

Novel Preprocessing Technique in the Computer Aided Detection of Breast Cancer

¹Senthilkumar Balasubramaniam and ²Umamaheswari Govindaswamy

¹Department of Electronics and Communication Engineering,
Tamilnadu College of Engineering, Coimbatore, Tamilnadu, 641 659, India

²Department of Electronics and Communication Engineering,
P.S.G. College of Technology, Coimbatore, Tamilnadu, 641 004, India

Received 2012-08-21, Revised 2012-09-19; Accepted 2012-11-08

ABSTRACT

Analyzing medical images is the most challenging task in medical image processing. Computer Aided Detection (CAD) tool is the aid for the radiologists in analyzing such images for the effective detection and diagnosis of the disease. Such a CAD tool consists of Preprocessing, Segmentation and detection processes. In this study we have improved the preprocessing by using the Selective Median Filter (SMF) for the noise removal and modified the Local Range Modification (LRM) as modified LRM (MLRM) for the contrast enhancement to detect the breast cancer. We have tested the performance of some preprocessing methods and compared with the proposed method (MLRM). These methods (SMF and MLRM) had been tested for over 30 mammogram images and found the accuracy of 97.9% which is better than the other existing methods.

Keywords: Computer Aided Detection (CAD), Preprocessing, Breast Cancer, Mammogram, SMF, MLRM

1. INTRODUCTION

Breast cancer is the most dangerous disease for women over the age of 40. Early detection and treatment is the only possibility of control it. In olden days, breast cancer has been identified by taking sample tissue from the suspected breast area (biopsy). In this modern world, many CAD has been designed and it has been used by the radiologists as an aid for the effective detection of breast cancer (Yang *et al.*, 2005; Cheng *et al.*, 2006; Retico *et al.*, 2006; Boccignone *et al.*, 2000). CAD consists of preprocessing (noise removal and contrast enhancement), segmentation and detection. Among this, Preprocessing is the main process. In this study we have discussed about the SMF for noise cancellation and the LRM, the MLRM for contrast enhancement.

2. MATERIALS AND METHODS

The database of mammograms used in this work is known as Mammographic Image Analysis Society

(MIAS) Mini Mammographic Database. The example image used in this study is mdb75 and it is shown as original image in Fig. 1. The entire method presented in this study was implemented in MATLAB 7.0 and makes extensive use of the Image Processing Toolbox. The methodology used consists of three main stages. First is the pre-processing stage and it consists of noise removal and enhancement. Second is the segmentation stage and the third is detection stage.

2.1. Preprocessing

The basic need for pre-processing in mammography is to remove the noise and to increase the contrast, especially for dense breasts. There are two possible approaches to enhance mammographic features. One is to increase the contrast of suspicious areas and the other is to remove background noise (Yang *et al.*, 2005). Removing background noise while preserving the edge information of suspicious areas can enhance a digital mammogram. In this approach four selective averaging schemes were used to modify the median

Corresponding Author: Senthilkumar, B., Department of Electronics and Communication Engineering,
Tamilnadu College of Engineering, Coimbatore, Tamilnadu, 641 659, India

filtering called selective median filtering. A selective median filter is, given a window $W(i, j)$, centered at image coordinates (i, j) , the output of the selective median filter is given in Eq. 1:

$$X_{i,j} = \text{median}\{X_{r,s} : (r,s) \in N(i,j), \text{ and } |X_{r,s} - X_{i,j}| < T\} \tag{1}$$

Where:

- $X_{i,j}$ = The image intensity at (i, j)
- $N(i, j)$ = The area in the image covered by the window $W(i, j)$
- T = A threshold

In computing the median, the set of pixel is restricted to those with a difference in gray level no greater than some threshold T . If T is small, the edge-preserving power of the filter is strong, but its smoothing effect is small. If T is large, the filtering behaves the other way around. And the resulted image is shown as selective median filtered image (Output of SMF). After the noise removal the mammogram image is further processed by contrast enhancement technique (MLRM).

2.2. Modified Local Range Modification (MLRM)

Local Range Modification (LRM) is a linear stretching approach follows the formula $y = ax + b$, where y is the enhanced image, x is the original grayscale image and a, b are parameter depending on the local contrast, which are computed by an interpolation procedure using overlapping image blocks (Papadopoulos *et al.*, 2008). The LRM algorithm processes twice the whole image. In the first pass, the local parameters are calculated and in the second contrast enhancement is performed. During the first pass, the maximum and minimum pixel values of non-overlapping 51×51 pixel sized blocks are computed. In the same way, the maximum and the minimum pixel values of half-overlapping blocks are calculated. The MLRM algorithm processes same as that of LRM (Papadopoulos *et al.*, 2008) but with two changes. The first is maximum and minimum pixel values of non-overlapping 48×48 pixel sized blocks are computed during first pass instead of 51×51 in LRM. And the second is estimation of regional maximum and minimum values based on the interpolation of eight surrounding grid points instead of four in LRM is shown in Fig. 1. The (H_i, L_i) and (M_i, N_i) are the minimum and maximum grayscale values:

$$\begin{aligned} \max = & \left[\frac{s_y}{s} M_4 + \left(\frac{s-s_y}{s} \right) M_1 \right] \left(\frac{s-s_x}{s} \right) + \\ & \left[\frac{s_y}{s} M_8 + \left(\frac{s-s_y}{s} \right) M_7 \right] \left(\frac{s-s_x}{s} \right) + \\ & \left[\frac{s_y}{s} M_6 + \left(\frac{s-s_y}{s} \right) M_9 \right] \left(\frac{s-s_x}{s} \right) \\ & + \left[\frac{s_y}{s} M_2 + \left(\frac{s-s_y}{s} \right) M_3 \right] \left(\frac{s-s_x}{s} \right) \end{aligned} \tag{2}$$

From Eq. 2, s is the size of the block, s_x and s_y are the horizontal and vertical distances of the examined point, respectively, from the M_5 grid point and $M_1, M_2, M_3, M_4, M_6, M_7, M_8$ and M_9 are the intensity values of the eight surrounding grid points. The output value of each pixel with coordinates $[m, n]$ is calculated by linear stretching in Eq. 3:

$$y[m,n] = \frac{L-1}{(\max-\min)} (x[m,n] - \min) \tag{3}$$

where, L is the number of grayscales (image depth), \max and \min are the margins of the local input grayscale range, respectively.

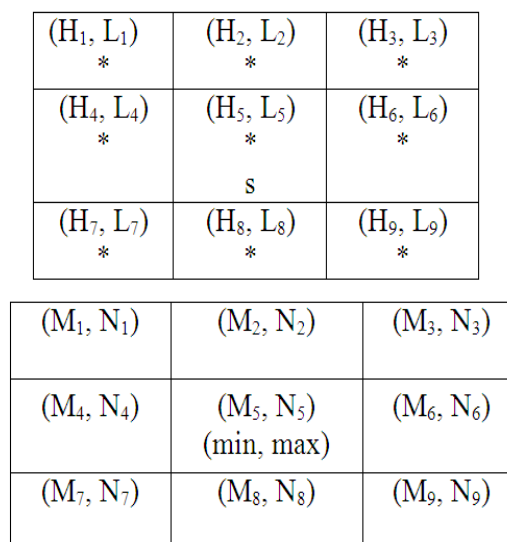


Fig. 1. Estimation of regional maximum and minimum values based on the interpolation of eight surrounding grid points ((H_i, L_i) and (M_i, N_i) are the minimum and maximum grayscale values for each grid point s)

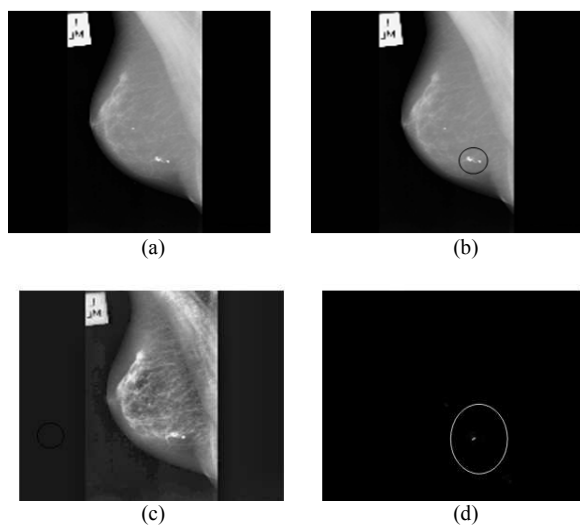


Fig. 2. Results of the mammogram image (mdb75) taken from MIAS database (a) Original Image (b) Output of SMF (c) MLRM Enhanced Image (d) Feature Segmented Image

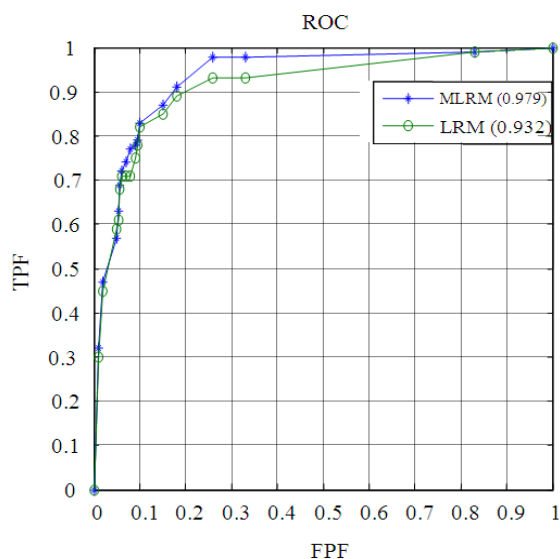


Fig. 3. ROC analysis

The critical parameter in the LRM algorithm (Papadopoulos *et al.*, 2008) is the local neighborhood length which corresponds to the block size. A size of 51×51 pixels was selected as block size.

It is noted that large local maxima or minima develop regions of low contrast around them and sharp differences between two regions of large gray differences result in

low-contrast regions. These modifications enhance the image better than the LRM method and shown in **Fig. 2c**.

3. RESULTS

The proposed preprocessing method combined with the fuzzy clustering of spatial patterns algorithm (Xia *et al.*, 2007) for segmentation and segments the feature, shown as segmented image in **Fig. 2d** (circled area).

The first image in **Fig. 2a** is the original image (mammogram mdb75), the second (b) is the selective median filtered image (Output of SMF) and the third (c) is the MLRM enhanced image. This image has already been tested and found that it is a fatty breast with malignancy. Further, the Receiver Operating Characteristics (ROC) analysis has been done for this CAD method and shown in **Fig. 3**.

4. DISCUSSION

We have tested over 30 sample images for the True Positive (TP): lesions called cancer and prove to be cancer, False Positive (FP): lesions called cancer that proves to be benign, False Negative (FN): lesions that are called negative or benign and prove to be cancer, True Negative (TN): lesions that is called negative and proves to be negative. On the basis of this terminology, performance of the proposed technique has been evaluated by calculating True Positive Fraction/sensitivity (TPF) and False Positive Fraction/specificity (FPF).

The average ROC values of LRM and MLRM has been compared with Contrast Limited Adaptive Histogram Equalization (CLAHE), Wavelet Linear Stretching (WLST), Wavelet Shrinkage (WSRK) and Wavelet Background Approximation (WBGK) are tabulated in **Table 1**.

Comparison of MLRM with the LRM in detecting the cancer is also tabulated in **Table 2**. From these results, it is found that this method works better than the other methods used in CAD and this has also been validated by expert radiologists.

5. CONCLUSION

This modified preprocessing method for CAD will provide good support to the radiologist in detecting the breast cancer. Selective median filter correctly removes the noise and provides the original image in a clear form. The result we found is good and satisfactory.

Table 1. Average ROC values of preprocessing techniques for several percentages of most contrasted pixel values (minimum number of pixel per object: 7 and the best % of most contrasted pixel value is 4)

Percentage of the most contrasted pixels	Enhancement technique						
	Without enhancement	CLAHE	WSRK	WLST	WBGK	LRM	Modified LRM
1	0.721	0.710	0.687	0.713	0.741	0.753	0.764
2	0.743	0.741	0.713	0.764	0.758	0.782	0.791
3	0.772	0.781	0.762	0.841	0.804	0.841	0.856
4	0.803	0.809	0.815	0.862	0.823	0.932	0.979
5	0.847	0.834	0.836	0.882	0.874	0.865	0.892

Table 2. Comparison based on Average ROC of LRM and MLRM

Enhancement TECHNIQUE	Average ROC	Comments
LRM method (Papadopoulos <i>et al.</i> , 2008)	93.2%	It can enhance image better than the other existing methods and the accuracy is poor comparing to proposed method.
Proposed Method (MLRM)	97.9%	Accuracy is high and it can enhance image much better than the LRM and all other existing methods

Further, the MLRM, with the fuzzy segmentation method provides better result on mammogram image and able to find the exact cancer area (Microcalcification). ROC analysis has been done for MLRM and this method provides the accuracy of 97.9%, which is better compared to the existing methods.

6. REFERENCES

- Boccignone, G., A. Chianese and A. Picariello, 2000. Computer aided detection of microcalcifications in digital mammograms. *Comput. Biol. Med.*, 30: 267-286. DOI: 10.1016/S0010-4825(00)00014-7
- Cheng, H.D., X.J. Shi, R. Min, L.M. Hu and X.P. Cai *et al.*, 2006. Approaches for automated detection and classification of masses in mammograms. *Patt. Recogn.*, 39: 646-668. DOI: 10.1016/j.patcog.2005.07.006
- Papadopoulos, A., D.I. Fotiadis and L. Costaridou, 2008. Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Comput. Biol. Med.*, 38: 1045-1055. DOI: 10.1016/j.combiomed.2008.07.006
- Retico, A., P. Delogu, M.E. Fantacci, A.P. Martinez and A. Stefanini *et al.*, 2006. A scalable computer-aided detection system for microcalcification cluster identification in a pan-European distributed database of mammograms. *Nucl. Instruments Methods Phys. Res.*, 569: 601-605. DOI: 10.1016/j.nima.2006.08.094
- Xia, Y., D. Feng, T. Wang, R. Zhao and Y. Zhang, 2007. Image segmentation by clustering of spatial patterns. *Patt. Recogn. Lett.*, 28: 1548-1555. DOI: 10.1016/j.patrec.2007.03.012
- Yang, S.C., C.M. Wang, Y.N. Chung, G.C. Hsu and S.K. Lee *et al.*, 2005. A computer-aided system for mass detection and classification in digitized mammograms. *Biomed. Eng. Appl. Basis Commun.*, 17: 215-228. <http://bme.ntu.edu.tw/abc/17.5/17-5-1.pdf>