EXTRACTION OF BOUNDARY & PECTORAL MUSCLE SUPPRESSION FOR EARLY DETECTION OF CANCER

Researcher:

Muhammad Ali
Reg No. 337-FET/MSEE/F13

Supervisor:

Prof. Dr. Ihsan ul Haq
DEE, FET

Department of Electronics Engineering,

Faculty of Engineering and Technology,

INTERNATIONAL ISLAMIC UNIVERSITY,

ISLAMABAD

Medical screening
Cancer - Diagnosis

EXTRACTION OF BOUNDARY & PECTORAL MUSCLE SUPPRESSION FOR EARLY DETECTION OF CANCER

Muhammad Ali
Reg No. 337-FET/MSEE/F13

Submitted in partial fulfillment of the requirements for the
Master of Philosophy in Electronics Engineering,
with Specialization in Signal Processing,
at the Faculty of Engineering and Technology,
International Islamic University,
Islamabad

Supervisor
Dr. Ihsan ul Haq

Month, Year
May, 2016

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

CERTIFICATE OF APPROVAL

Title of Thesis Extraction of Boundary & Pectoral Muscle Suppression for Early Detection of Cancer

Name of Student Muhammad Ali

Registration No. 337-FET/MSEE/F13

Accepted by the Faculty of Engineering and Technology, INTERNATIONAL ISLAMIC UNIVERSITY, ISLAMABAD, in partial fulfillments of the requirements for the Master of Philosophy Degree in Electronics Engineering with specialization in Signal Processing.

Viva Voce Committee

Dr. Ihsan ul Haq
Supervisor ----------------------------

Prof. Dr. Abdul Jalil
Internal Examiner ----------------------------

Dr. Abdul Basit
External Examiner ----------------------------

Prof.Dr. Muhammad Amir
ChairmanDEE: ----------------------------

Prof. Dr. Naveed Aqdas Malik
Dean FET: ----------------------------

DECLARATION

I, Mr. Muhammad Ali, Reg. No. **337-FET/MSEE/F13** student of MS Electronics Engineering in Session 2013-2016, hereby declare that the matter printed in the thesis titled **"EXTRACTION OF BOUNDARY & PECTORAL MUSCLE SUPPRESSION FOR EARLY DETECTION OF CANCER"** is my own work and has not been printed, published and submitted as research work, thesis or publication in any form in any University, Research Institution etc. in Pakistan or abroad.

M. Ali

Muhammad Ali

Dated:____________

ABSTRACT

Worldwide the most common and frequent diagnosed cancer in women is breast cancer. To save the life of cancer patient, the appropriate way is to detect it at early stages when either cancer cells are confined in small area or cancerous cells are few in numbers. Significance of early detection can be understood from the fact that 90% of all types of cancers including breast cancer are curable if it is detected before stage 2. For early detection of cancer, screening tests are recommended on regular basis for women who do not have any visible symptoms. Mammogram screening is one of such test used worldwide for detection of cancer in breasts.

Segmentation of the breast from the image is one of the preliminary and important step in early cancer detection in mammogram processing. An algorithm is proposed in this thesis for breast segmentation. The Hough transform is proposed for image alignment followed by the guided filter for contrast enhancement. Finally connected object labelling has been introduced for improved breast segmentation. Furthermore, pectoral muscle is suppressed by a new proposed technique based on geometry and contrast variance between the breast tissue and pectoral muscle.

Entire Mini-MIAS Database is used to evaluate the proposed algorithm on Mammogram Images. The results are compared with radiologist drawn boundary i.e. Gold standards and other latest techniques used for breast segmentation. The results of proposed algorithm are very promising which are not only accurate but also take comparatively less execution time.

ACKNOWLEDGEMENT

I am really thankful to Almighty Allah only because of His countless blessing made this possible. I sincerely thanks to my IIUI course Supervisor Dr. Ihsan ul Haq and my senior Dr. Mohsin Riaz whose technical advice, critical feedback on the thesis topic selection, proposal preparation and guidance throughout the thesis work really make it a success. It's been an honor to work with these two great personalities which are sea of knowledge. I am also thankful to my university colleagues, friends who shared their valuable opinion for improvement. Let me not forget the role of my study partner Ali Rauf Rao who keeps me motivated throughout the MS phase. Last but not the least, gratitude to my great Family, my father, my mother and my dear wife for the support, sacrifice and motivation throughout the phase of thesis along with the tough job.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

ACS:	American Cancer Society
CAD:	Computer Aided Design
CC View:	Cranio-Caudal View
CLAHE:	Contrast Limited Adoptive Histogram Equalization
DDSM:	Digital Database for Screening Mammography
DICOM:	Digital Imaging and Communications in Medicine
EPIC:	European Prospective Investigation into Cancer and Nutrition
GB:	Giga Bytes
GHz:	Giga Hertz
IARC:	International Agency for Research on Cancer
IEEE:	Institute of Electrical and Electronics Engineers
LCC:	Left Cranio-Caudal
LMLO:	Left Medio Lateral Oblique
MATLAB:	Matrix Laboratory (Math works, Inc.)
MIAS:	Mammographic Image Analysis Society
MLO:	Medio Lateral Oblique View
MRI:	Magnetic Resonance Imaging
PGM:	Portable Grey Map
RAM:	Random Access Memory
RCC:	Right Cranio-Caudal
RIMLO:	Right Medio Lateral Oblique
ROI:	Region of Interest
WHO:	World Health Organization

Chapter 1 Introduction

1.1 Cancer

Abnormal or uncontrollable growth of a cell is called cancer. Cells are the basic building block of all living being. Normal cells grow and die after producing new cells which replace the old or damaged cells and the cycle continues till the life time of living being. But sometimes it happen that cells keep on producing new cells but the damaged old cells won't die due to any reason or disease. These damaged cells keep on producing more new damaged cells and thus ultimately forms a mass of tissues which is called tumor or lump. These Cancerous cells or malignant cells can develop in almost any organ or tissue. Recent worldwide statistics, released by IARC (International Agency for Research on Cancer) warns that new cancer cases increases to 0.141 Billion per year and is continuously increasing worldwide with around 24 different types of cancer found in around 184 countries of all seven continents of the world. Same report also caution that if same trends is continued, by 2025, an estimated number of cancer patients will rise to 0.193 Billion new cases per year across the globe [1]. Data reveals that more than half of the deaths due to cancer were reported in less developed countries. From 24 cancer types, most frequently diagnosed cancers were those of the lung which is 13 percent (1.8 million), breast cancer share is 11.9 percent (1.7 million), and colorectal cancer 9.7 percent (1.4 million) reported cases per year. But when we limit these cancer cases to women gender Breast Cancer is most frequent diagnosed cancer.

1.2 Breast Cancer

Common and most frequent diagnosed cancer in women worldwide is Breast cancer. In year 2012 alone, 1.67 million women were diagnosed with cancer in breasts and there were 6.32 million alive cancer women patients who had been diagnosed with same in last 5 years. Moreover, Breast cancer alone was the cause of about 0.52 Million women deaths in 2012 worldwide [1]. The developing countries across the Global have major share of these deaths where the majority of cancer cases are diagnosed in late or incurable stages due to issues like poverty, limited government funds for health care, lack of professionals in health care sector. Unfortunately, Pakistan has the highest incidence of breast cancer in South East Asia. The last 18 years statistics(1994 to 2014) released by country largest cancer treatment hospital, Shaukat Khanum Memorial Cancer Hospital & Research Centre, reveals the fact that Breast cancer is still the highest malignancy reported with 45.4% of cancer in women over 18 years of age in Pakistan [3].

1.3 Significance of Research and Motivation

A Cancer mass can take years to develop by a reproduction of a single cancer cell. By the time a cancerous mass is perceived, it's estimated that 100 million to 1 billion cancer cells must be present, dividing for five years or even more [4]. To save the life of cancer patient, the appropriate way is to detect it at an early stages when either cancer cells are confined in small area or cancerous cells are few in numbers. Ninety percent of all types of cancers including breast cancer are curable if detected at stage 1 [5]. Stage 1 of the cancer is the stage when tumor size is two cm or less and have spread to only tissue itself or at most neighboring tissue [6].

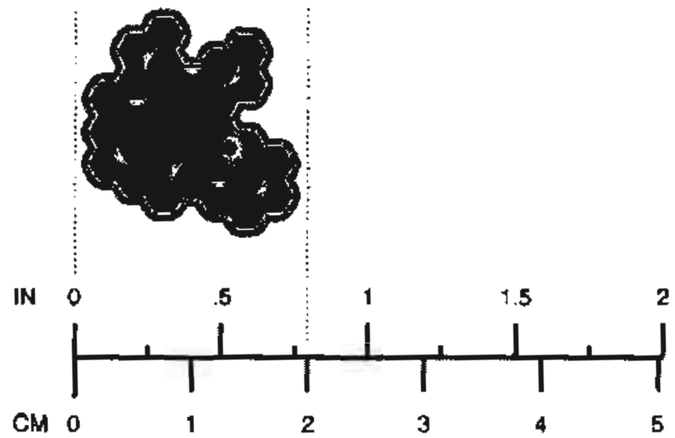


Figure 1: Size of Tumor Stage 1

For early detection of cancer regular screening tests are recommended for women who do not have any visible symptoms. The goal is to detect cancers before visible symptoms to tackle it early. The Table.1 shows the survival rate of patients with respect to the stage at which cancer is diagnosed [7].

Table 1: 5-Year Relative Cancer Patient Survival Rate

Stage	Relative Survival Rate (5 Years or More)
0	100 %
1	100 %
2	93 %
3	72 %
4	22 %

1.4 Significance of Digital Processing of Mammograms

Studies have shown that 10 to 20% cancers (False Negative results) are not diagnosed by Radiologists [8]. With the aid of digital Image processing, it has become possible that this rate can be significantly reduced by giving radiologists a better base to take their decision and reduce human margin of error.

1.5 Screening Methods for Cancer Detection

The two screening methods commonly used by radiologists are

1. Magnetic Resonance Imaging (MRI)
2. X-Ray or Mammography.

1.5.1 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging is one of the medical imaging technique or a test that make use of magnetic waves to create images of organs inside the body. MRI test is highly sensitive i.e. MRI detects cancer better than X-ray or mammograms, but on the other hand it has a higher False Positive (FP) rate (non- cancer cells can be confused with cancer). This can become the cause of unwanted biopsies and other confirmation tests in many of these misdiagnosed cases, causing worry, anxiety and resources waste [9].

Figure 2: MRI Scan of Breast

Another major reason due to MRI is not frequently used in developing countries is its cost [2]. Not only MRI equipment is expensive but also per test cost is higher and hence it is less preferred test in low income developing countries such as Pakistan.

Cancers found during initial screening process or early stages are expected to be of smaller size and is confined in a limited area. The tumor size and its spread are key factors in predicting the cancer stage and the level of threat the person may have.

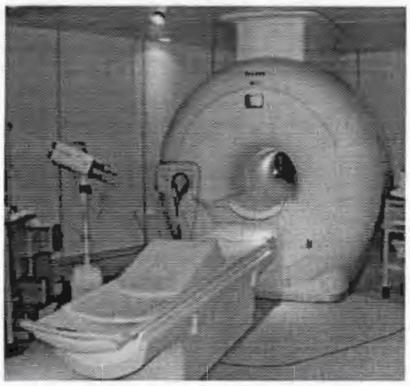


Figure 3: MRI Machine

1.5.2 Mammograms

A mammogram is a distinct type of medical imaging of the breasts that uses low dose x-rays to reveal tumors before they become large enough for women to feel through physical examination [10].

Figure 4: Mammogram Machine

The Gold Standard for Breast Cancer Detection fixed by American Cancer Society (ACS) for women age 40 and above is that they should have mammogram test once every year [11]. Detecting breast cancer early can increase survival rates and hence this test is recommended in order to fight cancer timely and minimize threat to human life [12].

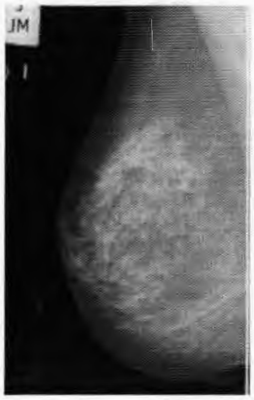


Figure 5: Mammogram of Left Breast

1.6 Digital Processing of Mammogram

Inter and intra observer errors are a common practice in medical images analyses and mammograms are no exception. The very same mammogram report is differently interpreted by different radiologists due to different interpretation.

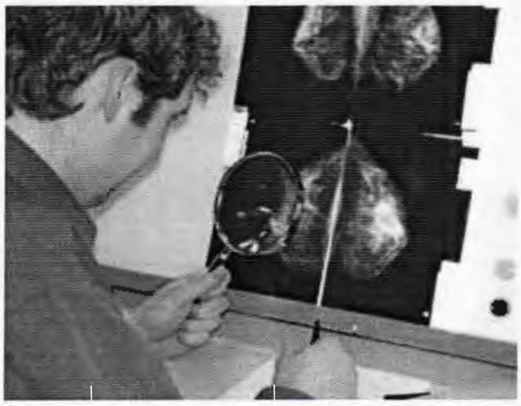


Figure 6: Manual Examination by Radiologist

Data shows that around 10 to 30% women who have cancer which are not picked by this mammogram test (False Negative). This may be due to fatigue of radiologist eye, poor image quality, lack of expertise and experience of radiologist which in turn increases the reliance on computer technology and CAD systems [6]. The mammographic screening sensitivity also varies with quality of the image. With limited number of expert radiologist and huge number of women patients, this task is not easy and demands computer aid to analyze such large population mammograms. Moreover computer results can also be used as a 2^{nd} opinion to confirm radiologist opinion and help them to detect abnormalities in breast especially masses and micro-calcification clusters more efficiently at an early stage.

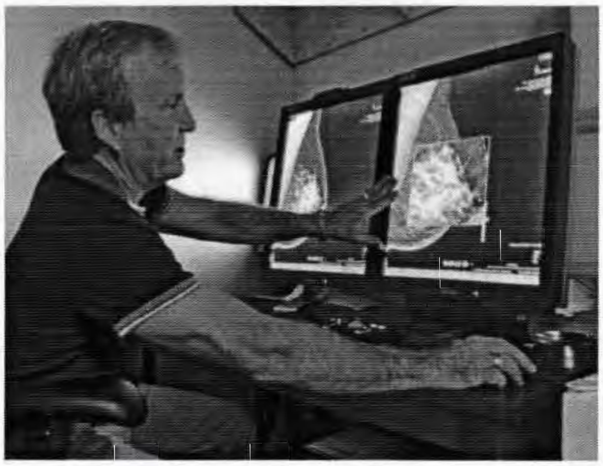


Figure 7: Digital Analysis of Mammogram

1.7 Standard Views for Mammogram

There are two standard types of view used for Mammograms.

1. Craniocaudal (CC) View.
2. Medio Lateral Oblique (MLO) View.

1.7.1 Cranio-Caudal (CC) View

It is the view taken from above. Cranial end means head and caudal end means tail. X ray beam enter from above and exits from the bottom of breast. Basically, CC View is Top down View of a breast.

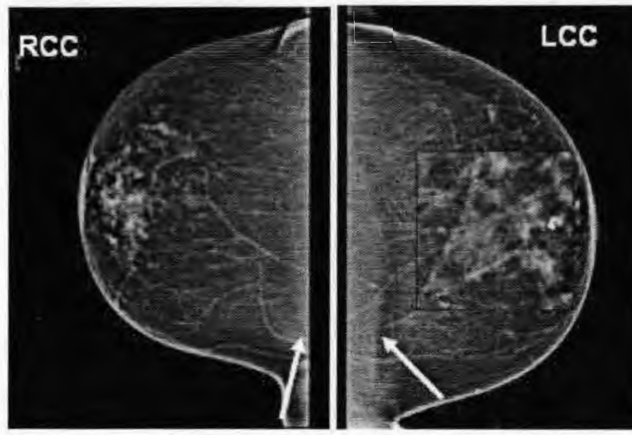


Figure 8: CC View Mammogram of Right & Left Breast

1.7.2 Medio Lateral Oblique (MLO) View

MLO view is the oblique or angled view in which the medial or middle portion is prominent. This view is preferred as it imaged the upper outer portion of the breast too. But this view also contain pectoral muscle which is not the part of the breast.

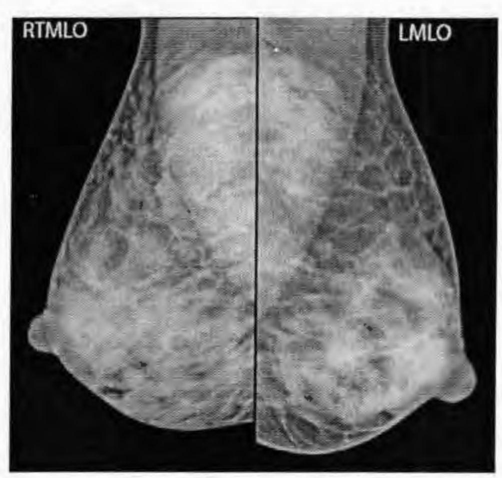
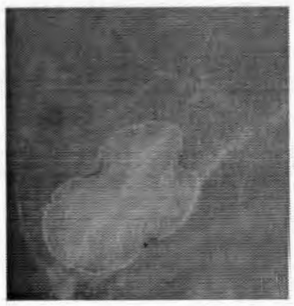


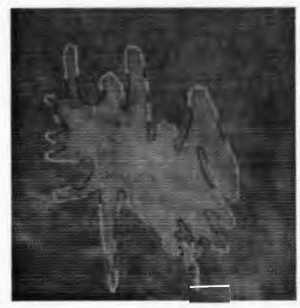
Figure 9: MLO View of Right & Left Breast

1.8 Categorization based on the Abnormality

Doctor or Radiologist after analyzing mammogram will categorized the patients in any one of the following three categories.

1. Normal.
2. Benign (Non-Cancerous Mass Detected).
3. Malignant (Cancerous Mass Detected).





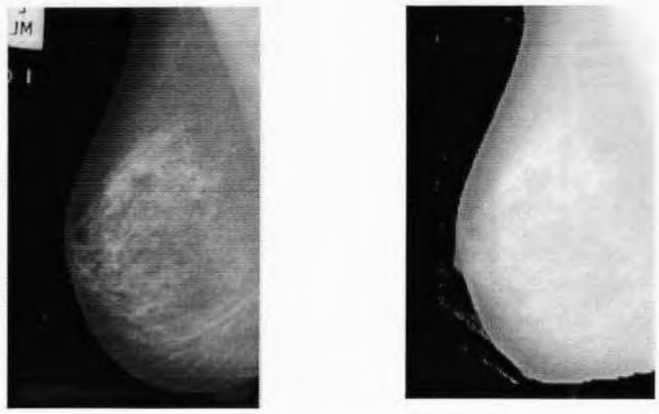
(a) **(b)**

Figure 10: (a) Benign, (b) Malignant

1.9 Breast Segmentation and Pectoral Muscle Suppression

The process of segmentation consists of two segmentations which are independent to one another. The first segmentation, fragments background region containing labels and annotations, whereas the second segmentation isolates the pectoral muscle portion present in Medio Lateral Oblique (MLO) views from the leftover portion of the breast.

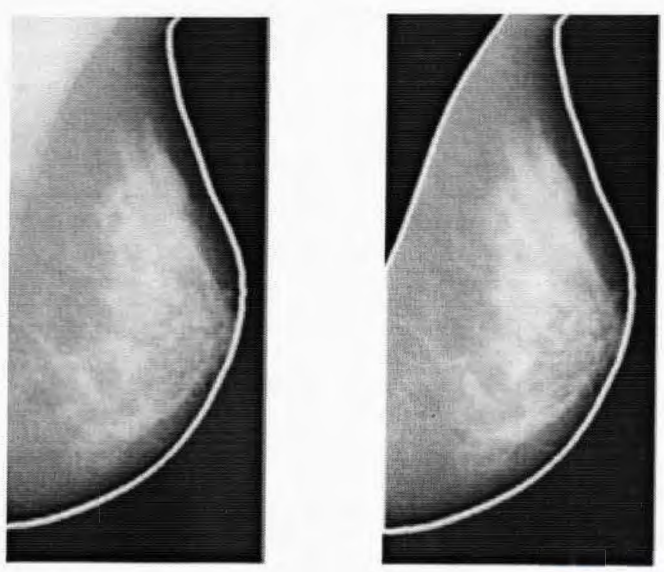
Breast Region segmentation is an important prerequisite in Computer Aided Design (CAD) analysis in which our target is to separate breast from its background of mammograms. The Fig. 11(a) is an image of Raw Mammogram and Fig. 11(b) is the resultant of breast segmentation



(a) (b)

Figure 11: (a) Mini MIAS Mammogram, (b) Segmented Breast

In MLO view, although it provide the better view but in this view, pectoral muscle is also framed during image capture. Removal of this pectoral muscle from the mammogram image is called the suppression of pectoral muscle. The Fig. 12(a) shows the unsuppressed and Fig. 12(b) represents suppressed pectoral muscle from the mammogram.



(a) (b)

Figure 12: (a) Unsuppressed, (b) Suppressed Pectoral Muscle

1.10 Problem Statement

In segmentation, region of interest is limited from the whole mammogram to the breast region only. In this way, segmentation helps fast execution of an algorithm for further cancer classification and detection.

The problem in segmentation with latest techniques is that breast is not properly segmented specially in low contrast images in which it is difficult to differentiate between breast skin line and the background. It is also due to presence of noise and other artifacts present in the mammogram. Breast air interface detection thus becomes difficult and demands digital processing and an aid of computer technology to help radiologist for further cancer classification and detection.

In MLO view of mammogram, pectoral muscle presence causes a problem in further cancer classification and detection. The intensity of the pectoral muscle resembles with that of the cancer cells and also with the fatty cells hence it is very essential to removal or suppress the pectoral muscle portion.

These are the two main preprocessing problems that this thesis addressed.

1.11 Objectives of Research

Our goal in this master research is to develop a new improved or hybrid technique for Breast Segmentation and Pectoral Muscle Suppression in Mammograms. Also we aim to reduce the computational complexity.

Our aim is to fight against Breast Cancer by aiding in preprocessing and breast segmentation and remove the pectoral muscle in mammograms and the same segmented breast may later be used in CADs for the early detection of cancer in breasts and save lives.

1.12 Research Methodology

The most commonly used datasets by the researchers worldwide for research on image processing are Mini Mammographic Image Analysis Society (MIAS) Database and Digital Database for Screening Mammography (DDSM). In this thesis, we have preferred entire Mini MIAS database to test our algorithm to remove any biasness in results by avoiding any pre selection of images. The Mini MIAS Database contains 322 mammogram images of 161 patients which includes all types of background tissues like fatty, fatty-glandular and dense glandular breast tissues. In addition to that, this is a comprehensive database that contain all three possibilities i.e. benign, malignant and normal cases.

The proposed algorithm to segment the breast and suppression of pectoral muscle was implemented in MATLAB simulation software and the results were compared with latest techniques and with the expert radiologist drawn masks or ground truths.

1.13 Organization of Thesis

In Chapter 1, as the name implies includes the introduction about the thesis. It highlights the importance of early detection of cancer and its effectiveness in saving precious human life of cancer patients. Problems faced during the digital processing of mammogram are discussed briefly. Breast Segmentation and pectoral muscle terminologies are also explained in this introductory chapter.

Chapter 2, is about the literature survey of latest techniques used for breast segmentation and pectoral muscle suppression. Strengths and loop holes of different latest techniques were debated. This chapter includes success rates of different algorithms to compare different algorithm effectiveness.

Proposed algorithm is discussed in detail in Chapter 3. The research mythology adopted to address the problem of breast segmentation and pectoral muscle suppression is explained in depth

The results and discussion were presented in chapter 4. The results are compared with other available techniques used by different authors. Quantitative and comparative analyses are made in this chapter.

Concluding remarks, recommendations for future work and summary are the point of focus in the last chapter 5 of this thesis.

Chapter 2 Literature Review: Segmentation

Literature survey was focused on the segmentation methods developed over time to address the same problems. It has been found during the studies that the two segmentations of breast and pectoral muscle are independent to one another. The different techniques proposed for the Extraction of Breast boundary are presented first followed by methods for the Removal of Pectoral Muscle are discussed in this chapter.

2.1 Breast Segmentation Methods

There are many methods that are used for breast segmentation. However there are some methods that are frequently used by majority of authors and these are discussed along with their merits and demerits in this section.

2.1.1. Thresholding Method

Hoyer et al. [13] in 1979 attempted Breast Region separation using simple histogram thresholding. Later in 1995, Bick et al. [14] presented an algorithm which used local histogram thresholding, region growing and morphological filtering. In 1998, Hein et al. [15] proposed global thresholding. Later in 2000 Masek et al. [16] proposed a new local thresholding method. Raba et al. 2005 [17] also used Global Threshold method but when compared to latest technique developed by Mustra et al. 2013 [18] show lower performance in comparison. Since thresholding is dependent on pixel intensity so these methods fail to give accurate results in low contrast images.

2.1.2. Gradient Based Techniques

Gradient based techniques were also proposed for Breast Region Extraction by other Researchers in parallel periods. In 1980, Breast boundary was obtained by Semmlow et al. [19] using combination of Sobel Edge detection and spatial filters. Later Mendez et al. 1996 [20] came with a better algorithm and used a two level histogram threshold to obtain breast region. They track breast boundary using gradient technique. Karssemeijer et al. 1998 [21] exploit multiresolution scheme and processed a low resolution version and then extrapolate to get desired results. They then used global thresholding technique to obtain an initial sketch of a region and then processed using Sobel operator of size 3 x 3. Mustra et al. 2013 [18] developed further improvement in this technique by introducing K-Means thresholding in algorithm and attains more efficient results at the cost of complexity.

2.1.3. Polynomial Fitting

Effective polynomial modeling was proposed by Chandrasekhar and Attikiouzel et al. 1996 [22]. Breast region is approximated with the aid of initial threshold. Through image translation and rotation method, a quadratic or cubic polynomial was fitted as suggested by Goodsitt et al. 1998 [23]. Chen et al. 2010 [24] further improved the algorithm but when compared by latest technique developed by Mustra et al. 2013 [18] show very low performance.

2.1.4. Active Contour or Region Growing Method

Active contours techniques were also suggested for breast segmentation and Ojala et al. 1999 [25] were the pioneer to use these techniques for breast segmentation. Later, in 2000 McLoughlin et al. [26] improved technique and use a global threshold to obtain preliminary region of interest of breast. The statistical model is used for breast pixels inside the region on which snake algorithm was applied to segment breast ad get final approximation for breast boundary. Wirth et al. 2004 [27] segmented the breast from the technique of active contour. Their technique works on two initial regions and form a convolution matrix and then further enhance the edges and boundary. Two regions have two thresholds from which they calculate control point for the s comparison of the two regions. Maitra et al. 2011 [28] proposed a modified version of seeded region growing algorithm and provide better results but slightly less performance than the technique proposed by Mustra at al 2013 [18].

Table 2: Comparison of Methods for Breast Segmentation

Method	Authors	Success (%)	Database	YEAR
Sobel, K-Means, CLAHE	Mustra [18]	91.61	Mini MIAS	2014
CLAHE + Seeded Region Growing	Maitra [28]	95	Mini MIAS	2012
Global Threshold	Raba [17]	86	Mini MIAS	2005
Polynomial Fitting Edge Detection	Chen [24]	62.5	EPIC	2010

Summarizing, the algorithms that used more than one techniques show better results over single algorithms because of the fact of varying nature of breast contour for different type of human.

2.2 Pectoral Muscle Suppression

The pectoral muscle consists of bright pixels and are presented in most of the Medio Lateral Oblique MLO views of Mammograms. Due to these bright intensity values present in the image, this can influence the image processing final results. Another issue is that the density of the Pectoral muscle is approximately the same as that of the dense cancer cells of our interest, so the suppression of pectoral muscle needs special care. Many detection and segmentation techniques of pectoral muscle have been suggested. Some of the notable techniques are presented in section followed to detect and segment the pectoral muscle.

2.2.1 Wavelet Transform

Mustra et al. 2009 [30] detect pectoral muscle by using a combination of bit depth reduction and wavelet decomposition. Ferrari et al. 2000 [29] suggested Gabor wavelet filter bank for the same purpose. They first convolved Gabor filters' group and then used Wavelet filters for pectoral muscle detection. The same authors (Mustra et al. 2013 [18]) further proposed different techniques in their latest paper by using polynomial fitting which is very useful in low contrast mammograms for pectoral muscle removal.

2.2.2 Region Growing

Raba et al. 2005 [17] proposed method for pectoral muscle detection was based on region growing. They grow region of pectoral muscle by threshold estimation. Maitra et al. 2011 [28] used CLAHE and seeded region growing algorithm, whereas Raba et at 2005 [17] used Region growing with Intensity threshold method but their performance was less than the polynomial fitting as proposed by Mustra et al. 2013 [18].

2.2.3 Polynomial Fitting

Chen et al. 2010 [24] proposed Polynomial edge fitting detection technique this was further improved by Mustra 2013 [18] in their recent proposed algorithm which used polynomial curve fitting for pectoral muscle suppression. The result comparison shows an overall improvement. Mustra et al. 2013 [18] results show poor segmentation of breasts having varying slope. It is further identified that the detected breast boundary was not regular and natural as can be seen in Fig. 13 which is one of the results of breast segmentation result of mammogram processing presented in their paper (white portion shows error of their algorithm).

Figure 13: Extracted and Actual Breast Boundary

Clearly visible in above extracted breast, the non-uniform breast boundary (leftwards).

Table 3: Comparison of Methods for Pectoral Muscle Suppression

Method	Authors	Success %	Database	Year
Polynomial Fitting	Mustra [18]	96.56	Mini MIAS	2013
Modified Region Growing, CLAHE	Maitra [28]	95.71	Mini MIAS	2012
Region Growing, Linear Estimation	Chen [24]	93.50	EPIC	2010
Region Growing, Threshold Estimation	Raba [17]	86.00	Mini MIAS	2005

2.3 Summary

Every algorithm has its pros and cons. Some provide better results than he rest. Global thresholding based methods are although fast and took less execution time but fails to provide accurate results for segmentation due to contrast variation between pectoral muscle, breast tissue and the background. On the other hand local thresholding techniques takes more computations but provide better results and is less affected due to difference in contrast and pixel value.in mammograms. For gradient based techniques, sobel edge detection do local averaging for three points, and hence more suitable for boundary detection. Although it is complex but easy to implement on hardware. Polynomial fitting methods are highly dependent on the initial seed value for the curve or polynomial calculation. A slight wrong selection of points for curve fitting have great impact on end results of segmentation. Active contour method accuracy depends on the convergence criteria. Small detail is frequently neglected during convergence and hence affect results. Wavelet transform works well when a large part of pectoral muscle but when visible portion of pectoral muscle is small, the efficiency of wavelet transform is also compromised. The current technique proposed by Mario et al. [18] although have better overall results. But their algorithm fails to properly segment non symmetrical breast. The reason is they have suppose a breast as a circular object with the center as one third of the breast height and use polar coordinates. Hence their algorithm fails in images which include larger portion of pectoral muscle or shape of breast is non circular or asymmetrical.

2.4 Other Methods for Segmentation

Some other methods used by different authors are mentioned in the Table.4.

Table 4: Methods used by Different Authors

Method	Paper	Authors
Boundary and Region Based	A novel approach for breast skin-line estimation in mammograms.	Y. Sun et al.[32].
Active Contours	Identification of the breast boundary in mammograms using active contour.	R.J.Ferrari, et al. [29].
Hough Transform	Segmentation of the Breast Region with Pectoral Muscle Suppression and Automatic Breast Density Classification.	J.S.Tomas [33].
Morphological Segmentation	Morphological segmentation,	F. Meyer, et al. [34].
CLAHE	Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the	E.D. Pisano et al.[35].

	Detection of Simulated Speculations in Dense Mammograms.	
Morphological + Polynomial of Degree 3	An automatic method for delineating the pectoral muscle in mammograms,	I.M. de Carvalho, et al.[36].
Fast Marching Method + Area Morphology	Breast skin-line estimation and breast segmentation in mammograms using fast-marching method.	R.D. Yapa et al. [37].
Wavelet Transform	Breast border extraction and pectoral muscle detection using wavelet decomposition.	M. Mustra et al. [30].
Hough Transform	Detection of pectoral muscle in mammograms using a mean-shift segmentation approach.	A. Sultana et al. [38].

Chapter 3 Proposed Algorithm

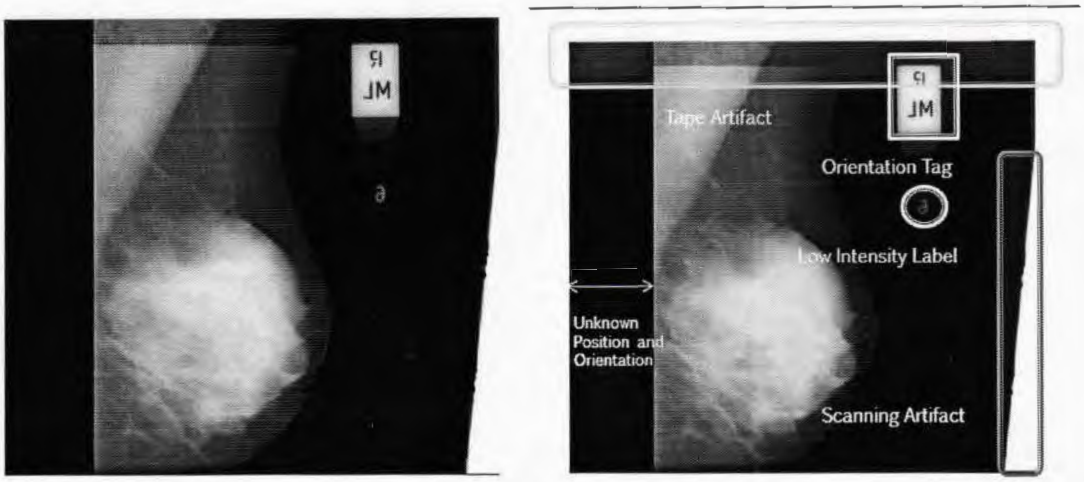
In this chapter, the algorithm for the breast segmentation is discussed followed by the pectoral muscle suppression. Breast alignment, which is also one of the preliminary process in segmentation, is therefore explained before segmentation.

3.1 Breast Alignment

Breast orientation and position is usually unknown and it is very essential for digital processing of mammogram to first solve this issue and aligned the image. Image alignment of proposed algorithm is discussed in this section.

Before alignment, mammogram are preprocessed in order to remove the noise and other artifacts present in the mammogram during the image registration. The common artifacts that a mammogram can have are stated below are also shown in Fig 14(b).

1. Tape artifact
2. Orientation Tag
3. Low Intensity Label
4. Scanning artifact
5. Unknown orientation and position of the breast.



(a) **(b)**

Figure 14: (a) MIAS Mammogram Image, (b) Labelled Artifacts

3.1.1 Steps for Image Alignment

The proposed algorithm for the mammogram alignment consists of the following steps.

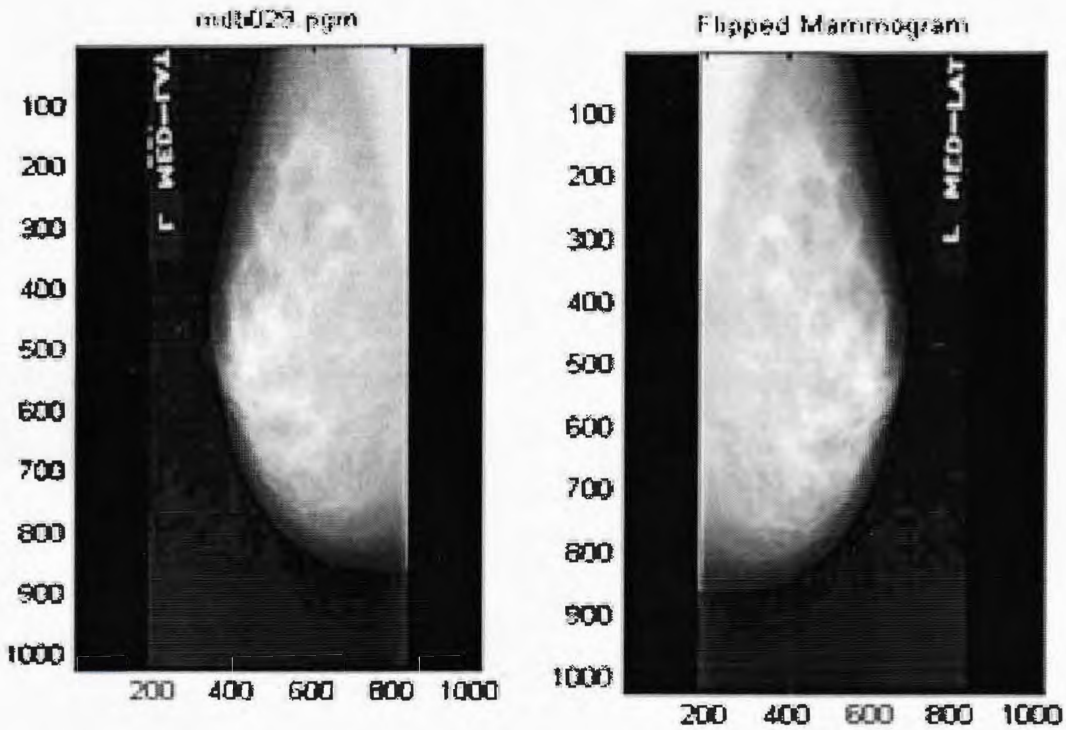
Step 1: Breast mammogram can be either left orientated or right oriented. Mammogram orientation has been determined by first dividing the mammogram image $I(x,y)$ of size $n \times m$ where 'n' is the number of rows and 'm' is the number of columns, equally in left $I_L(x,y)$ and right section $I_R(x,y)$and then decide on the basis of comparison of pixel intensity between the two. The pixel intensities are higher in section where pectoral muscle, major portion of breast tissue is present and pixel intensities are comparatively lower smaller portion of breast tissue and dark background exist. That is

$$\sum_{x=1}^{n}\sum_{y=1}^{m} I(x,y) = \sum_{x=1}^{n}\sum_{y=1}^{\frac{m}{2}} I(x,y) + \sum_{x=1}^{n}\sum_{y=\frac{m}{2}}^{m} I(x,y)$$

If	$\sum_{x=1}^{n}\sum_{y=1}^{\frac{m}{2}} I(x,y) > \sum_{x=1}^{n}\sum_{y=\frac{m}{2}}^{m} I(x,y)$
Then	Mammogram $I\ (x,y)$ is right oriented
Else if	$\sum_{x=1}^{n}\sum_{y=1}^{\frac{m}{2}} I(x,y) < \sum_{x=1}^{n}\sum_{y=\frac{m}{2}}^{m} I(x,y)$
Then	Mammogram $I\ (x,y)$ is left oriented

In order to simplify, our proposed algorithm is customized for right orientated mammograms $I_{RIGHT}(x,y)$. The left oriented breast mammograms $I_{LEFT}(x,y)$ are flipped for processing and then flipped again at last step of algorithm.

$$I_{RIGHT}(x,y) = Flip\ \{\ I_{LEFT}(x,y)\}$$



(a) (b)

Figure 15: (a) Left Mammogram mdb029, (b) Flipped Mammogram

Step 2: Very high intensity noise like orientation tag and low intensity noise are very common in images like X-ray and mammogram. In order to remove these two completely different types of noises, the approach used is the combination of two adoptive Binary masks obtained through thresholding process.

Threshold value τ_1 and τ_2 are used to create the lower mask $M_L(x,y)$ and upper mask $M_U(x,y)$ of mammogram $I(x,y)$, where $P(x,y)$ is the pixel intensity at location (x,y) in Image $I(x,y)$ and can be represented by the expression

$$M_L\,(x,y) = \begin{cases} 0, & P(x,y) \leq \tau_1 \\ 1, & P(x,y) > \tau_1 \end{cases}$$

And

$$M_U\,(x,y) = \begin{cases} 0, & P(x,y) > \tau_2 \\ 1, & P(x,y) \leq \tau_2 \end{cases}$$

Where

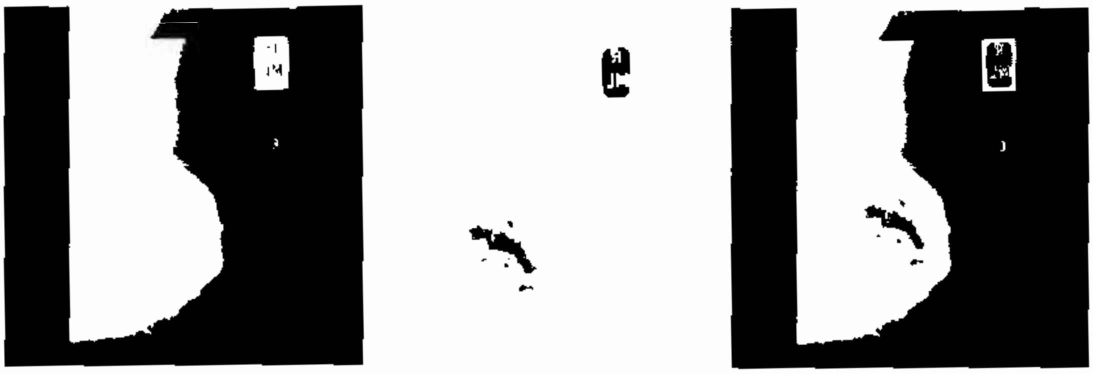
$$\tau_1 = 0.2 \times \max\ \{I(x,y)\}$$

And

$$\tau_2 = 0.93 \times \max\ \{I(x,y)\}$$

The final binary mask $M_F\,(x,y)$ is obtained by the *AND* operation on the lower mask $M_L\,(x,y)$ and the upper mask $M_L\,(x,y)$

$$M_F\,(x,y) = M_L\,(x,y).M_U\,(x,y)$$



(a) (b) (c)

Figure 16: (a) Lower Mask (b) Upper Mask (c) "AND" Result

Step 3: In order to remove the orientation tag from the image, a square structuring element, $S_{SQ}\ (x, y)$ of size 31 x 31 is used for morphological opening, and is convolved with the final mask $M_F\ (x, y)$ to obtain an image $I_L(x, y)$ in which label has been removed shown in the Fig. 17.

$$.I_L(x, y) = S_{SQ}\ (x, y) \circledast M_F\ (x, y)$$

Where

$$S_{SQ}\ (x, y) = 1 \qquad 1 \leq x \leq 31,\ 1 \leq y \leq 31$$

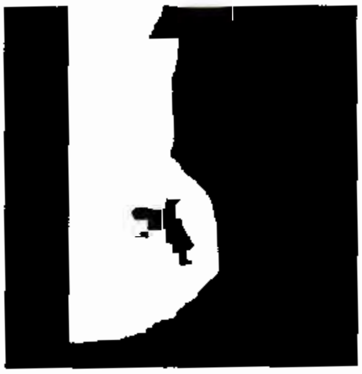


Figure 17: Result of Morphological Opening

Step 4: The next step is the edge detection. Gradient based Sobel filter ∇_{sobel} preserve the breast edge of the morphological open image $I_L(x, y)$ where the gradient of image is maximum to give an image $I_E(\mathrm{x}, \mathrm{y})$ as shown in Fig. 18.

$$I_E(x, y) = \nabla_{sobel} * I_L(x, y)$$

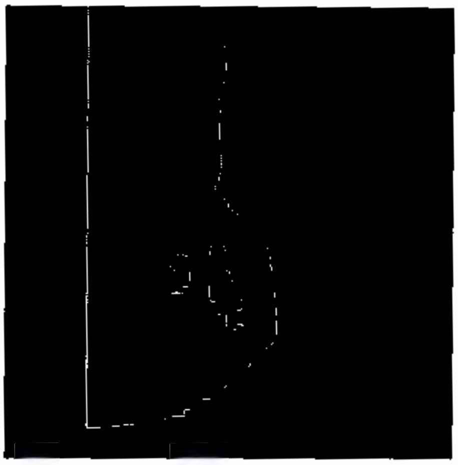


Figure 18: Result of Sobel Edge Detection

Step 5: Next step in our algorithm is the final step in image alignment. From sobel edge detection, the longest line is the start of the breast tissue which need to align with the first column of the image. Hough transform is applied to the image and longest line in the image is searched which is indicated as green line in the Fig. 19.

Once longest line is detected, by using Hough transform, it can be aligned with the first column of Image$I(x, y)$. Hence breast alignment objective has been achieved.

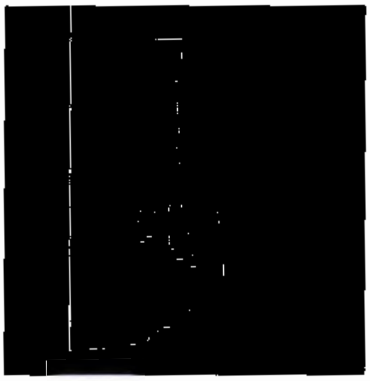
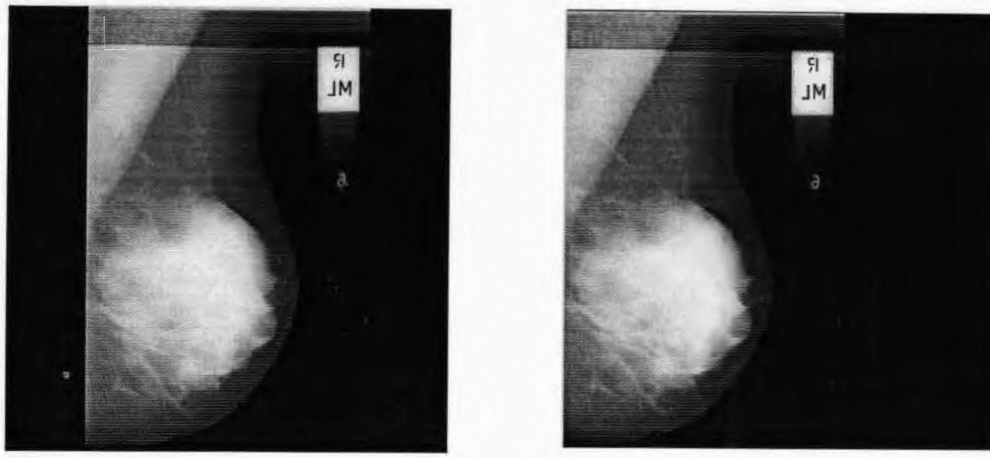


Figure 19: Longest Line of Edge Detection

$$I_{align}(x, y) = \Im_{HOUGH} \lfloor (I_E(x, y)) \rfloor$$

The final image $I_{align}(x, y)$, is aligned with known position of breast is shown in Fig. 20(b).



(a) (b)

Figure 20: (a) Mammogram, (b) Aligned Mammogram

Algorithm 1: Breast Alignment

Input: Mammogram Image $I(x,y)$ of size $n \times m$

Output: Aligned Mammogram, $I_{align}(x,y)$

1. **Check** Orientation

$$\sum_{x=1}^{n}\sum_{y=1}^{\frac{m}{2}} I(x,y) > \sum_{x=1}^{n}\sum_{y=\frac{m}{2}}^{m} I(x,y)$$

If true go to step 2,

$else \quad I(x,y) = Flip\{\, I(x,y)\}$

2. **Compute** Binary Mask

$$M_F\,(x,y) = and(M_L\,(x,y)\,, M_U\,(x,y))$$

Where

$$M_L\,(x,y) = \begin{cases} 0, & P(x,y) \leq \tau_1 = 0.2 \times \max\,\{I(x,y)\} \\ 1, & P(x,y) > \tau_1 = 0.2 \times \max\,\{I(x,y)\} \end{cases}$$

$$M_U\,(x,y) = \begin{cases} 0, & P(x,y) > \tau_2 = 0.93 \times \max\,\{I(x,y)\} \\ 1, & P(x,y) \leq \tau_2 = 0.93 \times \max\,\{I(x,y)\} \end{cases}$$

$P(x,y)$is the pixel intensity at location (x,y) in Image $I(x,y)$

3. **Calculate** label free image $I_L(x,y)$ through convolution of final mask $M_F\ (x,y)$ and structuring element $S_{SQ}\ (x,y)$

$$I_L(x,y) = S_{SQ}\ (x,y)\ \circledast\ M_F\ (x,y)$$

Where

$$S_{SQ}\ (x,y)\ = 1 \qquad 1 \le x\ \le 31,\ \ 1 \le y \le 31$$

4. **Do** Sobel Edge detection by convolution,

$$I_E(\mathrm{x},\mathrm{y})\ =\ \nabla_{sobel}\ *\ I_L(x,y)$$

5. **Detect** longest line for alignment by applying Hough transformation:

$$I_{align}(x,y) = \Im_{HOUGH}\ [(I_E(\mathrm{x},\mathrm{y})\ )]$$

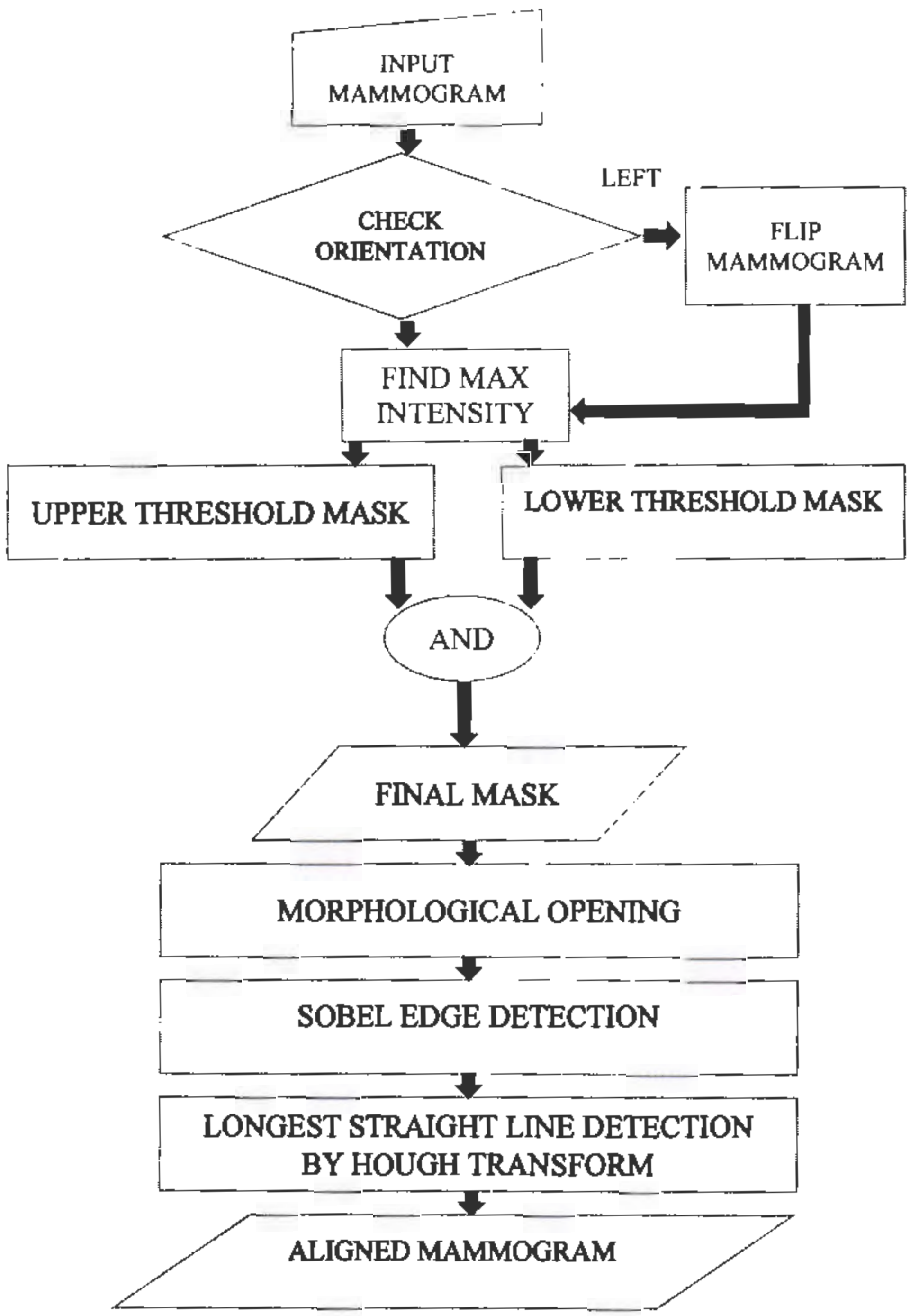


Figure 21: Breast Alignment Flow Diagram

3.2 Breast Segmentation

In MLO view of any mammogram, there are at least three major portions. These are breast tissue, pectoral muscle and background. The area or our Region of Interest (ROI) for the breast cancer detection is inside the breast tissue. In this section, breast segmentation proposed algorithm is discussed in detail.

3.2.1 Contrast Enhancement

Pixel intensities or brightness in the Mammograms varies due to different reasons. The light intensity is not the same throughout the mammogram. Moreover the exposure of the breast also varies from one person to other and also from machine to machine. The mammogram image brightness is also dependent on the angle between the light source and the breast object. In order to segment the breast accurately, the contrast variation must be enhanced in order to clearly differentiate between the breast border and the background.

The technique used in our proposed algorithm for enhancing contrast is "guided filter" in which the mammogram image itself used as the guiding image. The reason for choosing guided filter is that it is one of the fastest edge preserving algorithm currently available [31]. This aspect is very much desired in real time processing of mammogram. He et al. [31] proposed guided filter was used with below input parameters to get the output $I_{GFO}(x, y)$

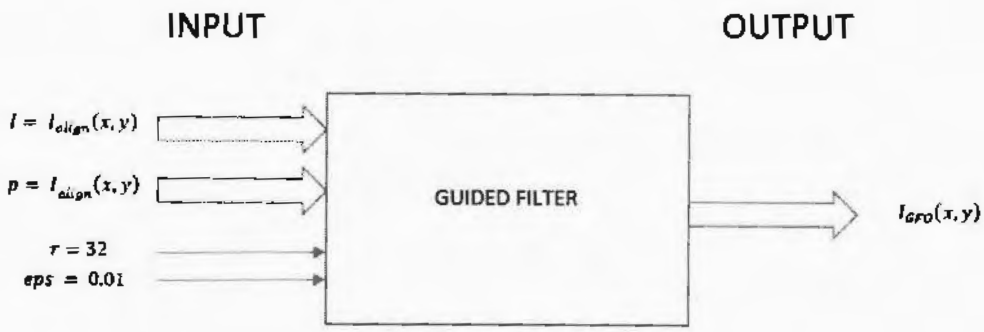
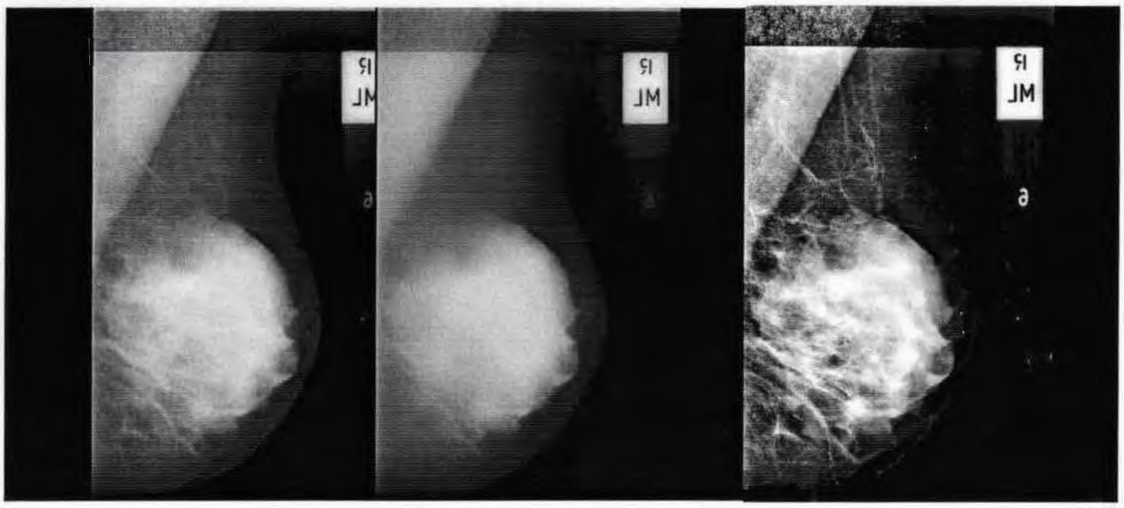


Figure 22: Guided Filter Block

Guided filter output $I_{GFO}(x,y)$ is then used to produce contrast enhanced image $I_E(x,y)$ using equation,

$$I_E(x,y) = (I_{align}(x,y) - I_{GFO}(x,y)) \times 5 + I_{GFO}(x,y)$$



(a) (b) (c)

Figure 23: (a) Mammogram, (b) Guided Filter Output, (c) Enhancement Result

3.2.2 Thresholding

Smoothed image $I_{GFO}(x,y)$ obtained as the result of guided filter is then thresholded with threshold value of $\tau = 0.07$ to obtain binary image $T(x,y)$ free from background noise. If $Q(x,y)$ is the percent pixel intensity at location

(x, y) of a grey scale image $I_{GFO}(x, y)$, then thresholded binary image $T(x, y)$ can be expressed as

$$T(x,y) = \begin{cases} 0, & Q(x,y) \leq \tau \\ 1, & Q(x,y) > \tau \end{cases}$$

Where

$$\tau = 0.07$$

The threshold value of $\tau = 0.07$ replaces all pixels of bottom 7 % intensities with 0 and sets all other as1. This value $of\ \tau$ remove the lower pixel values which are present in the smoothed image $I_{GFO}\ (x, y)$.

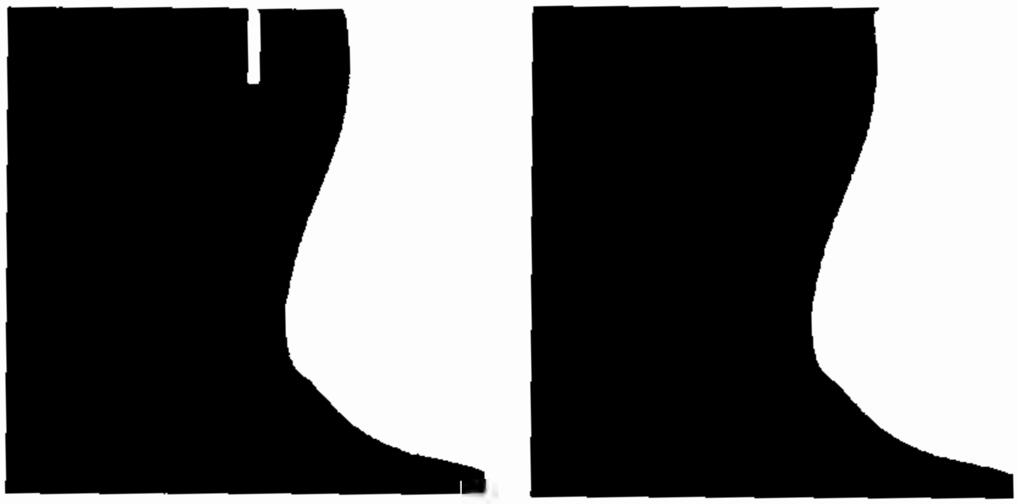
(a) **(b)**

Figure 24: (a) Smoothed Image, (b) After Thresholding

3.2.3 Morphological Opening

In order to give breast boundary regular natural shape, the disk or circular shape image structuring element $I_{SE\ disk}$of radius $r = 25$ is used. The morphological opening operation comprises of two operations. .i.e. image erosion and then image dilation. The structuring element for both operations is the $I_{SE\ disk}$

$$I_{MO}(x,y) = \langle\left(T(x,y) \circledast I_{ER-DISK}(x,y)\right) \circledast I_{DI-DISK}(x,y)\rangle$$



(a) (b)

Figure 25: (a) Binary Image, (b) After Image Opening

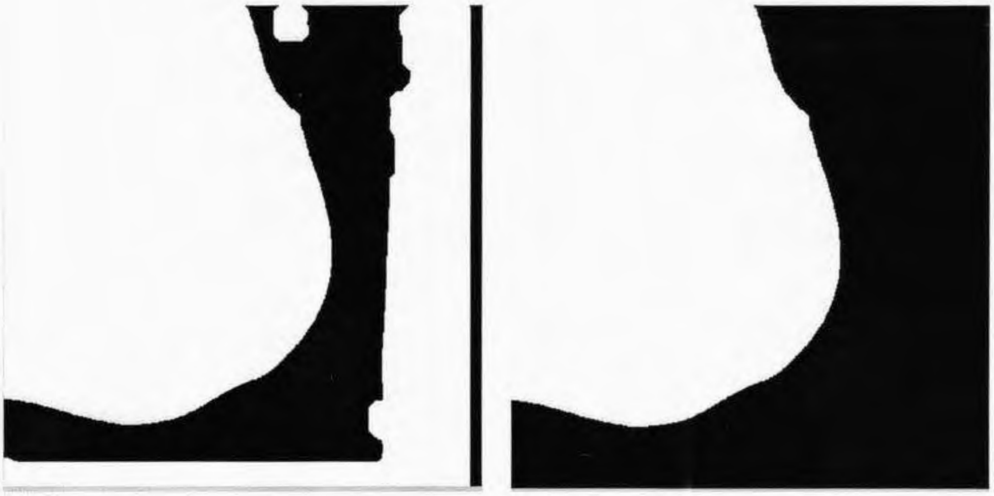
3.2.4 Image Cleaning

During Mammogram registration, a large slab is sometime present in the mammogram. In order to remove such unwanted object, objects of size smaller than 35000 pixel are removed from the binary mammogram image $I_{MO}(x,y)$. The resultant image $I_C(x,y)$ is shown in Fig. 26(b)

$$I_C(x,y) = \begin{cases} 1, & O_{area} \geq 35000\, pixels \\ 0, & O_{area} < 35000\, pixels \end{cases}$$

Where

O_{area} is the area of the object in image $I_{MO}(x,y)$

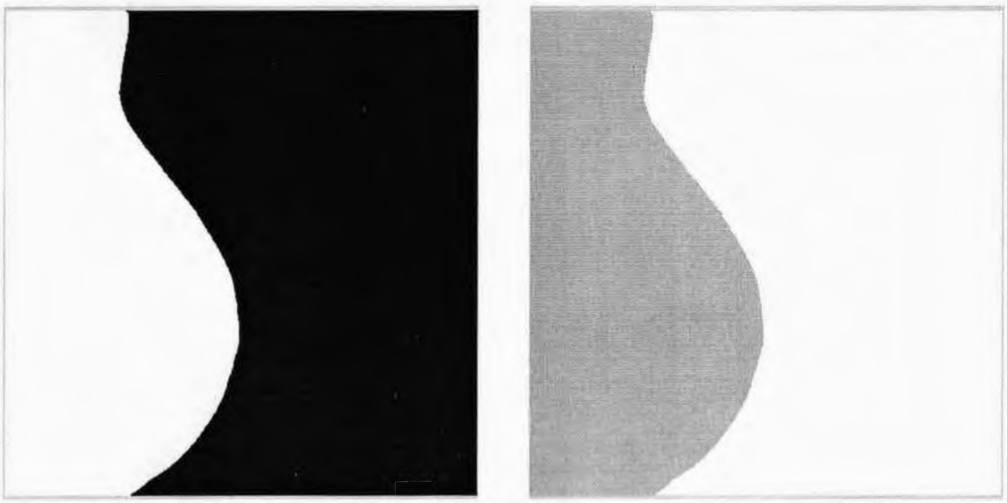


(a) (b)

Figure 26: (a) Before Image Cleaning, (b) After Image Cleaning

3.2.5 Connected Object Segmentation

In this process, 8-neighborhood N_8 of input pixel is searched for connectivity. If connectivity is established then those neighborhood pixels of N_8 which are of same intensity value as input pixel $P(x, y)$ are marked with same label. In this way all pixels of the connected object in the image are labelled as one value.



(a) (b)

Figure 27: (a) Cleaned Mammogram, (b) Connected Object Labeled Mammogram

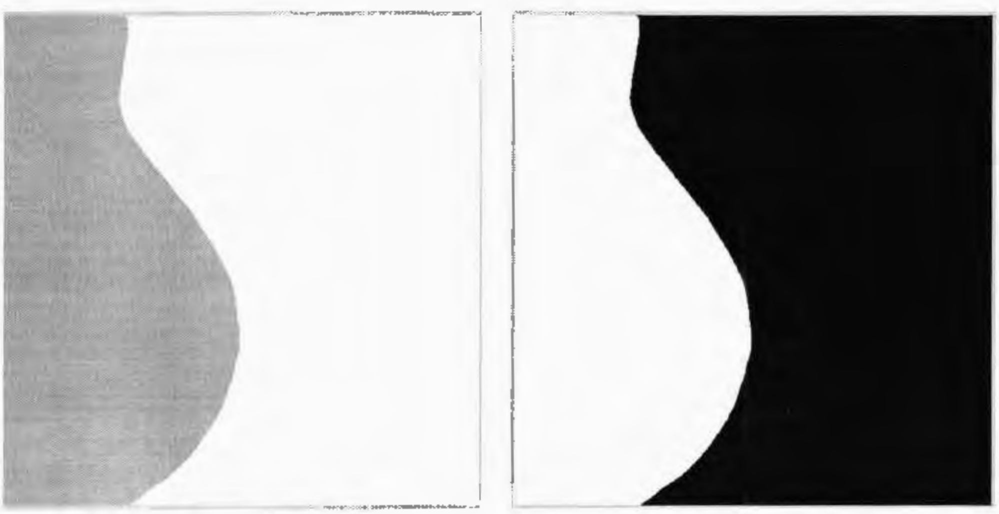
$$Label_{Tt(x,y)} = \{Label_{P(x,y)} \quad if \;\; Tt(x,y) = P(x,y), \forall\, (x,y) \in N_8$$

3.2.6 Breast Object Selection

In the previous section, all the same objects in the mammogram are labelled as one value. Since alignment is already done with the mammogram, from this known fact, the left most pixel $P(1,1)$ of the image $I(x,y)$ must be of the breast object. Hence all the pixels inside the breast having same label value are set to 1 and 0 is assigned to other label values.

In this way our final image $I_{BM}(x,y)$ only contain the breast object.

$$I_{BM}(x,y) = \begin{cases} 1, & Label_{(x,y)} = Label_{P(1,1)} \\ 0, & Label_{(x,y)} \neq Label_{P(1,1)} \end{cases}$$



(a) (b)

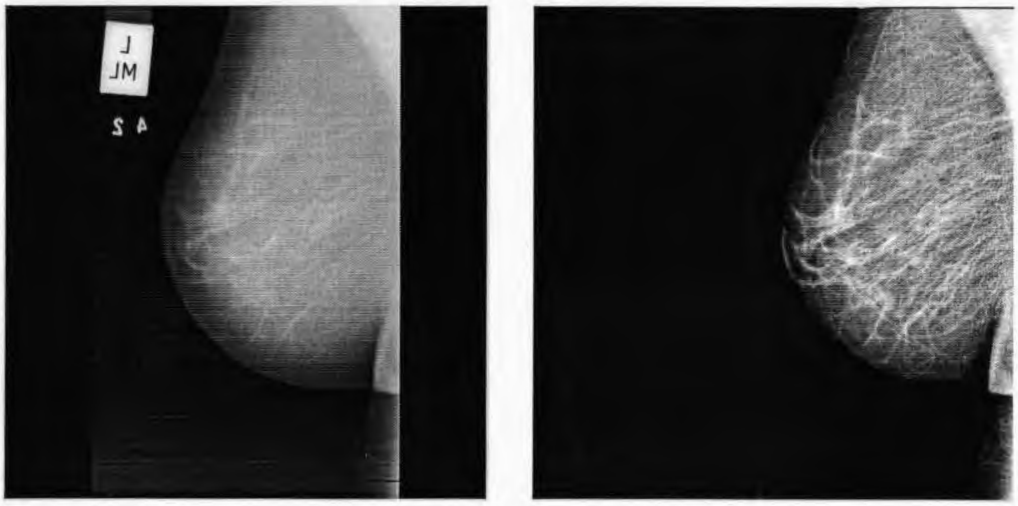
Figure 28: (a) Labelled Mammogram, (b) Segmented Breast Mask

3.2.7 Applying Mask and Finalizing Segmentation

Final segmented breast $I_{SB}(x,y)$ is obtained by applying the segmented breast mask $I_{BM}(x,y)$ on the enhanced image $I_E(x,y)$obtained by processing guided filter output $I_{GFO}(x,y)$. In this way the breast is segmented and rest of the unwanted objects are removed.

The final breast image $I_{SB}(x,y)$ is aligned, segmented and its contrast is also enhanced.

$$I_{SB}(x,y) = I_{BM}(x,y) * I_E(x,y)$$



(a) **(b)**

Figure 29: (a) Mammogram, (b) Final Segmented Breast

Algorithm 2: Breast Segmentation

Input: Aligned Mammogram, $I_{align}(x,y)$

Output: Enhanced, Segmented Breast Mammogram, $I_{SB}(x,y)$

1. **Compute** Guided Filter Output, $I_{GTO}(x,y)$

   Where

   $$I = p = I_{align}(x,y)$$

   $$r = 32$$

   $$eps = 0.01$$

2. **Compute** enhanced image $I_E(x,y)$

$$I_E(x,y) = (\,I_{align}(x,y) \;-\; I_{GFO}(x,y)) \times 5 + \,I_{GFO}(x,y)$$

3. **Calculate** Thresholding Image $T(x,y)$, using $\tau = 0.07$

$$T(x,y) = \begin{cases} 0, & Q(x,y) \le \tau \\ 1, & Q(x,y) > \tau \end{cases}$$

Where

$Q(x,y)$ is percent pixel intensity

4. **Evaluate** Morphological Opened image $I_{MO}(x,y)$

$$I_{MO}(x,y) = \langle\big(T(x,y) \odot I_{ER-DISK}(x,y)\big) \odot I_{DI-DISK}(x,y)\rangle$$

Where

$I_{ER-DISK}(x,y)$, $I_{DI-DISK}(x,y)$ are Erosion & Dilation Images by structuring element disk $I_{SE\ disk}$ of $radius\ =\ 25$

5. **Compute,** clean Image $I_C(x,y)$ by

$$I_C(x,y) = \begin{cases} 1, & O_{area}\ \ge 35000\,pixels \\ 0, & O_{area} < \ 35000\,pixels \end{cases}$$

Where

O_{area} is the area of the object in image $I_{MO}(x,y)$

6. **Label** N_8 connected neighbors in image $I_{MO}(x,y)$.

$$Label_{Tt(x,y)} = \{Label_{P(x,y)} \quad if\ \ Tt(x,y) = P(x,y), \forall\ (x,y) \in N_8$$

7. **Compute** $I_{BM}(x,y)$ by choosing pixels having label same as that of Breast tissue $Label_{P(1,1)}$

$$I_{BM}(x,y) = \begin{cases} 1, & Label_{(x,y)}\ =\ Label_{P(1,1)} \\ 0, & Label_{(x,y)} \ne\ Label_{P(1,1)} \end{cases}$$

8. **Calculate** Final Segmented Breast $I_{SB}(x,y)$

$$I_{SB}(x,y)\ =\ \ I_{BM}(x,y) * \,I_E(x,y)$$

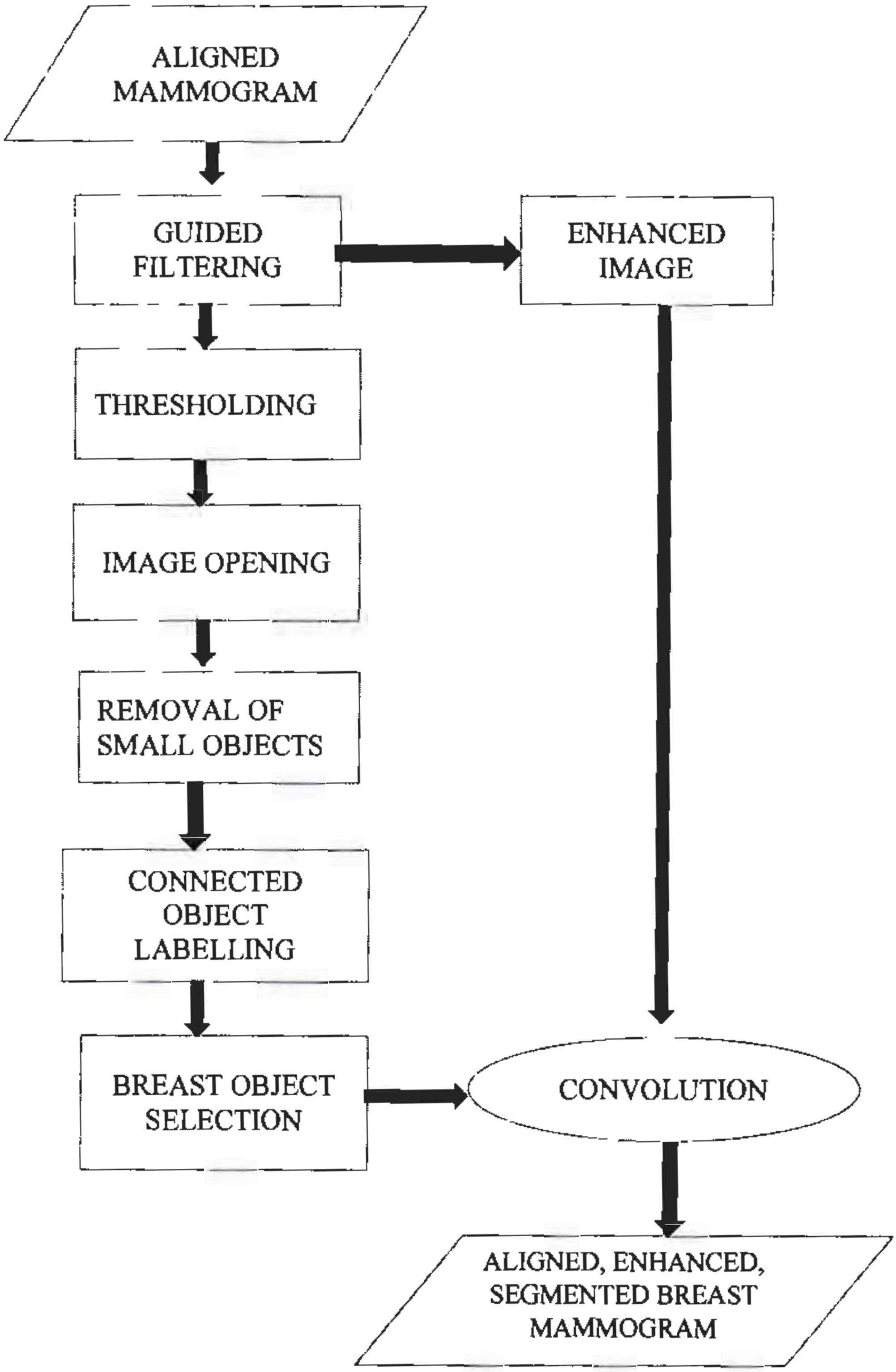


Figure 30: Breast Segmentation Flowchart

3.3 Pectoral Muscle Suppression

The location of the pectoral muscle in an MLO view is known unlike the breast boundary. Pectoral muscle is located in upper left most corner in right oriented breast mammogram and upper right most corner in the right oriented breast mammogram. The other vital information that is known about the pectoral muscle is that it contain brighter pixels as compared to its background which is breast tissue.

All this information laid the basis of our proposed algorithm for pectoral muscle suppression. In the section 3.1, for breast alignment, the left breast is flipped to right one. Same flipped left mammogram image or right oriented breast mammograms is used in algorithm for the suppression of pectoral muscle.

3.3.1 Region of Interest Extraction

Likelihood of the presence of Pectoral Muscle is always the left most top corner in MLO view (as left orientated are flipped). In our proposed algorithm pectoral muscle suppression is always followed by the breast segmentation process. Breast line already detected is used for the extraction of pectoral muscle. For pectoral muscle suppression, two points R_1 and R_2 are very significant on the breast boundary are shown in Fig. 31.

Point R_1, it is the last point on first row of segmented breast image $I_{SB}(x,y)$ whereas the R_2 is the point inside the segmented breast $I_{SB}(x,y)$ which has max value of Y-coordinate among all pixels of breast tissue

$$R_1(x',y') = P(1,y_{max}) \qquad x = 1$$

$$R_2\,(x'',y'') = P(x,y_{max}) \qquad \forall\, x = 1,2,\ldots.1024$$

From point R_1, a verical line is drawn, and from point R_2 horizontal line is drawn. Four quardrants are created as the results of these two perpendiculr

3.3 Pectoral Muscle Suppression

The location of the pectoral muscle in an MLO view is known unlike the breast boundary. Pectoral muscle is located in upper left most corner in right oriented breast mammogram and upper right most corner in the right oriented breast mammogram. The other vital information that is known about the pectoral muscle is that it contain brighter pixels as compared to its background which is breast tissue.

All this information laid the basis of our proposed algorithm for pectoral muscle suppression. In the section 3.1, for breast alignment, the left breast is flipped to right one. Same flipped left mammogram image or right oriented breast mammograms is used in algorithm for the suppression of pectoral muscle.

3.3.1 Region of Interest Extraction

Likelihood of the presence of Pectoral Muscle is always the left most top corner in MLO view (as left orientated are flipped). In our proposed algorithm pectoral muscle suppression is always followed by the breast segmentation process. Breast line already detected is used for the extraction of pectoral muscle. For pectoral muscle suppression, two points R_1 and R_2 are very significant on the breast boundary are shown in Fig. 31.

Point R_1, it is the last point on first row of segmented breast image $I_{SB}(x,y)$ whereas the R_2 is the point inside the segmented breast $I_{SB}(x,y)$ which has max value of Y-coordinate among all pixels of breast tissue

$$R_1(x',y') = P(1,y_{max}) \qquad x = 1$$

$$R_2\,(x'',y'') = P(x,y_{max}) \qquad \forall\, x = 1,2,\ldots.1024$$

From point R_1, a verical line is drawn, and from point R_2 horizontal line is drawn. Four quardrants are created as the results of these two perpendiculr lines as shown in Fig. 31. Quadrant 1 is Region of Interest (ROI) for Pectoral Muscle.

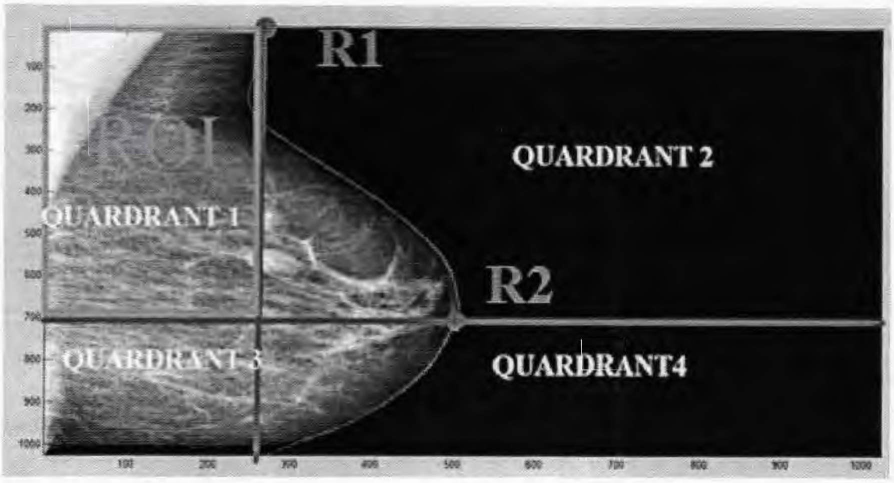


Figure 31: ROI for Pectoral Muscle

Extracted ROI from mdb012.pgm after image enhancement through guided filtering is shown in Fig. 32.

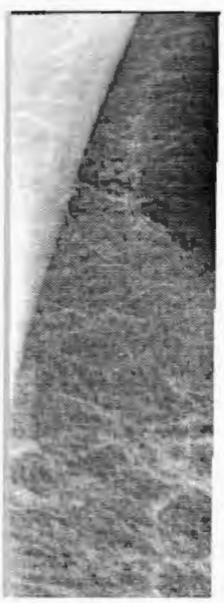


Figure 32: ROI after Enhancement mdb012

3.3.2 Seed selection for Pectoral Muscle

For the selection of initial seed value, three straight lines $\overline{P_1P_2}$, $\overline{P_1P_3}$ and $\overline{P_1P_4}$ are drawn. The line $\overline{P_1P_2}$ is drawn at first row of the image or 0°

angle, other line $\overline{P_1P_3}$ at the angle of 45 degrees and the third line $\overline{P_1P_4}$ on the first column of the image or at an angle 90 degrees by taking point P_1 or $P(1,1)$ as starting point.

$\overline{P_1P_2}$ and $\overline{P_1P_4}$ are the first row and the first column of the extracted ROI of the pectoral muscle. Calculation of coordinates for $\overline{P_1P_3}$is followed below.

Linear equation of line having Slope "m"can be mathematically expressed as,

$$y = mx \qquad \textit{Eq.1}$$

For 45° line, slope "m" is given by

$$m = Tan(45^\circ)$$

So Eq. 1 becomes

$$y = x \qquad \textit{For all values of x}$$

These values of x and y can be used to draw two straight line across the extracted region of interest as shown in Fig. 33.

Figure 33: 0°, 45° and 90° Lines in ROI mdb012

One these three straight line, pixel intensities are tracked individually by moving across these lines $\overline{P_1P_2}$, $\overline{P_1P_3}$ and $\overline{P_1P_4}$ starting from P_1 towards P_2, P_3 and P_4. There is a sudden fall in intensity value moving from brighter region of pectoral muscle to the darker region of breast tissue in all these three lines $\overline{P_1P_2}$, $\overline{P_1P_3}$ and $\overline{P_1P_4}$. These points are labelled in Fig.33 as S_1, S_2 and S_3. The sudden fall in the intensity level is shown in Fig. 34.

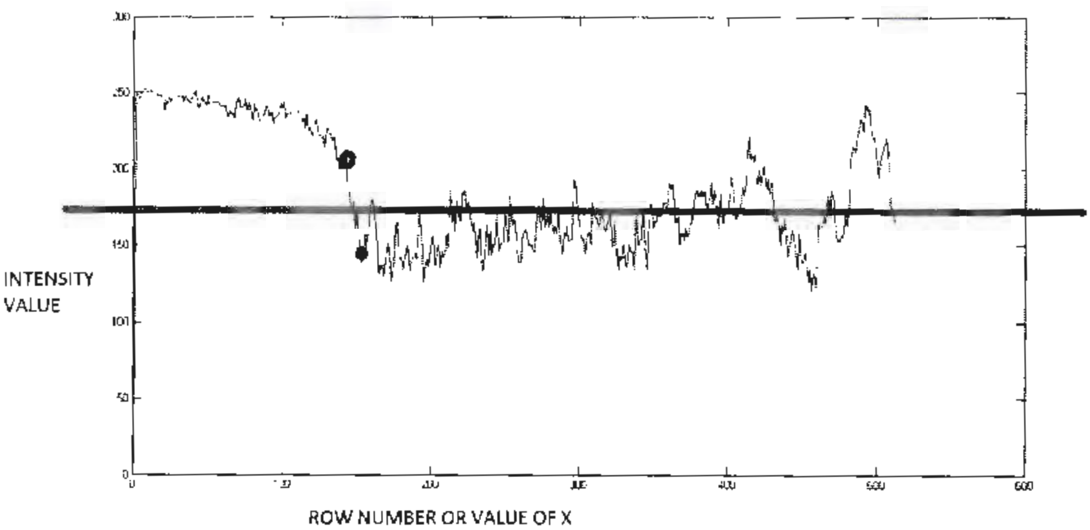


Figure 34: Image Intensities on Straight line

These three point of maximum dip S_1, S_2 and S_3 are considered as the seed value for the approximation of pectoral muscle area.

3.3.3 Triangle Estimation

As pectoral muscle shape resembles than that of the triangle in MLO view. It is assumed that this pectoral muscle is the combination or sum of two triangles and one side of both the triangles is shared or common.

In order to suppress pectoral muscle through these two triangle, the points P_1, S_1, S_2 and S_3 are used as shown in Fig. 35. Two new lines $\overline{S_1S_2}$

and $\overline{S_2S_4}$ are drawn. Two $\Delta\ P_1S_1\ S_2$ and $\Delta\ P_1S_2S_3$ are formed as the result of these two lines as shown.

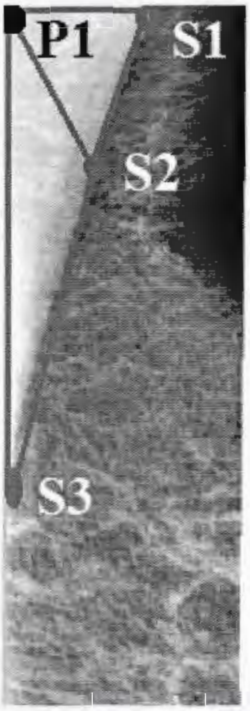


Figure 35: Triangle Formation Using Seed Points

3.3.4 Mask formation for Pectoral Muscle

Finally all the pixel, inside and on the $\Delta\ P_1S_1\ S_2$ and $\Delta\ P_1S_2S_3$ are set to zero for the creation of final mask for pectoral muscle suppression $I_{MASK\_PM}(x, y)$.

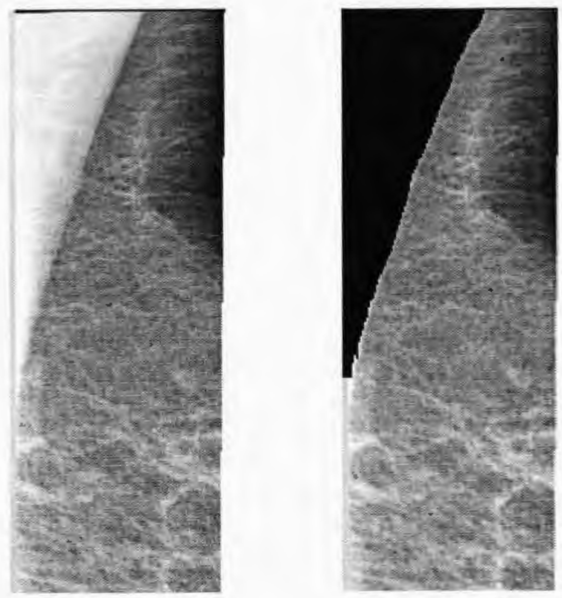
$$I_{MASK\_PM}(x, y) = \begin{cases} 0, & (x, y) \in \Delta\ P_1S_1\ S_2 \text{ and } \Delta\ P_1S_2S_3 \\ 1, & (x, y) \notin \Delta\ P_1S_1\ S_2 \text{ and } \Delta\ P_1S_2S_3 \end{cases}$$

3.3.5 Applying Mask and Finalizing Segmentation

Final segmented pectoral muscle $I_{SPM}(x, y)$ is obtained by applying the segmented pectoral muscle mask $I_{MASK\_PM}(x, y)$ on the enhanced segmented breast image $I_{SB}(x, y)$.

Final image $I_{SPM}(x, y)$ have aligned and segmented breast, enhanced contrast and suppressed pectoral muscle.

$$I_{SPM}(x,y) = I_{MASK\_PM}(x,y) * I_{SB}(x,y)$$



(a) (b)

Figure 36: (a) Pectoral Muscle ROI, (b) Suppressed Pectoral Muscle

---

Algorithm 3: Pectoral Muscle Suppression

Input: Segmented Breast Mammogram $I_{SB}(x,y)$

Output: Suppressed Pectoral Muscle Mammogram $I_{SPM}(x,y)$

1. **Extract** Region of Interest for Pectoral Muscle

   Determine points R_1, R_2 in image $I_{SB}(x,y)$

   $$R_1(x',y') = P(1, y_{max})$$

   $$R_2(x'',y'') = P(x, y_{max})$$

   Draw two perpendicular lines using R_1 and R_2.

   Quadrant 1 is Pectoral Muscle ROI.

2. **Compute** Seed Values S_1, S_2 and S_3.

   Draw lines $\overline{P_1P_2}$, $\overline{P_1P_3}$ and $\overline{P_1P_4}$

Track intensities on lines, determine max dip points S_1, S_2 and S_3.

3. **Form** $\Delta\, P_1S_1\, S_2$ and $\Delta\, P_1S_2S_3$

   Sketch lines $\overline{S_1S_2}$ and $\overline{\overline{S_2S_4}}$

4. **Evaluate** binary mask for pectoral muscle suppression $I_{MASK_{PM}}(x, y)$

$$I_{MASK_{PM}}(x, y) = \begin{cases} 0, & (x, y) \in \Delta\, P_1S_1\, S_2 \text{ and } \Delta\, P_1S_2S_3 \\ 1, & (x, y) \notin \Delta\, P_1S_1\, S_2 \text{ and } \Delta\, P_1S_2S_3 \end{cases}$$

5. **Compute** suppressed pectoral muscle $I_{SPM}(x, y)$

$$I_{SPM}(x, y) = I_{MASK\_PM}(x, y) * I_{SB}(x, y)$$

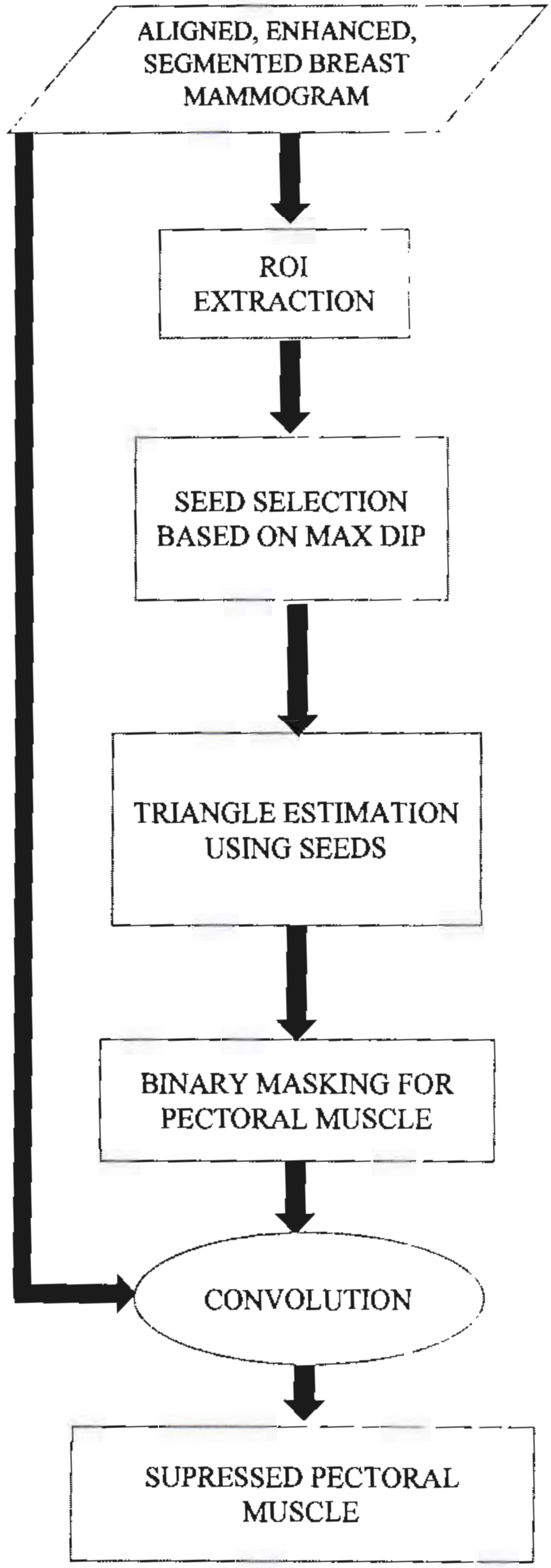


Figure 37: Pectoral Muscle Suppression Flow Diagram

Chapter 4 Results and Discussions

Results of our proposed algorithm have been presented in this chapter. For quantitative comparison and analyses the comparison has also been made with some methods and techniques discussed in chapter 2 that have used similar data set.

4.1 Selection of Database

The basis criteria for the selection of any database is that it should have sufficient number of images, easily available to researchers for comparison and further improvement of algorithm. For proposed algorithm testing, Mini MIAS database has been used. The reason of choosing this database is that majority of authors that are currently working in this field of breast cancer have used this database so for comparison with latest techniques, this database is the most suitable.

Results in research work are mostly manipulated by the preselection of images. In order to avoid such problem, mini MIAS database is used which contain 322 images and the implemented code can be run on the entire database without any biasness or pre selection of images to influence the results.

Hence for benchmarking algorithm performance, the obvious option is Mini MIAS database to test our proposed algorithm efficiency.

Never the less mini MIAS database contain scanned mammogram having many imperfections and noises which are very common to low cost mammograms. Hence, the algorithm which can work on this database can also be used for other database with reduce noises and other such problems with slight modifications in the algorithm.

On the other hand other, other database available of mammogram is Digital Database for Screening Mammography (DDSM). This database contain 2620 cases of size over 230 Gigabyte. The images are of high quality and resolution size varies from 12 bits to 16 bits. One of the major reasons of not using this database for mammogram image analyses is that the authors, that have used this database, have preselected some images from entire database which may result in biasness and hence comparison with other algorithm can easily be compromised.

4.2 Data Set for Results Generation

As discussed in previous section Mini MIAS database is used for mammogram image analyses in this study. This database contain 322 scanned mammogram images of 161 patients. Each patient right and left mammogram images are contained. Odd number in this database contain left breast mammogram and even number of this database contain right breast mammogram. Out of these 322, 208 are normal, 63 are benign and 51 are malignant or cancerous breast mammograms.

The size of the image is 1024 x 1024 pixel. Each pixel is 8 bit quantized and its pixel value varies from 0 to 255. The spatial resolution of mammogram is 200μm per pixel.

The database contain a mix of fatty tissue, fatty-glandular and dense glandular. Also benign and malignant cases made this database one of the most versatile databases available for research. All the images contain one or many of the artifacts Since the original mammogram are scanned and then digitalized so analog as well as digital type of noise exists in the images.

4.3 Simulation Environment

Software Used for the implementation of proposed algorithm was Matlab Student Version 8.1.0.604 (R2013a), Released on February 15, 2013.

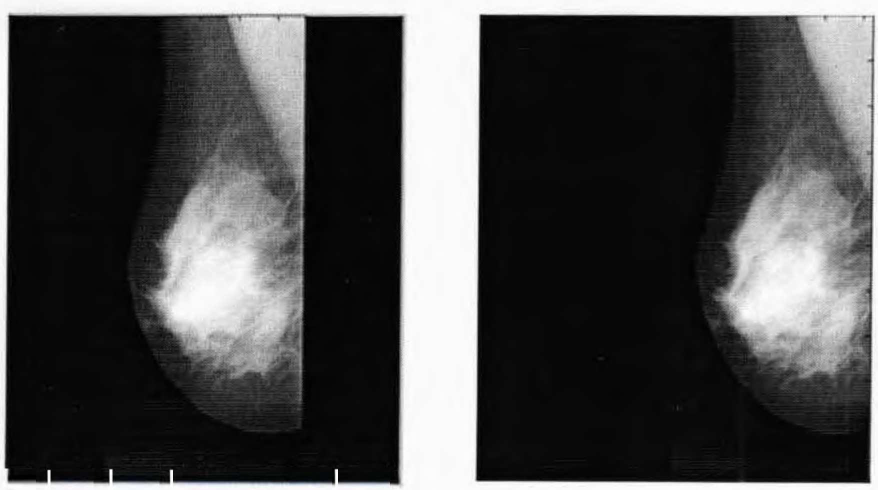
The code was run on Toshiba system having processor Intel Core I5, 2.5 GHz, RAM 6 GB and Microsoft Windows 7 Ultimate operating system.

4.4 Algorithm Result and Discussion

The Complete mini MIAS database of 322 images were taken as a test set and none of them is excluded. In addition to visual inspection, quantitative comparison was made by comparing radiologist drawn mask and the algorithm drawn boundary.

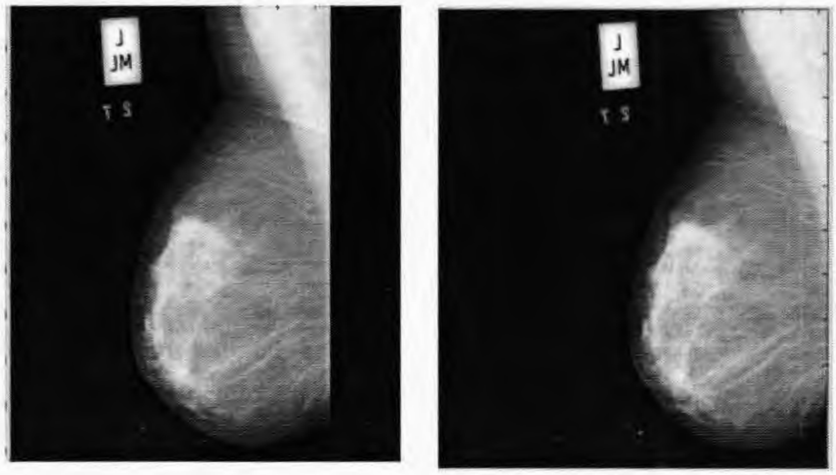
4.4.1 Alignment Results Comparison

The alignment results are of our proposed algorithm are 100% accurate and all 322 images of complete Mini MIAS database were perfectly aligned. Some of the alignment result are shown in this section.



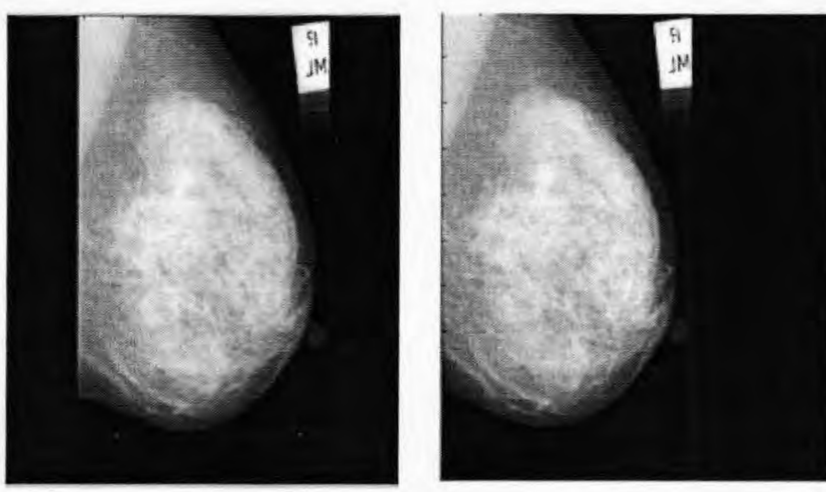
(a) (b)

Figure 38: (a) Mammogram mdb001, (b) After Alignment



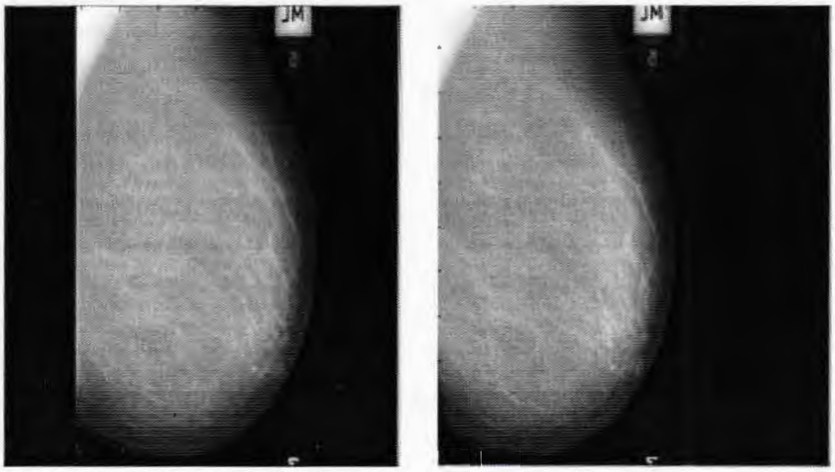
(a) **(b)**

Figure 39: (a) Mammogram mdb095, (b) After Alignment



(a) **(b)**

Figure 40: (a) Mammogram mdb126 (b) After Alignment



(a) (b)

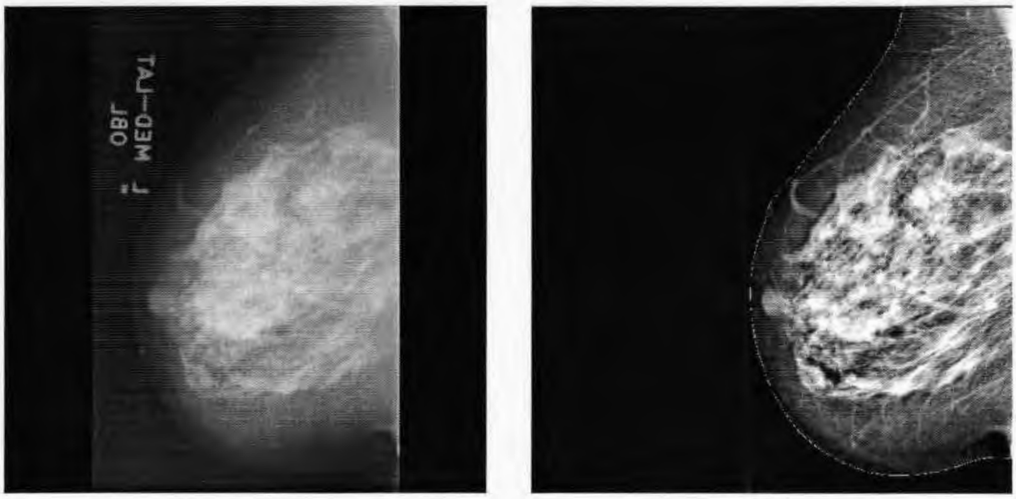
Figure 41: (a) Mammogram mdb310, (b) After Alignment

4.4.2 Breast Segmentation Results

For breast segmentation, as discussed in previous chapters, our goal was to removal of artifacts and breast boundary detection. On the basis of these two parameters we have classified our results into three major categories.

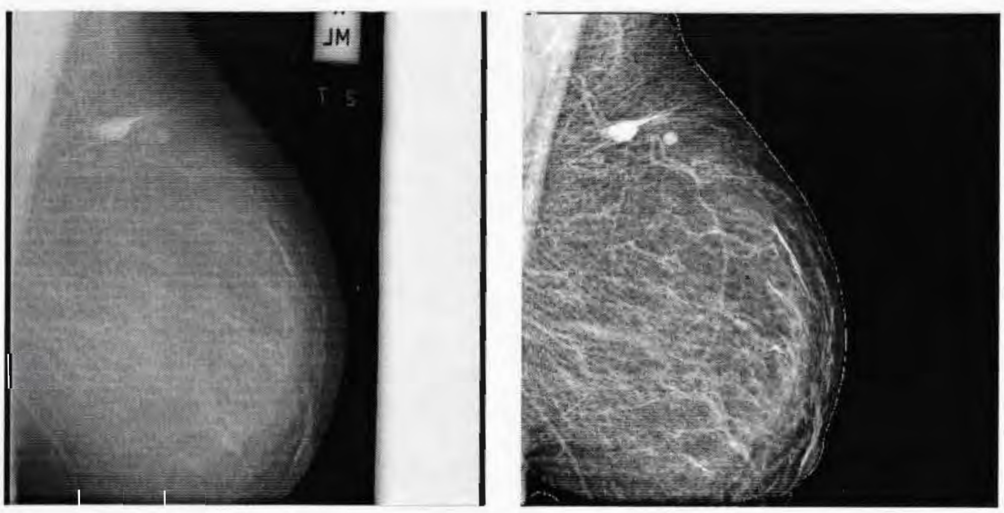
1. Successful category.
2. Acceptable category.
3. Unsuccessful and unacceptable category

1. **Successful category** included segmented breast that is free from the artifacts, does not contain any part of the artifact and also the detected breast line is in accordance with the ground truth.



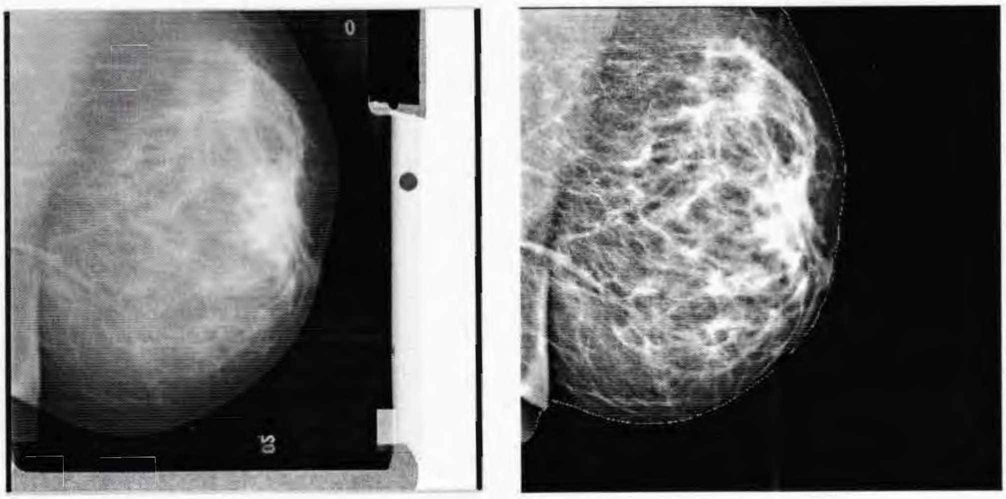
(a) (b)

Figure 42: (a) Mammogram mdb031, (b) Segmented Breast



(a) (b)

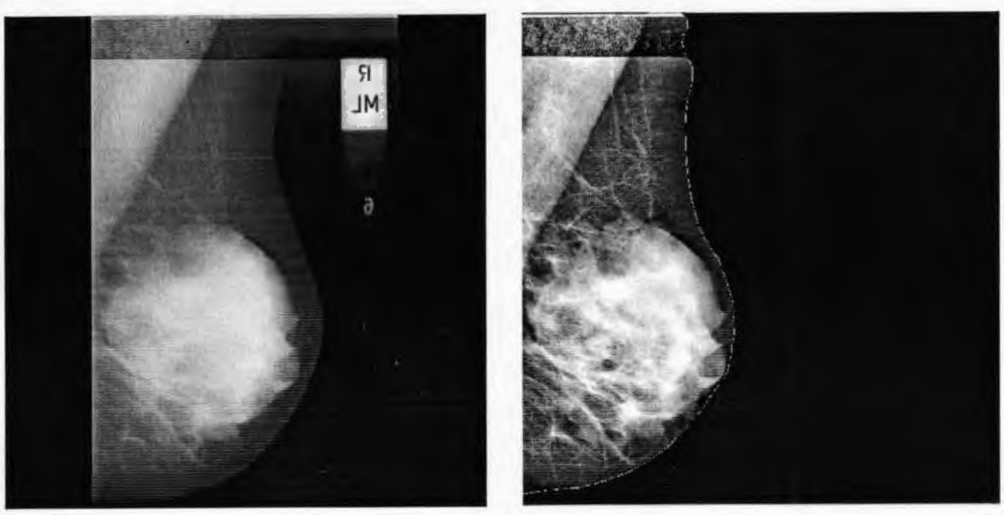
Figure 43: (a) Mammogram mdb132, (b) Segmented Breast



(a) (b)

Figure 44: (a) Mammogram mdb280, (b) Segmented Breast

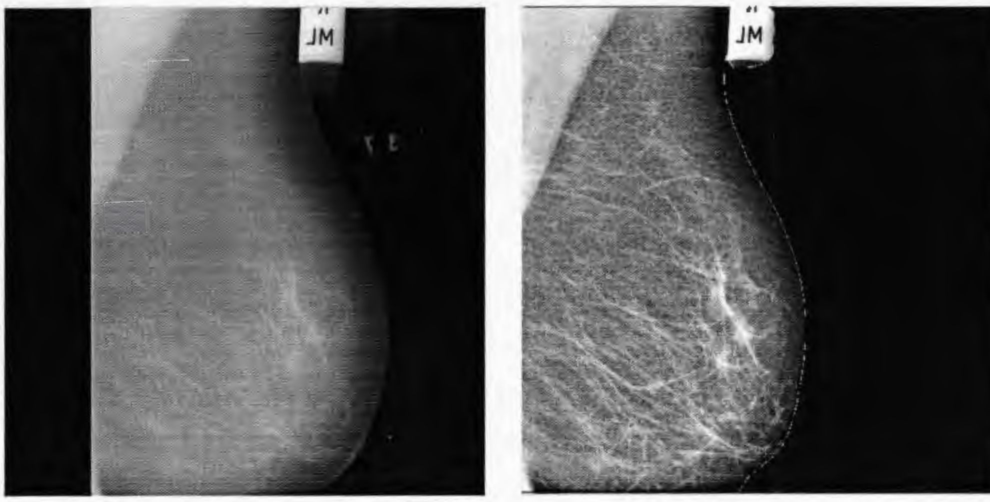
2. **Acceptable category** artifacts removed partially but part of low intensity beast pixels were cropped during breast segmentation but still segmented image can be used for further processing.



(a) (b)

Figure 45: (a) Mammogram mdb002, (b) Segmented Breast

3. Unsuccessful and unacceptable category artifacts are not removed or a major portion of breast is cropped during segmentation and segmented image cannot be used for further processing.



(a) **(b)**

Figure 46 :(a) Mammogram mdb006, (b) Segmented Breast

Table 5: Breast Segmentation Results

CATEGORY	**No of Images**	**Percentage**
SUCCESSFUL SEGMENTED	316	98.1
ACCEPTABLE SEGMENTED	03	0.9
UNSUCCESSFUL SEGMENTED	03	0.9

From Table.5, it is evident that our proposed algorithm works on 98.1% for the entire Mini MIAS database.

4.4.3 Results of Pectoral Muscle Segmentation

Likewise for pectoral muscle suppression, our objective was to remove pectoral muscle from the breast tissue. In mini MIAS database there are 322 images but pectoral muscle is visible in 318 images only. The four mammograms including mdb98, mdb137 mdb158 and mdb 236 are the one in which pectoral muscle is not visible. On the basis of visible inspection and comparison with the hand drawn mask for pectoral muscle, results were classified into two major categories.

1. Successful category.
2. Unsuccessful category

Successful category includes in which pectoral muscle is removed from the breast tissue. It also includes mammograms in which a small part of pectoral muscle is not removed but it will not affect segmentation and it implies no challenge for the cancer classification at the later stage.

Unsuccessful category includes in which algorithm fails to suppress pectoral muscle from its background.

Table 6: Pectoral Muscle Segmentation Results

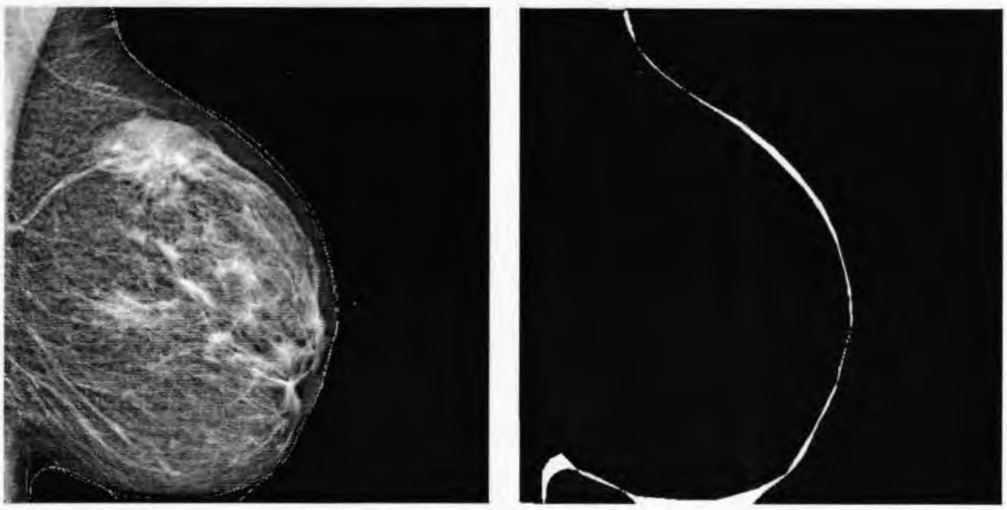
CATEGORY	No of Images	Percentage
SUCCESSFUL	280	88.1
UNSUCCESSFUL	48	11.9

In order to check our algorithm effectiveness, we quantify segmentation error. Segmentation error is the difference between the mask calculated by our proposed algorithm and the radiologist drawn mask for the same mammogram. Exclusive OR "XOR" logic is used for making this comparison. The truth table of "XOR" operation is represented in Table.7.

Table 7: Truth Table XOR

Radiologist Mask	Algorithm Mask	XOR Result
0	0	0
0	1	1
1	0	1
1	1	0

From Table.7 XOR truth table, if both the radiologist mask value and the proposed algorithm value are same, the result is zero and if there is a mismatch we get the "one". So number of ones in the end result image by "XOR" are in fact the "error" or "difference". The radiologist drawn boundary 'red' and algorithm drawn boundary 'green' are shown in Fig. 47(a).



(a) **(b)**

Figure 47: (a) Radiologist (Inner) Algorithm (Outer) Boundary mdb150, (b) Error

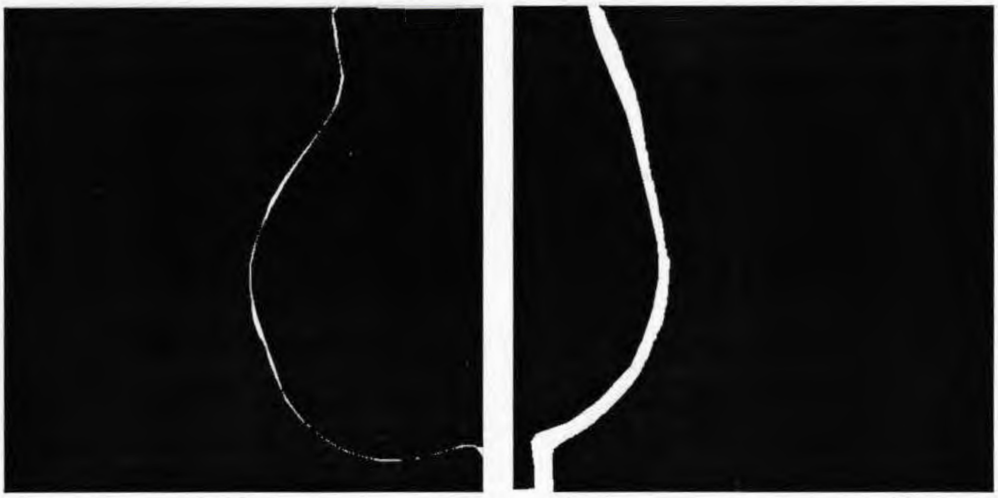
For percentage segmentation error, we used the below mathematical equation

$$\text{Segmentation Error} = \frac{\sum \text{Radiologist Mask} \oplus \text{Proposed Algorithm Mask}}{\sum \textit{Total Breast Area}} \times 100\%$$

Individual error for each of the image for the entire database of mini MIAS is calculated. Table.8. shows maximum and minimum error value in percentage.

Table 8: Error of Best and Worst Case and Average

Description	**Mammogram ID**	**Error Percentage**
Minimum Error	Mdb277	1.21
Maximum Error	Mdb318	14.9
Average Error	**3.68**	



(a) **(b)**

Figure 48: (a) Segmentation Best Case mdb277, (b) Worst Case mdb318

The error comparison for the entire Mini MIAS database of 322 images of our proposed algorithm with the gold standard available is provided in Fig. 49.

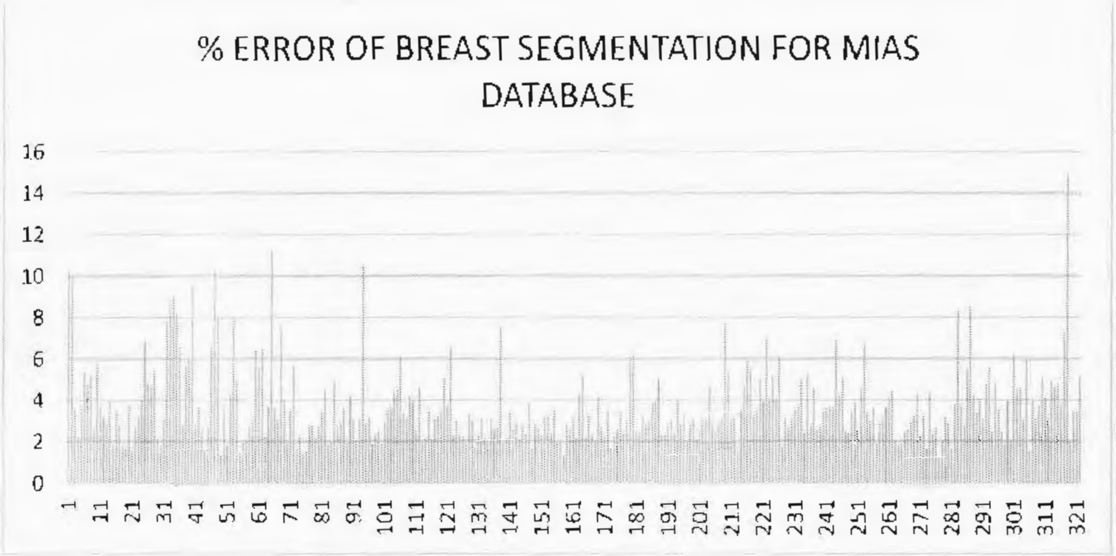


Figure 49: Percentage Segmentation Error for MIAS Database

Mario et al. [18] achieved 91% successful results. Maitra et al. [18] achieved 95.71% accuracy for the entire mini MIAS using CLAHE (Contrast Limited Adoptive Histogram Equalization) for contrast enhancement. Chen and Zwiggelaar achieved 66% accuracy but they have used only 240 out of 322 mammograms from the Mini MIAS database. Raba et al. [17] achieved 86% accurate results and their key was the use of selection of suitable threshold by the iterative or repetitive search which serve as a global threshold for that particular mammogram. The comparison of these with our proposed algorithm is provided in the Fig.50.

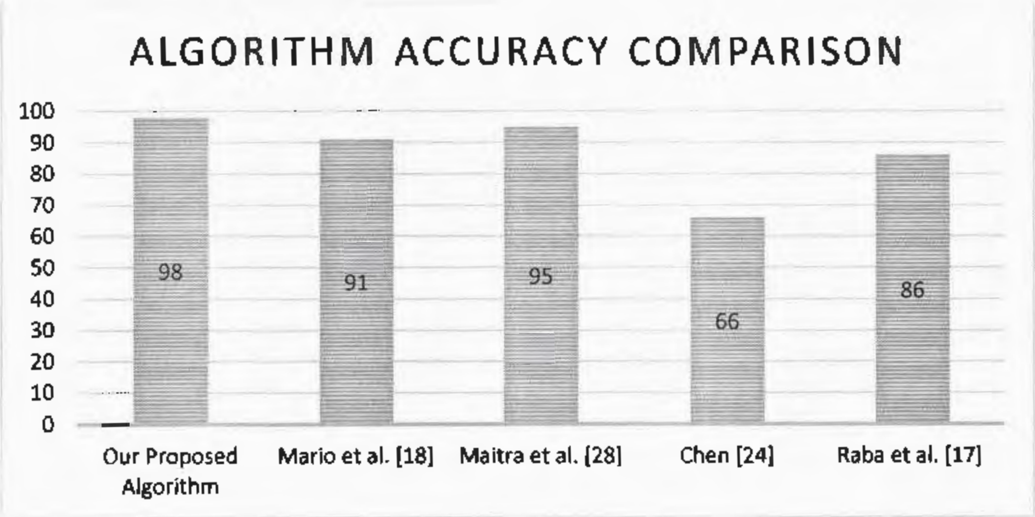


Figure 50: Algorithm Accuracy Comparison

Chapter 5 Conclusions and Future Work

In this concluding chapter, the main focus is on the final recap of the thesis. It also includes the major accomplishments and conclusion. In later part of this chapter, the, possible future developments were briefed.

5.1 Conclusions

In this thesis, we have proposed a hybrid algorithm which used a combination of Guided filter for image contrast enhancement, the proposed algorithm improves the segmentation results of the latest technique by 4%. Execution time of the algorithm is also few seconds for single input and is most suitable for real time CAD applications. Guided filtering technique for the very first time introduced in mammograms processing for image enhancement in our proposed algorithm which make it more efficient. Average segmentation error of our proposed algorithm is almost is 1.8% less than the latest available technique. However, there is still margin of improvement in our proposed algorithm.

5.2 Future Work

Selection of threshold, has a major impact on end results. For different mammograms, the values of optimized threshold is different. Hence in order to calculate optimized threshold, more research is required to be done to set the criteria for this threshold selection. Furthermore, feature extraction of breast and pectoral muscle from the mammogram for its segmentation need exploration.

For a future research, this algorithm should also be modified for other databases like DICOM and DDSM which because the lack of time, it has not been possible. Enhancing our new proposed technique and further refining it to give even better results will be also one of our future works.

References

[1] Lyon, Press Release 223, Geneva WHO, IARC (International Agency for Research on Cancer), DEC, 2013.

[2] John R, R. H.: "The Global Economic Cost of Cancer", American Cancer Society (ACS), 2010.

[3] SKMCH, "Collective Cancer Registry Report from Dec 1994 till Dec 2014", *Shaukat Khanum Memorial Cancer Hospital and Research Center*, 2015.

[4] SKMCH, "Information About Cancer", *Shaukat Khanum Memorial Cancer Hospital and Research Center*, 2014.

[5] C. A. Parise and V.Caggiano, "Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers", *Journal of Cancer Epidemiology*, 2014.

[6] "Breast Cancer Stage 0 & Stage1 ", *National Breast Cancer Foundation INC.*

[7] National Cancer Institute's SEER database. Available : http://seer.cancer.gov/

[8] "Mammograms Fact Sheet", *National Cancer Institute (NCI), USA*, 2015.

[9] S, D., Boetes, C., Burke, W., Harms, S., Leach, M. O., Lehman, C. D., & Russell, C. A.: "American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography", *A Cancer Journal for Clinicians*, Vol. 57, No.2, pp.75-89, 2007.

[10] Palmer, G. M., Zhu, C., Breslin, T. M., Xu, F., Gilchrist, K. W., & Ramanujam, N.: "Comparison of Multi-Excitation Fluorescence and Diffuse Reflectance

Spectroscopy for the diagnosis of Breast Cancer", *Biomedical Engineering, IEEE Transactions on*, Vol. *50*, No.11, pp.1233-1242, 2003.

[11] Smith, R. A., Cokkinides, V., & Eyre, H. J.: "American Cancer Society guidelines for the early detection of cancer", *A Cancer Journal for Clinicians*, Vol. 53, No.1, pp.27-43, 2003.

[12] Parkin, D. M., Bray, F., Ferlay, J., & Pisani, P.: "Estimating the world cancer burden", Globocon 2000. *International Journal of Cancer*, Vol.94, No.2, pp.153-156, 2001.

[13] Hoyer, A., Spiesberg, W.: "Computerized mammogram processing", Phillips Technical Review. Vol. 38, 1979.

[14] Bick, U., Giger, M.: "Automated segmentation of digitized mammograms", *Academic Radiology*, Vol 2, 1995.

[15] Hein, J., Kallargi, M.: "Multiresolution wavelet approach for separating the breast region from the background in high resolution digital mammography", Digital Mammography, Nijmegen, Kluwer, 1998.

[16] Masek, M., Attikiouzel, Y.: "Skin-air interface extraction from mammograms using an automatic local thresholding algorithm", *ICB, Brno*, 2000.

[17] D. Raba, A. Oliver, J. Martí, M. Peracaula, J. Espunya: "Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms", *Proceedings of the 2nd Iberian Conference, Pattern Recognition and Image Analysis*, Springer-Verlag, Hei delberg, 2005.

[18] Mustra, M., & Grgic, M.: "Robust automatic breast and pectoral muscle segmentation from scanned mammograms", *Signal processing*, Vol.93, No.10, pp.2817-2827, 2013.

[19] Semmlow, J., Shadagopappan, A.:" A Fully Automated System for Screening Xero-mammograms", *Computers and Biomedical Research*, Vol.13. 1980.

[20] M'endez, A., Tahoces, P.: "Automatic detection of breast border and nipple in digital mammograms", *Computer Methods and Programs in Biomedicine*. Vol. 49, 1996.

[21] Karssemeijer, N., Te Brake, and G.: "Combining single view features and asymmetry for detection of mass lesions", *IWDM*. 1998.

[22] Chandrasekhar, R., Attikiouzel and Y.: "Gross segmentation of mammograms using a polynomial model", *International Conference of the IEEE in Medicine and Biology Society*, Vol. 3. pp. 1056–1058, 1996.

[23] Goodsitt, M., Chan and H.: "Classification of compressed breast shapes for the design of equalization filters in X-ray mammography", *Medical Physics,* Vol. 25, pp. 937–947, 1998.

[24] Z. Chen and R. Zwiggelaar: "Segmentation of the Breast Region with Pectoral Muscle Removal in Mammograms", *Medical Image Understanding and Analysis (MIUA)*, pp. 71-76, 2010.

[25] Ojala, T., Liang and J.: "Interactive segmentation of the breast region from digitized mammograms with united snakes", *Technical Report 315, Turku Centre for Computer Science*, 1999.

[26] McLoughlin, K., Bones and P.: "Locating the breast-air boundary for a digital mammogram image", *Image and Vision Computing*, 2000.

[27] Wirth, M., Stapinski and A.: "Segmentation of the breast region in mammograms using snakes" *Canadian Conference on Computer and Robot Vision*, pp. 385–392, 2004.

[28] I.K. Maitra, S. Nag and S.K. Bandyopadhyay: "Technique for Preprocessing of Digital Mammogram", *Computer Methods and Programs in Biomedicine*, pp. 175–188, 2011.

[29] Ferrari, R., Rangayyan and R.: "Segmentation of mammograms: Identification of the skin boundary and the pectoral muscle", *IWDM*, Vol. 23, 2000.

[30] M. Mustra, J. Bozek and M. Grgic.: "Breast Border Extraction and Pectoral Muscle Detection Using Wavelet Decomposition", *Proceedings of the IEEE Eurocon 2009 conference*, pp. 1426-1433, 2009.

[31] He, K., Sun, J., & Tang, X.: "Guided image filtering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.35, No.6, pp. 1397-1409, 2013.

[32] Sun, Y., Suri, J., & Rangayyan, R, "A novel approach for breast skin-line estimation in mammograms, *18th Proceedings of IEEE Symposium on Computer-Based Medical Systems*, pp. 241-246, 2005.

[33] Sastre Tomàs, J.. "Segmentation of the breast region with pectoral muscle suppression and automatic breast density classification", 2011

[34] Meyer, F., Beucher, S., "Morphological segmentation", *Journal of visual communication and image representation*, Vol. 1, No. 1, pp. 21-46, 1990.

[35] Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K. Pizer, S. M., "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated speculations in dense mammograms", *Journal of Digital imaging*, Vol. 11, No 4, pp. 193-200, 1998.

[36] D Carvalho, I. M., Luz, L. M. S., Alvarenga, A. V., Infantosi, A. F. C., Pereira, W. C. A., & Azevedo, C. M, "An automatic method for delineating the pectoral muscle in mammograms", *Bioengineering Solutions for Latin America Health*, pp. 271, 2007.

[37] Yapa, R. D., & Harada, K., "Breast skin-line estimation and breast segmentation in mammograms using fast-marching method", *International Journal of Biological, Biomedical and Medical Sciences*, Vol.3, No.1, pp.54-62, 2008.

[38] Sultana, A., Ciuc, M., & Strungaru, R., "Detection of pectoral muscle in mammograms using a mean-shift segmentation approach", *8th International Conference on Communications*, pp. 165-168, 2010.