# Preprocessing and Enhancement for Mammogram Images using Unified Approach

A.Srikarthik * R.Sivakumar **
*Dept of Computer Science Engineering,*
*Tamilnadu College of Engineering,*
*Coimbatore.*

***Abstract*** – Breast cancer is one of the foremost causes for the increase in mortality among women, especially in developed countries. Micro-classifications in breast tissue is one of the most incident signs considered by radiologist for an early identification of breast cancer is one of the most common forms of cancer between women. Mammography has been shown to be the most successful and reliable method for early signs of breast cancer such as masses, bilateral asymmetry architectural distortion and calcifications. In this paper the thresholding method is applied for the breast boundary identification and a new proposed modified tracking algorithm is introduced for pectoral muscle fortitude in Mammograms.

*Keywords* – Mammography, CAD, modified tracking algorithm

## I. INTRODUCTION

Computer technology has had a incredible impact on medical imaging. The analysis of medical images is still almost absolutely the work of humans. In the next decades, the use of computers in image analysis is expected to increase vastly. Breast cancer is a malignant tumor that starts in the cells of the breast. A malignant tumor is a collection of cancer cells that can produce into (overrun) neighboring tissues or spread to distant areas of the body. The disease occurs around wholly in women, but in some cases men also. Mammography is widely used as a major breast cancer screening method mass screening is generating large number of images.

A mammogram is a mild x-ray exam of the breasts to look for changes that are abnormal. The results are recorded on x-ray film or directly into a computer for a doctor called a radiologist to inspect. Advantages of Digital mammography greater than conventional film mammography is faster image gaining, shorter Exams, and easier image storage, easy transmission of images to other physicians and computer processing of breast images for more accurate detection of breast cancer [1, 2].

The proposed tracking algorithm aims to simplify the process of preprocessing and nipple detection in breast
Mammograms for CAD systems.

Mammograms used in this study are digitized images at 400x400 DPI and 2036x2266 pixels in 24 bit, Bit Map (BMP), format Collected from different sources including MIAS data base.

## II. PREPROCESSING AND ENHANCEMENT

### *A. Image Preprocessing:*

Image processing modifies pictures to get better results (enhancement, restoration), dig out information (analysis, recognition), and alter their structure (work, image editing). Images can be process by optical, pictorial and electronic mean, but image processing using digital computers is the most common method, because digital methods are very fast, flexible, and precise. An image can be synthesized from a micrograph of various cell organelles by assigning a light intensity value to each cell organelle. The sensor signal is "digitized" converted to an array of numerical values, each value representing the light strength of a small area of the cell.

Digitized values are called picture elements, or "pixels," and are stored in computer memory as a digital image. A size for a digital image is an array of 512/512 pixels each pixel has value in the range of 0 to 255. The digital image is process by a computer to achieve the preferred result [3, 4, 5].

### *B .Image Enhancement:*

Image enhancement improves the class (clarity) of images for human viewing. Eliminating blurring of images and noise, increasing contrast, and enlightening details are examples of augmentation operations. For example, an image might be in use of an endothelial cell and image might be of low contrast and blurred. Plummeting the noise and blurring and increasing the contrast range could enhance the image. The original image capacities have areas of very high and very low concentration mask details. An adaptive enhancement algorithm peals out these details. Adaptive algorithms change their function based on the image particulars (pixels) being processed. In this case the mean intensity sharpness contrast and (amount of blur removal) could be adjusted based on the pixel strength statistics in various areas of the image [6. 7].

Tumors have higher X-ray shrinking effect than normal soft tissues means higher intensity in mammogram image. Present CAD systems rely heavily on complicated technique in machine learning to address the area of pattern recognition and categorization which has high computational load leading to longer time required for the analysis of a single case.

*C. Database (Image Acquisition)*:

To access the real medical images for carrying out the tests is a very difficult due to privacy issues and heavy technical hurdles. X-ray film mammogram is converted into digital mammograms. Laser scanners are used to digitize conventional film mammograms by measuring the Optical Density (OD) of small windowed regions of film and converting them to pixels with a grey level intensity. The size of the window determined the spatial resolution of the digitized image. The declaration is usually expressed in units of microns per pixel indicating the size of the square region of film that each pixel in the digitized image. Each pixel location on the file is accumulated with a beam of known intensity (photon flux density).

The exact pixel value is based on the range of optical densities that the scanner is adept of measuring and the number of bits used to store the grey level of each pixel. The accuracy of computer detection schemes on digital mammograms will depend partly on the spatial resolution and range of grey levels at which the images are digitized.

The Mammography Image Analysis Society (MIAS), which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography record (ftp://peipa.essex.ac.uk).
The data used in these experiments was taken from the MIAS. The X-ray films in the database have been carefully selected from the United Kingdom National

Breast Screening Program and digitized with a Joyce-Lobel scanning microdensitometer. The database contains left and right breast images for 161 patients are used. Its quantity consists of 322 images fit in to three types such as usual, benevolent and malignant. There are 208 normal, 63 benign and 51 malignant (abnormal) images. Figure 2.1 shows the input image.

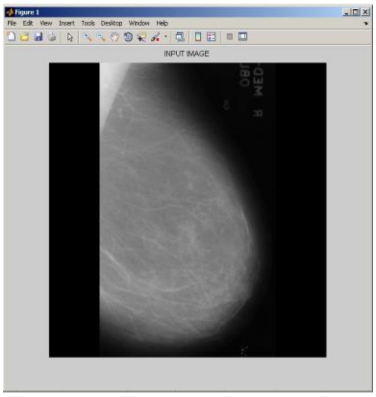


Figure 2.1 the input image.

*D. Thresholding Algorithm-Label Removal:*

First step of the algorithm is to identify breast boarder. In many mammograms back ground objects with high intensity values make breast boundaries identification a challenging task, especially for the scanned ones where the original film has some artifacts [8, 9]. The following algorithm is used to exclude background objects (noise) and identify breast boundaries.
-Using the histogram curve, the threshold point is obtained for the input image (10% from the bottom of the image).

- Translate the given input image into binary image using the threshold value from the histogram curve.

- Label is extracted and the binary image is obtained.

- Since breast object is the biggest object, the object with maximum number of pixels is recognized to be the breast.

Source mammogram is masked with this image and the resulting image is the denoised using median filtering process. The proposed algorithm aims to reduce the complexity for abnormality detection by identifying breast object and exclusive of pectoral muscle using modified tracking algorithm for the mammogram images.
Figure 2.2 shows the binary image and label removed image.

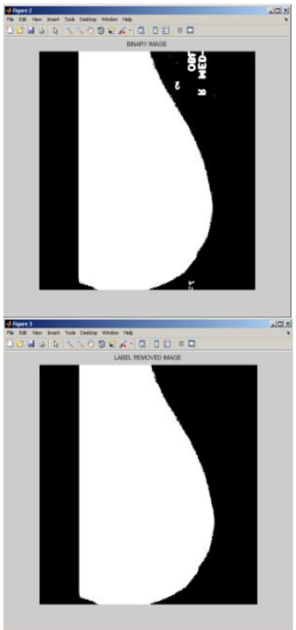


Figure 2.2 Binary image and label removed image.

*E. Median Filter*:

The median filter compares each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is part of its surroundings [10]. Instead of simply replacing the pixel value with the mean of neighboring pixel values. It replaces with the median of individuals values. The median is calculated by first sorting all the pixel values from the adjoining neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value. (If the locality under anxiety contain an even number of pixels the average of the two middle pixel values is used).

| | | | | |
|---|---|---|---|---|
| 23 | 22 | 20 | 26 | 24 |
| 25 | 22 | 16 | 33 | 32 |
| 28 | 27 | 34 | 26 | 15 |
| 18 | 17 | 29 | 13 | 35 |
| 19 | 21 | 31 | 30 | 11 |

*Table 2.1* Calculating the median value of a pixel Neighborhood

Neighborhood Values: 13, 16, 17,22,26,27,29,33,34
Median Value: 26

As can be seen, the central pixel value of 34 is rather misleading of the surrounding pixels and is replaced with the median value: 26. A 3×3 square neighborhood is used here larger neighborhoods will produce more severe smoothing.

By calculating the median value of a neighborhood the median filter has two main advantages. First, the median is a more robust average than the mean and so a single very deceiving pixel in a neighborhood will not affect the median value significantly.Second, the median value must actually be the value of one of the pixels in the neighborhood; the median filter does not create new improbable pixel values when the filter straddles an edge. For this reason the median pass through a filter is much improved at preserving sharp edges.

*F. Normalization:*

Normalization is a process that changes the range of pixel intensity values. An application includes photographs with poor distinction due to glare, for example. Normalization is as well as called contrast stretching. More general field of data processing such as digital signal processing are referred to as dynamic collection expansion. The purpose of dynamic range expansion in the various applications is usually to convey the image or other category of signal into a range that is more familiar or normal to the senses. Hence the term normalization [11].

Normalization is a linear process. If the intensity series of the image is 50 to 180 and the preferred range is 0 to 255 processes entails subtracting 50 from each of pixel intensity making the collection 0 to 130. Then each pixel intensity is multiply by 255 by 130, making the range 0 to 255. Auto-normalization in image processing software typically normalizes to the full dynamic range of the number system specified in the image file format. The normalization process will construct iris regions have the same constant dimensions. So that the 2 photographs of the similar iris under different conditions will have characteristic features at the same spatial location.
Figure 2.3 shows the normalized Image.

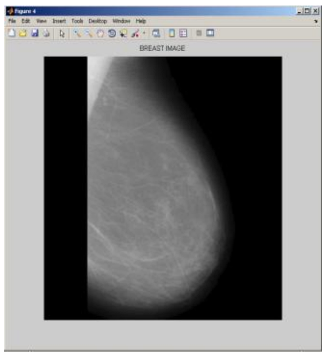


Figure2.3: Normalized Image.

## III. PROPOSED ALGORITHM

*A. Tracking Algorithm- Pectoral Muscle Removal:*

The pectoral muscle represents a major density region in most medio-lateral oblique (MLO) view of mammograms and can shape the outcome of image processing methods. Intensity-based methods, can present poor performance when applied to differentiate dense structures such as the fibro-glandular disc or small doubtful masses, since the pectoral muscle appears as more or less the same density as the dense tissues of interest in the image. To increase the consistency of boundary matching, the pectoral muscle can be removed from the breast region.

Another important need to identify the pectoral muscle lies in the possibility that the local information of its edge, along with an internal analysis of its region, may be used to identify the attendance of abnormal axillaries lymph nodes, which may be the only manifestation of occult breast carcinoma.

A modified tracking algorithm technique is applied to separate the pectoral muscle region [12]. The global optimum in the histogram is select as the threshold value (about 90% of the histogram curve value).This is used to alter the input gray image to the binary image. Then by using this binary image as masking image original is extracted. Pixel value of first line read from left corner and point where pixel difference is greater than 50 intensity values is noted.

This process is continued till in which line there is no such abrupt variation. Noted points are connected and pictorial muscle is selected. Thus pectoral muscle is extracted and median filtering is applied.

IV PROPOSED ALGORITHM FLOWCHART

Input gray scale mammogram

↓

Convert to binary image using the thresholding algorithm

↓

Modified Tracking Algorithm

↓

Tracking is done for the lines for abrupt changes in intensity values

↓

Reference value is obtained and Pectoral muscle region is identified

↓

Subtracting the pectoral muscle region From original image

↓

Pectoral muscle extracted image is obtained

**Algorithm: Removal of Pectoral Sways Section**

S → Read the mammogram image

[m,n] → Size of the image

th → thresholding algorithm used to convert The gray image to binary image

**Procedure:**

- Set threshold point form the histogram curve of input image (10% from the bottom of the image).
- Translate the specified contribution image into binary image using the threshold value from the histogram curve.
- Label is removed and the binary image is obtained.
- Since breast object is the largest object, then the object with maximum number of pixels is recognized to be the breast.

**Typical Procedure for the Removal of pectoral muscle region using Modified Tracking Algorithm:**

- The Brest region is mapped to original image.
- The first line image information is obtained and point at which intensity value is abruptly changed is noted.(i.e difference between consecutive pixel values are greater than 50 )
- This process is continued until line which has no abrupt changes.
- Obtain the location location of these connected components.
- Pectoral muscle will be detected by the connected components.
- The rest of the scrap detected using the 90% threshold values will be eliminated.
- By subtracting the pectoral muscle region from the original image.
- The required region for segmentation is obtained and filtration process is done using the median filter.

Figure 4 shows the threshold image and Removal of Pectoral Muscle Region in the mammogram image using modified tracking algorithm.

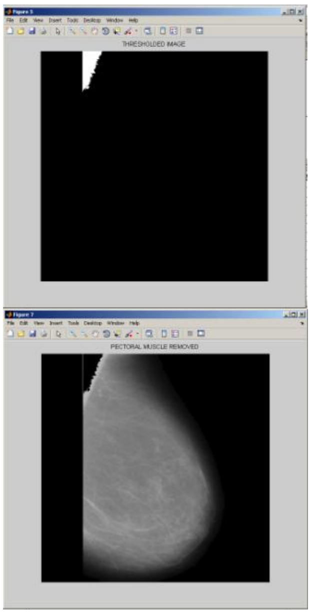


Figure 4. Threshold image and Pectoral Muscle Region removed from the mammogram image.

## V. CONCLUSION

In this paper a new mammogram pre-processing algorithm based on statistical features is introduced to detect breast boarder and the pectoral muscle segmentation. The thresholding algorithm is used to identify the breast boarder and the modified tracking algorithm is used to identify image orientation and then to remove the pectoral muscles. This outcome will be used to further analysis of breast tissue more accurately than other algorithms.

## VI.REFERENCES

[1] Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, K. Doi, Automated segmentation of digitized mammograms, Acad. Radiol. 2 (1995) 1-9.[2]

[2] R. Chandrasekhar, Y. Attikiouzel, Automatic breast bordersegmentation by background modeling and subtraction, in: M.J. Yaffe (Ed.), Proceedings of the 5th International Workshop on Digital Mammography, Medical Physics Publishing, Toronto, Canada, 2000, pp. 560-565.

[3] L.P. Clarke, M. Kallergi, W. Qian, H.D. Li, R.A. Clark, M.L. Silbiger, Tree-structured non-linear filter and wavelet transform for microcalcification segmentation in digital mammography, Cancer Lett. 7 (1994) 173-181.

[4] S. Detounis, Computer aided detection and second reading utility and implementation in a high volume breast clinic, Appl. Radiol. 3 (9) (2004) 8-15.

[5] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, A.F Frere, Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets, IEEE Trans. Med. Imag. 20 (9) (2001) 953-964.

[6] M. Karnan, K. Thangavel,"Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications"Computer Methods and Programs in bio medicine, Elsevier Ltd 8 7 ( 2 0 0 7 ) 12-20.

[7] Lobregt, S. & Viergever, M. A. (1995),"A discrete dynamic contour model", *IEEE Trans. Med. Imaging* 14(1), 12–24.

[8] A.J. Mendez, P.G. Tahocesb, M.J. Lado, M. Souto, J.L. Correa, J.J. Vidal, Automatic detection of breast border and nipple in digital mammograms, Comput. Meth. Prog. Biomed. 49 (1996)253-262.

[9] Griffiths & Tenenbaum,"From algorithmic to subjective randomness, In advances in neural information processing", 2004.

[10] H Sheshadri, A Kandaswamy,"Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms", *Medical Imaging and Information Sciences* vol. 23 2006.

[11] K. Thangavel, M. Karnan, R. Siva Kumar, A. Kaja Mohideen, Automatic detection of microcalcification in mammograms—a review, Int. J. Graph. Vision Image Process. 5 (5) (2005)31-61.

[12] K. Thangavel, M. Karnan, CAD system for preprocessing and enhancement of digital mammograms, Int. J. Graph. Vision Image Process. 9 (9) (2006) 69-74.