

# A Feature-based Tumor Detection for MRI Breast Imaging

Guo-Shiang Lin<sup>1</sup>, Sin-Kuo Daniel Chai<sup>2</sup>, Wei-Cheng Yeh<sup>3</sup>, Lin-Jie Cheng<sup>1</sup>

<sup>1</sup>Dept. of Computer Science and Information Engineering, Da-Yeh University, Taiwan, R.O.C.

[khlin@mail.dyu.edu.tw](mailto:khlin@mail.dyu.edu.tw)

<sup>2</sup>Dept. of Health Services Administration, China Medical University, Taiwan, R.O.C.

[skchai@mail.cmu.edu.tw](mailto:skchai@mail.cmu.edu.tw)

<sup>3</sup>Radiology dept., Nantou Hospital, dept. of Health, Executive Yuan, Taiwan, R.O.C.

[h0515@hotmail.com](mailto:h0515@hotmail.com)

## ABSTRACT

In this paper, we propose a feature-based scheme composed of the intra-slice, texture and inter-slice analyses to achieve tumor region identification in MRI breast images. Our intra-slice analysis evaluates the intensity and size information of candidate regions. To find a precise region, a region growing algorithm is proposed based on ellipse fitness. In the texture analysis, some texture features are extracted and combined with a neural network to reduce the false alarms. The inter-slice analysis is based on the continuity characteristic to verify the static behavior of tumor regions. The experimental results show that our proposed scheme can correctly identify tumor regions.

## 1. INTRODUCTION

The mortality from breast cancer has been greatly increasing in Taiwan over the past two decades, and it has become the fourth leading cause of cancer death. According to a statistics investigated by The Department of Health, Taiwan [2], the death rate of female breast cancer rose from 3.9% in 1985 to 12.9% in 2005. It is critical, therefore, to detect breast cancer at an early stage since it will greatly improve the therapy and decline mortality [3]. Detection of breast cancer is often carried out by inspecting ultrasonography, mammography, or MRI images [4]-[7]. Ultrasonography is a popular method for breast cancer diagnosis, but its performance is limited to detect cystic lesions or advanced breast cancer [4]. Mammography is another widely used tool for breast cancer detection, and various computer-aided methods for this have been developed [4]-[6]. However, it is still a great challenge for mammography to detect lesions at an early stage. At this stage, the lesion size is quite small, and its brightness in the image is similar to that of tissues such as fat or gland, especially for eastern young people. In order to improve the early detection rate of breast cancer, healthcare organizations have been working on promotion of self breast examination, and the government has provided financial for ultrasound and mammography screening over the past years. The progress, however, is unsatisfactory. A statistics shows that near about forty-five percent of breast cancer cases were found at stage 2 and one fourth of the breast malignancy cannot be removed surgically at the time when it is diagnosed.

While ultrasound and mammography have their weaknesses in screening ductal carcinoma in situ (DCIS), magnetic resonance imaging (MRI) has demonstrated its capability of early breast

cancer detection. Studies have shown that MRI has higher sensitivity than mammography does [8]. This is especially true for young women and for those with dense breasts. In addition, MRI can also accurately determine tumor regions. These reasons motivate us to develop a tumor detection scheme for MRI.

Even with great advancement in image processing techniques, it is still a difficult problem to precisely detect a tumor boundary especially when the background is spoiled by soft tissues and noises. Traditional methods, such as Sobel and Laplacian operators, texture analysis, such as moments, and de-noising techniques can be used for achieving the tumor detection. However, considerations in complexity and robustness are challenging for this kind of techniques. Dissimilar to most of previous schemes, we adopt not only the intra-slice and texture analyses in each MRI slice but also the inter-slice analysis to raise the detection performance, even there exists certain similarity in some tissue regions, e.g., vessels, in this paper. Experiments show that our scheme is capable of correctly detecting the tumor regions.

## 2. FEATURE-BASED TUMOR DETECTION

Basically, to design a high efficient mechanism for tumor region detection in MRI breast images, we should first analyze characteristics of tumor regions. As we know, a MRI breast slice is basically composed of tumors and soft tissues. Though the composing components may be complicated, the tumor regions are characterized of high intensity, small size, high texture, and continuity between MRI slices. In this paper, we measure these characteristics as features via the intra-slice, texture, and inter-slice analyses and then develop a tumor detection scheme. Figure 1 illustrates the block diagram of our proposed scheme for tumor detection. In the following, we will introduce the proposed algorithm in detail.

### 2.1 Intra-slice analysis

According to physicians' opinions, a tumor region has higher pixel values and the size of a tumor region is small at stage 0. Then two common features, intensity and size information, of tumor regions are extracted in the intra-slice analysis. Detailed steps are described below.

1. Morphological opening and automatic thresholding for tumor candidate selection

Because the tumor regions have generally high intensity, morphological opening [1] is used here to split the background part and foreground part which contains the candidates of tumor regions. The foreground part can be obtained as

$$\mathbf{I}^R = \mathbf{I} - (\mathbf{I} \circ S), \quad (1)$$

where  $\circ$  stands for the morphological opening operator,  $\mathbf{I}$  denotes the original image,  $\mathbf{I}^R$  represents the resulting version after morphological opening and subtraction, and  $S$  is the structure element in the morphological operator. After obtaining  $\mathbf{I}^R$ , the automatic thresholding for selecting the candidates of tumor regions can be performed as follows:

$$I^B(x, y) = \begin{cases} 1 & \text{if } I^R(x, y) > \max(I_1^L, I_2^L), \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where  $\mathbf{I}^B$  represents the output of automatic thresholding and  $\max(\cdot)$  is an operator to find the maximum value of input.  $I_1^L$  stands for a predefined threshold, and  $I_2^L$  denotes the threshold to select 10% of pixels with high value after morphological opening. Since the range of pixel values may not be similar for different MRIs, a predefined threshold  $I_1^L$  and another one  $I_2^L$  relative to image content are calculated to choose the candidates of tumor regions in this step.

## 2. Connected component labeling

The chosen regions are labeled for following processing. Actually, a labeled object may contain tumors and other tissues, e.g., vessels. Connected regions whose size satisfy a given constraint (e.g., smaller than  $T^A$ ) are identified for the following shape evaluation.

## 3. Shape refinement

Although, the tumor candidate regions can be obtained after morphological opening, thresholding, and labeling, their regions may not be correctly detected. In addition, because the shape of tumor regions can be approximated by using an ellipse, survived connected regions should encounter an ellipse-shape test. Referring to [1], a method based on ellipse fitness was employed to find the correct tumor regions and measure their properties like center of mass, best-fit ellipse, etc. Figure 2 illustrates the measurement of ellipse fitness.

To achieve shape refinement, a region growing algorithm is proposed and applied to each tumor candidate and described as follows.

(1) Compute the average intensity  $\bar{I}_{L_i}$  and center of mass  $(x_c, y_c)$  of  $L_i$  as follows:

$$x_c = \frac{1}{|L_i^l|} \sum_{(x,y) \in L_i^l} x, \quad y_c = \frac{1}{|L_i^l|} \sum_{(x,y) \in L_i^l} y, \quad (3)$$

where  $L_i^l$  denote the  $i$ -th candidate in the  $l$ -th iteration and  $|L_i^l|$  represents the area of  $L_i^l$ .

(2) Measure  $D(x, y)$  and  $\beta(x, y)$  as follows:

$$\beta(x, y) = \frac{1}{2a_L} \sum_{i \in \{1,2\}} \sqrt{(x - x_f^i)^2 + (y - y_f^i)^2}, \quad (4)$$

$$D(x, y) = \beta(x, y) \sqrt{(I_{neighbor}(x, y) - \bar{I}_{B_i})^2}, \quad (5)$$

where  $I_{neighbor}(x, y)$  is the pixel value of the neighboring pixel based on 8-connectivity,  $(x_f^i, y_f^i)$  denotes the coordinate of focus points of an ellipse. In Eq. (4), referring to [1], the major axe  $a_L$  is defined as

$$a_L = \left( \frac{4}{\pi} \right)^{\frac{1}{4}} \left[ \frac{(I_{\max})^3}{I_{\min}} \right]^{\frac{1}{8}}, \quad (6)$$

where

$$I_{\min} = \sum_{(x,y) \in B_i} [(y - y_c) \cos \theta - (x - x_c) \sin \theta]^2,$$

$$I_{\max} = \sum_{(x,y) \in B_i} [(y - y_c) \sin \theta + (x - x_c) \cos \theta]^2,$$

$\sin(\cdot)$  and  $\cos(\cdot)$  represent sine and cosine functions, respectively, and  $\theta$  denotes the orientation. According to Eqs. (4) and (5), as a pixel is near the focus points of an ellipse and its pixel value is closed to  $\bar{I}_{L_i}$ ,  $D(x, y)$  is small. This means the neighboring pixel with smaller  $D(x, y)$  is similar to those pixels in this candidate.

- (3) Include neighboring pixels as their  $D(x, y)$  are less than  $T^D$ , where the threshold  $T^D$  is adjusted according to image content.
- (4) Check whether the candidate overlaps the others. If this situation exists, then merge overlapped candidates, re-label all candidates, and jump to step (1).
- (5) Check whether there are neighboring pixels to be processed by calculating  $|L_i^D| = (|L_i^l| - |L_i^{l-1}|)$ . If  $|L_i^D|$  is less than  $T^L$ , repeat step (1) to step (4); otherwise, stop.

## 2.2 Texture analysis

Since texture information is useful for tumor detection, here we also think of texture information as features. We first divide a slice into several blocks of  $32 \times 32$  pixels and then each block covering the tumor candidate is determined as a candidate block. Referring to [9], to examine each candidate block, we measure several texture features in the spatial and wavelet domains. The spatial texture features include the average, standard deviation, and energy of pixel values in each candidate block. In addition, after wavelet transform, the energies of three subbands (LL, LH, HL) in each candidate block are computed as features. Here we describe the steps of feature extraction in the wavelet domain as follows.

- (1) Transfer this block into the wavelet domain to obtain LL, LH, and HL subbands.
- (2) Quantize the wavelet coefficients of each subband into 64 levels.
- (3) Normalize the quantized wavelet coefficients of each subband into the range [0 1].
- (4) Calculate the energy of each subband as features.

After extracting 6 texture features, a supervised multi-layer neural network is adopted as a classifier. In the training phase, we collect a large number of blocks classified by professional physicians as a training set. To training the neural classifier, we adopted the  $Q$ -fold cross-validation method [10]. That is, the training set is randomly divided into  $Q$  disjoint sets of equal size  $n_s/Q$ , where  $n_s$  is the total number of samples. The  $(Q-1)$  sets arbitrarily selected from  $Q$  disjoint sets are exploited to train the neural classifier and the remained one is employed to estimate the generalization error. The neural classifier is then trained  $Q$  times, each time with a different set as the validation set. After training, this neural classifier can be exploited to determine whether this block contains tumors.

### 2.3 Inter-slice analysis

In addition to the intra-slice and texture analyses, inter-slice analysis is also performed to raise the accuracy of tumor region identification in this paper. Generally, the position of a tumor region may not change significantly during MRI slices. Based on this observation, we compare the center of mass between the recognized tumor regions of two test slices via checking

$$CT = \begin{cases} 1 & \text{if } (dist(L_i^t, L_i^{t-1}) < T^D) \wedge (L_i^t \cap L_i^{t-1}) \wedge \left( \frac{|L_i^t \cap L_i^{t-1}|}{\min\{|L_i^t|, |L_i^{t-1}|\}} < T^O \right), \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where  $\wedge$  is the logical AND operator,  $\cap$  is the intersection operator,  $L_i^t$  and  $L_i^{t-1}$  denote the  $i$ -th tumor regions in the  $t$ -th slice and the  $(t-1)$ -slice, respectively,  $dist(L_i^t, L_i^{t-1}) = \sqrt{(x_c^{L_i^t} - x_c^{L_i^{t-1}})^2 + (y_c^{L_i^t} - y_c^{L_i^{t-1}})^2}$  is the Euclidean distance between their centers of mass,  $T^D$  and  $T^O$  are thresholds. For each tumor region candidate  $L_i$ , the  $CT$  test is performed for all of MRI slices.

Actually, the size, pixel value, and texture information of a tumor are unlike in different slices. It is possible to break the continuity property due to missing errors in one slice. To solve the problem, we utilize the output results of texture analysis in two previous slices and two following slices to re-check whether  $CT$  is true or not in the current slice. The rule is as follows: if all of  $CT$ s in two previous slices and two following slices are true, the  $CT$  of the current slice is set to be true.

For each MRI, if the number that a recognized tumor candidate region continuously exists is over a threshold  $T'$ , this tumor region is thus identified. Otherwise, the candidate region is considered as noises. Because the size of a tumor is from 0.2cm to 2cm and the thickness of one slice is 0.12cm,  $T'$  is set to be 6 in the proposed scheme.

## 3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed system, we use 2 MRIs captured in Taiwan and each MRI has 128 slices. The size of one slice is  $512 \times 512$  pixels. One MRI contains 1 tumor region and the other has two tumor regions. According to physicians' suggestions, the size of the structure element in the morphological opening is  $25 \times 25$  pixels, and  $T_1^L$ ,  $T^A$ ,  $T^O$ , and  $T'$  are set to be

5000, 625, 0.5, and 6, respectively. Figures 4 and 5 show parts of experimental results. Here a training set is composed of 15616 blocks and the size of each block is  $32 \times 32$  pixels. For supervised training, the training set is composed of two classes, one class whose size is 6158 covers tumor regions and the other whose size is 9458 does not contain tumors. Here,  $Q$  is set to be 3. The positive and negative detection rates for the validation set are 92.6% and 92.8%, respectively.

The first column of Figs. 4 and 5 show the original MRI slices and the red windows locate the true tumor regions. The second, third, and fourth columns show the experimental results after intra-slice, texture, and inter-slice analyses, respectively.

### 3.1 Intra-slice analysis

The second columns in both figures, Fig. 3(a2) and Fig. 3(b2), show the results of intra-slice analysis. Comparing with the first column of Fig. 4, as we can see in the second column, there are some false alarm errors. The misclassification comes from several aspects. First, false alarms often occur as vessels exist. In fact, the average pixel value in a vessel region is often higher and its shape is also ellipse-like. As the characteristics (e.g., size and intensity) of a vessel region satisfy these conditions of size and intensity, the vessel region is considered as the tumor region. As we can see in the second column of Fig. 5, similar results happen in the other case.

### 3.2 Texture analysis

The third columns, Fig. 3(a3) and Fig. 3(b3) show the results after intra-slice and texture analyses. Comparing with the second columns in both figures, as we can see in the third columns, some false alarms are removed after texture analysis. That means that these texture features can distinguish tumor regions from other ones and the trained neural classifier with texture features can reduce the number of regions mis-classified by the intra-slice analysis.

### 3.3 Inter-slice analysis

The last columns, Fig. 3(a4) and Fig. 3(b4) show the experimental results after the spatial, texture, and inter-slice analyses. As we can see in the last columns of two figures, the true tumors can be correctly identified in both MRIs. Comparing with the third columns in two figures, the false alarm errors are eliminated by using inter-slice analyses. That means that the continuity of the tumor region did exist and this property can be exploited to reduce interferences resulting from vessels or other tissues. The results show that the performance of our proposed combining the intra-slice, texture, and inter-slice analyses can be raised.

## 4. CONCLUSIONS

In this article, we propose a feature-based scheme which contains the intra-slice, texture, and inter-slice analyses to achieve tumor region identification for MRI breast images. In the intra-slice analysis, the intensity and size information is utilized to find the candidates of tumor regions. To find a precise tumor region for further inspection, a region growing algorithm is proposed according to the intensity and distance information. Some texture features are extracted and combined with a supervised neural

network to be classifier in the texture analysis. In the inter-slice analysis, the characteristic of continuity of a tumor region among slices is exploited to remove noise resulting from other tissues and noises. The experimental results show that our proposed scheme can effectively identify the tumor region. The future work is to use more MRIs for evaluating the performance of our proposed scheme.

### REFERENCES

[1] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[2] Health and Vital Statistics, <http://www.doh.gov.tw/statistic/>, Department of Health, December 2006.

[3] R. M. Rangayyan, L. Shen, Y. Shen, J. E. Leo Desautels, H. Bryant, T. J. Terry, N. Horeczko, and M. S. Rose, "Improvement of sensitivity of breast cancer diagnosis with adaptive neighborhood contrast enhancement of mammograms," *IEEE Trans. Information Technology in Biomedicine*, vol. 1, no. 3, pp. 161-170, Sep. 1997.

[4] T. Arodz, M. Kurdziel, T. J. Popiela, E. O.D. Sevre, and D. A. Yuen, "Detection of clustered microcalcifications in small field digital mammography," *Computer Methods and Programs in Biomedicine*, vol.81, pp.56-65, 2006

[5] S. Joo, Y. Seok, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features," *IEEE Trans. on Medical Imaging*, vol. 23, no. 10, pp.1292-1300, Oct. 2004.

[6] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Medical Imaging*, vol. 21, no. 12, pp.1552-1563, Dec. 2002.

[7] M. L. Essink-Bot, A. J. Rijnsburger, S. van Dooren, H. J. De Koning, and C. Seynaeve, "Women's acceptance of MRI in

breast cancer surveillance because of a familial or genetic predisposition," *The Breast*, vol. 15, pp. 673-676, 2006.

[8] E. A. Morris, "Screen for breast cancer with MRI," *Seminars in Ultrasound, CT, and MRI*, vol. 24, no. 1, pp.45-54, February 2003.

[9] Fang-Wei Hu, "A computer aided diagnosis system for microcalcifications in digital mammograms based on wavelet and independent component analysis", *Master thesis, National Chung Cheng University*, July 2006.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, 2001.

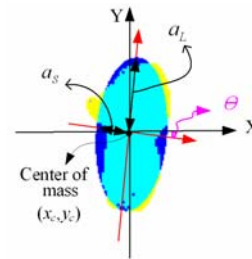


Fig. 1. An illustration for ellipse fitness

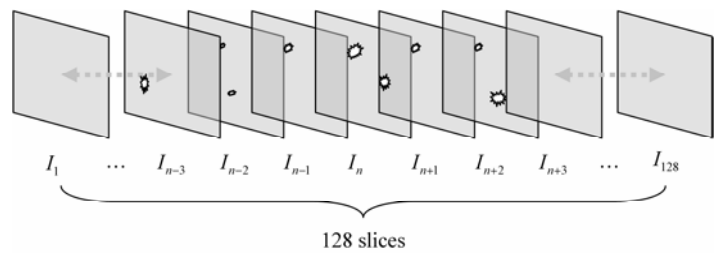


Fig. 2. An illustration of continuity in the inter-slice analysis

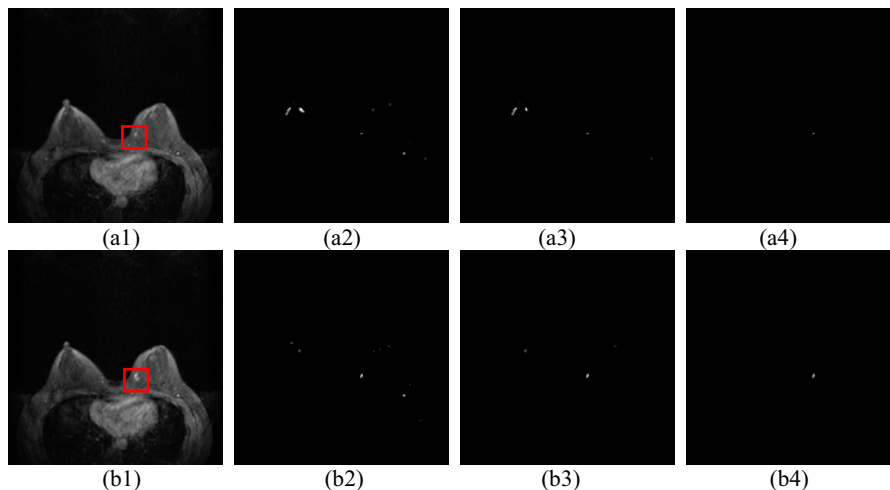


Fig. 3. Original the 75-th and the 76-th MRI slices and their corresponding detection results. (a1) and (b1) are original the 75-th and the 76-th MRI slices, respectively; (a2) and (b2) are the results after spatial analysis; (a3) and (b3) are the results after spatial and texture analyses; (a4) and (b4) are the final results.