

# Pre-Processing Algorithms on Digital Mammograms

Sara I. Basheer

Dept. of Computer and Info. Eng.  
University of Mosul, Iraq

Younis M. Abbosh, PhD

Dept. of Computer and Info. Eng.  
University of Mosul, Iraq

Majid Dh. Younis, PhD

Dept. of Computer and Info. Eng.  
University of Mosul, Iraq

## ABSTRACT

Breast cancer is the greatest challenging health complexities that medical science is facing. Presently, there are no active methods to avert breast cancer, because its cause is not yet completely identified. Screening mammography is the available method that is currently used for reliable detection of breast cancer. Computer Aided diagnosis (CAD) techniques are used to enhance the diagnostic accuracy and efficiency of screening mammography. The sensitivity of mammogram decreases due to some factors like density of breast, presence of labels, and artifacts or even pectoral muscle. Therefore, the preprocessing of mammograms is a very significant step in breast cancer analysis and detection since it might reduce the number of false positive.

In this paper, several procedures have been performed for preprocessing including (noise reduction, separate the breast from the artifacts and pectoral muscle, mammogram alignment).

## Keywords

Digital Mammogram, Breast Cancer, Preprocessing

## 1. INTRODUCTION

Breast cancer is the most common cancer among women, comprising 23% of the 1.1 million female cancers that are newly diagnosed each year [1,2]. Breast cancer ranks second to lung cancer as the leading cause of death in women diagnosed with cancer[3].

Mammography is the best method for early mass detection. By using mammography screening program the mortality rate of breast cancer have been reduced by 30-70%[4].

Mammograms are difficult to interpret, especially in the screening context. Variety of Computer Aided diagnoses(CAD) systems have been presented in recent years to improve the efficiency and accuracy of interpretation process which consist of two important process: Computer-Aided Detection (CADE) and Computer-Aided Diagnosis(CADi) [5]. Detection process involves identifying the abnormal part of breast and classifying regions of mammography as positive and negative regions, While Diagnosis process involves characterizing those abnormal regions into either benign or malignant. This process can be performed either manually by the radiologist [6] or automatically by CADi system, Before using CADE system pre-processing algorithm must be used which represent an important factor for success or failure of the system in order to limit the search for abnormalities without undue influence from background of the mammogram. pre-processing steps include: mammogram orientation, label and artifact removal, detecting and removing pectoral muscle and mammogram alignment.

In this paper, a new method was proposed to detect and remove pectoral muscle and new idea to perform mammogram alignment that uses the lowest number of points for alignment.

## 2. PREVIOUS WORKS

A number of image processing methods have been proposed to perform the detection of suspicious mass regions.

R. M. Haralick and L. G. Shapiro [8] enhance digital mammogram to remove noise in the pre-processing using contrast limited adaptive histogram equalization (CLAHE) method.

Georgsson [9] proposed method for detecting and removing pectoral muscle by using region growing that consider pectoral muscle as a triangle then finding the yields of the muscle to be extract.

H. Mirzaalian and M.R Ahmadzadeh and S. Sadri and M. Jafari .[5] suggest two preprocessing methods, one for pectoral muscle segmentation and the other for the breast contour extraction. achieved by the Mean of Absolute Error Distance Measure (MAEDM) and Hausdorff Distance Measure (HDM).

Lau and Bischof [10] used methods for mammogram alignment by choosing multiple point in the breast contour from one image as a reference then performing transition operation on another image.

Bick, U., Giger, M. [11] use combination of region growing, morphological filtering and local thresholding for preprocessing ..

Most papers have multiple problems in the pre-processing stage, especially in the detecting and removing pectoral muscle and mammogram alignment steps. The problems in detecting and removing pectoral muscle stage are:

1. Most papers in general remove pectoral muscle by considering pectoral muscle as a triangle but this do not match with the natural state because pectoral muscle not a necessary triangle
2. In the comparison, most authors consider pectoral muscle area lighter than the rest of the breast and that is not completely right, because in some cases there are less light area within the muscle.

On the other hand the problems in Mammogram alignment are:

1. Choosing multiple points in the breast region for alignment can generate different problems because these points may be depend on the previous steps and this can change the position of these points and thus will cause to misalignment.
2. Considering one image as a reference and performing transition operation on another image will result in having a normal image and the other one that suffers from artificial distortion.

In this paper , new methods are proposed to overcome these problems by making the cut-line of the pectoral muscle a piecewise method and the comparison in gray level is row by row, furthermore, the number of points, used for alignment

that do not depend on the previous steps, are reduced, also transition operation on both left and right images is applied.

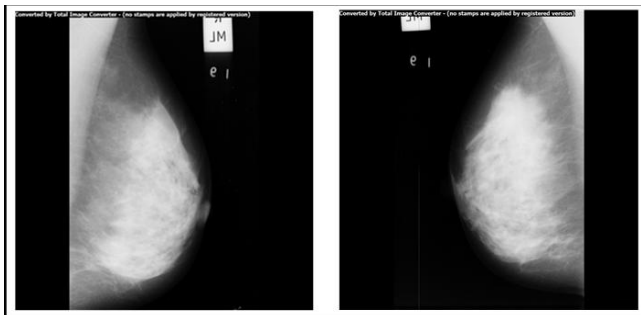
### 3. PRE-PROCESSING STEPS

#### 3.1 Mammogram Orientation

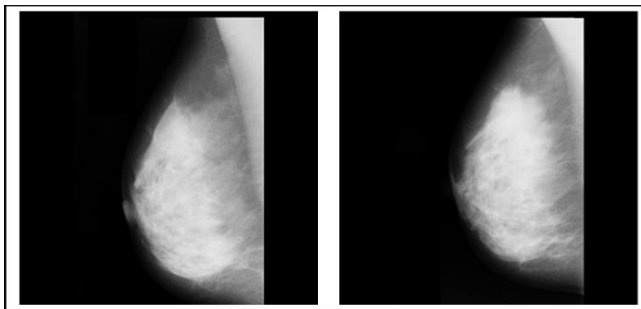
The mammogram images treated in this paper are taken from MIAS's database which contains images of both right and left breast. In some images, breast image is located in the left side, and in the other images, breast image is placed in the right side, this causes difficulties in the analysis of the breast to detect breast cancer. Therefore, breast image in all images must be moved so as to make the position of the pectoral muscle on the upper location of the image, this step requires the right breast flipped horizontally at 180° [12,13,14].

#### 3.2 Omit Noise and Unwanted Artifacts from Mammogram.

The next step is to remove noise and unwanted artifacts from mammogram. These artifacts may be caused by scanning error or they can be entered as meta information, then show the current view and side of the mammogram as shown in Fig.1. These artifacts are easily removed from actual image by thresholding the mammogram. After that, the breast appears as one object and the artifacts as separate objects. As a result, the largest connected component in each mammogram is kept and the rest are removed. The noise and uneven lighting are filtered from the mammogram image using average filter which is a type of linear filter.



Before steps A and B



After steps A and B

Fig.1: Left and Right mammogram before and after steps A&B

### 4. DETECTING AND REMOVING PECTORAL MUSCLE

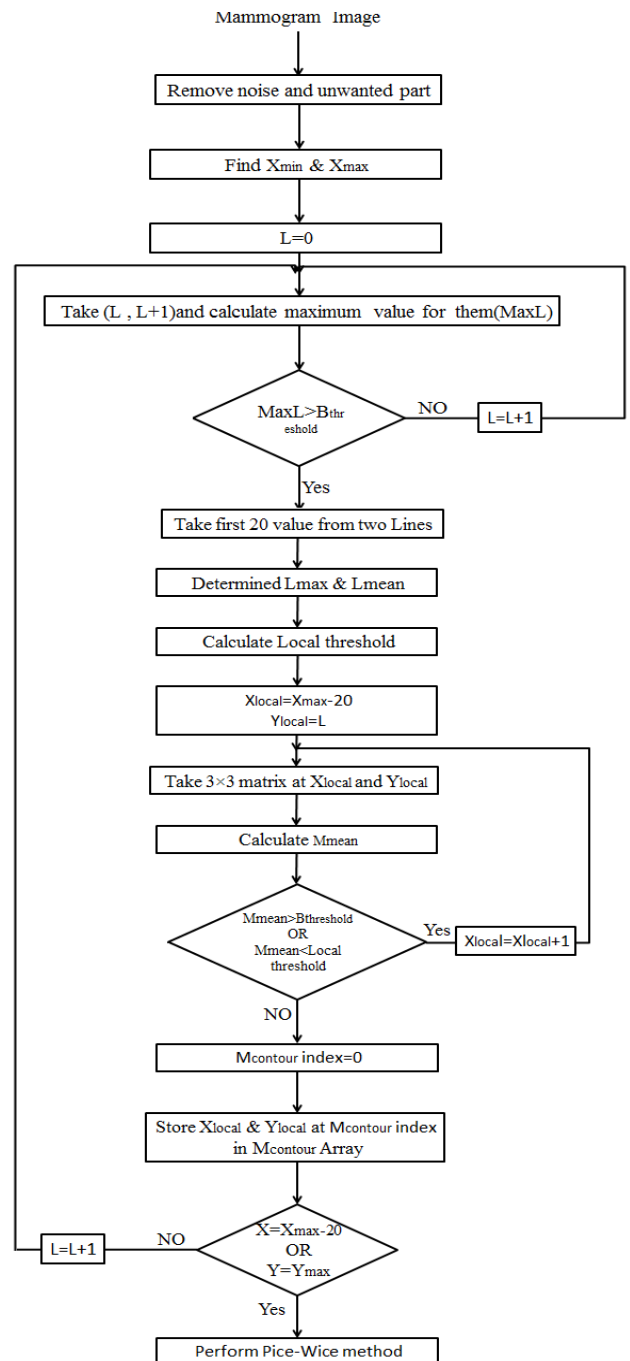
After removing noise and unwanted part from image,  $X_{min}$  and  $X_{max}$  are found from the breast area then we begin scanning the image from top to bottom and from right to left (from  $X_{max}$  to  $X_{min}$ ). we take two lines ( $L1, L2$ ) in order to find the maximum value and compare it with the background threshold value, if the value is less than background threshold value, we take the next two lines ( $L2, L3$ ) because the previous

two lines are part of background. If not, we take from both lines the first 20 value, calculate maximum value ( $L_{max}$ ) and the mean value ( $L_{mean}$ ) to be used in determining the local threshold that is calculated by the equation below:

$$Local\ threshold = 64 + \frac{(L_{mean} * 0.7 + L_{max} * 0.3)}{2}$$

After the first 20 value, we take 3x3 matrix from these lines and determine the mean for this matrix ( $M_{mean}$ ), we compare this value with Local threshold found above, if  $M_{mean}$  is greater than background threshold value and lower than Local threshold, we repeat this search until we found  $M_{mean}$  greater than local threshold or lower than background threshold value which represents the end of pectoral muscle in this line. We store X and Y of this point in array  $M_{contour}$  and repeat this method until we reach  $X_{max}$  or  $Y_{max}$ .

As shown in the diagram below:



Most points in the array  $M_{\text{contour}}$  will be in the edge of pectoral muscle and little of them will be inside the pectoral muscle or outside the muscle as shown in the figure below:

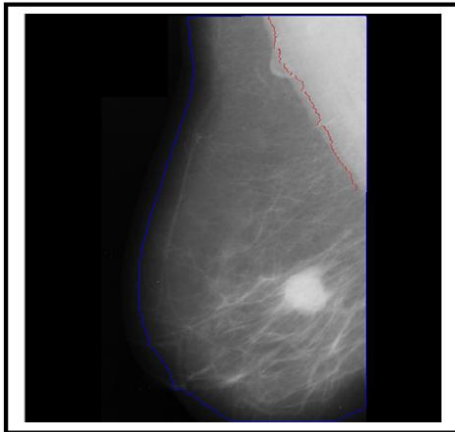
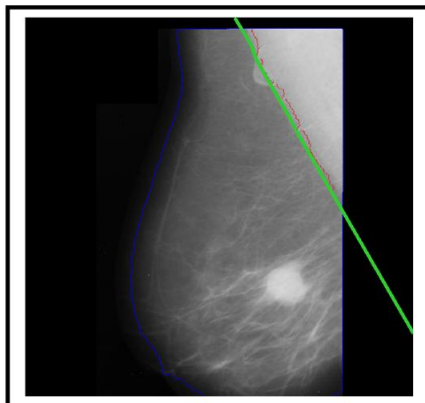


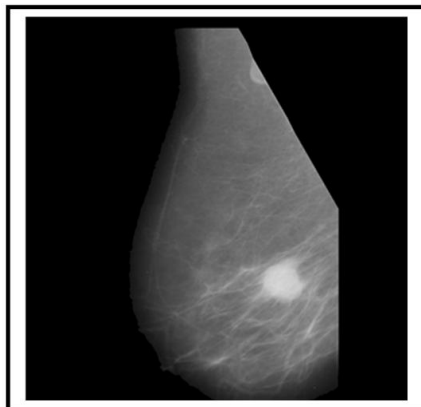
Fig.2: show points in array L

To enhance these points we use piecewise method which determines the average value for all points in the array  $M_{\text{contour}}$  representing the mid-point(mid1) between them. We store this point in array  $M_{\text{contourcount}}$ , then take the points from the beginning of the array to the (mid1), then we calculate and store average value for these points to find mid-point(mid2) between them and so on.

The cut line is the line that connects the mid-points stored in the array  $M_{\text{contourcount}}$ , the figure below show the cut line after perform piece-wise method .



(a)



(b)

Fig.3: (a) show cut line after perform piecewise method  
(b) show mammogram image after removing pectoral muscle

## 5. MAMMOGRAM ALIGNMENT

In the breast image, There are exist natural asymmetries, in the shape and structure between left and right breast in addition to distortion created through the mammographic recording procedure which makes the alignment very difficult.

In this paper the proposed method for alignment mammogram is based on translation and rotation only, ignoring any difference in the size and shape.

The nibble point( $N_p$ ), centroid point( $C_p$ ) and the average angle ( $\theta$ ) between the horizontal axis of  $C_p$ 's and the  $N_p$ 's are first found, and then those parameters are used to make the left and right images in the same slope.

$N_p$  had been used because it is the basic obvious point in the breast and can easily locate the position of this point;  $C_p$  is used because  $C_p$  is one of the basic points in the image that is less effected by the change even if there is noise in the image where noise has uniform distortion and this does not shift  $C_p$ .

To find  $N_p$ , the image is to be scanned from the far left pixel of the first row to the far right pixel of the last row, extract the points that represent the  $X_{\text{min}}$  and find all  $Y$  at  $X_{\text{min}}$ , increasing  $X_{\text{min}}$  by 1 ,when  $Y$  is in the first third of the image because the  $N_p$  cannot be in this part of the image. This operation is repeated until finding one or more points of  $Y$  in the last two thirds of the image, after that we consider  $X$  at this point of  $N_p$  and the average of  $Y$  points found in the last two thirds is  $Y$  at the  $N_p$ .

$C_p$  can be found using the following equations:

$$X_{\text{centroid}} = \frac{\sum_{x=1}^M \sum_{y=1}^N X * F(x, y)}{\sum_{x=1}^M \sum_{y=1}^N F(x, y)}$$

$$Y_{\text{centroid}} = \frac{\sum_{x=1}^M \sum_{y=1}^N Y * F(x, y)}{\sum_{x=1}^M \sum_{y=1}^N F(x, y)}$$

Where:

$F(x,y)$  is the image we used to find the  $C_p$  .

$M,N$  represent the image dimensions.

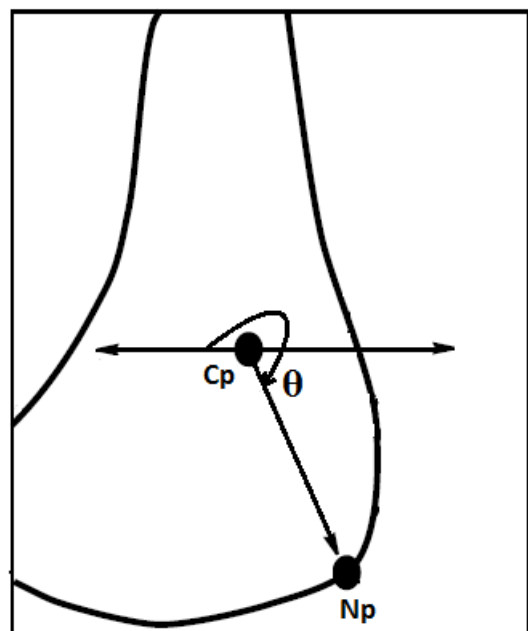


Fig.4: shows  $C_p$ ,  $N_p$  and  $\theta$

## 6. TRANSLATION AND ROTATION

After finding the  $N_p$  and  $C_p$ , the average angle between these two points for all images in the database( $\theta$ ) can be found by the following equation:

$$\theta = \frac{\sum_{i=0}^{i=K} \tan^{-1} (Np_y(i) - Cp_y(i)) / (Np_{yx}(i) - Cp_{xy}(i))}{K}$$

To achieve rotation without losing the details of the images we transmit each of the left and right images into square array(AL,AR) of  $\text{cil}\sqrt{((\max(W,H))^2 * 2)}$ , so apply CP point on the center of array. Then rotate the left and right images by ( $\theta$ ) that we found. As a result, we get two images in the same slope.

## 7. REJECTING UNWANTED REGION

Simple filter ( $3 \times 3$  mean filter) and thresholding are used to convert right and left array to binary images, making mask to left and right ( $M_L, M_R$ ) then apply AND operation to them as shown in following equation

$$Malign = M_L \text{ AND } M_R$$

To get matching among right and left images, we make AND function between  $M_{align}$  and the two images, finally we cut the left and right images according to the dimension of  $M_{align}$ . The mammogram alignment and rejecting unwanted region of the left and right image is shown in the fig. 5

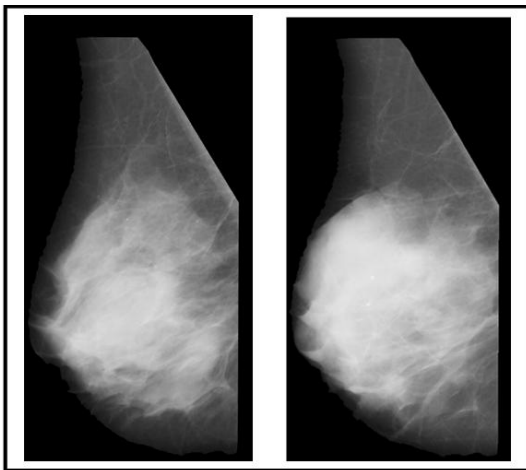


Fig.5: alignment of left and right mammogram

## 8. CONCLUSIONS

In this research, several automatic methods for preprocessing, which have been used to detect and remove pectoral muscle and mammogram alignment, have been proposed. The database used for testing these procedures is taken from a well-known and very referenced database, called MIAS. The noise and labels present in the mammogram images have been removed successfully with rate of 100%. Also all artifacts could be removed except the one that has one portion located inside the breast and the other inside the background and thus, the removal rate reaches (99.68%). The pectoral muscle extraction method gave less successful results due to the shape, the distortion introduced through the mammographic recording procedure and the variation in the intensity levels. The success in this step was (92.85%). With the mammogram alignment we obtained perfect results of (100%).

## 9. REFERENCES

- [1] Parkin DM et al. Global cancer statistics 2002, CA: A Cancer Journal for Clinicians, 2005,55:74–108.
- [2] Parkin DM, Fernandez LM. Use of statistics to assess the global burden of breast cancer. Breast,2006, 12(1 Suppl.):S70–S80.
- [3] J. Ferlay, F. Bray, P. Pisani and D.M. Parkin. GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide. Version 1.0. IARC CancerBase No. 5. Lyon, IARC Press, 2001.
- [4] Rangayyan, Rangaraj M., Fabio J. Ayres, and J. E. Leo Desautels. "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs." Journal of the Franklin Institute 344.3 (2007): 312-348.
- [5] Mirzaalian, Hengameh, et al. "Pre-processing Algorithms on Digital Mammograms." MVA. 2007.
- [6] M.L. Giger, R.M. Nishikawa, M. Kupinski, U. Bick, M. Zhang, R.A. Schmidt, D.E. Wolverton, C.E. Comstock, J. Papaioannou, S.A. Collins, A.M. Urbas, C.J. Vyborny, and K. Doi, "Computerized Detection of Breast Lesions in Digitized Mammograms and Results with a Clinically-Implemented Intelligent Workstation", presented at Computer Assisted Radiology and Surgery, Berlin, Germany, pp.325-330, 1997.
- [7] M. A. Wirth, Nonrigid Approach to Medical Image Registration: Matching Images of the Breast, Ph.D. Thesis, RMIT University, Melbourne, Australia, 2000.
- [8] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," Comput. Vis. Graph. Image Process., vol. 29, pp. 100-132, 1985.
- [9] F. Georgsson, "Differential analysis of bilateral mammograms", International Journal on Pattern Recognition and Artificial Intelligence, Vol. 17, July 2003, pp. 1207-1226.
- [10] T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," Comput. Biomed. Res., vol. 24, pp.273–295, 1991.
- [11] Bick, U., Giger, M.: Automated segmentation of digitized mammograms. In: Academic Radiology. Volume 2. (1995) 1-9.
- [12] Maitra, IndraKanta, Sanjay Nag, and Samir Kumar Bandyopadhyay. "Bilateral Breast Asymmetry Detection using Intensity Histogram and Landmark based Registration Technique." International Journal of Emerging Sciences 2.3 (2012).
- [13] J.Suckling et al, "The Mammographic Image Analysis Society Digital Mammogram Database", ExerptaMedica, 1994, Vol.1069, pp.375 – 378.
- [14] Indra Kanta Maitra et al, "Artifact Suppression and Homogenous Orientation of Digital Mammogram using Seeded Region Growing Algorithm", International Journal of Computer Information Systems, 2011, Vol. 3, No. 4, pp.32-38.
- [15] Bandyopadhyay, Samir Kumar. "Preprocessing of Mammogram Images." International Journal of Engineering Science and Technology 2.11 (2010): 6753-6758.