

# Preprocessing Technique for Mammographic Images

<sup>1</sup>Luqman Mahmood Mina, <sup>2</sup>Nor Ashidi Mat Isa

<sup>1,2</sup>School of Electrical and Electronic Engineering, Engineering Campus, University Sains, Malaysia,  
14300 Nibong Tebal, Penang, Malaysia

---

**Abstract:** Mammography is the best existing examination for the detection of early signs of breast cancer such as masses, calcifications, bilateral asymmetry and architectural distortion. In addition, mammograms are difficult to interpret, especially of cancer at their early stages. However, the presence of artifacts and noises can disturb the detection of breast cancer and reduce the rate of accuracy in the computer aided analysis (CAD). For this cause, pre-processing of mammogram images is very important in the process of breast cancer analysis because it could reduce the number of false positive. On the other hand, it could help radiologists to make a comparison between mammograms. The objective of this paper is to propose an algorithm of pre-processing on medio-lateral oblique-view (MLO) mammograms that involves two stages: the first is image enhancement of which the aim is to improve the interpretability or perception of information in images for human viewers, or to provide a better input for other automated image processing techniques. Eight techniques for the enhancement of overall mammograms inside the mini-MIAS database were implemented. The results were evaluated using Peak Signal to Noise Ratio (PSNR). The second stage is to extract the breast region from the rest of the image (background). For this purpose, a method based on automatic thresholding was used. Furthermore, the proposed pre-processing method was evaluated on 322 breast images collected from MIAS mini-database. The enhancement and extraction of the breast profile proved promising findings.

**Keywords:** Breast region extraction, Segmentation, Mammogram thresholding, Contrast stretching enhancement, Median filter and Computer aided design.

---

## 1. INTRODUCTION

Cancer occurs as a result of mutations, or abnormal changes, in the genes responsible for regulating the growth of cells and keeping them healthy. The genes are in each cell's nucleus, which acts as the "control room" of each cell. Normally, the cells in our bodies replace themselves through an orderly process of cell growth: healthy new cells take over as old ones die out. But over time, mutations can "turn on" certain genes and "turn off" others in a cell. The changed cell gains the ability to keep dividing without control or order, producing more and forming a tumor [1]. Breast screening programmes have been introduced in many countries, with the aim of detecting cancer at an early stage in asymptomatic women. X-ray mammography is the most common technique used by radiologists in the screening and diagnosis of breast cancer [2]. In mammography, low dose x-ray is used for imaging; the contrast of the image obtained with low dose X-ray machine is low. In the low contrast images, the minor difference between the normal tissue and the malignant disease is not discernable and makes the interpolation very difficult [3].

The use of computers in processing and analyzing biomedical images allows more accurate diagnose by a radiologist. Humans are susceptible to committing errors and their analysis is usually subjective and qualitative. Objective and quantitative analysis facilitated by the application of computers to biomedical image analysis leads to a more accurate diagnostic decision by the physician [4]. CAD, which integrates diagnostic imaging with computer science, image processing, pattern recognition, and artificial intelligence technologies, can be defined as a diagnosis that is made by a

radiologist who uses the output from computerized analysis of medical images as a “second opinion” in detecting lesions and making diagnostic decisions. In the past several years, CAD systems have drawn much attention from both research scientists and radiologists because of the associated challenging research topics and potential clinical applications [5].

Breast region extraction is a fundamental step in mammogram pre-processing that is important to find the skin-air interface, or the breast boundary. The aim of this step is to separate the breast from the rest of objects that could appear in a digital mammography; the black background, labels and tape artifacts. Enhancement also includes the pre-processing step. The aim of image enhancement is to improve the interpretability or perception of information in images for human viewers, or to provide a better input for other automated image processing techniques. The aim of this paper is to propose a method of pre-processing on mammograms that is composed of two stages: the first step is removing noise and enhancing images, while second step is suppression of background, labels and artifacts on the image. This step is referred as segmenting breast profile.

The remaining sections of this paper are organized in the following manner: Section 2 presents the proposed pre-processing algorithm. The methodology of the pre-processing enhancement and extraction of the breast profile are explained in detail in Section 3. Finally, Section 4 concludes this paper.

## 2. PROPOSED PRE-PROCESSING ALGORITHM

The pre-processing is based on an enhancement and segmentation steps of the mammogram. Image enhancement is the process of manipulating an image so that the result is more suitable than the origin for a specific application. The word **specific** is important because it establishes at the outset that enhancement techniques are problem oriented. In addition, removing noise is also important during the enhancement technique. On the other hand, breast segmentation is the second step of the pre-processing algorithm; segmentation step must be done to remove the background area (high intensity rectangular label, tape artifact and noise). Fig. 1 illustrates the flow chart of the proposed pre-processing to stages involved in this work.

## 3. PRE-PROCESSING METHODOLOGY

### 3.1- Image Data Set

The Mammographic Image Analysis Society (MIAS) database is the largest publicly available database of mammographic data. It contains approximately 322 screening mammography cases, where 207 images represent normal, while 64 and 51 images referred as benign and malignant cases respectively. This paper introduces an algorithm that used for the pre-processing of mammographic images, and it illustrates its performance using MIAS-mini database.

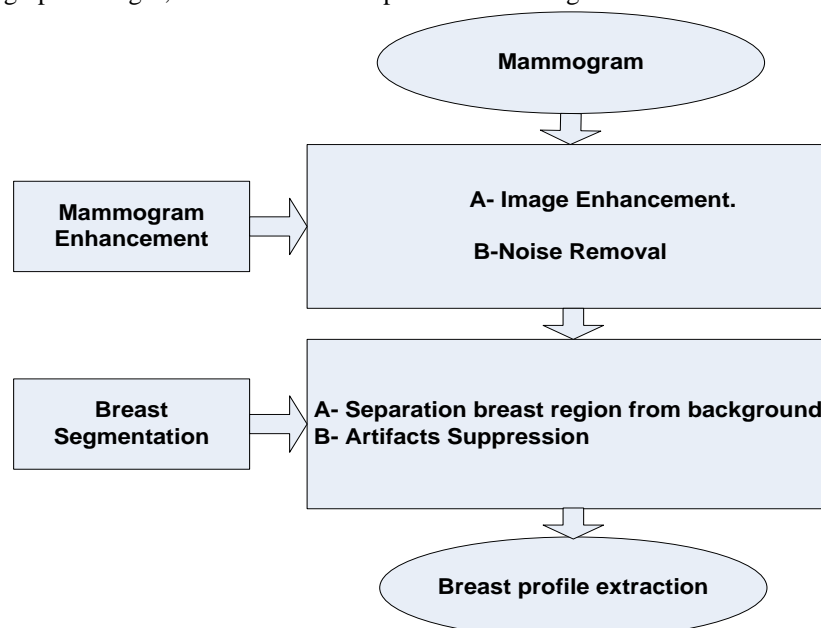


Fig. 1 Proposed framework for mammogram pre-processing

### 3.2- Image Enhancement

The aim of image enhancement is to improve the interpretability or perception of information in images for human viewers, or to provide a better input for other automated image processing techniques. Image enhancement transforms images to provide a better representation of the subtle detail. It is an indispensable tool for researchers in a wide variety of fields including medical imaging, art studies, forensics and atmospheric sciences. Image enhancement may be suitable for one problem and inadequate for another problem. For example, forensic images and videos employ techniques that resolve the problem of low resolution and motion while medical imaging benefits more from increased contrast and sharpness. In this paper we implemented contrast stretching enhancement.

Contrast stretching (often called normalization) is a simple image enhancement technique that attempts to improve the contrast in an image by 'stretching' the range of intensity values it contains to span a desired range of values, e.g. the full range of pixel values that the image type concerned allows. It differs from the more sophisticated histogram equalization in that it can only apply a linear scaling function to the image pixel values. As a result, the enhancement is less harsh. (Most implementations accept a gray level image as input and produce another gray level image as output.) Before the stretching can be performed, it is necessary to specify the upper and lower pixel value limits over which the image is to be normalized. Often these limits will just be the minimum and maximum pixel values that the image type concerned allows. For example, for 8-bit gray level images the lower and upper limits might be 0 and 255. The lower and the upper limits are represented by a and b respectively. The simplest sort of normalization then scans the image to find the lowest and highest pixel values currently present in the image. These values are represented by c and d respectively. Then each pixel P is scaled using the following function:

$$P_{out} = (P_{in} - c) \left( \frac{b-a}{d-c} \right) + a \quad (1)$$

Values below 0 are set to 0 and values above 255 are set to 255.

### 3.3- Digitization Noise Removal

Digitization noises such as straight lines present in the majority of acquired mammogram images are filtered using a two-dimensional (2D) Median Filtering approach in a 3-by-3 neighbourhood connection. Each pixel in the output contains medians value in the 3-by-3 neighbourhood around the corresponding pixel in the input images.

We implement eight enhancement techniques for the enhancement of overall mammogram inside the mini-MIAS database; the results were evaluated using PSNR as shown in table 1. The Peak Signal to Noise Ratio (PSNR) is measured in decibel (db), and can be calculated as shown in equation (2)[6]:

$$PSNR = 10 \log \left( \frac{255^2}{MSE} \right) \quad (2)$$

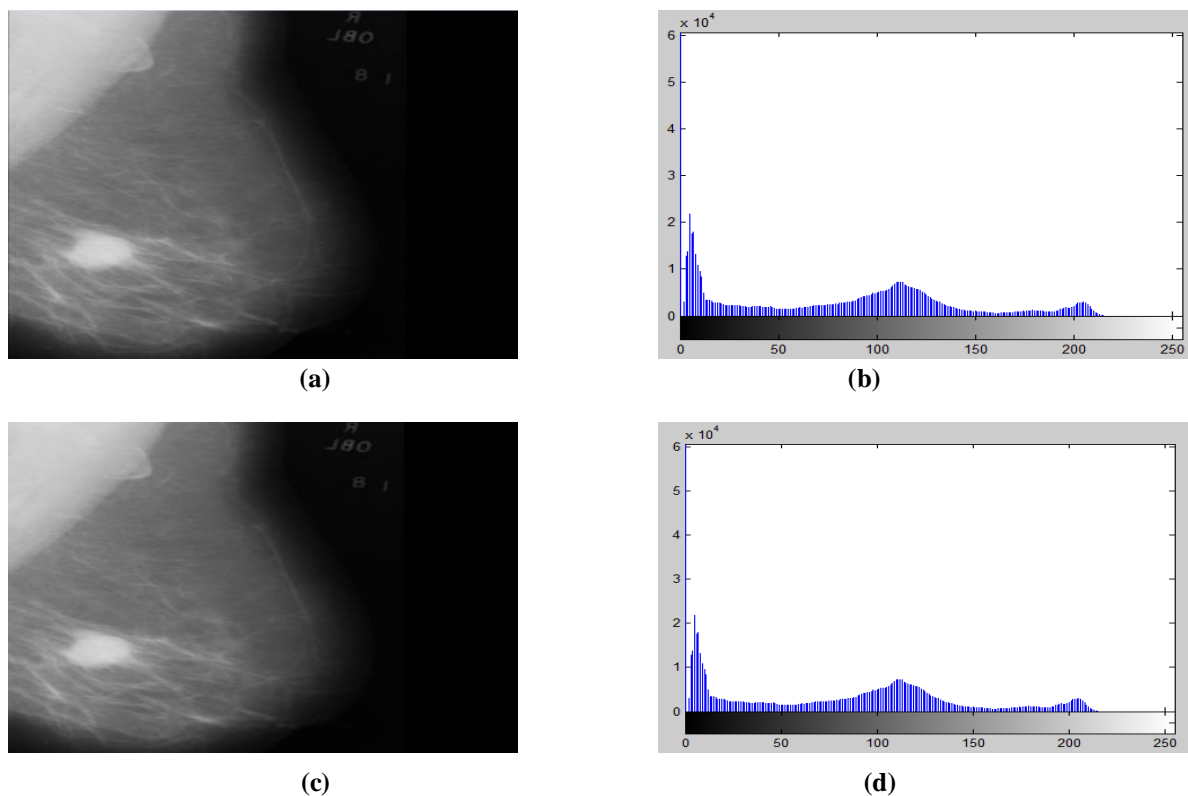
Where, the value 255 is the possible maximum value that can be attained by the image signal. Mean square error (MSE) is defined by a noise-free  $m \times n$  monochrome image I and its noisy approximation K as shown in equation (3). PSNR is most easily defined via the mean squared error (MSE). The higher the PSNR value is the better reconstructed image.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (3)$$

The experimental results figure out that the median filter yields maximum PSNR. In addition, it spends less time compared to other techniques. Because of these reasons, we apply the median filter for denoising and apply the contrast stretching technique for enhancement purposes. Fig.2 is a qualitative evaluation of denoising and enhancement.

**Table 1. Evaluation of Enhancement Techniques**

S.NO	Enhancement Technique	PSNR(dB)
1	Adaptive Histogram Equalization	23.32
2	Mean Filter	33.44
3	Contrast Stretching	23.63
4	Histogram Equalization	19.32
5	Logarithmic Image Enhancement	9.26
6	Median Filter	43.54
7	Classical Fuzzy Logic Enhancement	14.29
8	2-D adaptive noise-removal filtering	19.80



**Fig.2** (a) Original Image (b) Histogram of the Original Image, (c) Processed Image and (d) Histogram of the Processed Image.

### 3.4- Breast Segmentation

Image segmentation is the process of partitioning an image into multiple segments so as to change the representation of an image into something that is more meaningful and easier to analyze. In analyzing mammogram images it is essential to exclude background from breast profile. In this study, thresholding operation converts a multilevel image into a binary image i.e., it chooses a proper threshold  $T$ . The thresholding technique is employed to form a binary image by segmenting the mammogram image into two regions, namely a foreground represented by a white region and a background represented by a black region. Any pixel  $(x, y)$  is considered as a part of an object if its intensity is greater than or equal to the threshold value i.e.,  $f(x, y) \geq T$ . Otherwise the pixel belong to background [7, 8]. As per the selection of thresholding value, there are two types of thresholding methods that are available [9]; global and local thresholding. When  $T$  is constant, the approach is called global thresholding; otherwise, it is called local thresholding. Global thresholding methods can fail when the background illumination is uneven. In local thresholding, multiple thresholds are used to compensate for uneven illumination [10]. Threshold selection is typically done interactively however, it is possible to derive automatic threshold selection algorithms.

There are various approaches proposed to the task of segmenting the breast profile region in mammograms. Some of these have focused on using threshold [11][12], gradients [13], active contours[14].

#### 3.4.1. Background Separation

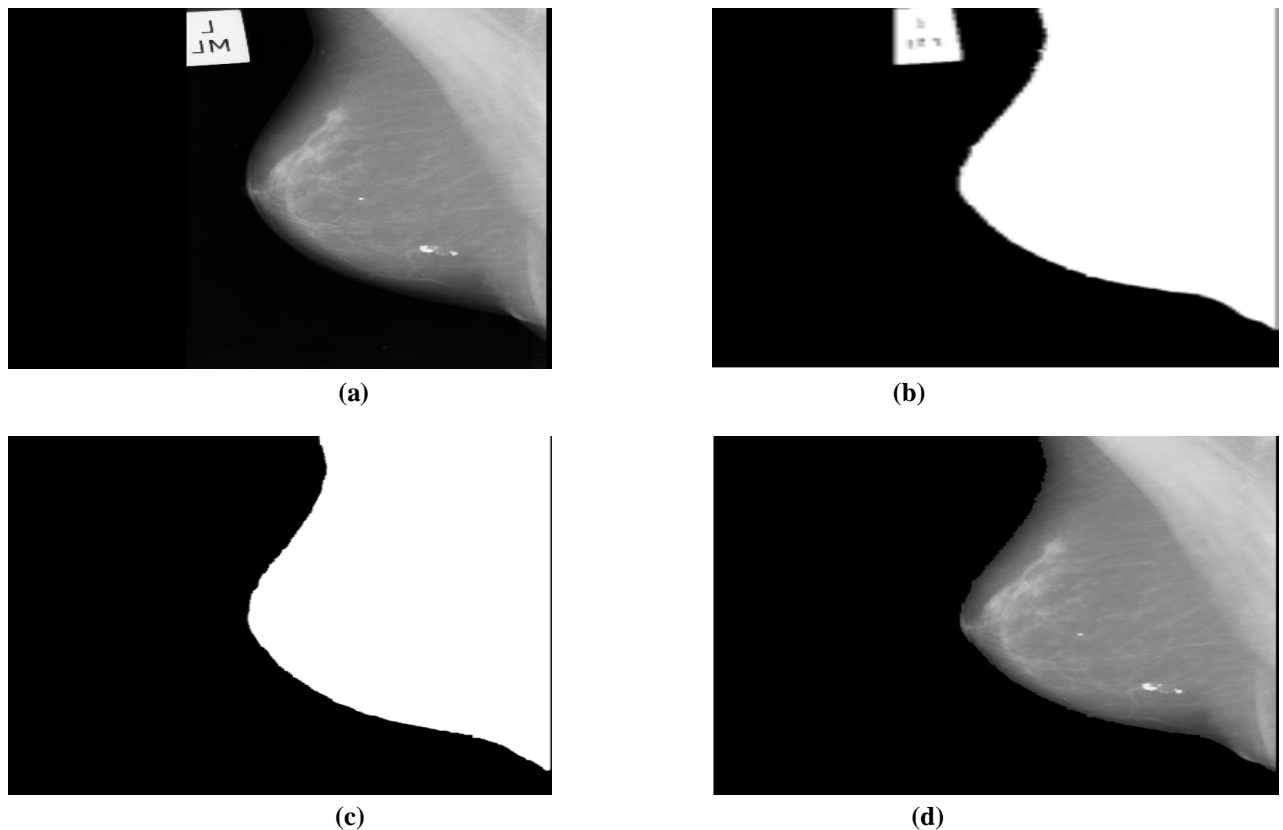
Radiopaque artifacts such as wedges and labels in the mammograms images are removed using threshold technique and morphological operations [15]. Fig. 3(a) shows a mammogram image with the existing radiopaque artifact. After converting the grayscale images into binary, Fig.3(b) will be formed.

#### 3.4.2. Breast Segmentation Algorithm Steps

The algorithm steps to find artifacts and labels, and to separate breast profile are as follows:

- Convert gray mammogram into binary using the threshold technique. The binary objects consist of the radiopaque artifacts and the breast profile region as indicated in Fig.3 (b).
- Binary image objects are labeled and number of pixels for each object in Fig.3 (b) is calculated.

- Select the largest object and remove other objects as shown in Fig. 3(c). This process morphologically opens binary images and removes all objects in the binary image, except the largest object (breast profile).
- Subtract the binary breast profile as shown in Fig.3(c) to the original mammogram image to form the final grayscale image in Fig.3(d).
- During artifact, wedge and label suppression, the breast profile region is also segmented from the background as shown in Fig.3 (c), and indicated in Fig.3 (d).



**Fig. 3.** Suppression of artifacts and labels from a mammogram. (a) Original image, (b) Threshold (c) Largest area (object) selected from threshold image 3(b), (d) Mammogram image with radiopaque artifacts suppressed.

#### 4. CONCLUSIONS

322 breast images were collected in the mini-MIAS open source database. We implemented the functional code for this prototype system using MATLAB 7.12 on a Pentium dual core, windows-based. The first step was to select an appropriate enhancement technique for enhancing the mammogram. The second step was apply two dimensional filter (Median Filter) to remove noise from mammogram images.

Finding an automated algorithm able of segmenting the breast region in mammograms has proven to be a difficult task. There are a number of factors which make it difficult to postulate the exact effect digital mammograms on a particular segmentation algorithm. The first of these relates to acquisition parameters, such as exposure time and energy level, which influence the quality of the image registered on film. Secondly, segmentation of the breast from the background is further hampered by the tapering nature of the breast.

From the results, mammography segmentation using the presented technique is efficient. Threshold and regions of segmentation algorithms work more efficient when images are preprocessed. There are two key contributions in this work. The first is valuation of enhancement technique. Eight techniques are used and by calculating PSNR was resolved which one has a better effect. The second is a fully automated segmentation algorithm which provides a breast contour representation of the breast profile region.

## REFERENCES

- [1] [http://www.breastcancer.org/symptoms/understand\\_bc/what\\_is\\_bc](http://www.breastcancer.org/symptoms/understand_bc/what_is_bc).
- [2] Amutha .S, D.R Ramesh Babu, Ravi Shankar .M, Harish Kumar. N” Mammographic Image Enhancement using Modified Mathematical Morphology and Bi-Orthogonal Wavelet”. In: IT in Medicine and Education (ITME), 2011 International Symposium on, 9-11 Dec. 2011. pp.548-553.
- [3] Arianna Mencattini, Marcello Salmeri “Mammographic Images Enhancement and Denoising for Breast Cancer Processing. IEEE transactions on instrumentation and measurement, vol. 57, no. 7, july 2008.
- [4] R.M. Rangayyan: Biomedical Image Analysis, CRC Press LLC, Boca Raton, Florida, USA, 2005.
- [5] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, “A review of computer-aided diagnosis of breast cancer: Toward the detection of early signs,” J. Franklin Inst., vol. 344, no. 3/4, pp. 312–348, 2007.
- [6] Lubin, J. 1995. A visual discrimination model for imaging system design and evaluation. David Sarnoff Research Center, February 1995.
- [7] K. G. Gunturk, “EE 7730- Image Analysis I”, Louisiana state university.
- [8] S. S. Varshney, N. Rajpal, R. Purwar, “Comparative Study of Image Segmentation Techniques and Object Matching using Segmentation”, Proceeding of International Conference on Methods and Models in Computer Science, pp. 1-6, 2009.
- [9] L.Aurdal, “Image Segmentation beyond thresholding”, Norsk Regnesentral, 2006.
- [10] Y. Zhang, H. Qu, Y. Wang, “Adaptive Image Segmentation Based on Fast Thresholding and Image Merging”, Artificial reality and Telexistence-Workshops, pp. 308-311, 1994.
- [11] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, and K. Doi, “Automated Segmentation of Digitized Mammograms”, Academic Radiology, vol. 2, no. 2, pp. 1–9, 1995.
- [12] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3 ed.: Pearson Prentice Hall, 2007.
- [13] A.J. Mendez, P.J. Tahoces, M.J. Lado, M. Souto, J.L. Correa, and J.J. Vidal, J.J, “Automatic Detection of Breast Border and Nipple in Digital Mammograms”, Computer Methods and Programs in Biomedicine, vol. 49, pp. 253–262, 1996.
- [14] M. A. Wirth, A. Stapinski, “Segmentation of the breast region in mammograms using active contours”, in Visual Communications and Image Processing, pp.1995–2006.
- [15] A. Sahakyan, "Segmentation of Mammography Images Enhanced by Histogram Equalization", Mathematical Problems of Computer Science 35,

## Author’s Biography:



**Luqman Mahmood Mina** was born in Dec, 02, 1976 in Sulaimanyah, Iraq. In 2002, he received the B.Eng degree in electrical and electronic engineering from University of Salahadden, Erbil, Iraq. He received his MSc degree in communication and signal processing from University of Sulaimani, Iraq in 2010. He is currently PhD student from University since Malaysia. His research interest includes image enhancement, pattern recognition, image segmentation, classification, and neural network.



**Dr. Nor Ashidi Mat Isa**, B. Eng (USM), PhD (USM) received the B. Eng in Electrical and Electronic Engineering with First Class Honors from Universiti Sains Malaysia (USM) in 1999. In 2003, he received his PhD degree in Electronic Engineering (majoring in Artificial Neural Network) from the same university. He is currently lecturing at the School of Electrical and Electronics Engineering, USM. His research interests include intelligent system, image processing, neural network, Biomedical Engineering (i.e. intelligent diagnostic system) and algorithms. As of now, he and his research team (Imaging and Intelligent System Research Team, ISRT) managed to publish their works national and internationally. Their contribution can be found in numerous international and national journals, chapters in books, international and national proceedings.