Pectoral Muscle Elimination on Mammogram Using K-Means Clustering Approach

Nashid Alam*, Mohammed J. Islam

Department of Computing Science and Engineering
Shahjalal University of Science and Technology, Sylhet, Bangladesh

Abstract

The performance of the Computer Aided Detection (CADe) system, to detect breast cancer, can be decreased due to some factors like presence of labels or other artifacts or pectoral muscle. Detection and segmentation of pectoral muscle can also help in image registration for further analysis of breast abnormalities such as bilateral asymmetry. In this paper, we proposed an algorithm based on K-means clustering to eliminate the triangular area of pectoral muscle. The reduction of irrelevant noise and unwanted artifacts are also performed using morphological preprocessing and seeded region growing (SRG) techniques. The method suggested for detection was tested over 322 images of 161 women taken from mini-MIAS database, out of which the proposed algorithm is able to eliminate pectoral muscle, showing 94.4% true positive value, from 291 images successfully. Results of pectoral muscle elimination are divided into three groups: Good (90.37%), Acceptable (8.07%) and Unexpected (1.5%).

Keywords: Breast Cancer, Digital Mammogram, X-ray label deduction, Pectoral muscle, Segmentation

1. INTRODUCTION

Pectoral muscle detection is necessary to limit the search for breast abnormalities only inside the breast region, and to reduce the image size without losing anatomic information, which improves the accuracy of the overall CADe system. As the line between the pectoral muscle and breast tissue in certain regions is not so obvious, the breast is needed to be segmented before the mammograms are subjected to analyze. Therefore, accurate segmentation of the pectoral muscle is essential to improve computer-aided diagnosis for mammography, which is already proven as the most effective and dependable technique for early detection of breast cancer. Medio-Lateral Oblique (MLO) view mammograms are characterized by high spatial resolution that allows identifying subtle scale signs such as microcalcifications, masses, tumors, lumps and lesions. The pectoral muscle appears approximately at same density as the dense tissues inside breast region on MLO view and cause problems by affecting the results of different image analysis methods. Intensity-based methods provide poor results when applied to differentiate dense structures such as

suspicious masses or fibro-glandular discs, as the pectoral muscle has a similar opacity as tumors, mass, microcalcifications. Thus exclusion of pectoral muscle from the mammogram with a minimum loss of breast tissues facilitates the search of abnormalities and modeling of parenchyma tissue. Sivakumar and Karnan [1] and Stephen et al. [2] shows that none of the existing imaging systems including MRI, ultra-sounds, nuclear medicine and thermal imaging can ensures the fine details which are obtained by conventional mammography. Craniocaudal (CC) is the other type of mammographic view, which has not been considered in this paper because the pectoral muscle is only seen in about 30%- 40% of CC images [3]. An adaptive algorithm is proposed by Kwok et al. [4] that use knowledge about the position and shape of the pectoral muscle on mediolateral oblique views. Mirzaalian et al. [5] presented an approach for pectoral muscle segmentation based on nonlinear diffusion algorithm. In order to segment the pectoral muscle region Aylward et al. [6] used gradient magnitude ridge traversal algorithm at a small scale and at multiple initial points. They resolved the resulting multiple edge definitions via a voting scheme. Ferrari et al. [7] presented a technique to detect the pectoral muscle based upon the Hough transform, which was a modification of the method proposed by Karssemeijer [8], but difference here from the

---
*Corresponding author
  Email addresses:* annanya_cse@yahoo.co.uk (Nashid Alam),
jahir-cse@sust.edu (Mohammed J. Islam)

Karssemeijer's method is the use of geometric and anatomical constraints in order to reduce the number of unlikely pectoral lines instead of using threshold value. Wirth et al. [9] and Timp [10] emphasized on suppressing image noises by performing image enhancement. Raba et al. [11] presented an automated technique for segmenting a digital mammogram into breast region and background with pectoral muscle suppression. They used region growing intensity threshold estimation to locate the pectoral muscle.

In this paper, we proposed a novel approach for pectoral muscle exclusion based on K-means clustering segmentation technique. This study focuses on the problems of eliminating pectoral muscles from the mammogram to confine further processing to the breast region alone. Moreover, subtraction of background from the image for facilitating further post- processing is done involving histogram equalization, contrast enhancement and Otsu's thresholding method with some image morphology.

2. PROPOSED METHOD

This study deals with pre-processing and enhancement activities such as removal of film artifacts and labels, filtering the image, normalization and removal of pectoral muscle region. Our proposed method consists of several image processing steps. In the first step, the images are identified as left or right and the film artifacts such as labels and X-ray marks are removed. Noise and artifacts are then removed by applying image morphology and image filtering techniques. Histogram equalization and contrast enhancement is done to get high contrast image and to prevent losing image details within the region of interest. After wards, thresholding technique is applied for image binarization. This binary image undergoes image morphology to find the biggest blob, which will contain the pectoral muscle, fatty tissues, ligaments, sinus, lobules and ducts of the original image. The biggest blob is kept and the rest are extracted from the original image. Some image morphologies are applied on the biggest blob region to prevent the loss of interesting image details. Contrast equalization technique is applied to get high contrast image with fine details. The pectoral muscle is then excluded from the X-ray image, using image morphology and K-means clustering algorithm [12]. The proposed method to detect and remove the pectoral muscle is summarized in Fig. 1.

K-means clustering algorithm is used to cluster different regions of the image. In our approach, a number (K=3) of clusters of the dataset contains a set of N entities, I, and M measurements, V, and a quantitative entity-to-feature, , such that represents the value of feature at entity . This algorithm attempts to optimize the clusters by minimizing the sum of the within-cluster distances of the N entities to their respective centroid where k=1, 2, ... K: it produces a partition, Each partition is non-overlapping, so each entity i is assigned to only one partition, a non-empty subset of a cluster, S.
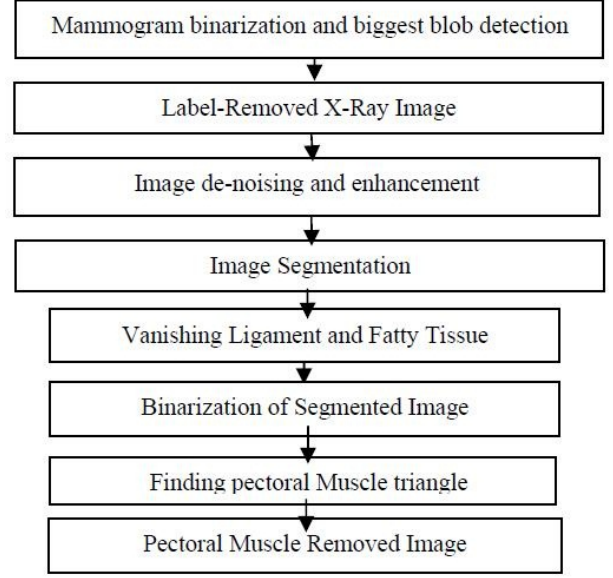

Figure 1: Block diagram for pectoral muscle removal

$$W(S,C) = \sum_{k=1}^{K} \sum_{i \in s_k} d(i, c_k) \qquad (1)$$

here the distance measure is represented by d, which is a squared Euclidean distance, and can be re-written as follows:

$$W(S,C) = \sum_{k=1}^{K} \sum_{i \in I} \sum_{V=1}^{M} S_{ik} (y_{iv} - c_{kv})^2 \qquad (2)$$

where S_{ik} is a Boolean variable representing the cluster membership of i to the cluster k. Formally, S_{ik} will be equal to one if S_{ik}. In clustering, any chosen score function will impact on the homogeneity of the clusters. K-means is guaranteed to converge, as demonstrated by Manning et al. [13], but the minimization of its score function is known to be NP-hard [14].

The fatty tissue and ligament area between the main breast counter and pectoral muscle are made completely black to separate the pectoral muscle. The segmented image is then binarized. To find the pectoral muscle pixel value intensity, scanning is started from the first row on the image, and continued until the first white seed point is found. The pixel value intensity scanning ends after getting five consecutive black points and the last white point in the first row of the image is stored. By following the same procedure the third point of the triangle is found. Joining these three points the triangle indicating pectoral muscle area in the mammogram is found. This triangle is flooded using the flood-filling approach and the flooded area of the triangle is made black on the original image. The removal of pectoral muscle results in less computational post-processing and analysis of the mammograms.
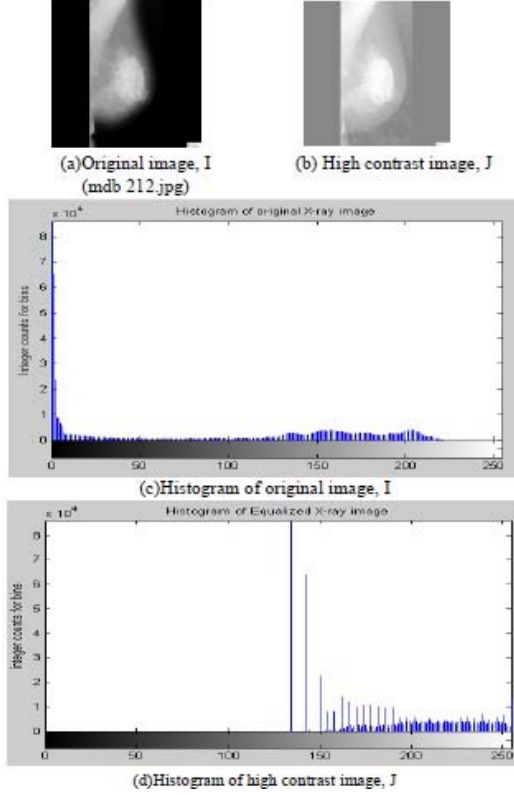
(a)Original image, I
(mdb 212.jpg)

(b) High contrast image, J

(c)Histogram of original image, I

(d)Histogram of high contrast image, J

Figure 2: Histogram equalization of mammogram X-ray

(a)After Histogram equalization of high contrast image (mdb212.jpg)

(b)Histogram equalization of high contrast image

Figure 3: Contrast enhancement of mammogram
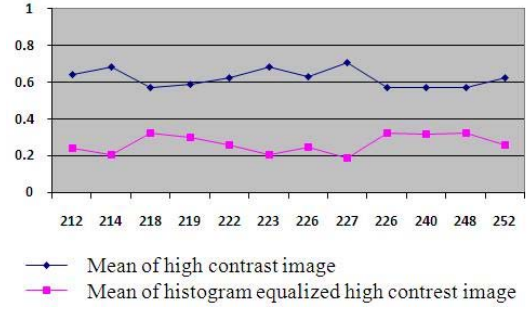
Figure 4: Quantitative measurement, decrease of mean values due to contrast enhancement

Mammograms preprocessing is done to accentuate certain features and to eliminate noise or error besides enhancing the appearance of X-ray image quality. Thresholding of pixels is performed for background correction, which standardizes the images by making use of an empty image. We use matlab function $histeq$ to enhance contrast (Fig. 2). Unnatural enhancement is observed in most of the image. This is a direct consequence of the excessive change in brightness by histogram equalization when the image has a high density over high gray levels.

Further image enhancement is done for attenuation and sharpening of image features such as edges and boundaries, as well as contrast to make the processed image more useful for analysis. This includes gray-level and contrast manipulation, noise reduction, background removal and filtering. We used $imadjust$ function on equalized intensity image for increasing the contrast of the image by mapping the values of the input intensity image to new values such that, by default, 1% of the data is saturated at low and high intensities of the input data (Fig. 3).The grey level normalization formula in equation 1 is used to bring the mammograms into a range that is more normal to the senses.

$$g(x,y) = g_{min} \frac{(g_{max} - g_{min})X(g_o(x.y) - g_{omin})}{(g_{omax} - g_{omin})} \qquad (3)$$

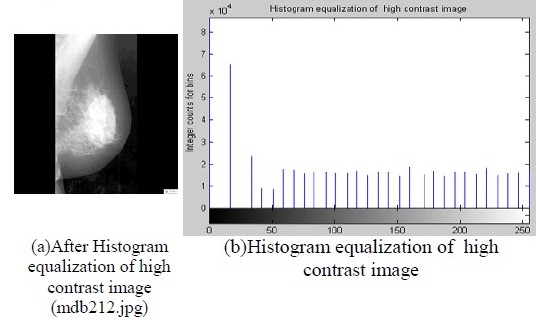where g_{omax} and g_{omin} are the maximum and minimum intensity level of original mammogram. Moreover,

g_{min} and g_{max} are the minimum and maximum intensity levels of the normalizes image, and $g_o(x,y)$ and $g(x,y)$ are the gray-levels of (x,y) coordinates before and after normalization.

It is obvious by looking the histogram Fig. 3(b), that $imadjust$ stretched the histogram of the mammogram, while histogram equalization spread it almost evenly. This is why the result of $imadjust$ looks more natural. The high contrast image has greater mean value then the histogram-equalized high contrast mammogram (Fig. 4). High contrast image has less variance then the histogram-equalized high contrast mammogram (Fig. 5).This will prevent the loosing of fine details from the mammogram X-ray image (Fig. 6).

3. RESULTS AND DISCUSSIONS

3.1. MIAS database

The proposed work is done using mini-MIAS Database, a digital mammography database, produced by an organization of UK research groups; The Mammography Image Analysis Society [15]. Mini-MIAS is composed by a set of 322 MLO digitized mammograms corresponding to left and right breast of 161 women, which include 208 normal, 63 benign and 51 malignant (cancerous) images. Every image is 1024 x 1024 pixels in size. For each image, experienced radiologists have given the type, location, scale, and other useful information. A read me file is included to the
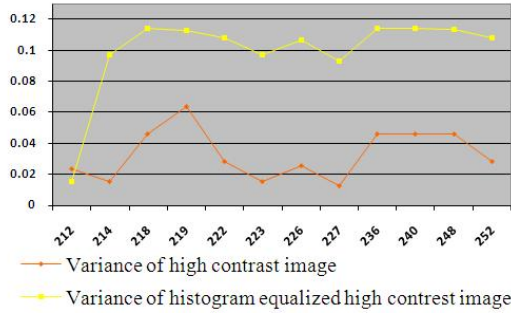
Figure 5: Quantitative measurement, increase of variance values due to contrast enhancement
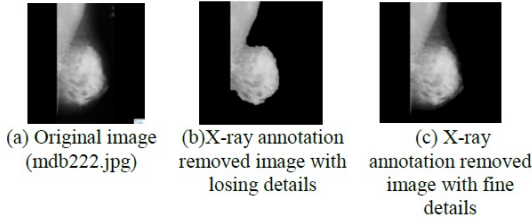

Figure 6: Prevention of image details losing

Figure 7: Effect of image binarization

database, which details the type of abnormality, whether it is a radial lesion, circumscribed mass, or microcalcification. The class of the abnormality i.e. benign or malignant and the location of the center of the abnormality and its diameter are also provided. The X-ray films are digitized with a Joyce-Lobel scanning microdensitometer to a spatial resolution of 50 micrometer and representing each pixel with an 8-bit gray depth.

3.2. X-ray annotation removing

The high and low intensity X-ray annotations and artifacts are removed as they can affect the quality of a segmentation process. To discard X-ray label Image binarization and biggest blob detection are done after performing histogram equalization and image contrast enhancement (Fig. 6). Otsu's *graythresh* Thresholding Method [16] is applied to binarize the image (Fig. 7(b)) by minimizing the intra class variance of the black and white pixels. Here, normalized intensity value is laid in the range [0, 1]. The function *graythresh* is used to get the effectiveness metric, EM that indicates the effectiveness of the thresholding of the input image. The lower bound is attainable only by images having a single gray level, and the upper bound is attainable only by two-valued images.

3.2.1. Biggest blob detection

A blob is a region of a mammogram in which some properties are constant or vary within a prescribed range of values; all the points in a blob can be considered similar. One main reason for using blob detection is to provide complementary information about regions, not obtained
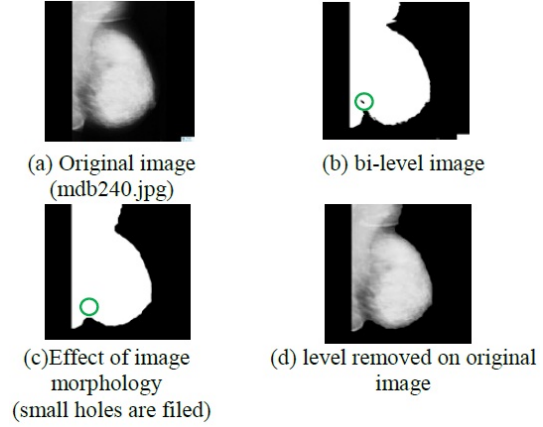
from edge detectors or corner detectors. These regions indicate the presence of objects in the image. The areas of all blobs exist in the mammogram are calculated and the largest blob is found by comparing all the blobs areas on the mammogram. The biggest blob contains the pectoral muscles, fatty tissues, lactiferous sinus, areolas, lactiferous ducts, breast lobules, coopers ligaments etc. The basic morphological operators used are erosion and dilation to avoid the loosing of image details. Narrow breaks are fused and small holes are filed by using disk shaped 18*18 structuring element musk, Fig. 7(c).The holes cannot be fully filled by using less than 18 sized structuring elements. Bigger structuring element will cause more computational cost without showing better result. In the end, the pixels in the largest blob are kept white and the rest of the pixels are made zero. Finally, in the originally mammogram, only the pixels having the same coordinate value are kept and all other pixels are made black in order to remove the x-ray labels, Fig. 7(d).The algorithm to find the blobs and to fill the holes inside the blob is described at Algorithm-01.

```
Algorithm-01:X-ray label removing by
largest blob detection
Input:  Original X-ray image
Begin
1.  Histogram equalization of the original
X-ray image
2.  Adjust image contrast
3.  Apply Otsu's Thresholding Method and
find bi-level the image which has several
blobs in it.
4.  Find the Largest blob and fill holes
inside the blob using image morphology
5.  Keep the true pixel value covering
only the area of largest blob and discard
other features from the original image.
6.  X-ray label is successfully removed
End
```
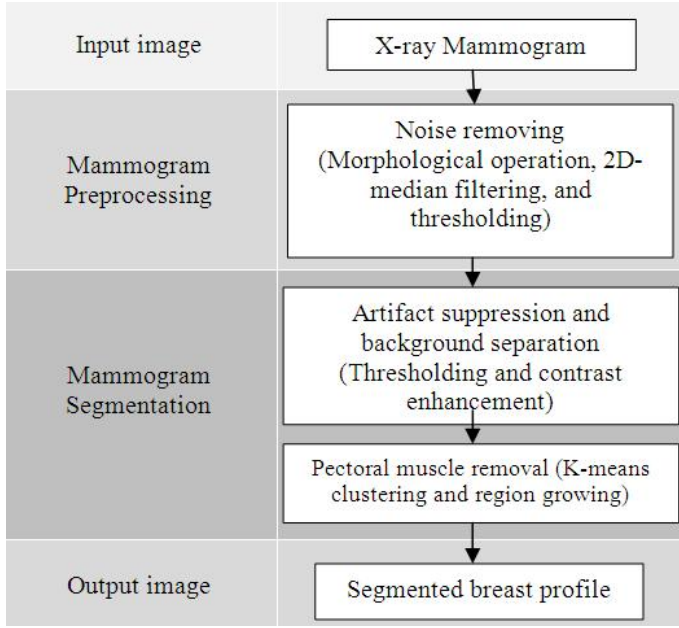
Figure 8: Breast profile segmentation framework

3.2.2. Breast Segmentation

Breast segmentation is used for assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Segmentation process is applied on mammograms to isolate pectoral muscle from the background by dividing the breast x-ray into distinct regions. The framework proposed for breast profile segmentation is indicated in Fig. 8. The result of image segmentation is a set of segments that collectively cover the entire mammogram. This will simplify and change the representation of X-ray image into something that is more meaningful and easier to analyze. A median filter of 5x5 is used to remove the noise without disturbing the edges, as structuring element less than 5x5 will not cause any change and structuring element bigger than 5x5 will cause the same change in the output. For each pixel a window of 5x5 neighborhood pixels are extracted, and the median value is calculated to smooth the local variations and to reduce noise [17].This procedure is done for all the pixels in the image to smooth the mammogram image [18]. This strengthens the relation between pixels in the same window and cause more strength layer that we will use this relation in these layers in segmentation of mammogram.

Thresholding technique is used to identify the pectoral muscle using Algorithm-02. Table-02 represents several thresholding values to detect the pectoral muscle in segmented image. Sometimes we can get false output in case of setting an improper thresholding value (Fig. 9).

```
Algorithm-02: Finding the edge of
pectoral muscle (Approach-01)
Input:  Label-removed x-ray image
Begin
1.Scan pixel value intensity at each points
```

```
2.Find out the sudden big intensity change
at the edge location
3.Mark the pixels at edge location
4.Estimate a straight line depending on the
marked edge points
End
```

The predefined threshold values, found based on image histogram, applied on benign and malignant images are given in the tabulated format, Table 1 and Table 2 respectively.

Table 1: Predefined threshold value to detect the edge of pectoral muscle (benign image)

Image No.	Lower Threshold	Upper Threshold
mdb212	150	200
mdb214	130	205
mdb218	150	210
mdb219	120	200
mdb222	150	210
mdb223	150	225
mdb226	110	210
mdb227	150	230
mdb236	160	210
mdb240	150	200
mdb248	150	210
mdb252	140	210

Table 2: Predefined threshold value to detect the edge of pectoral muscle (malignant image)

Image No.	Lower Threshold	Upper Threshold
mdb209	140	210
mdb211	140	210
mdb213	140	210
mdb216	140	210
mdb231	140	210
mdb233	140	210
mdb238	140	210
mdb239	140	210
mdb241	140	210
mdb245	140	210
mdb249	140	210
mdb253	140	210
mdb254	140	210
mdb256	140	210

But the problems faced in the above Algorithm-02 are:
i. Finding appropriate thresholding value that will work on every image.
ii. The threshold value must be found in an unsupervised manner
iii. Any predefined threshold value will not produce desired output for all images, Fig. 9.

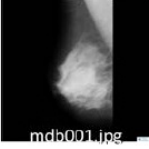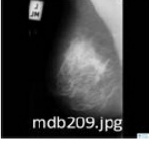A region-based segmentation approach automatically
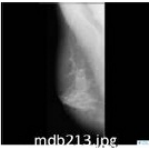
Main image	Output Image	Image containing	Missing part
mdb001.jpg		Duct, Lobules, Sinus *Predefined threshold value* lower 140 Upper 210	Fatty tissue area, Ligaments
mdb209.jpg		X-ray Labels *Predefined threshold value* lower 140 Upper 210	Fatty tissue area, Duct, Lobules, Sinus, ligaments
mdb213.jpg		Pectoral muscle *Predefined threshold value* lower 140 Upper 210	Fatty tissue area, Duct, Lobules, Sinus, ligaments

Figure 9: Output of using improper pre-defined thresholding value

clusters the pixels according to their color information using the k-means algorithm, where these clusters are assigned to either 'region' or 'non-region' class and the textural features of pixels are extracted using auto-correlation function [19]. Algorithm-03 is used to find the edge of pectoral muscle using K-means segmentation.

```
Algorithm-03: Finding the edge of
pectoral muscle (Approach-02)
Input: Label-removed x-ray image
Begin
1.  Segment the image using Algorithm- 01
2.  Separate the pectoral muscle from the
Duct, Lobules, Sinus region
3.  Make all the pixels black (zero) resides
in the fatty tissue and ligament area.
3.  Find the binary image of image found
in step 2(it will be used as outer image)
4.  Erode the image found in step-3 (it
will be used as inner image)
5.  Subtract the inner image from the
outer image to get the edge
End
```

The steps used in Algorithm-03 are depicted in Fig. 10.

But the problems faced in Algorithm-03 are:

i. Pectoral muscle and ligaments in fatty tissue area merged, Fig. 11.

ii. Discontinuity in Pectoral muscle edge, Fig. 12.

iii. Same thresholding value (130-210) does not work well on all the images and produces improper output (complete black image as output), Fig. 13.

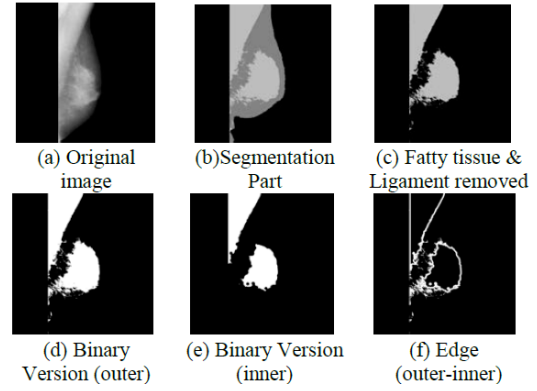The pectoral muscle and the area covered with fatty tissue and ligaments merge as depicted in Fig. 14(c).This

(a) Original image   (b)Segmentation Part   (c) Fatty tissue & Ligament removed

(d) Binary Version (outer)   (e) Binary Version (inner)   (f) Edge (outer-inner)

Figure 10: The steps followed to extract the region of interest (Approach-02)

1. Original image (mdb240.jpg)   2. Segmentation Part   3. Fatty tissue & Ligament removed

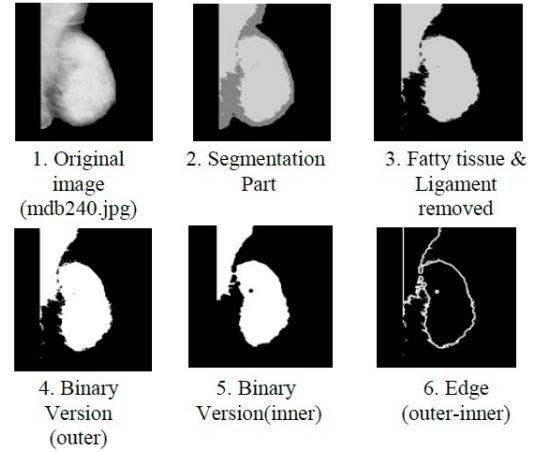4. Binary Version (outer)   5. Binary Version(inner)   6. Edge (outer-inner)

Figure 11: Pectoral muscle and ligaments in fatty tissue area merge

problem is resolved by using morphological analysis on images, Fig. 15(c).

This operation works very well for separating pectoral muscle from the main breast counters. The method requires no training data or prior knowledge of the image contents. However, this method alone is not suitable for images with more complex regions, Fig. 16.

3.3. K-means Clustering

K-means clustering algorithm was used to separate pectoral muscle from the main breast region. The mammogram is divided into 3 main clusters, one containing pectoral muscle, another containing fatty tissues, and ligaments, and the rest containing lactiferous ducts, sinus and breast lobules etc. (Fig. 17(b)). The region containing the fatty tissue and ligaments is deducted from the clustered image to separate the pectoral muscle region (Fig. 17(c)). K-Means unsupervised clustering algorithm [12] is used to classify the input data points into multiple classes based on their distance from each other. This algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are
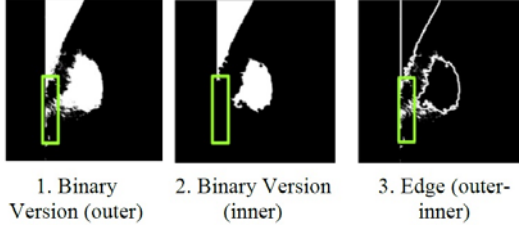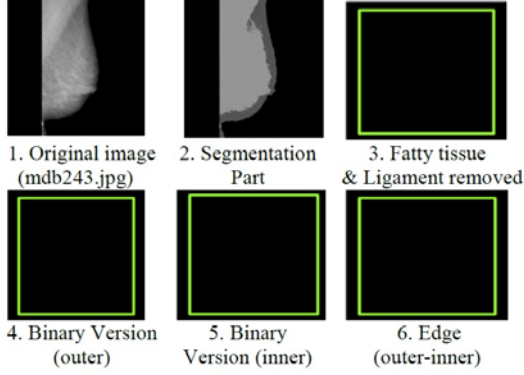
Figure 12: Discontinuity in Pectoral muscle edge

Figure 13: Thresholding value (130-210) does not work well on all the images(produces completely black image)

clustered around centroids $\mu, \forall = 1..k$, which are obtained by minimizing the objective:

$$v = \sum_{i=1}^{k} \sum_{x_j \in s_i} (x_j - \mu_i)^2 \qquad (4)$$

Where there are k clusters $s_i, i = 1, 2..k$ and μ_i is the centroids or mean point of all the points $x_j \in s_i$. As a part of this project, an iterative version of the algorithm was implemented. The algorithm takes a two-dimensional image as input. Steps in the algorithm are:
1. Compute the intensity distribution (also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster labels of the image do not change any more.
4. Cluster the points based on distance of their intensities from the centroid intensities.
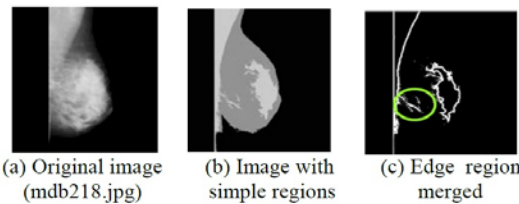

Figure 14: Edge of adjacent region merged
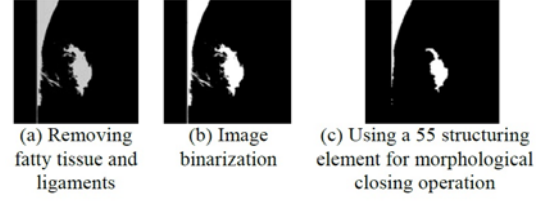
Figure 15: Separation of pectoral muscle using image morphology
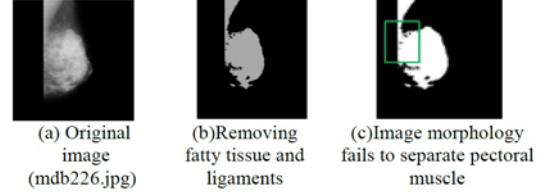
Figure 16: Complex scenarios where image morphology fails to separate pectoral muscle

$$c^i = arg_j min \left\| x^{(i)} - \mu_j \right\|^2 \qquad (5)$$

5. Compute the new centroids for each of the clusters.

$$\mu_i = \frac{\sum_{i=1}^{k} 1\{c(_j)^j\} x^{(i)}}{\sum_{i=1}^{k} 1\{c(_i) = j\}} \qquad (6)$$

Where k is a parameter of the algorithm (the number of clusters to be found), i iterates over the all the intensities, j iterates over all the centroids and μ_i are the centroids intensities.
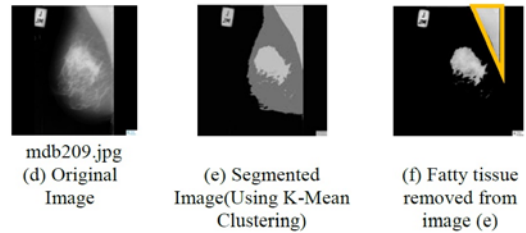
Figure 17: Applying K-means Clustering Algorithm to separate pectoral muscle

3.4. Triangle detection of pectoral muscle

From the k-means clustering segmentation approach, it becomes clear that the pectoral muscle is a triangular area, Fig. 18(b).

Based on the fact that the pectoral muscle is a triangular area we moved on an approach to detect the triangle and make the triangle portion black on the main X-ray image to discard the pectoral muscle region. The Fig. 19

mdb212.jpg   mdb214.jpg   mdb209.jpg
a) Original image

mdb212.jpg   mdb214.jpg   mdb209.jpg
b) Triangular area of muscle

Figure 18: Separation of triangular pectoral muscle area

Figure 19: Steps followed to detect and remove the triangular area of pectoral muscle.

shows the steps followed to detect and remove the triangular area of pectoral muscle.

Algorithm-04 is used to find the triangular region of pectoral muscle. We have used 8-neighborhood connected component labeling method for breast region extraction and removal of pectoral muscle. Region growing starts with a set of ẅeedṗoints and region grows by appending to each seed those neighboring pixels that have properties similar to the seed.

```
Algorithm-04: Finding the triangular
region of pectoral muscle
Input:  Label-removed X-ray image
Begin
1.  Contrast stretching of label-removed
x-ray image
2.  Binarize the contrast stretched image
3.  Find white seeding point
4.  Find the 1st black point of 1st row
after getting a white seeding point
5.  Draw a horizontal line between these
two points.
6.  Find the 1st black point of 1st
column after getting a white seeding point
```
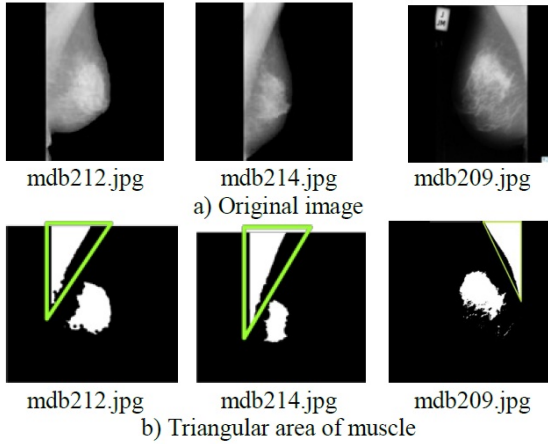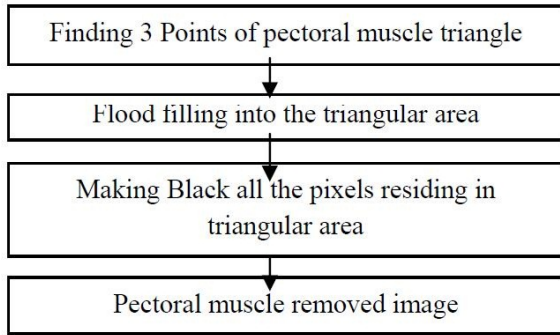
```
7.  Draw a vertical line and angular line.
8.  Make all the pixels black (zero) in
the pectoral muscle area
End
```

The result of Algorithm-04 is depicted in Fig. 20.

1. Original image   2. Contrast   3. Binary of
                        stretching   contrast image

4. Triangle   5. Triangle Filled   6. Muscle removed
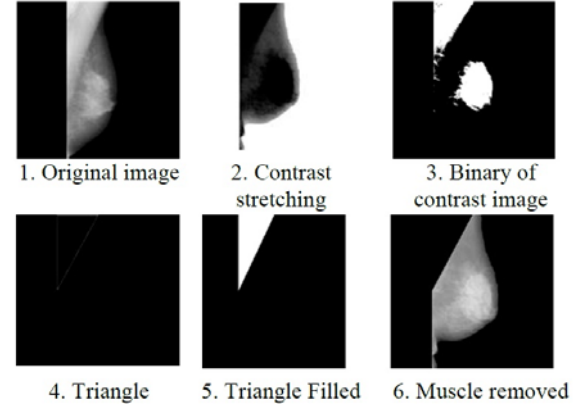
Figure 20: Pectoral muscle removing steps

But the problems faced in removing pectoral muscle are:
i. The triangle does not always indicate the proper pectoral muscle area duo to the discontinuity in edges in first 3 or 4 rows and columns, Fig. 21(a).This situation can be handled by replicating the first 4 rows from the top of the mammogram Fig. 21(b).
ii. In case of flood filling, leakage in angular line cause false output, Fig. 22(a).This problem is solved by creating a thick angular line while creating the triangle Fig. 22(b).

1. Triangle   2.Triangle Filled   3.Muscle partially
                                         removed
a. Before replicating

1. Triangle after   2.Triangle Filled   Muscle fully
replicating   removed
b. After replicating

Figure 21: Removing defects of the mammogram by replication rows and columns for removing pectoral muscle (mdb222.jpg)

Table 3 shows the time to eliminate pectoral muscle from the mammogram.

1.Triangle   2.whole image is   3.False output
(thin angular line) flood-filled
a. Lick age in angular line cause false output

1.Triangle after   2.Triangle Filled   3.Muscle fully
replication(thick   removed
angular line)   (mdb222.jpg)

Figure 22: Leakage in angular line cause false output

Table 4: Result of pectoral muscle detection

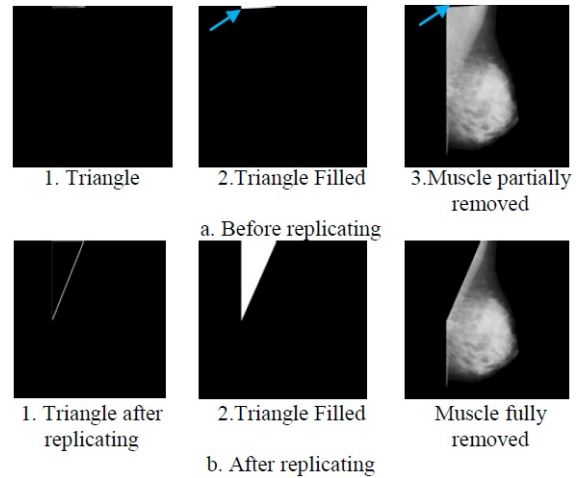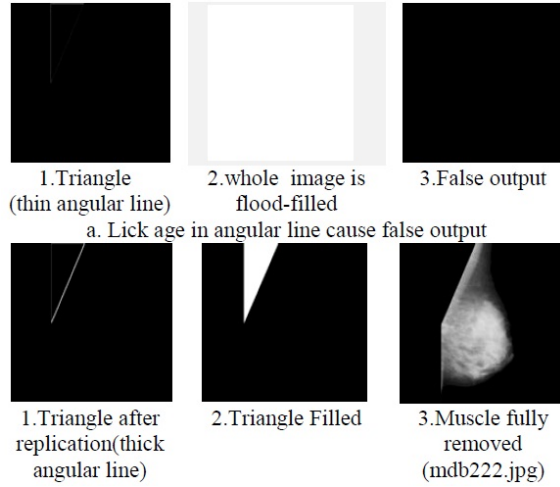|	Rate (%)
True positive	94
True negative	0
False positive	5
False negative	1

Figure 23: Pectoral muscle removing

Table 3: Time to eliminate pectoral muscle by performing three vertices detection of the triangle, row and column replication and applying seeded region growing method.

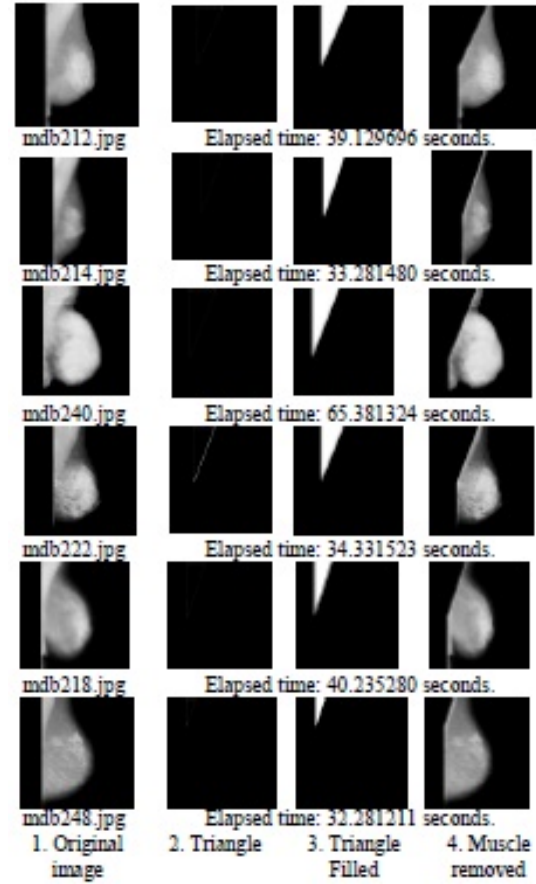| Type of abnormality	Image No.	Elapsed Time (Seconds)
	mdb212	39.129696
	mdb214	33.281480
	mdb218	40.235280
	mdb219	31.281223
	mdb222	32.249163
Benign Image	mdb223	53.357321
	mdb226	32.241184
	mdb227	52.975222
	mdb236	28.311229
	mdb240	65.381324
	mdb248	32.281211
	mdb252	45.925482
	mdb209	34.893824
	mdb211	29.927583
	mdb213	30.759374
	mdb216	45.745826
	mdb231	44.946283
	mdb233	45.828972
	mdb238	46.947010
Malignant Image	mdb239	32.938303
	mdb241	36.153729
	mdb245	48.383173
	mdb249	41.973910
	mdb253	53.925852
	mdb254	43.927183
	mdb256	38.098381

The experimental results showed a strong indication of pectoral muscle removal, Fig. 23. For the breast background segmentation 70.5% are accurate, 13% are nearly accurate, 12.9% are acceptable, and 3.6% are unacceptable. For the pectoral muscle removal, we have obtained 68.9% accurate, 18.4% nearly accurate, 7.2% acceptable, and 5.5% unacceptable segmentations. In some cases (1.6% of the breast background segmentations and 6.5% of the pectoral muscle segmentations are classified as unacceptable) the method does not obtain what could be considered an acceptable segmentation. For the breast region segmentation, those are mainly related to the extremely low contrast between the breast tissue near the boundary and the background region resulting in an inaccurate binary mask. Furthermore, a significant amount of noise in the image leading to a poor placement of the initial seed

Table 5: Comparative analysis of pectoral muscle removal

Author reference	Methods	Number of Images used for experiments	Good/Acceptable (%)
Mustara et al.[20]	Hybrid with Bit depth reduction and wavelet decomposition	40	34(85%)
Li Liu et al.[21]	AD method	100	81(81%)
Raba et al.[11]	Adaptive histogram approach	322	278(86%)
Chen et al.[22]	Histogram thresholding, edge detection in scale space, contour growing and polynomial fitting	86	80(93.5%)
Boss et al.[23]	Histogram based 8-neighborhood connected component labeling method	322	288(89.5%)
Proposed method	K-means clustering based method	322	291(90.3%)

point of contour growing and the non-uniform breast intensity distribution yield under-segmented results. For the 322 mammograms evaluated, the mean values of accuracy and error are 0.894 and 0.106 respectively.

Bick et al. [24] tested their algorithm on 740 digitised mammograms, and 97% of the segmentation results were visually rated as acceptable. Mndez et al. [25] tested their algorithm on 156 digitised mammograms, and segmentation results were deemed to be accurate or nearly accurate in 89% of the mammograms. The method presented by Chandrasekhar and Attikiouzel [26] was tested on all the images from the MIAS database, and it provided about 94% acceptable segmentation results that is shown in Table 4. We compared the results with previous studies where similar visual evaluation criteria were used, Table 5.

4. CONCLUSIONS

This paper presents a new and efficient method for pectoral muscle detection in MLO view mammograms by combining k-means segmentation with some pre-processing parameters calculated from the regions obtained in the segmentation step. These measures are based on localization of pectoral muscle in mammograms and the increased gray-level value of pectoral muscle. A set of preprocessing steps including image morphology, contrast enhancement and noise filtering techniques have been applied to discard X-ray labels. The preprocessing segmentation processes reduce noise and edge-shadowing effect, accurately detect region of interest (ROI) for pectoral muscle. The objective of this work is to reduce the high computational cost or further processing of mammograms by discarding the triangular pectoral muscle and labels, as breast cancer symptoms can never be existed in these areas. The performance was evaluated using a set of 322 MLO digitized mammograms with 51 cases of pathologically proven malignant, 208 normal and 63 benign images of Mini-MIAS database. The results obtained over Mini-MIAS database have shown generally good behavior. From the computational point of view, the detection step is most time consuming. Running time of the program is about 40 sec-

onds on Intel Core 2 Solo 1.40GHz CPU with 2.96GB.The computational cost was acceptable since the detection rate were good. The success rate of our proposed algorithm indicates 94.4% true positive, 5% false positive and 1% false negative results. The implementation and qualitative evaluation of experimental results was carried out using MATLAB 7.12.0.635 (R2011a) on Windows XP Version 5.1.The processed mammogram can be used for the automated abnormalities detection of human breast like calcification, circumscribed masses, speculated masses and other ill-defined masses speculated, circumscribed lesions, asymmetry analysis etc.

References

[1] R. Sivakumar, K. Karnan, Diagnose breast cancer through mammograms using eabco algorithm, International Journal on Artificial Intelligence and Machine Learning 4 (2005) 302–307.

[2] F. S. Stephen, D. J. Winchester, D. P. Winchester, E. Barrera, M. Bilimoria, E. Brinkmann, E. Alwawi, S. Rabbitt, D. H. Schermerhorn, Survival rates for breast cancers detected in a community service screening mammogram program, The American Journal of Surgery 191 (3) (2006) 406–409.

[3] G. W. Eklund, G. Cardenosa, W. Parsons, Assessing adequacy of mammographic image quality, Radiological Society of North America 190 (2) (1994) 297–307.

[4] S. M. Kwok, R. Chandrasekhar, Y. Attikiouzel, M. T. Rickard, Automatic pectoral muscle segmentation on mediolateral oblique view mammograms, IEEE Transactions on Medical Imaging 23 (9) (2004) 1129–1140.

[5] H. Mirzaalian, M. R. Ahmadzadeh, S. Sadri, Pectoral muscle segmentation on digital mammograms by nonlinear diffusion filtering, Proceedings of the IEEE Pacific Rim Conference on Communications Computers and Signal Processing, 2007, pp. 581–584.

[6] S. R. Aylward, B. M. Hemminger, E. D. Pisano, Mixture modeling for digital mammogram display and analysis, Digital Mammography, Computational Imaging and Vision 13 (1) (1998) 305–312.

[7] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, A. F. Frre, Segmentation of mammograms: Identification of the skin - air boundary, pectoral muscle, and fibroglandular disc, in: Proceedings of the 5th International Workshop on Digital Mammography, Medical Physics Publishing, 2000, pp. 573–579.

[8] N. Karssemeijer, Automated classification of parenchymal patterns in mammograms, Physics in Medicine and Biology 43 (2) (1998) 365–378.

[9] M. Wirth, D. Nikitenko, J. Lyon, Segmentation of the breast region in mammograms using a rule-based fuzzy reasoning algorithm, International Journal on Graphics, Vision and Image Processing 5 (2) (2005) 387–96.

[10] S. Timp, Analysis of temporal mammogram pairs to detect and characterise mass lesions, Ph.D. dissertation, Radboud University Nijmegen, Netherlands (2006) 315–323.

[11] D. Raba, A. Oliver, J. Mart, M. Peracaula, J. Espunya, Breast segmentation with pectoral muscle suppression on digital mammograms, in: Proceedings of the 2nd Iberian Conference on Pattern Recognition and Image Analysis, Springer-Verlag, 2005, pp. 471–478.

[12] J. A. Hartigan, M. A. Wong, A k-means clustering algorithm. journal of the royal statistical society, Blackwell Publishing 28 (1) (1979) 100–108.

[13] C. D. Manning, P. Raghavan, H. Schutze, Introduction to information retrieval.

[14] P. Drineas, A. Frieze, R. Kannan, S. Vempala, , V. Vinay, Clustering large graphs via the singular value decomposition, Machine learning 56 (2) (1999) 1–3.

[15] J. Suckling, J. Parker, D. R. Dance, S. Astley, I. Hutt, C. R. M. Boggis, I. Ricketts, E. Stamatakis, N. Cernaez, S. L. Kok, P. Taylor, D. Betal, J. Savage, The mammographic image analysis society digital mammogram database, in: Proceedings of the 2nd International Workshop on Digital Mammography, York, England, 10-12 July 1994, Elsevier Science, Amsterdam, The Netherlands, 1994.

[16] N. Otsu, A threshold selection method from gray-level histogram, IEEE Transactions on Systems, Man, And Cybernetics 9 (1) (1979) 62–66.

[17] R. C. Gonzlez, R. E. Woods, Digital image processing, New Jersey: Pearson Prentice Hall, 2008.

[18] K. Thangavel, M. Karnan, Computer aided diagnosis in digital mammograms: Detection of microcalcifications by meta heuristic algorithms, International Journal on Graphics, Vision and Image Processing 5 (7) (2005) 41–45.

[19] O. D. Faugeras, W. K. Pratt, Decorrelation methods of texture feature extraction, IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (1) (1980) 323–332.

[20] M. Mustra, J. Bozek, M. Grgic, Breast border extraction and pectoral muscle detection using wavelet decomposition, EUROCON, IEEE, St. Petersburg (2009) 1426–1433.

[21] L. Liu, J. Wang, T. Wang, Breast and pectoral muscle contours detection based on goodness of fit measure, in: IEEE 15th internasnal conference on bioinformatics and bio-medical engineering, 2011, pp. 1–4.

[22] Z. Chen, R. Zwiggelaar, Segmentation of the breast region with pectoral muscle removal in mammograms. medical image understanding and analysis (miua), The University of Warwick, Coventry, UK (2010) 71–76.

[23] R. S. C. Bose, T. Tangaval, D. A. P. Daniel, Automatic mammogram image breast region extraction and removal of pectoral muscle, International Journal of Scientific Engineering Research 4 (5) (2013) 1–8.

[24] U. Bick, M. L. Giger, Automated segmentation of digitized mammograms, Academic Radiology 2 (1) (1995) 1–9.

[25] A. Z. Mndez, Automatic detection of breast border and nipple in digital mammograms. computer methods and programs in biomedicine, IEEE Transactions on Medical Imaging 49 (3) (1996) 253–262.

[26] R. Chandrasekhar, Y. Attikiouzel, Automatic breast border segmentation by background modeling and subtraction, In IWDM (2000) 560–565.