust, not sensitive to data incompleteness or noise, and can identify deformation or partially occluded object. To use this method to estimate the initial profile, the initial contours can be made more accurately and positioned in the vicinity of edge; curve evolution converges to the correct outline to overcome the influence of initial contour segmentation of geometric active contour model. Segmentation of left ventricle profiles, edge-based image segmentation models are often less than ideal results, so, a lot of prior knowledge need to be introduced to improve the robustness of the model. Shape constraint based on prior knowledge and information of the target area with more myocardial physiological constraints can be used to control the relative position of the left ventricular internal and external contours, and regional information can provide category information for image segmentation.

Incomplete K-means clustering algorithm is a clustering method based on squared error, and is simple and fast clustering method. While MR image grey is not consistent and histogram of image region is not according to left ventricular internal area, myocardial region and background region form ideal three peak distribution, therefore, the incomplete K-means algorithm for classification is better than the Gaussian mixture model. This algorithm is simple and practical, and our experimental results show that the algorithm is highly accurate.

Acknowledgments

This paper is supported by Science and Technology Plan of GanSu Province (No.1308RJZA266) and Gansu Radio & TV university youth fund project(NO.QN201501). This support is gratefully acknowledged.

References

1. A.F. Frangi, W.J. Niessen, M.A.Viergever. Three-dimensional Modeling for Functional Analysis of Cardiac Images: A Review. *IEEE Transactions on medical imaging*, 2001, 20(1):2-5.
2. American Heart Association. Heart and Stroke Statistical Update, 1998. Website: www.Americanheart.org.
3. LiPei-hua, ZhangTian-wen. Review on active contour model (Snake model) [J]. *Journal of Software*, 2000, 11(6): 751-757.
4. C.L.Poh, R.I.Kitriey, R.B.K. Shrestha. Visualisation of cardiac wall motion using MR images. *Computers in Cardiology*, 2005, 17-19.
5. N.Paragios.A variational approach for the segmentation of the left ventricle in MR cardiac image analysis. *International Journal of Computer Vision*, 2002, 50(3): 345-362.
6. X.Zeng, L.Staib, R.Schukz, and J. Duncan. Volumetric Layer Segmentation Using Coupled Surfaces Propagation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara,USA,23-25 Jun.1998,708-715.
7. Singh C, Bhatia N, Kaur A. Hough transform based fast skew detection and accurate skew correction methods. *Pattern Recognit* 2008;41:3528-46.
8. Gonzalez RC, Woods RE. *Digital image processing*. Prentice Hall; 2008.
9. Canny J. A computational approach for edge detection. *IEEE Trans Pattern*
10. E. Baeker, A. Jain. A clustering Performance measure based on fuzzy set decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligenece*, 1981,3(1):66-75
11. GOLDBERG N. Colour image quantization for high resolution graphics display [J] . *Image and Vision Computing*, 1991,9(5) : 303 -312.
12. MIGNOTTE M. Segmentation by fusion of histogram-based k-means clusters in different color spaces [J].*IEEE Transactions on Image Processing*, 2008,17(5) : 780 -787.
13. LINDE Y, BUZO A, GRAY R. An algorithm for vector quantizer design [J]. *IEEE Transactions on Communications*, 1980, 28(1) : 84 - 95.
14. AL-DAOUD M, ROBERTS S .New methods for the initialisation of clusters [J]. *Pattern Recognition Letters*, 1996, 17(5) : 451 - 455.
15. TURNBULL D, ELKAN C. Fast recognition of musical genres using RBF networks [J] . *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(4) : 580 -584.
16. MARTIN D, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C] // *Proceedings of 2001 International Conference on Computer Vision* . Washing-ton , DC: IEEE Computer Society, 2001: 416 -423.
17. K.P. PhiliP. Automatic detection of myocardial contours in cine computed topographic images. PhD thesis, University of Iowa,1991
18. Da-meng DAI, De-jun MU, "A Fast Approach to K-means Clustering for Time Series Based on Symbolic Representation", *International Journal of Advancements in Computing Technology*, vol. 4, no. 5, pp. 233-239, 2012.