Classification of Breast Masses Using Color Doppler Flow Imaging

Yingtao Zhang[1], H. D. Cheng[1,2], Yan-Hui Guo[1],Jiawei Tian[3] ,Jianhua Huang[1], Bo Liu[1], Yanxin Su[3]

[1]School of Computer Science and Technology, Harbin Institute of Technology. Harbin 150001. China

[2]Department of Computer Science, Utah State University. Utah . U.S.A

[3]Department of Ultrasound, Second Affiliated Hospital of Harbin Medical University. Harbin 150086. China

Abstract

Color Doppler flow imaging takes a great value in diagnosing and classifying benign and malignant breast tumors. However, previous computer aided diagnosis (CAD) systems have not analyzed blood flow information. In this paper, we present a new approach to classify benign or malignant masses based on two dimensional (2D) ultrasound plus Color Doppler flow imaging in CAD system. In the proposed method, geometric features, textural features and blood flow features are extracted automatically. The experimental results show that the proposed system can improve true positive (TP) and decrease false positive (FP) detection rates greatly, and it will useful for breast cancer control.

Keywords: Breast cancer, Color Doppler Flow Imaging (CDFI), blood flow features, support vector machine (SVM).

1. introduction

Breast cancer is the most prevalent cancer among women [1]. Ultrasound examination is non-invasive, painless, real time and safe for the patient [2]. Thus, there is an increased interest in the use of ultrasound in breast cancer diagnosis [3].

In medical ultrasound image, flow data is usually recorded with B-mode data. Tumor vascularity particularly has played an important role in promoting cancer growth, invasion, and metastasis [4]. Most malignant tumors have copious blood supply. The formation of new blood vessels is the vital for rapid growth of solid malignant tumors. Tumors with volume of 1-2mm^3 or more have to acquire a vascular supply in order to grow [5]. Doppler spectral analysis is a valuable indicator for distinguishing benign and malignant breast neoplasm.

In this paper, Color Doppler flows are employed to differentiate the benign and malignant lesions. A novel breast mass classification system is proposed, whose block diagram is illustrated as Figure 1.

A detailed description of this paper is organized as follows. In Section 2, the procedure of preprocessing is discussed,

which including mass segmentation as described in (Huang et al. 2008) [6], for segmenting the suspicious areas from 2D ultrasonography and blood detection from color Doppler flow imaging. In Section 3, we analysis and extract of three kinds of features: geometric features, textural features and blood flow features. In Section 4, we briefly introduce the procedure of classification using SVM. Finally, the results and the conclusions are summarized in Section 5.

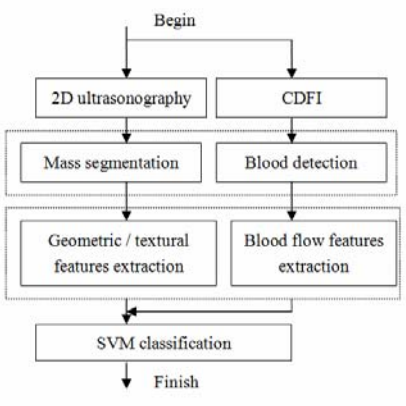
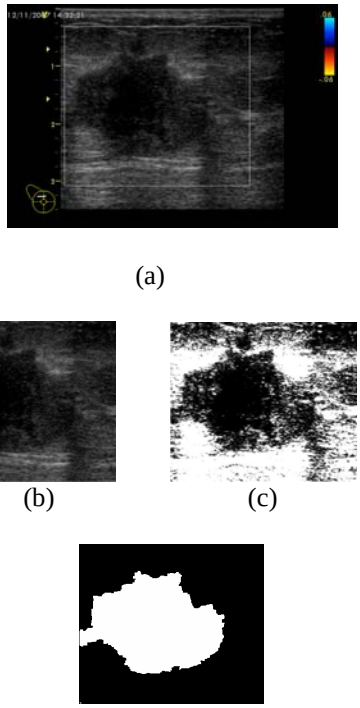


Fig 1. The diagram of the proposed system

1. Preprocessing

Here, we briefly introduce the preprocessing methods used in our CAD system. The preprocessing consists of two stages. The first stage is segmentation of candidate lesions from 2D ultrasonography; and the second one is detection of blood flow information from color Doppler sonography.

The method of breast ultrasound image segmentation is based on homogeneity histogram. Texture and edge features are used to compute homogeneity. Both global and local information is considered, which has not been achieved by previous algorithms. The image is divided into homogeneity subset and the non-homogeneity subset according to the threshold computed from the maximum entropy principle. The two subsets are segmented separately thereafter. Figure 2 illustrates an example of mass segmentation.



(a)

(b) (c)

(d)

(a)An image containing a malignant mass. (b) Region of interest (ROI) of (a) .(c) The image after segmentation of (b). (d) Binarized result of (c), which is after adaptive opening and closing.

Fig 2. An example of mass segmentation.

Doppler color flow systems assign a given color to the direction of flow; red is flow toward, and blue is flow away from the transducer, as shown in Fig 3. Therefore it is very easy to detect the blood flow.

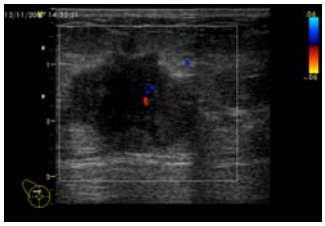


Fig 3. Color flow image of fig 2. (a)

3. Feature Extraction

A key stage of CAD (computer-aided diagnosis) systems is feature analysis and extraction. Many useful features of masses have been employed in previous CAD systems [7]. In this study, we have utilized several blood flow features, which reflect the blood-supplying characteristic of the masses. The typical features by proposed system are listed in table1.

Table 1 The typical features used by proposed system

Feature Sub-Space	Features
Geometric Features	1. compactness 2. border roughness 3. aspect ratio
Textural Features	1. Gray Asymmetry 2. Grads Asymmetry 3. Energy 4. Gray Mean 5. Grads Mean 6. Gray Variance 7. Grads Variance 8. Correlation 9. Gray Entropy 10. Grads Entropy 11.Interia

In color Doppler, progressively increasing velocities are encoded in varying hues of either red or blue. The more dull the hue, the slower the velocity. The brighter the hue, the faster the relative velocity is. High velocities away from the transducer will appear as lighter shades of blue, and higher velocities toward the transducer will be represented by lighter shades of red, or even yellow. Low velocity flow will be represented by darker shades of these colors.

The flow spectrum features include peak systolic velocity ($V\max$), diastolic terminal velocity ($V\min$), mean velocity (MV), resistant index (RI) and pulsitile index (PI). It is proved that $V\max$, PI, RI and MV are significantly higher in the malignant lesions than those in the benign ones. Their definitions are given as follows.

1) Peak systolic velocity

$$V\max = C \max_{k=2,3,\ldots,N-1} \left(\frac{\sqrt{\sum_{u\in R}\sum_{v\in R}(H_k(u,v)-H_{k-1}(u,v))^2}}{\Delta t} \right) \quad (1)$$

Where R is the suspected region of blood flow in CDFI video, and $H_k(u,v)$ is hue value of (u,v) in the frame k. Generally, the velocity of human artery is below 0.22m/s. So we set a coefficient C. N is the number of frames in one heart period, and Δt is delay between two frames.

2) Diastolic terminal velocity

$$V\min = C \min_{k=2,3,\ldots,N-1} \left(\frac{\sqrt{\sum_{u\in R}\sum_{v\in R}(H_k(u,v)-H_{k-1}(u,v))^2}}{\Delta t} \right) \quad (2)$$

3) Mean velocity

$$MV = \frac{C'}{k-1}\left(\sum_{k=2}^{k=N-1} \frac{\sqrt{\sum_{u\in R}\sum_{v\in R}(H_k(u,v)-H_{k-1}(u,v))^2}}{\Delta t} \right) \quad (3)$$

4) Resistant index

$$RI = \frac{V\max - V\min}{V\max} \quad (4)$$

5) Pulsitile index

$$PI = \frac{V\max - V\min}{MV} \quad (5)$$

4. Classification

SVM has been shown to provide higher performance than traditional learning machines [8]. In order to overcome the small sample size problem, we choose SVM and use the leave one-out method to evaluate the performance of classification. Validation set targets was used in training and model selection. The final choice of model, including the choice of kernel, use of a bias term, use of weighting in training and model selection and the choice of model selection criterion were all determined by minimizing the leave-one-out balanced error rate.

5. Results and Discussions

In the experiment, all of the ultrasound images and CDFI videos were acquired with an ultrasonic scanner ATL 3000 unit (GE, VIVID7), using a 7.5-14MHz linear probe, and captured directly from the video signals. The data consists of 65 cases, including 29 malignant solid masses and 36 benign solid masses. The natures of the lesions were all confirmed by pathology (either with surgical excision or with US-guided percutaneous core-needle biopsy). One case consists of at least three B-mode images and one Doppler video.

The results are expressed in terms of four parameters, True Positive (TP); False Positive (FP); True Negative (TN); and False Negative (FN). Table 2 shows the results of this research, and the comparisons with radiologist assessments.

The experimental results show that the proposed CAD system can produce a high accuracy of mass classification. The proposed approach is valuable to radiologists to improve the accuracy of the diagnosis of breast cancer and to reduce the number of unnecessary biopsies.

Table 2 Results of proposed system and radiologist assessments.

	Radiologist	Proposed system
TP	24	25
FP	13	2
TN	25	36
FN	3	2
Accuracy	75.4%	93.8%
Sensitivity	88.9%	92.6%
Specificity	65.8%	94.7%

In the future, the proposed system needs more test cases to prove its reliability and accuracy.

6. Acknowledgement

The work was supported, in part, by Natural Scientific Research Innovation Foundation in Harbin Institute of Technology, Project HIT.NSRIF.2008.48, and Natural Science Foundation of China NO.60873142 and No. 30670546.

Reference

[1] S.H. Landis, T. Murray, S. Bolden, and P.A. Wingo,"Cancer statistics, 1998," *CA Cancer J. Clin.*, vol. 48, no. 1,pp. 6-29, Jan. 1998.

[2] G. Georgion, F. S. Cohen, "Is early detection of liver and breast cancers from ultrasound scans possible" ,Pattern Recognition Letters **24** (2003) 729–739.

[3] PM Shankar, Vishruta A Dumane, Thomas George, Catherine W Piccoli, John M Reid, Flemming Forsberg and Barry B Goldberg, "Classification of breast masses in ultrasonic B scans using Nakagami and K distributions", Phys. Med. Biol. **48** (2003) 2229–2240.

[4] J. Folkman, “What is the evidence that tumors are angiogenesis dependent?” , *J. Nat. Cancer Inst.*, vol. 82, no. 1, pp. 4–6, Jan. 1990.

[5] Sohn C, Blohmer J, Hamper U. Breast ultrasound: A Systematic Approach to Technique and Image Interpretation 1999.

New York, USA: Thieme; 1999. pp. 104.
[6] Jianhua Huang, Yingxing Yu, Yingtao Zhang, Xianglong Tang, Yanhui Guo. "A method for breast ultrasound image segmentation based on homogeneityhistogram", Journal of Harbin Institute of Technology. vol. 40, no. 7, pp. 682–685, July. 2008.
[7]Huai Li, Yue Wang, K. J. Ray Liu, Shih-Chung, B. Lo, and Mattew, T. Freedman, "Computerized radiographic mass detection ", IEEE Trans. Med. Imag. , vol. 20, no. 4, pp. 302-313, April. 2001.

[8]C. Burges, "A Tutorial on support vector machines for Pattern Recognition ", Data Mining