A NEW METHOD FOR EXTRACTING REGION OF INTEREST IN MAMMOGRAMS

Wen Lu*, Ruhai Dou, Guangyu Zhang

Radiological Department of Taishan Medical College, Shandong, China
E-mail: aluwen@163.com

Abstract: In mammography computer-aided diagnosis, the automatic extraction of interesting region is one of the most difficult problems. This paper presents a method based on two-dimensional principal component analysis (2DPCA) to extract the region of interest (ROI) automatically. First, preprocess the mammograms, then, extract mammography features by 2DPCA method and edge-detection algorithm. Finally, extract ROI by neural network classifier. 60 cases were analyzed and 100 images which from Shandong medical imaging research institute were used in this investigation. The results show that a better positive detection ratio is obtained with this method. This approach can obtain better extraction accuracy by integrating 2DPCA, edge-detection algorithm and neural networks.

Key Words: Mammography, Region of Interest, Feature Extraction

1 INTRODUCTION

The extraction of interesting regions in mammograms is pre-process of the computer aided diagnosis for the microcalcifications of breast. In order to improve the speed and reduce the workload of the algorithm which finds the suspicious microcalcifications, the interesting regions must be extracted first in the mammogram [1,2]. Principal Component Analysis (PCA) is one of the most common methods for feature extraction. It automatically selects the PCA based on the eigenvalue. There are lots of disadvantages to process image matrix by using the algorithm of the PCA. Yang (2004) developed a new method for feature extraction: 2DPCA [3]. It directly extracts the feature based on the two-dimensional image matrix. In compare with the PCA, it takes less time than the PCA algorithm to compute the eigenvector. This paper will present a method for the extraction of the IR in a mammogram by integrating the mammography feature extraction which uses the improved 2DPCA and image edge characters as inputs to the neural network classifier [4,5].

2 METHODS

2.1 Preprocessing of Digital Mammogram

In this paper, 50 digital mammograms which contain microcalcifications are selected for the extraction of the breast region. Each sample image of the extracted breast region is divided into size 128*128 sub-images. 50 interested sub-images are chosen as training samples. A weight matrix D with the same dimensions as the image is created for each image training sample A (size 128 * 128) by using image statistic characteristics. The statistic average value is calculated as the image statistic characteristic:

$$V(x,y)=\frac{1}{(2m+1)\bullet(2m+1)}\sum_{i=-m}^{m}\sum_{j=-m}^{m}f(x+i,y+j) \tag{1}$$

The average gray value is computed based on the Eq.1 for a close region of 2m + 1 pixels. The coefficient value d_{ij} of the image weight matrix is determined by the difference between the pixel gray value and the average gray value, and $0 < d_{ij} < 1$. If the difference between the pixel gray value and the average gray value is big, the coefficient value will be big; otherwise, it will have a small value. The preprocessing of digital mammogram is completed by the multiple of the image training sample A and the image weight matrix D as the equation 2. It can increase the identifiable strength of some important pixels which can affect the accurate of the classification of the image.

$$B = A \otimes D = (a_{ij} \otimes d_{ij}) \tag{2}$$

2.2 Principal Component Feature Extraction of Digital Mammogram

The LR-2DPCA Method is used to extract the principal component features in this research. The procedure is as follows: Assuming there are N training sample mammograms with size m * n, an n * d dimensional characteristic projected matrix P is constructed by using the algorithm of 2DPCA, and an m * e dimensional characteristic projected matrix $P^{'}$ is built by using the algorithm of L-2DPCA. Then, the feature extraction is processed for an m * n dimensional training sample image A. The characteristic matrix of the image can be calculated by equation 3:

$$Y = (P^{'})^{T} AP \tag{3}$$

Where Y is the characteristic matrix of the image, will be used as input of the neural network classifier in following sections, its size is e * d.

978-1-4673-5887-3/13/$31.00 ©2013 IEEE

2.3 Feature Extraction from Image Edge

The algorithm of image edge detection presented in referenc e [6] is used to extract the edge image of the training sample image. The total pixels N_i of the edges were calculated, and the maximum gray value M_i of the edge pixels was selected based on the sample original image gray values. In a sampl e mammogram, the gray value of the edge pixel for microcal cification is generally larger than the value of other textures of breast, and the edges of all microcalcifications can be cla ssified. Therefore, the two features, the total pixels of the m icrocalcification boundaries and the maximum gray value of the edge pixels, are important for the extraction of the micr ocalcification in the mammogram. They are used as inputs f or the neural network classifier in this research. Because th e value region of the edge feature is different for each sampl e, it not only reduces the trained speed of the neural network, it also easily causes the neural network training to fail. So, t he feature data can be unified by:

$$N_i^{'} = N_i / k_1 \tag{4}$$

$$M_i^{'} = M_i / k_2 \tag{5}$$

where $N_i^{'}$、 $M_i^{'}$ are the total pixels of the microcalcificatio n boundaries and the maximum gray value of the edge pixel s after the unification, *i* is the index of the input samples.

2.4 Neural Network Classifier

In this paper three layer BP neural network is used to extract the ROI in the mammogram. There are 6 nodes in the input layer of the neural network. They are 4 principal component features and two edge features respectively. Two nodes were selected in the output layer [7]. Twelve nodes were considered in the hidden middle layer. There are two steps for the algorithm: training process and measuring process. 100 digital mammograms were selected as measuring samples. Each digital mammogram was divided into many sub-images with size 128 * 128. The image pre-process was made for each sub-image based on the Eq2. The image principal component features and the image edge features were extracted for the preprocessed images. The trained neural network classification algorithm was utilized to classify the samples. The parameters were selected as e = 2, d = 2 in this research [8].

3 RESULTS

The developed algorithm of this paper was verified by classifying 100 mammograms provided by Shandong Medical Imaging Research Institute. Some experiment results are shown as Figure 1. Where (a) is the original image, (b) is the suspicious area containing benign and malignant microcalcifications, and (c) is the enlarged image of the cell areas (ROI) of the picture (b).

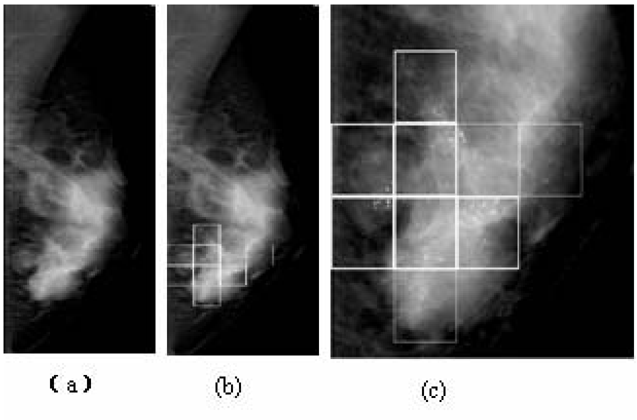

Fig.1. the Results of ROI Extraction

In order to verify the accuracy of this algorithm, the current general used diagnosis experimental verification ROC of the world is employed. In this research, a total of 100 mammograms were used as experimental samples. Under the guidance of expert radiologists, the ROC curve was obtained by the method of modifying the important points or threshold. From this diagram, the misdiagnosis rate of this algorithm is about 14%. The positive classification rate is about 95%.

4 CONCLUSIONS

This paper has provided the extraction method of ROI in mammogram based on the two-dimensional principal component and feature analysis. The whole procedure of the classification is automatic, and has high intellectual power. Since the image matrix is directly used by the two-dimensional principal component analysis, the training speed of the neural network is fast, the calculation work is reduced sharply.

This paper combines the multiple characteristic parameters for the classification of the neural network, which make the classification more efficient and accurate. Extracting microcalcification region in the mammogram reduces the computing work of the computer aided diagnosis system for later breast cancer, improves the intellect of the computer aided diagnosis system for breast cancer, and builds the basis of digital hospital system.

5 ACKNOWLEDGEMENTS

We would like to thank for support of Shandong Province Natural Science Fund.

REFERENCES

[1] S. Bouyahia, J. Mbainaibeye, N. Ellouze, Wavelet based microcalcifications detection in digitized mammograms, ICGST International Journal on Graphics, Vision and Image Processing, GVIP. Vol.8, No.5, 23-31, 2009.

[2] Sophie Paquerault, Paul T. Hardy, Nancy Wersto, Investigation of Optimal Use of computer Aided Detection Systems, Academic Radiology. Vol.17, No.9, 1112-1120, 2010.

[3] Yang J,Zhang D, Two-dimensional PCA:a new approach to appearance-based face representation and recognition, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.26, No.1, 2004.

[4] Ruiping Wang, New computer aided diagnosis method of mammograms: [ph.D, Tianjing:Tianjing University, 2003.].

[5] Sankar Deepa, Thomas Tessamma, Fractal modeling of mammograms based on mean and variance for the detection of microcalcifications, Proceedings-International Conference on Computational Intelligence and Multimedia Applications ICCIMA 2007, Vol.2, 334-338, 2008.

[6] Guangyu Zhang, Guangzhen Guong, Weiluo Zhu, New color image edge detection algorithm based on the clone algorithm, Electronic Journal, Vol.34, No.4, 702-707, 2006.

[7] Junli Li, Yanqin Hou, Ping Wei, et.., A removing noise method of medical image based on fuzzy variation and wavelet transform, China Biomedical Engineering Journal, Vol.26, No.3, 342-348, 2007.

[8] Ryohei Nakayama, Yoshikazu Uchiyama, Koji Yamamoto, et al., “Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms,” IEEE Transactions on Biomedical Engineering, Vol.53, No.2, 273-283, 2007.