PAPER • OPEN ACCESS

Segmentation of MR Breast Cancer Images based on DWT and K-means algorithm

To cite this article: Gaoteng Yuan et al 2019 J. Phys.: Conf. Ser. 1229 012025

View the article online for updates and enhancements.

You may also like

- Automatic delineation of body contours on cone-beam CT images using a delineation booster

G Stippel, D C van Rooijen, J Crezee et al.

- A rapid review of influential factors and appraised solutions on organ delineation uncertainties reduction in radiotherapy Sogand Sadeghi, Zahra Siavashpour, Alireza Vafaei Sadr et al.

- Validation of a new open-source method for automatic delineation and dose assessment of the heart and LADCA in breast radiotherapy with simultaneous uncertainty estimation Robert Finnegan, Ebbe Laugaard Lorenzen, Jason Dowling et al.





DISCOVER how sustainability intersects with electrochemistry & solid state science research



This content was downloaded from IP address 3.148.102.90 on 10/05/2024 at 20:32

IOP Publishing

Segmentation of MR Breast Cancer Images based on DWT and K-means algorithm

Gaoteng Yuan¹, Yihui Liu^{1, *} and Wei Huang²

¹School of Information, Qilu University of Technology (Shandong Academy of Sciences), Shandong Jinan 250353, China ²Department of Radiation Oncology, Shandong Cancer Hospital affiliated to Shandong University, Shandong Academy of Medical Sciences, Jinan, China

*yxl@qlu.edu.cn

Abstract. Breast-conserving surgery followed by radiotherapy to the whole breast and boost irradiation to the lumpectomy cavity (LC) is the standard strategy for the early stage breast cancer patients. Accurate segmentation of the target volume is a prerequisite for accurate radiotherapy, which directly affects the success or failure of tumor treatment. The current delineation of target is mainly done by manual drawing, which is time-consuming, laborious and easy to be affected by subjective factors. To solve this problem, we enhance the MR breast images using DWT (discrete wavelet transform) to get more detail of MR image feature firstly. Secondly, we use K-means algorithm to classify the feature vectors and establish the image segmentation model. Finally, compared with the traditional threshold segmentation method, the model is most suitable for automatic delineation of radiotherapy target area and the setting of optimal parameters are obtained. This method can realize the accurate automatic delineation of target area basically, and solve the problem of lack of accuracy and standardization in current tumor bed delineation.

1. Introduction

The incidence of breast cancer are the top of women's cancers and shows an increasing trend year by year [1]. More and more attention has been paid to the diagnosis and treatment of breast cancer. Currently, radiotherapy after breast conserving surgery is a standard treatment for early breast cancer. The main purpose of the postoperative radiotherapy is to improve the local control rate and breast preservation rate. Due to the recurrence area of breast in breast conserving patients, it is mostly located in the tumor bed. Furthermore, its surrounding tissues and preoperative delineation of the tumor area are very important for postoperative radiotherapy [2]. Preoperative segmentation of the tumor region is helpful to determine the location and shape of the tumor bed after surgery, and provides a basis for three-dimensional reconstruction of the tumor region and target delineation.

Therefore, it has a great clinical significance for the segmentation of tumor sites in breast cancer MR images. At present, many machine learning approaches are utilized in this research area. The paper [3] proposes a deep learning called CNN, which use convolutional neural network to MR image segmentation of brain tumors, but this method requires a large number of MR images for training [4], and it is not suitable for MR image segmentation of small samples. Wang et al [5]. applied Active contour models (Snake) [6] to segmentation of brain MR images, but the model required pre-set initial contour positions and low segmentation efficiency. Ganesh et al [7].used fuzzy k-means

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd 1

CMVIT2019	IOP Publishing
IOP Conf. Series: Journal of Physics: Conf. Series 1229 (2019) 012025	doi:10.1088/1742-6596/1229/1/012025

[8]clustering to segment brain MR images, but the initial clustering values of this method are uncertain and not suitable for small region segmentation. Zhang et al [9]. use the support vector machine (SVM)[10]method to segment the brain nuclear magnetic image, but the segmentation accuracy of this method is very low. There are few studies on tumor segmentation in breast cancer MR images. Due to the small range and the unfixed position of breast cancer, it is impossible to use the above method for segmentation treatment directly. At present, the research on breast cancer still has problems such as poor segmentation accuracy and low efficiency.

To solve this problem, we propose a feature extraction based on discrete wavelet and an image segmentation model using K-means algorithm. At first, we use discrete wavelet analysis to analyze the location of breast tumor and establish the feature matrix. Secondly, we use k-means algorithm to classify and recognize feature vectors and establish classification model to complete MR image segmentation. Finally, we compare it with the traditional threshold segmentation method to get the most suitable model for automatic target delineation and the setting of the optimal parameters. And, we compare it with the traditional threshold segmentation method to get the most suitable model for tumor segmentation and the setting of the optimal parameters. The experimental results show that this model can achieve the automatic segmentation of breast cancer MR images.

2. MR image data

The research data comes from Shandong cancer hospital treated 12 breast cancer patients. Each patient has detailed medical records. We use Philips Achieva 3.0T MRI scanner, fat suppression sequence, transverse (Axi) plain scan and enhanced MR images were selected. The scanning parameters were TE: 2.2ms, TR: 4.4ms with 1 mm slice spacing and 352*352 image matrix size. Figure. 1 is an MR image of the patient. The high signal (white) area in the red rectangular frame on the right side of the image is the focus area needs to be segmented accurately.



Figure 1. Enhanced sequence MR image of breast cancer.

3. Methods

Although compared with CT images, MR can provide a good soft tissue image, and can clearly distinguish breast tumors and surrounding tissues, MR images have blurred boundaries, uneven gray distribution, noise and other shortcomings, which makes the processing of MR images is very difficult. Before the segmentation of breast cancer MR images, we need to preprocess it and understand its performance on MR images, so as to establish an appropriate automatic segmentation model.

Compared with the size of the tumor, MR images display a very large range, so first we need to manually tailor the MR images, retaining only the interested side of the tumor (ROI) image. Secondly, we use discrete wavelet [11] to extract feature matrix and get feature matrix. Finally, we use k-means algorithm [12] to classify the features, so as to get the best automatic segmentation model of breast cancer. The specific segmentation process is shown in Figure 2.



Figure 2. The process of MR image segmentation in breast cancer.

3.1. Discrete wavelet analysis

Discrete Wavelet Transform (DWT) [10] is a kind of signal analysis tool with the ability of multiresolution in time domain and frequency domain, which adopts scaling and translation wavelet transform on a specific subset [13]. In signal analysis, DWT is widely used in multi-scale edge detection, weak signal extraction and signal-to-noise separation, and boundary processing and filtering. The basic idea of wavelet transform is to decompose the original signal into a series of sub-band signals with different spatial resolution, different frequency characteristics and direction characteristics by stretching and translation [14]. For a function f(x), the basic expression of the continuous wavelet transform is:

$$w_f(a,b) = \langle f, \psi \rangle = \frac{1}{\sqrt{a}} \int_R f(x) \overline{\psi}(\frac{x-b}{a}) dx_1 \tag{1}$$

Where ψ is a wavelet function and satisfies the admissibility condition $c_{\varphi} = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|}{|\omega|} d\omega < \infty$, The continuous wavelet transform has translation characteristics, that is, $f(x-x_0) \leftrightarrow W_f(a,b-x_0)$. When the values of *a*, *b* are discretized, the discrete wavelet transform is obtained. The commonly used discrete wavelet is binary wavelet. At present, the discrete wavelet transform is generally used Mallat fast algorithm [15]. For one-dimensional discrete wavelet transform, the algorithm matrix expression is:

$$CA_{i} = HCA_{i-1}, CD_{i} = GCD_{i-1}$$

$$\tag{2}$$

Where, $H = (H_{m,k}) = (h_{k-2m}), G = (G_{m,k}) = (g_{k-2m})$. CA_j and CD_j are discrete approximation and discrete detail of discrete signal f(x) at resolution 2^j , respectively. Generally, CA_0 is regarded as a signal to be decomposed for one-dimensional signal. The schematic diagram of decomposition is shown in Figure 3.



Figure 3. Fast wavelet transform decomposition.

The algorithm divides the signal into different frequency channel components. In the process of decomposition, the signal is filtered by high-pass and low-pass filters first, and then down-sampling is carried out at even points for each channel. The main advantage of this method is that the total number of data points after decomposition remains unchanged, but the data of each channel becomes 1/2 of the original. It helps to improve the speed of operation, so it is widely used at present. However, due to the problem of sampling points down-sampling, the translation invariance of wavelet transform is lost. The translation invariance of wavelet is very important for the application of statistical signal processing, so it is not conducive to the application of wavelet transform to extract edge, texture and other features. In this paper, the discrete stationary wavelet transform is used. Compared with the classical discrete orthogonal wavelet transform, the main characteristics of the discrete stationary wavelet transform are redundancy and translation invariance.

3.2. Discrete wavelet filtering results

In the process of MR image processing, due to the display range is too large, it is necessary to preprocess the image to obtain the region of interest (ROI) for the subsequent wavelet feature extraction. The image ROI area is shown in Figure 4.



Figure 4. MR image ROI area of breast cancer.

According to the characteristics of image texture which mainly displays in details, the image feature extraction is concentrated on the high-frequency channel of image wavelet transform [14]. The image is decomposed into approximate LL and high-frequency parts LH, HL, HH by two-dimensional discrete stationary wavelet decomposition. The low-frequency channel LL is further decomposed into {LL1, LH1, HL1, HH1}. Energy of wavelet coefficients is defined as follows:

$$E_{(i,j)} = \frac{1}{n^2} \sum_{k=i-n/2}^{i+n/2-1} \sum_{l=i-n/2}^{l+n/2-1} w_{(k,l)}^2$$
(3)

IOP Publishing

Where n is the size of the image window. We use the function to find the energy of wavelet coefficients respectively. When the window moves, the feature vectors of each pixel in the image are obtained, and then clustering is carried out according to these features to achieve the purpose of image segmentation. Two dimensional discrete wavelet transform is shown in Figure 5.



Figure 5. Discrete wavelet decomposition. (a) High frequency component decomposition (b) LL component continued decomposition.

In order to ensure the influence of each feature vector is similar and reduce the computational complexity of fuzzy clustering, the eigenvalues of each feature vector are normalized according to the following formula before classification:

$$F_{m}(i,j) = \frac{E_{m}(i,j)}{\max(E_{m}(i,j))}$$
(4)

Where $E_m(i, j)$ represents the eigenvalue of an image block in coarse segmentation, *m* represents the eigenvalue of a pixel in subdivision, and $F_m(i, j)$ represents the normalized eigenvalue.

3.3. K-means classification model

K-means is a specific clustering algorithm (data in one cluster cannot exist in another cluster). This algorithm is a typical unsupervised clustering algorithm, because it has the advantages of easy to understand, simple operation, high accuracy, so it has been widely used once proposed. The key idea of K-means algorithm is easy to understand, that is, according to the size of the distance between the experimental sample sets, the sample set is divided into K clusters as the initial samples, and then the K is adjusted by iteration until the distance between the points in the cluster is as small as possible and the distance between the points in the cluster is as small as possible and

The general process of K-means algorithm is as follows:

1) Firstly, K initial clustering centers are selected randomly.

2) The Euclidean distances from each point $x_i (i = 1, 2, ..., n)$ to each cluster center $m_i (i = 1, 2, 3..., p, ..., k)$ are calculated. If $D(x_i, m_p)$, the x_i point is assigned to the p class.

3) Recalculate the central position of each cluster, $m_i = \frac{1}{N_i} \sum_{x \in x_i} x$, N_i is the number of samples in class i = 1, 2, ..., k.

4) Convergence judgment is made, and the clustering partition is completed by cyclic calculation of steps 2 and 3 until the clustering center m_i is no longer changed.

K-means algorithm for image segmentation is to cluster the pixels. The traditional RGB color image has three channels, K-means is easy to cluster. But because the gray image is a single channel, the classification effect is poor. In order to improve the clustering effect, we use discrete wavelet to

extract the ROI region features, and cluster the feature components together with the ROI region to improve the clustering effect. The classification model is shown in Figure 6.



Figure 6. The process of establishing the model.

4. Experiments and Results

In order to verify the DWT and K-means algorithm, a lot of comparative tests are carried out in this paper. According to the algorithm processing steps, the MR images are experimented with MATLAB 2017b on the windows 10 operating system platform. The computer is basically configured with Intel (R) Core (TM) i7-6700 CPU @2.60GHz, and 16GB of running memory. The experimental results are shown in Figure 7.



Figure 7. Segmentation results of DWT and K-means models. (a) ROI region, (b) The result of 3 classification, (c) The result of tumor segmentation.

4.1. K value choice

In order to further verify the effectiveness of the proposed method, several other groups of MR images are selected and tested with different K values. The experimental results are shown in Figure 8.



Figure 8. Segmentation results with different K values. (a) (f)(k)ROI region;(b)(g)(l) Segmentation results with a K=3; (c) (h)(m) Segmentation results with a K=4;(d)(i)(n) Segmentation results with a K=5;(e)(j)(o) The result of tumor segmentation.

As shown in the figure, we choose three MR images for segmentation experiments, the results show that the segmentation effect is better when k = 4.

4.2. Compared with the traditional K-means algorithm

In order to evaluate our segmentation model and the traditional K-means algorithm, we choose several groups of MR images for experiments. The experimental results are as follows.



Figure 9. Comparison of two segmentation models, k=4. (a) (d)(g)ROI region;(b)(e)(h) Segmentation of traditional K-means algorithm;(c)(f)(i) Improved segmentation results of DWT and K-means model.

As shown in Figure 9, graphs b, e, h are the result of traditional K-means algorithm segmentation, and graphs c, f, I are the result of segmentation using our proposed model. The results show that the segmentation effect of clustering model based on DWT feature extraction is better than that of K-means algorithm directly.

5. Discussion

In traditional pixel-based image segmentation methods, the RGB channel of the image is often classified as a feature to complete the image segmentation. This method of classification and segmentation is simple and easy to implement, but it has great limitations. When dealing with single channel grayscale images, the segmentation effect will be worse because of fewer features. In order to solve the problems of medical image, single channel and blurred edge, we propose to use wavelet to extract features instead of pixel features. We have done a lot of comparative experiments to verify our method.

5.1. Experimental results of contrast threshold segmentation

In order to evaluate the segmentation effect, we try to compare with traditional thresholding segmentation and adaptive thresholding (Otsu) segmentation [16], and use multiple breast cancer MR images for repeated experiments. The experimental results are shown in Figure 10.





IOP Publishing

Figures a to d, e to h, and i to l are the segmentation results of three groups of breast cancer MR images using different methods. a, e, i are the original ROI region, b, f, j are the segmentation using threshold, c, g, k are the segmentation using adaptive threshold (Otsu) method, d, h, l are the segmentation models proposed in this paper. Through comparison, it can be seen that the segmentation model based on wavelet feature extraction is better than the traditional threshold segmentation model.

5.2. Experimental results of contrast with Snake and LeveSet segmentation

In order to further evaluate the effect of the model, we choose to compare the results with the segmentation algorithm which has been studied more. Figure 11 is the comparison of the experimental results, followed by Snake model and Level Set model [17].



Figure 11. The result graph of common segmentation algorithm. (a) (e) (i) Different sliced ROI areas; (b) (f) (j) Snake model segmentation; (c) (g) (k) Level Set model segmentation; (d) (h) (l) Improved segmentation results of DWT and K-means model.

Figures a to d, e to h, and i to i are the segmentation results of three MR images using different methods. a, e, i are the original ROI regions, b, f, j are the segmentation results using Snake, c, g, k are the results of Level Set segmentation model, d, h, l are the segmentation results using this paper. The segmentation model. The comparison of the above images shows that the effect of using discrete wavelet to extract features to build a segmentation model is better than the current common segmentation model.

6. Conclusions and future work

In this study we propose feature extraction method based on discrete wavelet transformation (DWT) to improve segmentation performance of breast cancer MR images. The discrete wavelet transform is used to extract features, which avoids the shortcomings of traditional image segmentation methods which only use pixels as features. The experimental results show that the segmentation effect of wavelet feature extraction combined with K-means clustering algorithm is better than the current common segmentation model. Therefore, it is a very good choice to use discrete wavelet to extract features and K-means algorithm for automatic segmentation of single-channel breast cancer MR images.

In the future work, we will try to registration MR images and CT images to complete the accurate localization of breast cancer.

Acknowledgments

This work was supported partly by National Natural Science Foundation of China (81773232).

References

- [1] Wang J, Du L, Li J. The current status and progress of imaging diagnosis of breast cancer [J]. Chinese Journal of Clinicians, 2016
- [2] Speers C, Pierce L J. Postoperative Radiotherapy After Breast-Conserving Surgery for Early-

Stage Breast Cancer: A Review[J]. Jama Oncol, 2016, 2(8):1075-1082

- [3] Pereira S, Pinto A, Alves V, et al. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images[J]. IEEE Transactions on Medical Imaging, 2016, 35(5):1240-1251
- [4] Girshick R, Donahue J, Darrell T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(1):142-158
- [5] Wang G, Wang D. Segmentation of Brain MRI Image with GVF Snake Model[C]// First International Conference on Pervasive Computing, Signal Processing and Applications. IEEE Computer Society, 2010:711-714
- [6] Kass A. Snake: Active contour models[J]. Proc.first Intl Conf.computer Vision, 1988, 4
- [7] Ganesh M, Naresh M, Arvind C. MRI Brain Image Segmentation Using Enhanced Adaptive Fuzzy K-Means Algorithm[J]. Intelligent Automation & Soft Computing, 2017, 23(2):325-330
- [8] Lee T H, Fauzi M F A, Komiya R. Segmentation of CT Brain Images Using K-Means and EM Clustering[C]// International Conference on Computer Graphics, Imaging and Visualisation. IEEE, 2008:339-344
- [9] Zhang N, Su R, Lebonvallet S, et al. Multi-kernel SVM based classification for brain tumor segmentation of MRI multi-sequence[C]// IEEE International Conference on Image Processing. IEEE Press, 2009:3337-3340
- [10] Joachims T. Making large-scale SVM learning practical[J]. Advances in Kernel Methods-Support Vector Learning, 2006, 8(3):499-526
- [11] Li D, Bissyande T F, Klein J, et al. Time Series Classification with Discrete Wavelet Transformed Data[J]. International Journal of Software Engineering & Knowledge Engineering, 2016, 26(09n10)
- [12] Wang Y, Li W, Gao R. An improved k-means clustering algorithm[C]// World Automation Congress. IEEE, 2012:44-46
- [13] Yao M, Zhao Z, Gao L, et al. Image Sketch Generation Algorithm Based on Discrete Wavelet Transform[J]. Computer & Digital Engineering, 2017
- [14] WU Jian. Study on blind source separation and its application in brain signal analysis [D]. Dalian: Dalian University of Technology, 2006: 21-22
- [15] Tian L, Ahmed J, Du Q, et al. Multi Focus Image Fusion using Combined Median and Average Filter based Hybrid Stationary Wavelet Transform and Principal Component Analysis[J]. International Journal of Advanced Computer Science & Applications, 2018, 9(6)
- [16] Xing T, Hu Q, Li J, et al. Refined SAR image segmentation algorithm based on K-means clustering[C]// Cie International Conference on Radar. IEEE, 2017:1-3
- [17] Gharsallah M B, Braiek E B. Image segmentation for defect detection based on level set active contour combined with saliency map[C]// International Conference on Sciences and Techniques of Automatic Control and Computer Engineering. IEEE, 2016:388-392