International Journal of the Physical Sciences Vol. 7(3), pp. 471 - 477, 16 January, 2012
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.1339
ISSN 1992 - 1950 © 2012 Academic Journals

Full Length Research Paper

Preprocessing and pectoral muscle separation from breast mammograms

Muhammad Talha[1,2]* and Ghazali Bin Sulong[1]

[1]Department of Computer Graphics and Multimedia, Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Malaysia.
[2]Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia.

Accepted 20 December, 2011

Computer aided diagnosis (CAD) systems can be used as a second opinion to the radiologists for diagnosis of breast cancer from mammogram images. In this paper, we have proposed preprocessing method to remove noise from mammogram images. Then, enhancement has been performed. After that, background has been removed. Finally, pectoral muscle separation has been performed. It has been noted that results are very much satisfactory. This can be used further to improve the accuracy of diagnosing breast mammogram. We have used MIAS data set for experimentation purpose.

Key words: Breast cancer, mammogram, enhancement, pectoral muscle.

INTRODUCTION

Breast cancer is considered to be one of the leading causes of deaths among females on a global level. In Netherlands for example, approximately 10000 women are diagnosed with this disease per annum and approximately 3500 of these women die from this type of cancer. American National Cancer Institute reported that the population of the estimated new breast cancer cases for the 2006 in USA is round about 214640, while the estimation of deaths is more than 41,000 (Broeders and Verbeek, 1997). Cancer statistics claim that breast cancer got the third position of appearance in diagnosed new cases following genital organs and digestive systems cancer as compared to other forms of cancer. Over the past decades, it has become alarming that breast cancer incidence rates are increasing steadily. Changes in risk factors seem to contribute to the rising incidence. However, the mortality rates for breast cancer have remained relatively constant due to more effective treatment and earlier diagnosis (Nawazish et al., 2011). According to American Cancer Society 2007, United States has the highest figure in the world about crude and age-standardized breast cancer incidents. About 178,480 women suffer from invasive breast cancer and 62,030 from *in situ* breast cancer. 85% of the total *in situ* breast were ductal carcinoma *in situ* 00 (DCIS). 40,460 women in American died because of this disease in 2007. An increment is seen in breast cancer death rate between 1975 and 1990, by 0.4% annually. But due to some good treatment and mammographic diagnosis from 1990 to 2002, this is going down by an average of 2.3% per year. Infection rate is different in African black and American white woman. American white women have 20% more chances to have breast cancer than African black women. Although, in early 1980's it was higher. Computerized detection of lesions is also being done for mammograms. Computer aided diagnosis (CAD) is being developed for the detection and diagnosis of breast cancer and for the assessment of breast cancer risk (Wallis et al., 1991). Computerized image analysis in screening mammography has already yielded many fruitful results. Food and Drug Administration (FDA) shows that computer aided system are very helpful in screening method of mammogram images.

There are a number of well-known and potential risk factors for breast cancer. These can be divided into seven broad categories: age, hormonal factors, family history of breast cancer, proliferate breast disease, irradiation of the breast region at an early age, lifestyle factors and personal history of malignancy. In reality, estimates indicate that between 10 to 30% of breast cancers are missed by radiologists during routine screening. The penalty of errors in detection or classification is very high.

*Corresponding author. E-mail: mtalhanaseem@yahoo.com.

Mammography itself cannot prove that a suspicious area is malignant or benign. To decide that, the tissue has to be removed for examination using breast biopsy techniques. A false positive detection may cause an unnecessary biopsy. Statistics show that only 20 to 30% of breast biopsy cases are proved cancerous. In a false negative detection, an actual tumor remains undetected that could lead to higher costs or even to the cost of a human life. With the growth of computer technology, radiologists have a chance to improve their image interpretation using computer capabilities that can improve the image quality of mammograms (Tang et al., 2009). In order to develop the accuracy of interpretation, a variety of computer-aided diagnosis (CAD) systems like Wallis et al. (1991) have been proposed. CAD plays an important role in diagnosis of breast cancer and defining the extent of breast tumors. In previous twenty years, much effort has been made by computer scientists to support the radiologists in detection and diagnosis of cancerous masses by developing computer-aided tools for mammography interpretation. Image processing and intelligent systems are two important mainstreams of computer technologies that have been continuously explored in the development of computer-aided mammography systems.

Major contributions

1. A fully automatic and robust technique has been proposed.
2. Strong preprocessing technique and automatic abnormality type detection method is used.
3. No prior knowledge of the mammogram is required about its feature, type and contents.
4. This is a supervised method for diagnosing breast cancer.
5. Proposed system achieved quite good accuracy for the classification of mammograms as malignant and benign.

RELATED WORK

Several works have been done to develop computer aided breast cancer detection and diagnosis tools. Tang et al. (2009) gave an overview of recent advances in the development of such tools and related techniques. Kom et al. (2007) proposed a technique for the automated detection of malignant masses in screening mammography. The technique is based on the presence of concentric layers surrounding a focal area with suspicious morphological characteristics and low relative incidence in the breast region. Malignant masses were detected with 92, 88 and 81% sensitivity of 5.4, 2.4 and 0.6 false positive per image. Eltonsy et al. (2007) introduced an algorithm for detection of suspicious masses in mammographic images that shows a sensitivity of 95.91% for mass detection, with receiver operating characteristics (ROC) area of 0.946 when the enhancement of the original image was performed before detection and 0.938 otherwise.

Other approaches that are less dependent of the contrast may be more useful, like template-matching, a method used in some of the earlier papers in this field (Nawazish et al., 2011; Wang et al., 1999). A model is made of the appearance of a mass, and the mammogram is searched for regions that resemble this model. This approach is more related to the shape, and less to the contrast of the region. Especially for hard to detect low contrast masses, this method may outperform convolution based approaches. Most recent methods for mass detection focus on the analysis of the gradient patterns in an area of interest. The appearance of masses in mammograms varies and therefore the earlier described rigid approaches are not very successful. In an area with a central mass, the orientation of the gradients will be towards the center of the mass. Statistical analysis of this pattern can be used to discriminate masses from other structures. Timp (2006) used a generalized Hough transform for circles. The strongest edges in an area of interest are accumulated in a Hough space where each location relates to a center and a radius. Masses will yield peaks in this space.

Mavroforakis et al. (2004) applied a one-dimensional recursive median filter over a number of different angles to each pixel. Based on the variations in scale for various angles, they can determine whether the structure is a blob or has a more linear shape. Sometimes, a mass looks very much like normal glandular structure, and is only detectable due to asymmetry between the left and right breasts. A few papers have been published describing approaches for mass detection based on differences in left and right mammograms. These approaches perform some kind of image subtraction, and can also be used to detect temporal changes when a mammogram is compared with an older mammogram of the same breast. Matching two breasts is a complicated procedure, because there is only an approximate correspondence between the normal tissue in the two breasts, and due to variations in compression and positioning, the variation in appearance is even made larger. Detection of the spicules is another parameter of estimating the severity of disease. When a mass is surrounded by spicules, it is likely to be malignant. Many stellate lesions are easier to detect by their spicules than by their central mass, and for architectural distortions it is the only sign.

Histogram equalization (HE)

Histogram equalization is another method to enhance the contrast of an image. A new enhanced image with uniform histogram is created by histogram equalization. This is attained by using a normalized cumulative histogram as a gray scale mapping function.

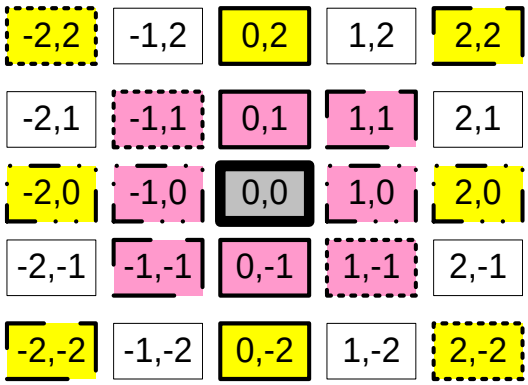

Figure 1. Four directions used for detecting noisy pixel.

Contrast-limited histogram equalization (CLAHE)

The contrast limiting procedure has to be applied for each neighborhood for enhancing the local contrast of an image from which a transformation function is derived. CLAHE was developed to prevent the over amplification of noise that adaptive histogram equalization can give rise to. The method has three parameters.

Block size

Block size is the size of the local region around a pixel for which the histogram is equalized. This size should be larger than the size of features to be preserved.

Histogram bins

Histogram bin is the number of histogram bins used for histogram equalization. The implementation internally works with byte resolution, so values larger than 256 are not meaningful. This value also limits the quantification of the output when processing 8 bit gray or 24 bit red/green/blue (RGB) images. The number of histogram bins should be smaller than the number of pixels in a block.

Max slope

Max slope limits the contrast stretch in the intensity transfer function. Very large values will let the histogram equalization do whatever it wants to do, that is, result in maximal local contrast. The value 1 will result in the original image.

PROPOSED SYSTEM

The proposed system is divided into four major parts:

1. Preprocessing for noise removal
2. Enhancement by using CLAHE
3. Background removal
4. Pectoral muscle separation.

The detail of these four steps is described subsequently one after the other.

Preprocessing for noise removal

Salt and pepper noise model

With this kind of noise, one pixel is assigned either minimum or maximum intensity value. In case of impulse noise, this type is considered to be most simple and most widely used. Other pixels can have any value from allowed dynamic limit when we use random values impulse noise model. This kind of noise is not easy to detect and separate as compared to simple salt and pepper noise. In our work, our main area of attention is the separation of both kind of noises from 8 bit gray scale images.

Let $x(i, j)$ and $y(i, j)$ be the pixel values at position (i, j) of the original and noisy image, respectively. Where p is the probability of impulse noise model. This can be described in this way.

$$x(i,j) = \begin{cases} o(i,j) & 1-p \\ \eta(i,j) & p \end{cases} \quad (1)$$

where $\eta(i, j)$ is the noisy pixel at position (i, j). The noisy pixel $\eta(i, j)$ can get value between 0 to 255 for 8 bit grayscale image.

Noise removal

We have used directional weighted median filter (Dong and Xu, 2007) which works as shown in Figure 1. In this approach, there are two major steps:

1. Detect noisy pixel using new impulse detector
2. Utilize weighted directional information to calculate the median for removing impulse noise and preserve details

Let DIR_j (j = 1…4) denotes a set of coordinates aligned with the jth direction centered at (0, 0), that is;

$Dir_1 = [Pix_{-2}^1$ (-2, -2), Pix_{-1}^1 (-1, -1), Pix_0^1 (0, 0), Pix_1^1 (1, 1), Pix_2^1 (2, 2)],
$Dir_2 = [Pix_{-2}^2$ (0, -2), Pix_{-1}^2 (0, -1), Pix_0^2 (0, 0), Pix_1^2 (0, 1), Pix_2^2 (0, 2)],
$Dir_3 = [Pix_{-2}^3$ (2, -2), Pix_{-1}^3 (1, -1), Pix_0^3 (0, 0), Pix_1^3 (-1, 1), Pix_2^3 (-2, 2)],
$Dir_4 = [Pix_{-2}^4$ (-2, 0), Pix_{-1}^4 (-1, 0), Pix_0^4 (0, 0), Pix_1^4 (1, 0), Pix_2^4 (2, 0)],

In Pix_i^k, the value of j specify the pixel in the k direction where $-2 \le j \le 2$. Now consider 5 × 5 window centered at (i, j), we calculate the sum of all the absolute weighted differences of gray level values in a specific direction, Diff(k) is used to define the differences where k specify the direction. The weights are multiplied at the time of calculating the differences for each pixel with the centered pixel in a particular direction, and the value of the weights depends on the closeness of the pixel Pix_i^k from the center pixel Pix_0^k. If the spatial distance for two pixels is small then their gray level values should be close to each other. Thus, here is an equation that represents the sum of the weighted difference between current and neighboring pixels.

$$\mathrm{Diff}^k = \sum_{i=-2, i\neq 0}^{2} (w^i * |\mathrm{PIX}^i - \mathrm{PIX}^0|), -2 \leq i \leq 2 \quad (2)$$

where

$$W_i = \begin{cases} 2, & i = -1, +1 \\ 1, & i = -2, +2 \end{cases}$$

This shows the weights according to the neighborhood. Diff^k has been used as a direction index and each direction index is receptive to the edge aligned with a given direction.

To identify the impulse noise, we use minimum value from all four direction indexes, if the selected value is greater than predefined threshold T, as its value is 495 in our case, then pixel is noisy, otherwise noise-free, as shown in Equation 3.

$$R = Min(Diff^k \mid 1 \leq k \leq 4) \quad (3)$$

and

$$PIX_0 \; is \begin{cases} Noise\ Free & If\ R > T \\ Noisy & If\ R \leq T \end{cases}$$

where T is threshold, Min is the operator to identify the minimum value from all four Diff^k values. Now we can determine the noise by employing a threshold T, no matter if we are dealing with an edge, flat region or thin line.

After detecting the impulse noise, many researchers apply the standard median filter for the reduction of the noise. The details preservation is not possible with standard median filter, so to overcome this problem, a new directional weighted median filter in which information of the four directions is incorporated to effectively preserve the detail and remove the impulse noise. We all know that the standard deviation is used to determine how tightly all values are clustered to a specific value. The steps of directional weighted median filter are as follows:

1. The standard deviation is calculated for each direction and we choose the direction where the standard deviation is minimum.
2. Pick pixel values from this direction and, add them to the existing window, add them twice to the existing window to increase the possibility of the nearest to the exact median value.
3. After that, we apply the standard median filter to the new updated window.

Results have been shown in Figure 2.

Enhancement of mammogram

In this step, contrast limited adaptive histogram equalization (CLAHE) technique has been applied (Antonis et al., 2007). In CLAHE, the pixel's intensity is transformed to a value within the display range proportional to the pixel intensity's rank in the local intensity histogram. The enhancement is condensed in flat areas of the image, which prevent over enhancement of noise. It also reduces the edge shadowing effect. The CLAHE operates on small regions in the image called tiles rather than the entire image. Each tiles contrast is enhanced, so that the histogram of the output region approximately matches the uniform distribution or Rayleigh distribution or exponential distribution. Distribution is the desired histogram shape for the image tiles. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries.

First of all, input image is divided into equal size of number of non-overlapping regions. Then, the histogram of each region has been calculated. Clip limit has to be set for clipping histograms. In our case, we have set t = 0.002. Each histogram has been redistributed in such a way that its height does not exceed the clip limit. All histograms were modified by the transformation function of normal histogram. Then, using bilinear interpolation, neighboring tiles has been combined. At the end, image gray scale values have been altered according to the modified histograms. Results have been shown in Figure 3.

Background removal

Automatic cropping of breast from the mammograms is a very critical task and it is known as region of interest (ROI).

For the symmetry, images which were having breast on the opposite side are flipped for nipple pointing to the right (Wallis et al., 1991). Then, the process of cropping is performed and its purpose is to focal point the process absolutely on the appropriate breast region, which reduces the possibility for erroneous classification by areas which are not of interest (Nawazish et al., 2011). We have used the following method to remove background. First of all, we have generated a matrix equal to the size of input image and this matrix will be used for labels. We have performed an operation to scan the image. We have to track the pixels whether already visited or not. We have to differentiate the background and foreground pixels. We have to scan the image till foreground pixels would not been found. If foreground pixel is found, then, we will check whether it is already visited or not. In this way, we will store all the positions of foreground pixels and labels these pixels in matrix that has been generated at the start. At the end, initial matrix will show all labels of foreground pixels and we will again scan the image. We retail all those pixels which have been labeled and remove all other pixels. In this way, we will get a new cropped image in which it background has been removed. The output of this algorithm has been shown in results part, subsequently. The results have been shown in Figure 4.

Pectoral muscle separation

For the detection of pectoral muscle, we have used modified seeded region growing algorithm. It separates the pectoral muscle from the breast. Modified seeded region growing algorithm will give us the desired output only if seed is taken from pectoral muscle. We have selected this seed point randomly at the start. This is an iterative process. We have taken a simple seed point and compare that pixel with the neighboring pixels in eight neighborhood way. We grow this region by adding these neighbor pixels which are similar to seed point. When this region is having no sufficient matching points, we stop this region. We have taken again a seed point which has not been used already or traversed. We start the growing process again in eight neighborhoods and grow these regions. At the end, we end up with all those regions which show the pectoral muscle part in breast. An initial set of small areas are iteratively merged according to similarity constraints. The results are as shown in Figure 5.

RESULTS AND DISCUSSION

We have used publically available databases MIAS (Suckling et al., 1994). The dataset was taken from the Mammographic Institute Society Analysis (MIAS). Each mammogram is of size 1024 × 1024 pixels, and resolution of 200 micron. There are 322 mammograms of right and left breast of 161 patients in this dataset. 69 mammograms were diagnosed as being benign, 54 malignant and 207 normal.

First of all, salt and pepper noise has been added and

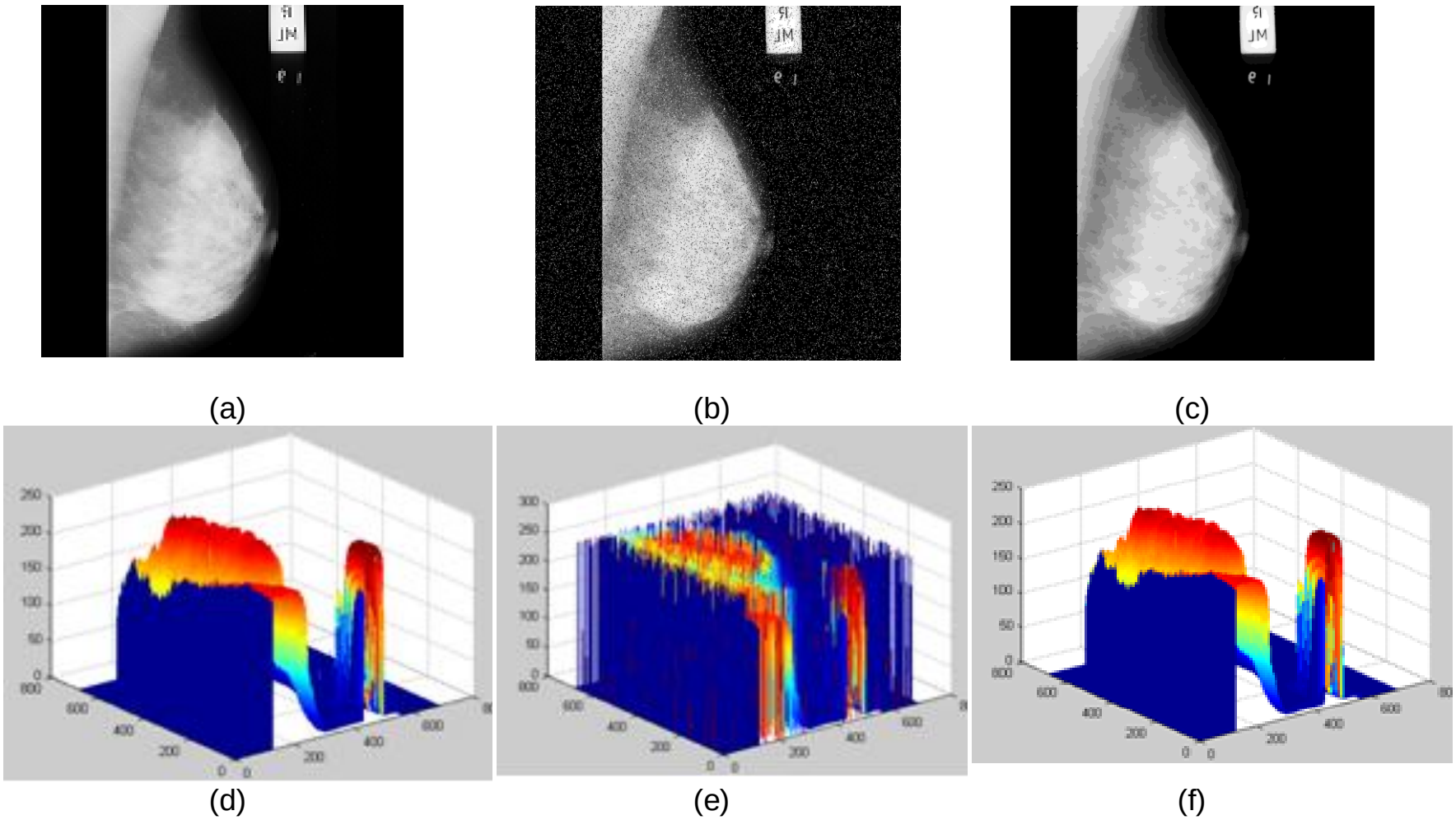

Figure 2. Noise removal from breast mammogram images. (a) Original Image(b) Noisy Image (0.1) (c) Restored Image(d) Original Image (e) Noisy Image (0.1) (f) Restored Image.

directional weighted median filter has been applied. Results have been shown in Figure 2. It has been shown that directional weighted median filter has restored image good. Mesh shows that restored image is closely related to the original image.

After noise removal, enhancement has been done by CLAHE. Results have been shown in Figure 3. Results show that CLAHE performs well as compared to histogram equalization. Visually results show that CLAHE is good. Histograms show that CLAHE is closely related to the original image.

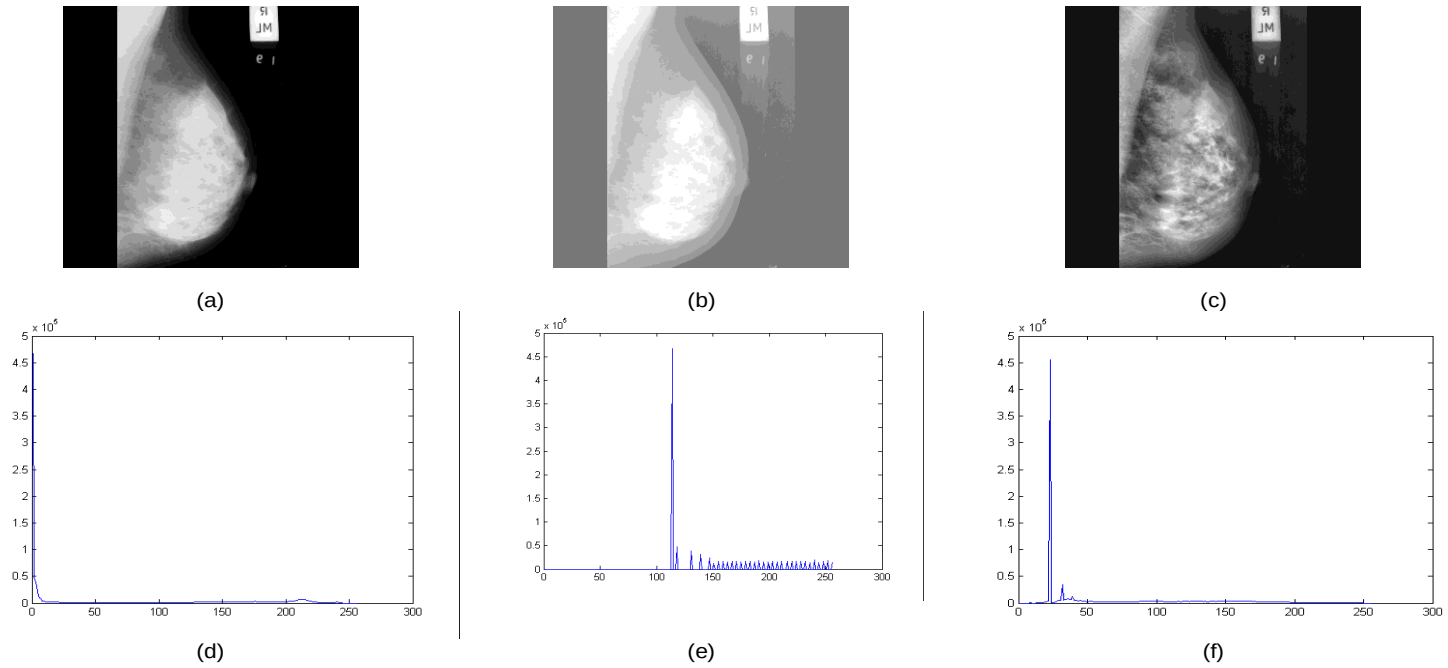

Figure 3. Enhancement of breast mammogram images. (a) Original Image (b) Histogram Equalization (c) CLAHE (d) Original Image Histogram (e) Histogram Equalization Histogram (f) CLAHE Histogram.

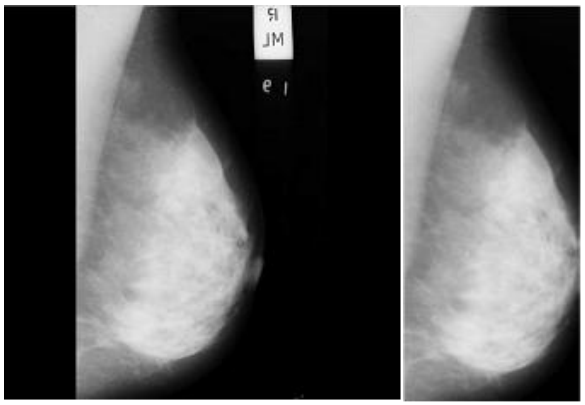

Figure 4. Background removal of breast mammogram images (a) Before background removal (b) After removal.

CONCLUSION AND FUTURE WORK

Proposed system is developed for diagnosing the breast cancer from mammogram images. This system performs this diagnosis in multiple phases.

In the first phase, preprocessing on mammogram image is done to remove noise. Directional median filter has been used to remove noise.

In the second phase, enhancement has been per-formed using CLAHE. In the third step, back-ground has been removed.

In the fourth step, modified seed region growing has been used for the separation of pectoral muscle from the breast.

All experiments show that the proposed system gives exception-ally good results. In future, we will perform classification of these breast images into benign and malignant to diagnose it.

ACKNOWLEDGEMENTS

The authors are thankful to the Faculty of Computer Science and Information Systems at the Universiti Teknologi Malaysia for providing research facilities and encouraging in conducting this research. The authors extend their appre-ciation to the Deanship of Scientific Research, King Saud University for funding this research work.

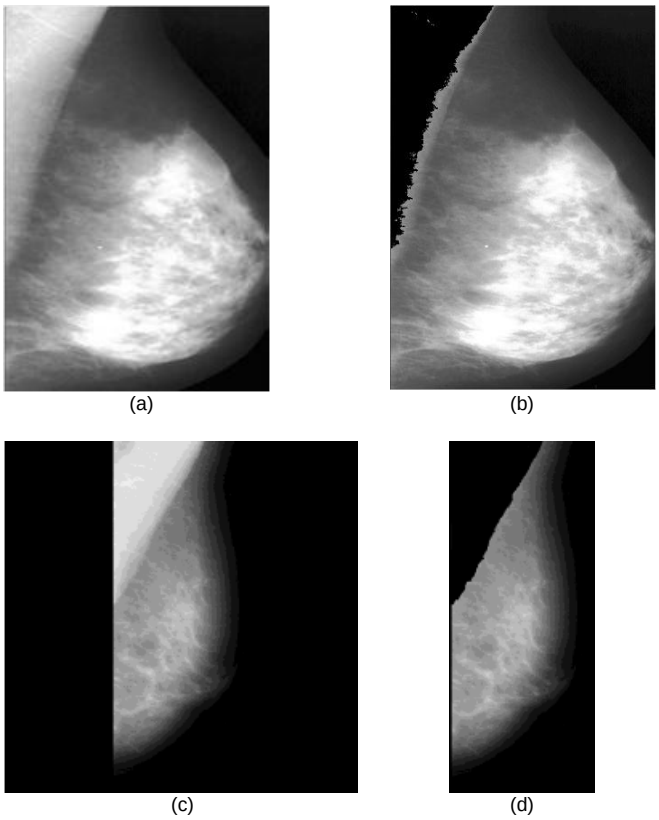

Figure 5. Pectoral muscle separation. (a) Image with pectoral muscle (b) Pectoral muscle removed (c) Image with pectoral muscle (d) Pectoral muscle remove.

REFERENCES

Antonis D, Dionisis C, Panagiotis B, Spiros K, Pantelis G, Ioannis K, George N (2007). An Efficient CLAHE-Based, Spot Adaptive, Image Segmentation Technique for Improving Microarray Genes' Quantification", 2nd International Conference on Experiments/Process/System Modelling/Simulation & Optimization, Athens.

Broeders MJM, Verbeek ALM (1997). Breast cancer epidemiology and risk factors. *Q*-J. Nucl-Med., 41(3): 179–188.

Dong Y, Shufang XU (2007). A New Directional Weighted Median Filter for Removal of Random Valued Impulse Noise, IEEE Signal Processing Letters, p. 14.

Eltonsy N, Tourassi G, Elmaghraby A (2007). A concentric morphology model for the detection of masses in mammography, Medical Imaging, IEEE Trans. On, 26(6): 880–889.

Kom G, Tiedeu A, Kom M (2007). Automated detection of masses in mammograms by local adaptive thresholding, Comput. Biol. Med., 37(1): 37–48.

Mavroforakis M, Georgiou H, Dimitropoulos N, Cavouras D, Theodoridis S (2004). Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines, European J. Radiol., 54(1): 80–89,

Nawazish N, Tae SC, Arfan MJ (2011). Malignancy and abnormality detection of mammograms using DWT features and ensembling of classifiers, Int. J. Phys. Sci., 6(8): 2107–2116.

Nawazish NM, Arfan J, Tae SC (2011). Segmentation and texture-based classification of breast mammogram images, Microscopy Research and Techniques, DOI: 10.1002/jemt.21070.

Suckling J, Parker J, Dance DR, Astley S, Hutt I, Boggis CRM, Ricketts I, Stamatakis E, Cerneaz N, Kok SL, Taylor P, Betal D, Savage J, Gale AG, Astley SM, Dance DR, Cairns AY (1994). The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series, 1069: 375-378.

Tang J, Rangayyan R, Xu J, El Naqa I, Yang Y (2009). Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances, Information Technology in Biomedicine, IEEE Trans. On, 13(2): 236–251.

Timp S (2006). Analysis of Temporal Mammogram Pairs to Detect and Characterise Mass Lesions. PhD in medical sciences, Radboud University Nijmegen, pp. 53-55.

Wallis M, Walsh M, Lee J (1991). A review of false negative mammography in a ymptomatic population," Clin. Radiol., 44: 13-15.