Automatic Segmentation and Detection of Mass in Digital Mammograms

NOR ASHIDI MAT ISA[*], TING SHYUE SIONG
Imaging and Intelligent Systems Research Team (ISRT)
School of Electrical and Electronic Engineering
Engineering Campus, Universiti Sains Malaysia
14300 Nibong Tebal, Penang,
MALAYSIA
*Tel: +604 5996051, Fax: +604 5941023, Email: ashidi@eng.usm.my

Abstract: - This paper presents an automated system for mass segmentation and detection in mammograms. Initially, breast segmentation is applied to separate the breast and non-breast area. Then, image enhancement is employed to improve the contrast of the tissues structure in mammograms. Finally, constraint region growing based on local statistical texture analysis is applied to detect and segment out the mass from the mammograms. The system is develop and evaluated with 322 mammograms from Mammographic Image Analysis Society Database. The verification results show that the proposed technique has a sensitivity of 94.59% and the number of false positive per image is 3.90.

Key-Words: - Image processing; image segmentation; mammogram images; medical imaging; mass detection

1 Introduction

Breast cancer is one of the common causes of death among women [1]. It has been reported that early detection of breast cancer in asymptomatic patients can reduce breast cancer mortality. Currently, mammography is one of the most reliable and effective ways of detecting breast cancer in early stage. Therefore, various computer aided diagnosis (CAD) systems have been developed for assisting doctor to identify the symptoms in early stage by using mammography.

CAD system has been developed either for detection of mass or calcification clusters. Detection of mass in mammography is the most challenging for CAD development because mass is always poor in contrast, ambiguous shape margins, complex shapes and indistinguishable from surrounding parenchyma tissues. Petrick et al. [2] employed density-weighted contrast enhancement (DWCE) segmentation combined with an object-based region growing technique algorithm for extracting potential mass from their surrounding tissues. Li et al. [3] extracted mass based on adaptive thresholding followed by a modified Markov random field model based method. Alfonso et al. [4] proposed the detection of mass in mammograms via statistically based enhancement by multilevel-thresholding and region selection.

In this paper, the method for segmentation and detection of mass in mammograms is proposed. The proposed method is divided into three steps as shown in Fig. 1. Initially, breast segmentation is applied to exclude the background information. Then, contrast enhancement is employed to improve the distinct structures of the breast tissue, which could increase the visualization of the mass from the background of breast tissues. Finally, mass segmentation using constraint region growing (CRG) [5] technique is applied to detect and segment the mass from the mammogram.

Breast Segmentation	→	Image Enhancement	→	Mass Segmentation

Fig. 1. The proposed method for mass segmentation in digital mammogram.

The database of mammograms used in this study is taken from the Mammographic Image Analysis Society (MIAS) Mini Mammographic Database [6].The database contains 322 mammograms including normal, mass, and micro calcification cases. It provides different classes of abnormality such as calcification, well-defined circumscribed mass, speculated mass, architectural distortion, asymmetry and normal.

In this study, the proposed detection of mass is based on the local statistical analysis, which provides the information of suspicious regions to be segmented. The implementation of the proposed image processing technique for mass detection and

segmentation will be done using MATLAB version R2010a.

2 Methodology

2.1 Breast Segmentation

Breast segmentation is essential to exclude the background information in analyzing mammogram image. In this study, a manual thresholding is proposed to separate the foreground (i.e. breast) and background (i.e. non-breast) regions in mammogram image. This technique is an easy and convenient way of performing the breast segmentation. Suppose that the original mammogram image is $p(x,y)$. The thresholding technique is employed to form a binary image by segmenting the mammogram image into two regions, namely foreground (i.e. represented by white region) and background (i.e represented by black region). The thresholded image $q(x,y)$ is defined as

$$q(x,y) = \begin{cases} 1 & \text{if } p(x,y) \geq M \\ 0 & \text{if } p(x,y) \leq M \end{cases} \tag{1}$$

where M is the threshold value that separates the foreground and background regions, which is manually set by user.

2.2 Image Enhancement

The purpose of image enhancement is to improve contrast between the mass structure and surrounding texture of the breast tissues. The proposed image enhancement method is based on the concept of contrast limited adaptive histogram equalization (CLAHE). This method has been used to reduce the noise produced in homogeneous areas for medical imaging [7]. In the Matlab toolbox, the function for CLAHE technique is given by:

$$f = \text{adapthisteq } (q) \tag{2}$$

where f is the enhanced image and q is the thresholded image. This function has been used for our image enhancement technique to improve the visualization subtle signs of abnormalities in mammogram image.

2.3 Mass Segmentation

CRG region algorithm based technique is introduced as one of the mass segmentation methods. Region growing is defined as adding pixels to an initial set, known as seed point. The task of region growing is to segment mammogram image into two sets: mass pixels set, and surrounding background pixels set. In the region growing segmentation, the initial seed point is first determined. The selection of seed point in mammogram image for region growing is vital for the segmentation result. In the proposed method, a seed point is obtained automatically from the statistical texture analysis. Since masses are generally radiograhycally denser than surrounding tissues, the statistical texture analysis of intensity histogram is employed on the brighter region. In this study, the smoothness descriptor data is used to determine the seed point for region growing process by selecting the smallest smoothness descriptor value. The formulae for smoothness descriptor can be defined as:

$$R = 1 - \frac{1}{(1+\sigma^2)(N-1)^2} \tag{3}$$

where σ^2 is the variance of gray level intensity and N is the number of possible intensity levels.

Conventional region growing defined the mass segmentation solely based on gray level information in the image as long as the connectivity of pixel is satisfied. Thus, the region being grown can sometimes includes neighboring pixels that belong to the breast tissue, producing wrong segmentation. To prevent the inclusion of the pixels that differs significantly from the reference shape or located away from the mass region, the mass region will be multiplied by the constraint function before region growing process is applied. Pixels with the same gray level but different locations are modified by an appropriate constraint function so that the pixels close to seed point are differed from those pixels far to seed point. From the study of mass segmentation, a well known constraint function is an isotropic Gaussian function centered on the seed point location (μ_x, μ_y). The function is derived based on the assumption that most of the mammographic mass is approximately circular in shape. The function $h(x, y)$ is derived from the multiplication of the enhanced image $f(x, y)$ with the constraint function is given by:

$$h(x,y) = f(x,y)N(x,y;\mu_x\mu_y,\sigma_c^2) \tag{4}$$

where $N(x, y; \mu_x, \mu_y, \sigma_c^2)$ is a circular normal distribution centre at (μ_x, μ_y) with variance σ_c^2, which is related to the size of the object to be segmented.

In order to segment the ground truth of mass, we propose to evaluate a series of mass regions M with different threshold value, t_i; $i=0, 1, \ldots, T$. In this study, the initial threshold value, t_0 is set to the intensity of predetermined seed point. $t_1, t_2, \ldots, t_T$ represent the threshold value which are set to (t_0-1), $(t_0-2), \ldots, (t_0-T)$ respectively. In this case, the region will grow by reducing the gray level with one (1)

intensity value. T is the maximum gray level intensity with reference to seed pixel intensity level. Therefore, the series of mass regions M_i are defined as:

$$M_i = \{(x,y) : h(x,y) > t_i\} \qquad (5)$$

In this study, another statistical texture analysis called the skewness is employed on the different detected mass regions to decide the preferable mass region representing the ground truth of mass in mammogram. As shown in Fig. 2, the skewness of one suspicious mass region (i.e. as an example, which is taken from one mammogram image among the tested mammogram images) is calculated for every M_i. Based on Fig. 2, the range of different gray level interval with reference to seed point intensity level for representing the mass region is from 0 to 130 intervals. For this example, the value of T is set to 122 as highlighted in Fig. 2. In this analysis, the mass region with the minimum skewness value is returned as final gray level interval T, to represent the ground truth of mass in mammogram.
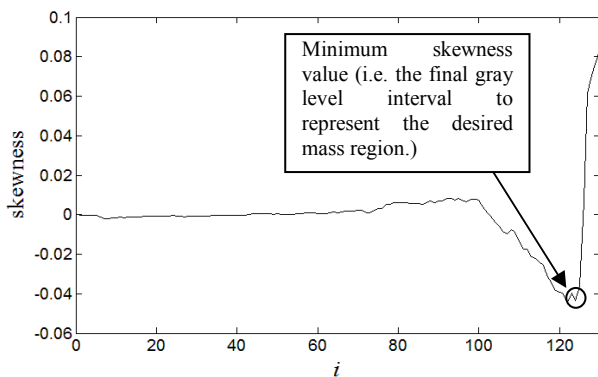

Fig. 2. The skewness value for different M_i of image $f(x, y)$ versus the different gray level interval i

In mass segmentation, the generated mass region is produced from the image $h(x, y)$ while the skewness measurement is from the enhanced mammogram image $f(x, y)$. In order to produce a round and mimic shape of mass region, morphological operation namely dilation based on structuring element is applied on the segmented mass region to fill image region and hole. The segmentation of the proposed technique will proceed to search for other suspicious regions in the same image and converted to binary images. For each iteration, the binary image is produced until all suspicious masses are segmented in the mammogram image.

3 Results and Discussion

The method for mass segmentation and detection is tested with a total of 322 mammogram images

from Mini-MIAS database. The total number of mammograms for well-defined and ill-defined masses in the database is 37. The results for mass segmentation and detection using the proposed technique on the two selected mammogram images are shown in Fig. 3.

Based on Fig. 3, second column shows examples of the original mammograms with suspicious mass region as marked by arrow. When performing the contrast enhancement technique as shown in the third column of Fig. 3, it is found that the mass is more visualized and thus could increase the sensitivity of subsequent process. For the resultant segmented images as shown in fourth column of Fig. 3 for both mammogram images, four different initial seed points are determined in each image which results in detection of four different mass regions. However, in most cases, the proposed technique leads to an accurate delineate the ground truth of the mass detected. As shown in Fig. 3, the location, shape and size of mass regions are successfully detected and preserved. In addition, it is expected that the promising segmentation performance obtained will result in more meaningful for shape features extraction and ultimately, for better classification of mass and normal tissues.
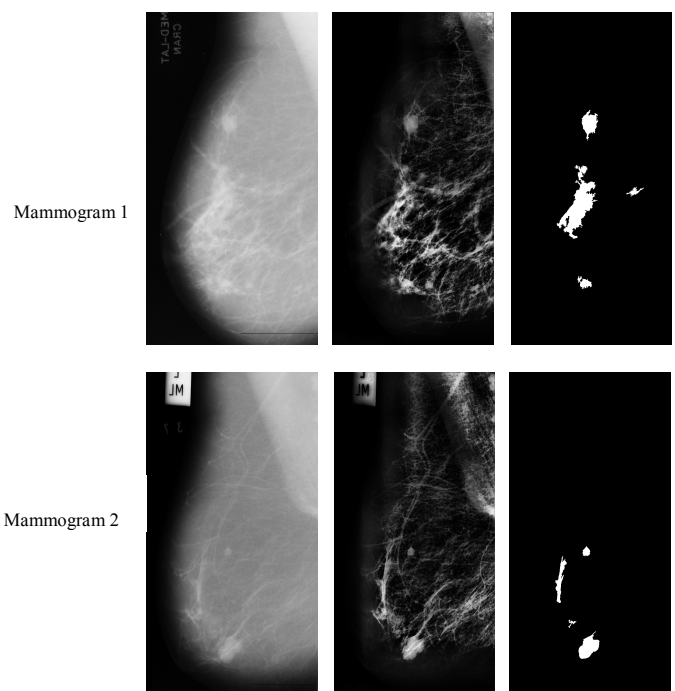

Fig. 3. Results on segmentation and detection of mass regions for Mammogram 1 and Mammogram 2 images. Second column: The original image. Third column: The output of contrast enhancement. Fourth column: The result of CRG segmentation.

After applying the proposed algorithm on all MIAS database, quantitative analysis is carried out by employing statistical analyses as follows:

- Average number of suspicious regions per image

$$\frac{\text{Total number of suspicious region}}{\text{Total number of images}} = \frac{1290}{322} = 4.00 \quad (6)$$

- Number of false negatives per image

$$\frac{\text{Total number of undetected positive regions}}{\text{Total number of images}} = \frac{2}{322} = 0.006 \quad (7)$$

- Number of false positives per image

$$\frac{\text{Total number of false alarms}}{\text{Total number of images}} = \frac{1290 - 35}{322} = 3.90 \quad (8)$$

- Sensitivity

$$\frac{\text{Detected true positives}}{\text{Real number of positives}} = \frac{35}{37} \times 100 = 94.59\% \quad (9)$$

The results show that the proposed algorithm could obtain good performance for mass segmentation in mammogram image. Both of the well-defined and ill-defined masses are detectable. From 322 tested mammogram images, two mammogram images are missed by the proposed algorithm because one of the mass is heavily mixed with dense tissues and another is very small in size with low gray level after processed by the proposed contrast enhancement. For future research, the investigation needs to be done to reduce the false positive rate. This can be done by incorporating features extraction or classifier in classification of mass and normal tissues.

4 Conclusion

Digital mammograms are among the most difficult medical images to be read due to the low tissue image contrast and slight perceptible differences. Therefore, an automated computer-aided method for the detection and segmentation of mass in mammograms has been presented. The results show that the proposed technique effectively detects and segments out the mass from the mammogram. It successfully produces excellent segmentation results with 94.59% of sensitivity. The high sensitivity rate can greatly assist the radiologists to identify the mass from mammography.

Acknowledgement

References:
[1] World Health Organization. Fact sheet No. 297: Cancer. Retrieved on 26 April 2007.
[2] N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, An Adaptive Density-Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection, *IEEE Trans. Med. Imag.*, Vol. 15, 1996, pp. 59–67.
[3] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, Markov Random Field for Tumor Detection in Digital Mammography, *IEEE Trans. Med. Imag.*, Vol. 14, 1995, pp. 565–576.
[4] A. Rojas, A. K. Nandi, Detection of Masses in Mammograms via Statistically Based Enhancement, Multilevel-Thresholding Segmentation, and Region Selection, *Computer Med. Imag. and Grap.*, Vol. 32, 2008, pp. 304-315.
[5] M. A. Kupinski, M. L. Giger, Automated Seeded Lesion Segmentation on Digital Mammogram, *IEEE Trans. Med. Imag.*, Vol. 17, 1998, pp. 510-517.
[6] Mammographic Image Anal. Soc., Manchester, U.K.. [Online]. Available: *http://s20c.smb.man.ac.uk/services/MISA/MISA com.html.*
[7] X. Wanga, B. S. Wong, T. C. Guan, Image Enhancement for Radiography Inspection, *International Conference on Experimental Mechanics*, 2004, pp. 462-468.