

An Approach for Pre-processing Mammogram for Early Breast Cancer Detection

Fasna T A ¹, Dr. Rekha Lakshmanan ²

P.G. Student, Department of Computer Science and Engineering, KMEA Engineering College, Edathala, India¹

HOD, Department of Computer Science and Engineering, KMEA Engineering College, Edathala, India²

ABSTRACT: Mammography is defined as the process of using low-energy X-rays for examining the human breast, which is used as a tool for diagnostic and screening. Pre-processing of mammogram is necessary to reduce the presence of artefacts and noises which would otherwise adversely affect the detection of breast cancer and reduce the rate of accuracy in computer aided analysis. The objective of pre-processing operation is to improve the image quality in order to make it ready for further processing. The proposed approach of pre-processing is done in four stages. The noise present in the mammogram is filtered using the wiener filter in the first stage. In the second stage a histogram based threshold selection is done for segmentation. The third stage includes removal of small objects from binary image in order to remove the label. The image is then cropped in the final stage to get the actual region of interest. The proposed approach was tested on mini-MIAS database. The results suggest that the proposed method is well suited for pre-processing of mammography images.

KEYWORDS: Mammogram, Pre-processing, Threshold, Segmentation.

I. INTRODUCTION

Mammography is a special type of breast imaging modality that uses low-dose of x-rays in early breast cancer detection. Since the causes of the disease are still unknown, early detection is the only best way to improve breast cancer. Screening can be done through the techniques like digital mammogram, ultrasound, and magnetic resonance imaging (MRI). Digital mammogram has been proven to be the most effective method among others for detecting breast cancer [1]-[3]. Computer-aided detection (CAD) can act as a second reader where the final decision of the diagnosis can be made by the Radiologist.

Most mammograms are usually large in size with high resolution which may require specialized computing facilities for efficient processing. The main aim of the pre-processing step is to improve the quality of the image by removing or reducing the unrelated parts in the background of the mammographic images. A pre-processing step is necessary for the further analysis of mammograms in early breast cancer detection.

The types of noise observed in the mammogram image are marked in the Fig 1. Noise and artefact removal are the aim of preprocessing [4]. There are several methods for pre-processing mammography images because of its influences in cancer detection [5-6]. The common techniques in literature includes adaptive mean filter, median filter, mean filter, histogram equalization, histogram modified local contrast enhancement, CLAHE technique and morphological operations [7]. The proposed algorithm, removes all the noises from the mammogram image.

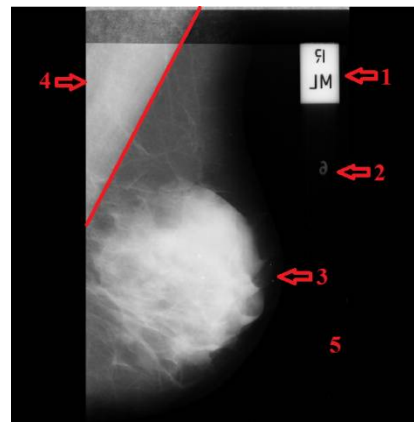


Fig 1 .Types of noises observed in original image and marked with numbers as 1. High Intensity,2. Low Intensity Tumor, 3.Breast Part, 4. Pectoral Muscle,and 5. Background.

Image segmentation is the process in which the digital Image is partitioned into multiple segments. The aim of segmentation is to simplify the representation of an image into many meaningful and easier to analyse segments. The extent to which segmentation is carried out depends on the problem to be solved. The main approaches used for the segmentation of intensity images of digital mammograms are namely, region- based methods, boundary-based methods threshold techniques, and hybrid techniques which combine boundary and region criteria.

The proposed method uses Wiener filter for removing the noise present in the mammogram. Image segmentation is done using the Thresholding technique. The threshold is calculated using a histogram based approach [5].All the small objects present in the binary image is removed and only the relevant breast portion is retained. 100 images of Mammographic Image Analysis Society (mini-MIAS) database were randomly selected for testing and analysis of the proposed approach.

II. RELATED WORK

Preprocessing step is done in medical image processing to improve the quality of image for further processing like feature extraction, segmentation etc. Poor image quality, noise, high and low intensity labels exist in the image which makes it difficult to process. Several approaches for preprocessing mammogram images are available. The commonly used techniques are adaptive mean filter, mean filter, median filter, histogram equalization, histogram modified local contrast enhancement, CLA HE technique ,breast region and pectoral muscle extraction and morphological.

In [8], P. K. Sahoo,et al describes the enhancement on digital mammogram using the CLAHE approach. Instead of considering the entire image, CLAHE operates on small regions in the image called tiles. Since the contrast of each tile is enhanced, the histogram of the output region looks like that of exponential, uniform or Rayleigh distribution. To eliminate unwanted boundaries, the neighbouring tiles are combined.

Farhan Akram, et al in [9] uses Active contour technique for preprocessing of mammographic images. The approach is explained in two stages. First is to remove high and low intensity labels, patient name, and scanning artifacts. An algorithm based on thresholding is used for this approach. In order to find the boundary of the muscle part in the second stage, active contour algorithm is applied. Finally the preprocessed image is obtained.

K. Santle Camilus, et al in [10] proposed a work to identify pectoral muscle in mammogram. Watershed transform was applied to the gradient of the ROI. The pectoral muscle boundary was correctly identified with a set of properties. There is a unique continuous watershed line starting from the top and ending at the left of the image .The width, gradually decreases and becomes zero as it reaches the left end. The watershed line of interest forms a curved shape triangular region containing the left top region of the image. The pectoral muscle is found to be oversegmented.

Morphological operations are an efficient preprocessing technique for mammograms. Dilatation, erosion, opening and closing are the basic morphological operations [11]. Erosion is used to fill small gaps or holes contained in mammography. A structuring element like disc is applied starting with dilation followed by erosion, opening, closing . It was actually defined for binary images, and later on extended to grayscale images. Morphological open operation followed by reconstruction operation removes the artefacts from the image.

In [12], Sara Dehghani, et al proposed a method to limit the processing to the breast region and omit excessive parts. Their work was described in three phases. The first phase is omitting the excessive image parts which are present in either sides of the image. The second phase includes the distinction of the breast direction and put all images in one direction. The third phase is the segmentation of the breast region from the background.

III. PREPROCESSING

The main goal of pre-processing is to improve the quality of image to make it ready for further processing by reducing or removing the unrelated parts in the background of the mammogram images. Pre-processing phase is necessary to improve image quality. The noise and other high frequency components are removed by filters and prepare the mammogram for further processing.

The main purpose of a breast cancer detection system is to aid the doctors and radiologist to take accurate decisions quickly. Providing the exact ROI will ease the identification of abnormality. In the proposed approach, the preprocessing is performed in 3 steps as shown in Fig 2.

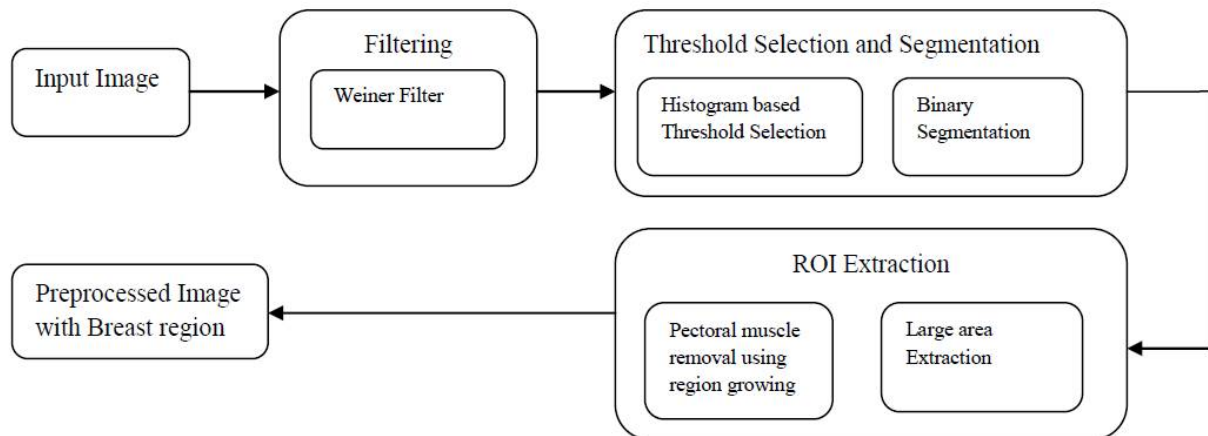


Fig 2. Block diagram of the proposed method

A. FILTERING

The filters are used to remove noise and high frequency components present in the mammogram. The proposed approach uses a wiener filter. The objective of a wiener filter is to minimize the mean square error. It is an optimum filter and has the capability of handling noise. A minimum mean square error constraint between estimate and original image produces an optimal estimate of the original image.

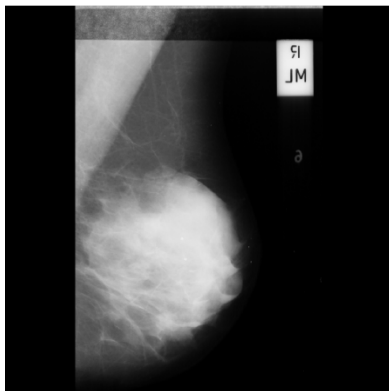
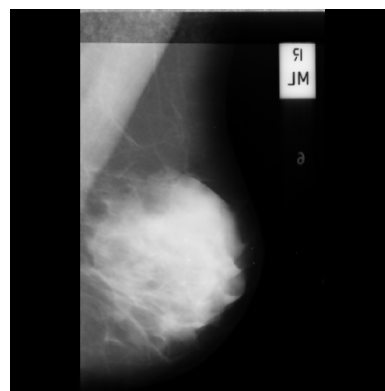


Fig 3. (a) Input Image



(b) Filtered Image

B. THRESHOLD SELECTION AND SEGMENTATION

Mammography image is segmented based on a threshold that is automatically computed from the histogram. The image is divided into different segments such as breast region, high and low intensity labels, pectoral muscle and background. Segmentation simplifies the image representation by highlighting the breast part. Binary thresholding is applied which creates binary images from a grayscale image.

The proposed approach uses a threshold selection method based on the slope difference distribution of the smoothed histogram for segmentation [8]. The normalized histogram distribution of the image is computed by arranging the grayscale values of the image in the interval [1,255]. This distribution is then filtered in frequency domain by Fast Fourier transform. The distribution is filtered back into spatial domain by removing the high frequency components. To compute the slope difference, two slopes for each point is computed by fitting a line model. Valleys and peaks are

obtained by setting the derivative of the slope difference at each point to zero. The selected threshold is the valley position of the slope difference distribution that has maximum absolute value.

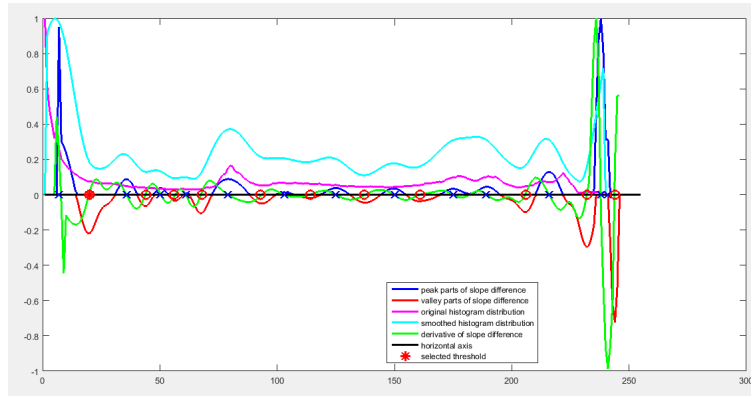


Fig 4. Threshold Selection

The image is segmented based on the threshold obtained from the histogram. The thresholding replaces each pixel in the image with a black pixel if the image intensity is less than selected threshold or a white pixel if the image intensity is greater than that constant.



Fig 5. Segmentation Result

C. ROI EXTRACTION

The small connected components in the image smaller than a specific size is being removed. Thus the large portion in the image is retained thereby removing the high and low intensity labels.

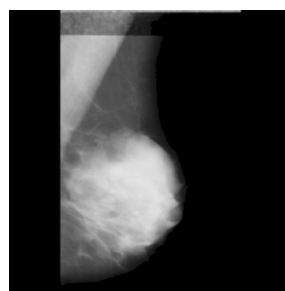


Fig 6. Label removed image

The pectoral muscle present in the above image is removed by region based active contours technique. The muscle is segmented out from the image to get the breast image.

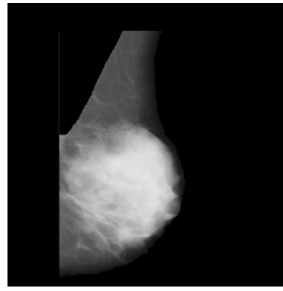


Fig 7.Pectoral muscle removed image

The entire image as such is not needed for examination. The background of the image doesn't contain any relevant data and thus can be removed. The image is undergone cropping to get the exact region of interest.

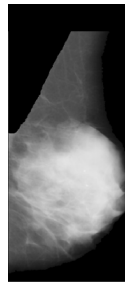


Fig 8.Region of interest extracted

IV. EXPERIMENTAL RESULTS

For testing and analysis of the proposed approach, images of Mammographic Image Analysis Society (mini-MIAS) database, was randomly selected. The database consists of 322 digitized mammograms of which 202 are normal and 120 abnormal images. It also includes markings by radiologists on the locations of abnormalities if present. Sample result is shown in Fig.8. First row shows three input images namely mdb007, mdb042, and mdb064. The preprocessed image by removing noise, labels and pectoral muscle is figured in the second row. The last row is the cropped image showing the exact ROI. The approach was tested by 100 sample images from mini-MIAS and proved 80% accuracy.

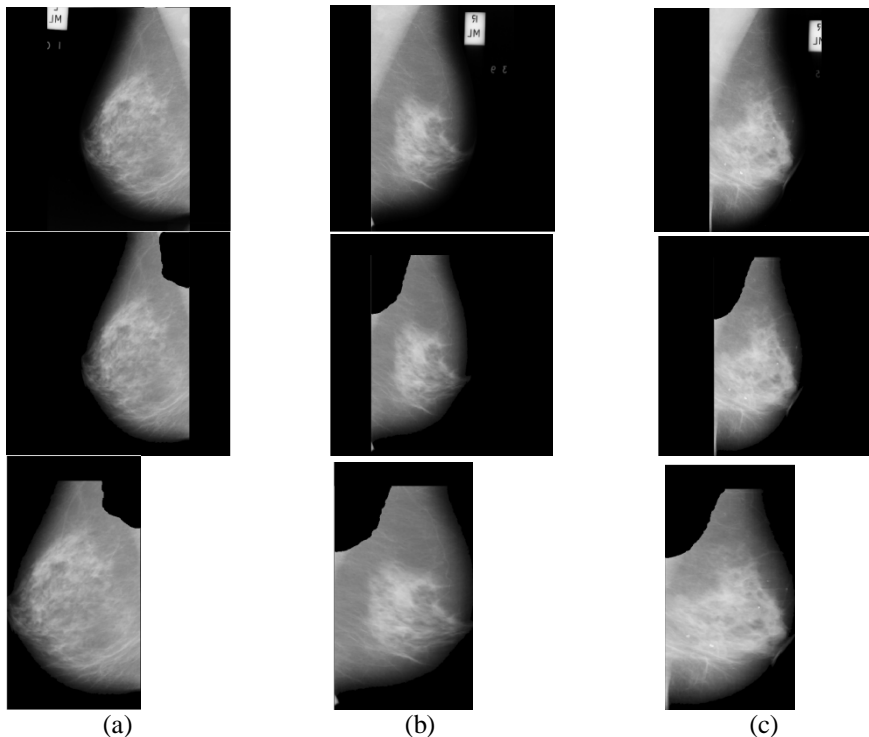


Fig 9. Analysis of the results by the proposed approach on three different mammograms

V. CONCLUSION

Pre-processing is an important for enhancing the content of medical images. Removing special markings and noise present in medical images will increase the quality of image segmentation. It will also improve the accuracy and efficiency of content based medical image classification and retrieval systems. In this paper, we have proposed a new histogram based approach for computing the threshold for segmentation. It is a robust and highly flexible automatic threshold computation approach. Small labels, special markings and pectoral muscles are then removed and the exact region of interest is considered for further processing. It is found that testing 100 sample images gives 80 correctly preprocessed.

REFERENCES

- [1] Michaelson J, Satija S, Moore R, et al. The pattern of breast cancer screening utilization and its consequences. *Cancer*. Jan 1 2002;94(1): 37-43.
- [2] Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist: Estimates of interobserver agreement and potential delay in cancer detection in the national screening study. *Invest Radiol* 1990, 25: 971-6.
- [3] Wallis MG, Walsh MT, Lee JR. A review of false negative mammography in a symptomatic population. *Clin Radiol* 1991, 44: 13-5.
- [4] Aziz Makandar , Bhagirathi Halalli ,Pre-processing of Mammography Image for Early Detection of Breast Cancer, *International Journal of Computer Applications* (0975 – 8887) Volume 144 – No.3, June 2016
- [5] Z.Z.Wang,"A new approach for segmentation and Quantification of cells or nanoparticles",*IEEE transactions on industrial informatics*,vol.12,no.3,june2016.
- [6] Ferrari RJ, Rangayyan RM, Desautels J E, Borges RA, Frere AF, Automatic identification of the pectoral muscle in mammograms, *IEEE Transactions on Medical Imaging* 2004;23(2), pp.232-245.
- [7] D.NarainPonraj, M.Evangelin Jenifer, P. Poongodi, J. Samuel Manoharan, A Survey of the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer, *Journal of Emerging Trends in Computing and Information Sciences*, VOL. 2, NO. 12, December 2011, pp.656-664.
- [8] P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMIIO,pp.167-192, 1988
- [9] Farhan Akram, Jeong Heon Kim, Inteck Whoang, and Kwang Nam Choi, " A Preprocessing Algorithm for the CAD System of Mammograms Using the Active Contour Method," *Applied Medical Informatics* Vol. 32, No.2, pp: 1-13,2013.
- [10] K. Santle Camilus, V. K. Govindan, P.S. Sathidevi, " Pectoral muscle identification in mammograms", *Journal of Applied Clinical Medical Physics*, Vol. 12, No. 3, pp. 215-230, 2011.
- [11] R.C., Gonzalez, R.E. Woods, *Digital image processing*, 3rd edition, pp. 104-289, 462-776, 2013.
- [12] Sara Dehghani and Mashallah Abbasi Dezfooli," A Method For Improve Preprocessing ImagesMammography", *International Journal of Information and Education Technology*, Vol. 1, No. 1, April 2011