Preprocessing of Digital Mammogram Image Based on Otsu's Threshold

Ashgan M. Omer[a*], Mohammed Elfadil[b]

[a]Biomedical Engineering Department, College of Engineering Science, Sudan University (SUST),74 Khartoum, Sudan

[b]College of Medical Radiological Science, Sudan University (SUST), 407 Khartoum, Sudan

[a]Email: ashganomer@yahoo.com

[b]Email: mohamedelfadil@gmail.com

Abstract

Mammograms are difficult images to interpret. Hence, a preprocessing stage is very important to standardize the mammogram image along with the reduction of its size, and improve the quality of image in order to produce reliable image for CAD system. The proposed technique of preprocessing involves removal of unwanted parts from background of the mammogram, removal of pectoral muscle, and image enhancement. Binarization based on Otsu's threshold is a main process in all preprocessing steps. Multi-level thresholding applied to segment the pectoral muscle, and level three shows perfect results of pectoral muscle segmentation. A propose method applied on 160 images from MIAS database. Using of level-three multi-thresholding technique, the success rate was 96% in mammogram preprocessing stage.

Keywords: Mammogram; Binarization; Otsu's Threshold; Multi-level Threshold.

1. Introduction

Preprocessing is very important stage aid to limit the search for abnormalities on only breast region without influence from background. Preprocessing helps in reduction of mammogram size, and improves the quality of the images to make the feature extraction phase more reliable.

----

* Corresponding author.

220

A mammography images involves excessive parts, these parts are not important in processing of mammogram image. Some parts of images have some notes and labels which consist of information like name, date, etc., these labels are bright as gray level of some breast tissues which will influence the classification and will not give the desired results [1]. The presence of noises and artifacts can disturb the detection of breast cancer and reduce the rate of accuracy in analysis [2]. Removing the excessive parts, all irrelevant information, noises and artifacts are the aim of preprocessing. As the mammogram images are texture in nature, it is very important to get proper enhancement for these images. In low contrast images, the minor difference between normal and the malignant tissue is not discernable and make the interpretation very difficult [3]. Hence, preprocessing enlarge the intensity difference between objects and background to produce reliable image for processing step. In mammograms, the pectoral muscle has nearly homogeneous gray level values and exhibits as a high intensity region. Failure in segmenting the pectoral muscle may cause a higher number of false positives for breast cancer detection [4]. So a method represented to automatic pectoral muscle removal from mammogram images. Numerous papers introduced different approaches for preprocessing mammogram images. Researchers used various methods to segment breast region and eliminate the background and pectoral muscle. In [5] are introduced a preprocessing technique based on threshold to omitting excessive sides and put all images in one side, then eliminated labels and background based on region growing method. The result obtained 99% by applied on 60 images of MIAS database.

In [6], pectoral muscle removed by using watershed transformation and merging algorithm. The method applied on 84 mammograms from MIAS database, the results obtained are 0.85% and 4.88% for mean false positive and mean false negative rates respectively comparing with manually identified pectoral muscle.

Researchers in [7] applied three steps on MIAS database. First step is contrast enhancement then pectoral muscle detection and suppression, they used CLAHE technique and seeds region growing. Out of 322 mammogram images, fourteen images failed in output; the accuracy obtained was 95%.

In 2016, a study applied on all MIAS database for pectoral muscle segmentation. Pectoral muscle detected by using the morphological method and the random sample consensus (RANSAC) algorithm. The results showed 92.2% accuracy [8].

2. Materials and Methods

To achieve the preprocessing, the proposed method consists of three stages, unwanted regions removal, pectoral muscle removal, and image enhancement. However, the binarization consider as a main process in all preprocessing stages.

2.1. Image Database

The mammogram images used for this study are from the Mammographic Image Analysis Society (MIAS) [9]. The MIAS database includes 320 films only mediolateral oblique (MLO) views taken from the UK National Breast Screening Program and digitized originally to 50µm/pixel and a gray-scale resolution to 8 bits per pixel (bpp), with a Joyce-Loebl scanning microdensitometer SCANDIG-3. The original mammograms are 1024×1024

pixels. The mammogram images are investigated and labeled by expert radiologists.

2.2. Image Binarization

Binarization is a process, which converts the grayscale image to binary scale image. A binary images (I) is an image in which the pixels are only allowed two possible intensity values, nominally zero and one as in equation (1). Binarization of the image is very important for separating the foreground object from the background. The binarization methods can be categorized in different groups according to the threshold calculation method [10].

$$I : (x, y) \rightarrow \{0,1\} \qquad (1)$$

2.2.1. Thresholding

Thresholding is used to extract an object from its background by assigning an intensity value for each pixel, so that pixel is either classified as an object or a background [11]. Global thresholding use an appropriate threshold (T), as in (2).

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) > T \\ 0, & if\ f(x,y) \leq T \end{cases} \qquad (2)$$

Threshold-based algorithms are divided into single-level and multi-level thresholding categories.

2.2.1.1. Otsu's thresholding

Otsu's method [12] is the most successful global thresholding method. It automatically performs histogram shape based image thresholding for the reduction of a gray level image to a binary image. The algorithm based on image variance, Otsu's method chooses optimal threshold by maximizing the between class variance (3).

The optimal threshold value (t*), satisfies:

$$\sigma_B^2(t^*) = max_{0<t<L-1}\ \sigma_B^2(t) \qquad (3)$$

Where: σ_B between class variances, L intensity levels $[0, ..., L-1]$.

2.2.1.2. Multiple thresholding

 The multilevel threshold segments the pixels into several distinct groups in which the pixels of the same group have gray levels within a specific range (4). The Otsu's method can be applied for multiple thresholding segmentation [13], [14], the optimal thresholds t_1^* and t_2^* can be computed as equation (5).

$$g(x,y) = \begin{cases} a, & if\ f(x,y) > T_2 \\ b, & if\ T_1 < f(x,y) \leq T_2 \\ c, & if\ f(x,y) \leq T_1 \end{cases} \qquad (4)$$

$$\sigma_B^2(t_1^*, t_2^*) = \max_{0 < t_1 < t_2 < \text{L}-1} \sigma_B^2(t_1, t_2) \qquad (5)$$

2.2.2. Morphological Operation

Morphological operations are applied into binary image to affect the form, structure or shape of an object. A structuring element (S) is a shape mask used in the basic morphological operations [15]. The two principal morphological operations are dilation and erosion.

2.2.2.1. Dilation and Erosion

Dilation (6) expands the connected sets of a binary image; it can be used for expanding shapes and filling holes, gaps and gulfs. Erosion (7) shrinks the connected set of a binary image; it can be used to shrinking shapes, removing bridges, branches and small protrusion.

$$Dilation : I \oplus S \qquad (6)$$

$$Erosion : I \ominus S \qquad (7)$$

2.2.2.2. Closing and Opening

Closing and opening [16] are derived from the fundamental operations of erosion and dilation. Closing (8) consists of dilation followed by erosion; it used to fill in holes and small gaps. Opening (9) consists of erosion followed by dilation, it used to eliminate all pixels in regions that are too small to contain structuring element. In closing and opening the compound operation have same structuring element.

$$Closing: I \bullet S = (I \oplus S) \ominus S \qquad (8)$$

$$Opening: I \circ S = (I \ominus S) \oplus S \qquad (9)$$

2.2.3. Connected Components

A connected component [17] in an image is a maximal subset of the white pixels in the image such that for any pair there is a path between them that crosses only white pixels. Recursively, any white pixel is a connected component, and any white pixel adjacent to another white pixel is part of the latter's connected component (10), (11). Connected components labeling scans an image and group its pixels into components based on pixel connectivity.

$$I(x_0, y_0) = 1 \ \rightarrow (x_0, y_0) \in C \qquad (10)$$

$$(x_0, y_0) \in C \ and \ I(x_1, y_1) = 1 \ and \ |x_1 - x_0| + |y_1 - y_0| = 1 \rightarrow (x_1, y_1) \in C \qquad (11)$$

2.3. Image Preprocessing

Preprocessing stage can be described by the flowchart in Figure. 1. From the total number of images included in the MIAS database, a total of 160 images were randomly selected for testing and analysis of the proposed approach.
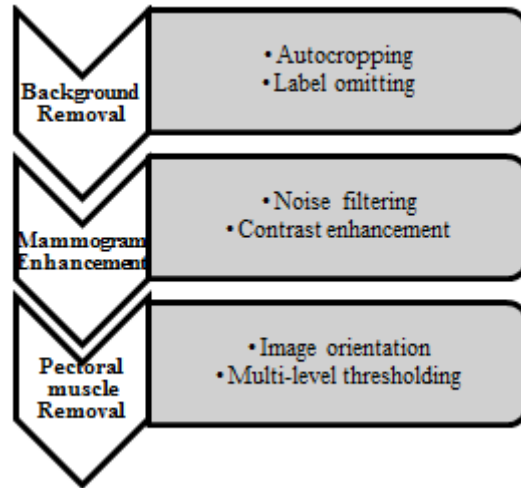

Figure 1: A flowchart for preprocessing process

2.3.1. Auto-Cropping & Labels Omitting

In general, MIAS images have blank spaces in the left and right edges. For omitting extra image parts, we get a double precision image from the original image and multiply them, then we get binary image. Binary opening is applied with a disk type structuring element, we sweep in the first row of image from right corner to left and left to right until get first white points (non-zero values). Then we cropped the image from right corner to first right point and cropped it from left corner to first left point, details as in Figure 2.
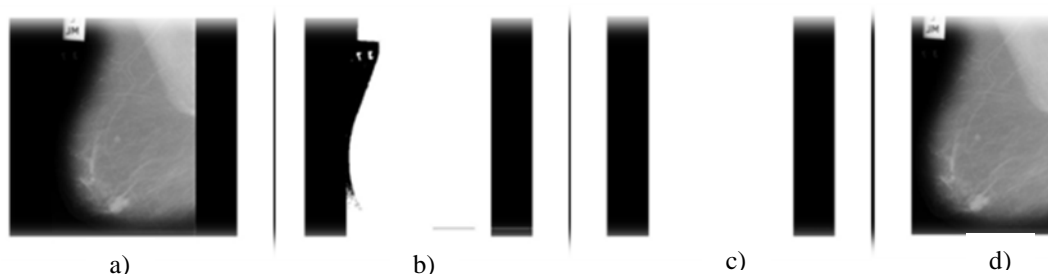

Figure 2: Auto-Cropping, (a) Original image, mdb005. (b) Double precise image. (c) Result of Multiply images on a) and b). (d) Cropped image

For removing all useless information, a previous cropped image converted to binary image. A morphological operation (opening) applied; a disk type structuring element is used here. Connected components are identified as white regions, the largest region which represents our region of interest (breast region) is considered as final mask image and we discard all other regions. See Figure 3.
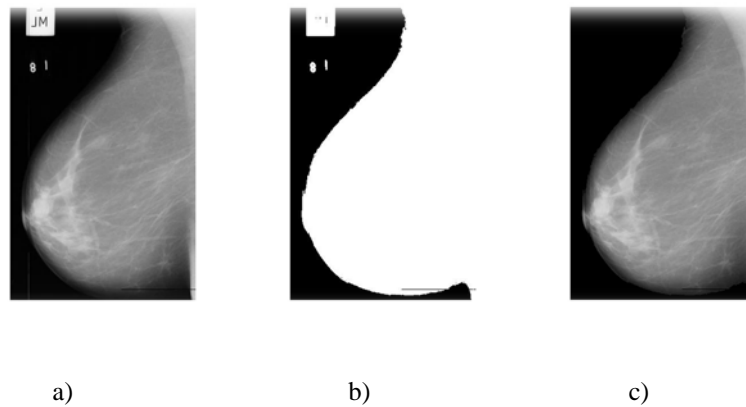
a)   b)   c)

Figure 3: Label Omitting: (a) Cropped image, mdb195. (b) Binary image. (c) Label free image

2.3.2. Mammogram Enhancement: Noise filtering and Contrast enhancement

To remove an impulse noises in mammogram images, a median filter [18] implemented assuming a window size (3×3). We used the contrast-limited adaptive histogram equalization (CLAHE) technique [19]. The CLAHE filter originally developed for medical imaging to reduce the noise and edge shadowing effect produced in homogeneous area. See Figure 4.
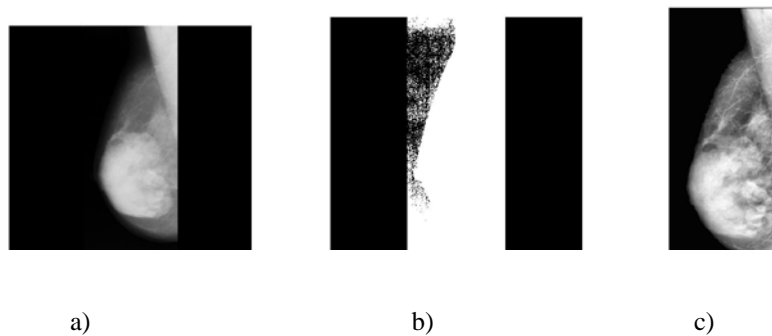

a)   b)   c)

Figure 4: (a) Original image 037. (b) Binary image with noises. (c) Noise filtered and contrast enhanced image

2.3.3. Pectoral muscle removal

In this work, we flip all right breast images (odd images number) to left side, so get unidirectional images and the pectoral muscles in upper left corner.

Mammogram image segmented by used multi-level threshold technique which based on Otsu method. Level (3) selected for segmentation. Using connected component labeling; the pectoral muscle identifies then removed according to its edge shape.

For non-straight line edge shape, pectoral muscle directly cropped as in Figure 5(1). For straight line edge shape, we applied a straight line refinement by starting from upper left margin point and search horizontally and

vertically to points where the pixel brightness change, then a triangle shape connected. Pectoral muscle lies within this triangle. Finally the demarked rectangle is cropped out from the original mammogram, Figure 5(2)
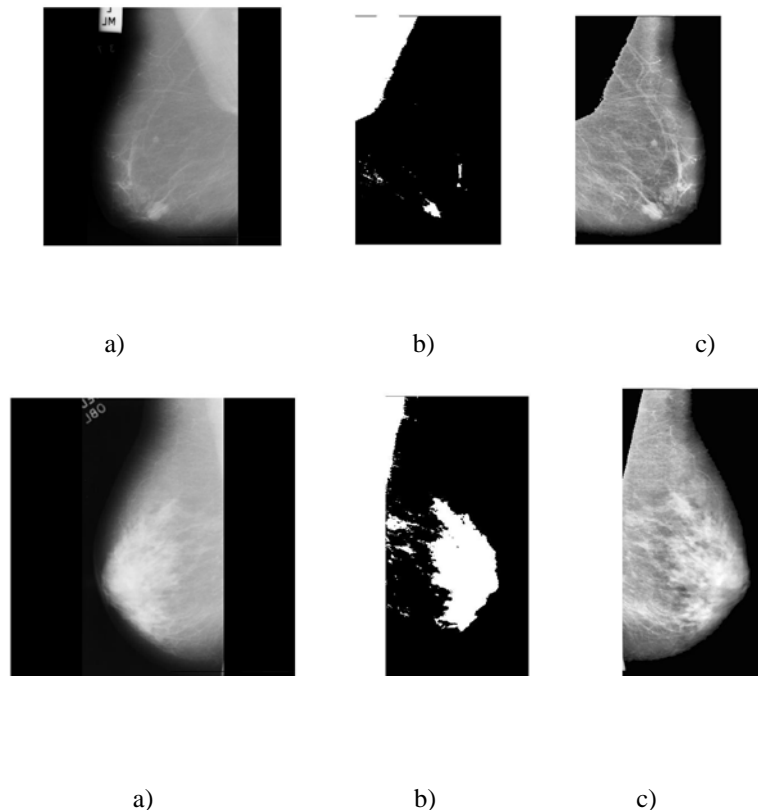

a)   b)   c)

Figure 5: **(1)** Automatically pectoral muscle removal: a) Original image 005. b) Level 3 threshold segmented image. c) Output image; **(2)** Straight line refinement: a) Original image 321. (b) Level 3 threshold segmented image. (c) Output image

3. Results and Discussion

The algorithm is implemented in Matlab. A set of samples of input images, and output images, are shown in Figure 6. In our approach, all steps of preprocessing based on binarization technique using Otsu's threshold. Even in multi-thresholding for pectoral muscle segmentation.

In general, muscles have high intensity than tissue, so pectoral muscle have maximum threshold, and we obtained the perfect result by using level -three in multi-thresholding for pectoral muscle segmentation. The accuracy of proposed approach was 96%. Pectoral muscle edges have different curvatures (convex, straight line, and concave). So we found the straight line technique efficient on straight line edges while directly cropped technique efficient on convex and concave edges.

The algorithm showed the ability to eliminate background, remove pectoral muscle and enhance the image contrast without losing any information from the image. The results show reducing of image size and consequently, will reduce the computational time of processing stage.
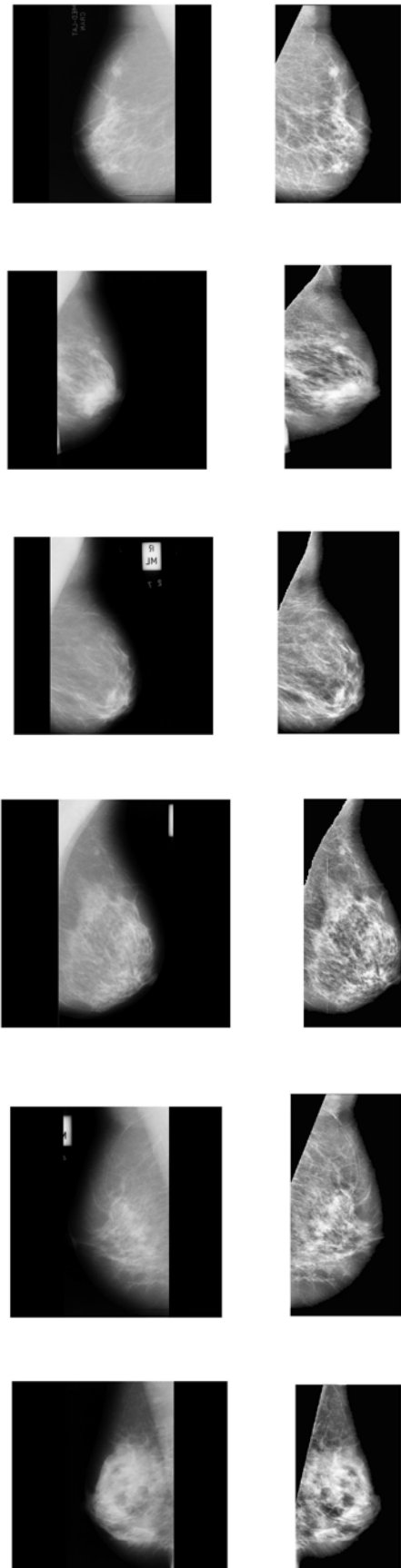
Figure 6: Preprocessing samples results, from top to bottom, mdb023, mdb072, mdb096, mdb100, mdb103, mdb225

4. Conclusion

The proposed technique for preprocessing mainly based on binarization process and Otsu's thresholding. The technique has been presented to remove unwanted regions with high intensities as labels, scanning artifacts and the pectoral muscle. Moreover, the proposed method enhanced the mammogram image by removed the noises and enhance the image contrast using CLAHE method which used widely in medical image applications. The results obtained on 160 images of MIAS database have shown excellent output. The resultant mammogram can be used further for processing stage to assist on mammogram image classification and breast cancer detection.

5. Recommendations

Further work may be conducted to reach an optimum level of Otsu's threshold and smoothing the pectoral muscle segmentation. Because of the simplicity of our technique, it has the potential for further development to adapt on the other mammography (MLO view) database.

References

[1] D. N. Ponraj, ME. Jenifer, P. Poongodi, and J. S. Manoharan , "A survey on the preprocessing techniques of mammogram for the detection of breast cancer", Journal of Emerging Trends in Computing and Information Science, vol. 2, no. 12, pp. 656–664, 2011

[2] A. C. Chaabani, A. Boujelben, A. Mahfoudhi, and M. Abid, "An Automatic Pre-processing Method for Mammographic Images", International Journal of Digital Content Technology and its Applications, vol.4, no.3, pp. 190-201, 2010

[3] M.Sundaram, K. Ramar, N. Arumugam, and G. Prabin, "Histogram Modified Local Contrast Enhancement for mammogram images", Applied Soft Computing, vol. 11, no. 8, pp.5809-5816, 2011

[4] Y. Li, H. Chen, Y. Yang, N. Yang, "Pectoral muscle segmentation in mammograms based on homogenous texture and intensity deviation", Pattern Recognition, vol. 46 , no. 3, pp. 681-691, 2013

[5] S. Dehghani and M. Abbasi, "A Method for Improve Preprocessing Images Mammography", International Journal of Information and Education Technology, vol.1, no.1, pp. 90-93, 2011

[6] K. S. Camilus, V. K. Govindan, and P. S. Sathidevi, "Pectoral muscle identification in mammograms", Journal of Applied Clinical Medical Physics, vol. 12, no. 3, pp. 215-230, 2011

[7] I. K. Maitra, S. Naj, S. K. Bandyopadhyay, "Technique for preprocessing of digital mammogram", Computer Methods and Programs in Medicine, ELSEVIER Journal, vol. 107, pp. 175-188, 2012

[8] WB. Yoon, JE. Oh, EY. Chae, HH. Kim, SY. Lee, KG. Kim, "Automatic detection of pectoral muscle region for computer aided diagnosis using MIAS mammograms", Biomed Research International, DOI:10.1155/2016/5967580, 2016

[9] MIAS database http://peipa.essex.ac.uk/info/mias.html

[10] N. Chaki, S. H. Shaikh, K. Saeed, "Exploring image binarization techniques", Studies in computational intelligence , Springer, DOI: 10.1007/978-322-1907-1_2, 2014

[11] C. A. Glasbey, "An analysis of histogram based thresholding algorithms", Graphical models and image processing, vol. 55, no. 6, pp. 532-537, 1993

[12] N. Otsu, "A threshold selection method from gray-level histogram", IEEE Trans. Syst. Man Cybern, vol. 9, no. 1, pp. 62-66, 1997

[13] Ch. H. Bindu, K. S. Prasad, "An efficient medical image segmentation using conventional OTSU method", International Journal of Advanced Science and Technology, vol. 38, pp. 67-74, 2012

[14] S. Arora, J. Acharya, A. Verma, P. K . Panigrahi, "Multilevel thresholding for image segmentation through a fast statistical recursive algorithm", Pattern Recognition Letters, vol. 29, no. 2, pp. 119-125, 2008

[15] K. Sreedhar and B. Panlal, "Enhancement of Image Using Morphological Transformation", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 4, No. 1, pp. 33-50, 2012

[16] R. Haralick and L. Shapiro, "Computer and robot vision", vol. 1, Addison-Wesley Publishing Company, 1992, pp. 174-185

[17] R. D. Yapa and H. Koichi, "A Connected Component Labeling Algorithm for Grayscale Images and Application of the Algorithm on Mammograms", ACM symposium on Applied Computing, pp. 146-152, 2007

[18] T. S. Huang, G. J. Yang, and G. Y. Tang, "A fast two dimensional median filtering", IEEE Transaction on Acoustics Speech, and Signal Processing, vol. 27, no. 1, pp. 13-18, 1979

[19] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization", Graphics GemsIV, P Heckbert Ed. Academic Press, 1994, pp. 474-485