Technique for preprocessing of digital mammogram

Indra Kanta Maitra, Sanjay Nag, Samir Kumar Bandyopadhyay*
Dept. of Computer Science & Engg., University of Calcutta, 92 A.P.C. Road, Kolkata 700009, West Bengal, India

Abstract

Digital mammogram has emerged as the most popular screening technique for early detection of breast cancer and other abnormalities in human breast tissue. It provides us opportunities to develop algorithms for computer aided detection (CAD). In this paper we have proposed three distinct steps. The initial step involves contrast enhancement by using the contrast limited adaptive histogram equalization (CLAHE) technique. Then define the rectangle to isolate the pectoral muscle from the region of interest (ROI) and finally suppress the pectoral muscle using our proposed modified seeded region growing (SRG) algorithm. The proposed algorithms were extensively applied on all the 322 mammogram images in MIAS database resulting in complete pectoral muscle suppression in most of the images. Our proposed algorithm is compared with other segmentation methods showing superior results in comparison.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Cancer is a group of diseases that cause cells in the body to change and grow out of control. Most types of cancer cells eventually form a lump or masses called a tumor, and are named after the part of the body where the tumor originates. Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple. The remainder of the breast is made up of fatty, connective, and lymphatic tissue [1].

Breast cancer is a leading cause of cancer deaths among women. For women in US and other developed countries, it is the most frequently diagnosed cancer. About 2100 new cases of breast cancer and 800 deaths are registered each year in Norway [2]. In India, a death rate of one in eight women has been reported due to breast cancer.

Early and efficient detection is the most effective way to reduce mortality, and currently a screening programme based on mammography is considered one the best and popular method for detection of breast cancer. Mammography is a low-dose X-ray procedure that allows visualization of the internal structure of the breast. Mammography is highly accurate, but like most medical tests, it is not perfect. On average, mammography will detect about 80–90% of the breast cancers in women without symptoms [3].

An increasing number of countries have started mass screening programmes that have resulted in a large increase in the number of mammograms requiring interpretation [4]. In the interpretation process radiologists carefully search each image for any visual sign of abnormality. However, abnormalities are often embedded in and camouflaged by varying densities of breast tissue structures. Estimates indicate that between 10% and 30% of breast radiologists miss cancers...
during routine screening [4,5]. The images provided by different patients have different dynamics of intensity and present a weak contrast. Moreover the size of the significant details can be very small. Several research works have tried to develop computer aided diagnosis tools that could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis [6–8].

Digital mammography is a technique for recording X-ray images in computer code instead of on X-ray film, as with conventional mammography. The first digital mammography [9] system received U.S. Food and Drug Administration (FDA) approval in 2000. Digital mammographies have advantage over conventional mammography as the images can be stored and retrieved electronically.

Imaging techniques play an important role by helping to perform digital mammogram, especially of abnormal areas that cannot be physically felt but can be seen on a conventional mammogram [10]. Before any image-processing algorithm can be applied on mammogram, preprocessing steps are very important in order to limit the search for abnormalities without undue influence from background of the mammogram. Breast segmentation consists of breast border contour extraction, pectoral muscle extraction, nipple identification etc. On images obtained directly from the digital mammography devices segmentation process is much easier. Previous work from many authors used mammography image databases including this paper, especially MiniMIAS [11] and DDSM [12]. In this paper we have proposed three steps of preprocessing of raw digital mammogram to isolate pectoral muscle from digital mammogram.

In the second section of this paper we have discussed different image segmentation techniques and their role in image analysis. In the third section we have discussed region growing algorithm and more specifically seeded region growing algorithm. The fourth section we have reviewed several works performed on seeded region growing algorithm by different authors. The fifth section contains the core discussion of our proposed work. This section has three subsections that describe the enhancement of digital mammogram (CLAHE) followed by defining the rectangle for pectoral muscle isolation. The final subsection we perform suppression of pectoral region. The sixth section we define the dataset we have used for evaluation. The section seven and eight cites the evaluation and result analysis. We have performed failure assessment in section nine and we have compared our results with the results obtained by other authors in section ten. Finally we have concluded our discussion in section eleven.

2. Image segmentation

The paper is based on the image segmentation method, which refers to the major step in image processing, the inputs are images and, outputs are the attributes extracted from those images. Segmentation divides image into its constituent regions or objects. The level to which segmentation is carried out depends upon the problem being solved i.e. segmentation should stop when the objects of interest in an application have been isolated. Image segmentation refers to the decomposition of a scene into its components. Image segmentation techniques can be broadly classified as into five main classes threshold based, Cluster based, Edge based, Region based, Watershed based segmentation [13].

Segmentation plays an important role in image analysis. The goal of segmentation is to isolate the regions of interest depending on the problem and its characteristics. Many applications of image analysis need to obtain the regions of interest before the analysis can start. Therefore, the need of an efficient segmentation method has always been there. A gray level image consists of two main features, namely region and edge.

Segmentation algorithms for gray images are generally based on two basic properties of image intensity values, discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity, such as edges in an image. The principle approaches in the second category are based on partitioning image into regions that is similar according to a set of predefined criteria. Thresholding, region growing, region splitting and merging are examples of the methods in this category.

3. Region growing

For the segmentation of intensity images like digital mammograms, there are four main approaches [14,15], namely, threshold techniques, boundary-based methods, region-based methods, and hybrid techniques which combine boundary and region criteria.

Threshold techniques [16] are based on the postulate that all pixels whose value (gray level, color value, or other) lies within a certain range belong to one class. Such methods neglect all of the spatial information of the image and do not cope well with noise or blurring at boundaries. Boundary-based methods [17] use the postulate that the pixel values change rapidly at the boundary between two regions. The complement of the boundary-based approach is to work with the regions [18]. Region-based methods rely on the postulate that neighbouring pixels within the one region have similar value. This leads to the class of algorithms known as region growing of which the “split and merge” technique [19] is probably the best known. The general procedure is to compare one pixel to its neighbour(s). If a criterion of homogeneity is satisfied, the pixel is said to belong to the same class as one or more of its neighbours. The fourth type is the hybrid techniques, which combine boundary and region criteria. This class includes morphological watershed segmentation [20] and variable-order surface fitting [14]. The watershed method is generally applied to the gradient of the image.

We use a method known as “seeded region growing” (SRG), which is closer to that of the watershed [21] with some necessary change in proposed technique, which is based on the conventional region-growing postulate of similarity of pixels within the regions. For seeded region growing (SRG), seed or a set of seeds can be automatically or manually selected. Their automated selection can be based on finding pixels that are of interest, e.g. the brightest pixel in an image can serve as a seed pixel. They can also be determined from the peaks found in an image histogram. On the other hand, seeds can also be selected manually for every object present in the image.
The method is employed to segment an image into different regions using a set of seeds. Each seeded region is a connected component comprising of one or more points and is represented by a set S. The set of immediate neighbours bordering the pixel is calculated. The neighbours are then examined and if they intersect any region from set S, then a measure \( \delta \) (difference between a pixel and the intersected region) is computed. If the neighbours intersect more than one region, then the set is taken as that region for which difference measure \( \delta \) is maximum. The new state of regions for the set then constitutes input to the next iteration. This process continues until the entire image pixels have been assimilated into regions. Hence for each iteration the pixel that is most similar to a region that it borders is appended to that region. The SRG algorithm is inherently dependent on the order of processing image pixels. The method has the advantage that it is fairly robust, quick, and parameter free except for its dependency on the order of pixel processing.

Gambotto proposed an algorithm that combines the region growing and edge detection methods for segmenting the images. The method is iterative and uses both of these approaches in parallel. The algorithm starts with an initial set of seeds located inside the true boundary of the region. The pixels that are adjacent to the region are iteratively merged with it if they satisfy a similarity criterion. A second criterion uses the average gradient over the region boundary to stop the growth. The last stage refines the segmentation. The analysis is based on cooperation between the region growing algorithm and the contour detection algorithm. During, adding segments to a region, some pixels that belong to a different region may be misclassified. Such erroneous regions may participate in the growing process. A nearest neighbour rule is then used to locally reclassify them [24].

Hojjatoleslami and Kittler proposed a new region growing approach for image segmentation, which uses gradient information to specify the boundary of a region. The method has the capability of finding the boundary of a relatively bright/dark region in a textured background. The method relies on a measure of contrast of the region, which represents the variation of the region gray-level as a function of its evolving boundary during segmentation. This helps to identify the best external boundary of the region. The application of a reverse test using a gradient measure then yields the highest gradient boundary for the region being grown. The unique feature of the approach is that in each step at most one candidate pixel will exhibit the required properties to join the region. The growing process is directional so that the pixels join the grown region according to a ranking list and the discontinuity measurements are tested pixel by pixel. The algorithm is also insensitive to a reasonable amount of noise. The main advantage of the algorithm is that no a priori knowledge is needed about the regions [25].

4. Literature review

Mehnert and Jackway improved the above seeded region growing algorithm by making it independent of the pixel order of processing and making it move parallel. Their study presents a novel technique for improved seeded region growing (ISRG). ISRG algorithm retains the advantages of seeded region growing (SRG) such as fast execution, robust segmentation and no parameters to tune. The algorithm is also pixel order independent. If more than one pixel in the neighbourhood has same minimum similarity measure value, then all of them are processed in parallel. No pixel can be labelled and no region can be updated until all other pixels with the same priority have been examined. If a pixel cannot be labelled, because it is equally likely belong to two or more adjacent regions, then it is labelled as ‘tied’ and takes no part in the region growing process. After all of the pixels in the image have been labelled, the pixels labelled ‘tied’ are independently re-examined to see whether or not the ties can be resolved. To resolve the ties an additional assignment criterion is imposed, such as assigning a tied pixel to the largest neighbouring region or to the neighbouring region with the largest mean. ISRG algorithm produces consistent segmentation because it is not dependent on the order of pixel processing. Parallel processing ensures that the pixels with the same priority are processed in the same manner simultaneously [22].

Beaulieu and Goldberg proposed a hierarchical stepwise optimisation algorithm for region merging, which is based on stepwise optimisation and produces a hierarchical decomposition of the image. The algorithm starts with an initial image partition into a number of regions. At each iteration, two segments are merged provided they minimise a criteria of merging a segment to another. In this stepwise optimisation, the algorithm searches the whole image context before merging two segments and finds the optimal pair of segments. This means that the most similar segments are merged first. The algorithm gradually merges the segments and produces a sequence of partitions. The sequences of partitions reflect the hierarchical structure of the image [23].

5. Proposed method

Digital mammograms are medical images that are difficult to be interpreted, thus a preparation phase is needed in order to improve the image quality and make the segmentation results more accurate. The main objective of this process is to improve the quality of the image to make it ready to further processing by removing the irrelevant and unwanted parts in the background of the mammogram.

It is important to determine the orientation of the mammogram. The mammogram image is transformed, so that the chest wall location, i.e. the side of the image containing the pectoral muscle, is on the upper left corner of the image. In order to determine the chest wall location, the decreasing pixel intensity of the breast tissue near the skin-air interface is used. By estimating the first derivatives of these pixels using the appropriate masks, we can determine the chest wall location. After determining the chest wall location we determine the top of the image. We extract the vertical centroid of the image and then we assume that the asymmetric region closest to the right side of the vertical centroid is the tip of the breast. The image is flipped horizontally if needed to place the asymmetric region below the vertical centroid, resulting in an image that is right way up.
There are different types of noises, which appear in MIAS images. We need to estimate these regions and exclude them from the remaining process. High intensity noise is characterized by high values of optical densities, such as labels or scanning artifacts. Tape artifacts are markings left by tapes, or other shadows presenting themselves as horizontal running strips. Such noise must be suppressed.

Here we have proposed three distinct phases of preprocessing before actual algorithm of automatic analysis of digital mammogram can be conducted. The proposed method is as follows:

6. Enhancement of digital mammogram

The contrast enhancement phase is done using the contrast limited adaptive histogram equalization (CLAHE) technique, which is a special case of the histogram equalization technique [26] that functions adaptively on the image to be enhanced. The pixel's intensity is thus transformed to a value within the display range proportional to the pixel intensity's rank in the local intensity histogram. CLAHE [27] is a refinement of Adaptive Histogram Equalization (AHE) where the enhancement calculation is modified by imposing a user-defined maximum, i.e. clip level, to height of the local histogram and thus on the maximum contrast enhancement factor. The enhancement is there by reduced in very uniform areas of the image, which prevent over enhancement of noise and reduces the edge-shadowing effect of unlimited AHE [28].

The CLAHE method seeks to reduce the noise and edge-shadowing effect produced in homogeneous areas and was originally developed for medical imaging [29]. This method has been used for enhancement to remove the noise and reduces the edge-shadowing effect in the pre-processing of digital mammogram [30].

The CLAHE operates on small regions in the image called tiles rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the uniform distribution or Rayleigh distribution or exponential distribution. Distribution is the desired histogram shape for the Image tiles. The neighbouring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise and reduce edge-shadowing effect that might be present in the image; the CLAHE technique is described below [31]:

Step 5: All histograms were modified by the transformation function.

\[ t(r_k) = \sum_{j=0}^{k} p_i(r_j) \]  \hspace{1cm} (1)

where \( p_i(r_j) = \frac{n_j}{n} \)  \hspace{1cm} (2)

is the probability density function of the input mammogram image grayscale value \( j \), \( n \) is the total number of pixels in the input mammogram image and \( n_j \) is the input pixel number of grayscale value \( j \).

Step 6: The neighbouring tiles were combined using bilinear interpolation and the mammogram image grayscale values were altered according to the modified histograms.

In our experiment, we defined tiles size i.e. the rectangular contextual regions to \( 8 \times 8 \). The optimal number of tiles depends on the type of the input image, and it is best determined through experimentation. We have conducted extensive experimentation on MIAS mammographic images and found that \( 8 \times 8 \) shows best contrast enhancement. So we have used \( 8 \times 8 \) tiles to perform CLAHE.

Contrast factor that prevents over-saturation of the image specifically in homogeneous areas is restricted to 0.01 here to get the optimized output. The number of Bins for the histogram building for contrast enhancing transformation is restricted to 64 and the distribution of histogram is uniform’ i.e. flat histogram for our experimentation. The range is not specified in the experiment to get the full range of output image.

7. Define rectangle to isolate pectoral muscle from ROI

The two common projections of mammogram are medio-lateral oblique (MLO) and cranio-caudal (CC) view. The advantage of the medio-lateral oblique projection is that almost the whole breast is visible, often including lymph nodes. The main disadvantage is part of the pectoral muscle is shown in upper left and upper right corner of left and right breast mammogram respectively, which is superimposed over a portion of the breast. In contrast the cranio-caudal view is taken from above, resulting in an image that sometimes does not show the area close to the chest wall. In our research work we consider the former view for its complete visibility though it introduces a new challenge to detect and isolate pectoral muscle.

Accurate segmentation is essential to obtain a correct computer aided diagnosis. Intensity based method may produce erroneous result when applied to dense structure i.e. abnormal masses, fibroglandular disc, because that has the nearly same opacity. So, it is important to detect the pectoral muscle and isolate it from the region of interest (ROI), for further analysis. The pectoral muscle have a slightly higher intensity compared to the rest of the breast tissue and appear in upper left corner of MLO view of mammogram (the orientation of right breast mammograms are flipped horizontally to left).

Step 1: Mammogram was divided into a number of non-overlapping contextual regions of equal sizes, experimentally set to be \( 8 \times 8 \).

Step 2: The histogram of each contextual region was calculated.

Step 3: A clip limit, for clipping histograms, was set (\( t = 0.001 \)). The clip limit was a threshold parameter by which the contrast of the image could be effectively altered; a higher clip limit increased mammogram contrast.

Step 4: Each histogram was redistributed in such a way that its height did not exceed the clip limit.
The most important landmark in mammogram image is the vertical straight line that separates background of the mammogram and the left side of breast region. We consider this line as horizontal reference and demarcate the line as $AB$, from top to bottom of the mammogram. In the following step we search for the last pixel of breast region at the top margin of mammogram. This point we refer as $C$. Now we connect point $C$ to the bottom left corner point of the mammogram namely $D$ as $CD$. The line $CD$ intersects horizontal reference $AB$ at the point $E$. We plot a parallel line with respect to the top boundary and passing through $E$. We plot another line parallel to horizontal reference passing through $C$. These two parallel lines intersect at point $F$. So, a rectangle is formed $ACFE$. We then consider the inverted right angle triangle $ACE$ to isolate the pectoral muscle and the other half of the rectangle $CEF$, the intensity values are changed to black. Our experimental results show that except very few mammograms of MIAS database, pectoral muscle lies within the defined triangle. Finally to reduce the computational complexity, demarked rectangle is cropped out from the original mammogram for further processing. The steps are depicted in Figs. 1–5.

8. **Suppression of pectoral muscle**

In this phase, we apply seeded region growing (SRG) algorithm on the extracted rectangle, to suppress the pectoral muscle. ‘Region growing’ is a procedure that groups pixels or sub regions into larger regions based on predefined criteria. The basic approach is to start with a “seed” point and from this it grows into regions by appending to each seed those neighbouring pixels that have properties similar to that seed. Selection of the seed depends on the nature of the problem. A problem in seeded region growing algorithm is the formulation of a stopping rule. Basically, growing a region should stop when no more pixels satisfy the criteria for inclusion in that region.

In our proposed method we have used the basic rules of SRG algorithm and introduced some new ideas to make it more efficient, problem specific and less time consuming. Instead of the seed being selected automatically by the system, we have provided a path to find the next seed. The left top to right bottom diagonal has been considered to select the seeds, specifically up to the cross point of right top to left bottom diagonal in the cropped image. The double arrow dotted line in Fig. 6 is indicating the line of consideration.

The Cartesian slope-intercept equation for the line may be chosen for traversing the line of consideration with end points $(x_1, y_1)$ and $(x_2, y_2)$ in cropped rectangle image (see Fig. 7)

$$y = mx + b$$  \hspace{1cm} (3.1)

The $m$ is representing the slope of the line and $b$ as the $y$ intercept. If we specify the two end points $(x_1, y_1)$ and $(x_2, y_2)$

**Fig. 1 – Original mammogram no mdb160 (Left).**

**Fig. 2 – Original mammogram after CLAHE.**

**Fig. 3 – Mammogram showing lines passing through points A, C, F and E.**
in the cropped rectangle image, we can determine value of the slope and y intercept as following:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$  \hspace{1cm} (3.2)  

$$b = y_1 - mx_1$$  \hspace{1cm} (3.3)  

For any given x interval $\delta x$ along a line, we can calculate corresponding y interval $\delta y$ from the following equation:

$$\delta y = m \cdot \delta x$$  \hspace{1cm} (3.4)  

Similarly, we can obtain the x interval $\delta x$ corresponding to a specified $\delta y$ as

$$\delta x = \frac{\delta y}{m}$$  \hspace{1cm} (3.5)  

All pixels along the line of consideration is read one after another from left top to right bottom and calculates the minimum, maximum and average intensity value of the pixel.

Now a selection criterion of pixel for region growing has been introduced. Read the inverted right angle triangle area from left top corner. The selection method is obtained by subtracting average intensity from that pixel intensity and then dividing it by the difference between maximum intensity and average intensity. If the pixel intensity is greater than 0 and less than equals to 1, pixel will be merged to the region growing and the intensity value will be 0 else it will be remain unchanged.

$$0 < \frac{l_{(x,y)} - l_{avg}}{l_{max} - l_{avg}} \leq 1$$  \hspace{1cm} (4)  

Here the $l_{(x,y)}$ is the intensity of a pixel, $l_{avg}$, $l_{max}$ is average intensity and maximum intensity calculated earlier from the traversed seeds. The region growing will be restricted with in the boundary of the inverted right angle triangle area of the cropped image. Using this algorithm pectoral muscle, that is an unwanted part of the mammogram, will be eliminated. Superimposing the eliminated pectoral region on the original mammogram will provide us the breast image containing the breast tissues only (ROI), which will be helpful for further investigation of mammogram properly.

9. Database

The mammogram images used in this experiment are taken from the mini mammography database of MIAS [11]. The database contains 322 mammogram images in MLO. The original MIAS Database (digitized at 50 µm pixel edge) has been reduced to 200-µm pixel edge and clipped/padded so that every image is 1024 pixels x 1024 pixels. All images are held

---

**Fig. 4** – Cropped rectangle ACEF.

**Fig. 5** – Inverted right angled triangle (ACE) and rest (CEF) masked.
as 8-bit gray level scale images with 256 different gray levels (0–255) and images are stored as portable gray map (.pgm) format. The Mammographic Image Analysis Society has provided the image database for the purpose of research.

The list is arranged in pairs of mammograms, where each pair represents the left and right breast of a single patient. In our experiment we have consider all types of breast tissues i.e. fatty, fatty-glandular, dense-glandular and the abnormalities like calcification, well-defined or circumscribed masses, speculated masses and other ill-defined masses. Three hundred twenty two, MLO view of bilateral pairs of mammogram images are used as test cases.

10. Execution result

The algorithm proposed by us automatically suppresses the pectoral muscle from the breast tissue. We have applied this algorithm over all the 322 mammograms in MIAS database that covers every type of breast shape, size and type including deformities, asymmetries and abnormalities.

10.1. Experiment with fatty tissue

Results obtained by applying the proposed segmentation algorithm on mammogram of MIAS Image 246.L predominantly comprised fatty tissues shown in Figs. 8–15.

10.2. Experiment with fatty fibro-glandular tissue

Results obtained by applying the proposed segmentation algorithm on mammogram of MIAS Image 014.L predominantly comprised fatty-fibro glandular tissues shown in Figs. 16–23.

10.3. Experiment with dense fibro-glandular tissue

Results obtained by applying the proposed segmentation algorithm on mammogram of MIAS Image 099.R predominantly comprised dense-fibro glandular tissues shown in Figs. 24–31.
11. Result analysis

Performance evaluation in algorithm design is an important step that is commonly neglected. What constitutes an “acceptable” result differs significantly, and is often based on visual subjective opinion with very little quantitative endorsement. The accuracy of this technique was evaluated through quantitative measures derived through the comparison of each segmented pectoral “mask” with its corresponding “gold standard”. The gold standard is generated by manually segmenting the pectoral region from each mammogram. The boundary of the Pectoral Region is then manually traced using general purpose image processing software to extract the real pectoral region from mammogram image. The generated Ground Truth (GT) obtained for each pectoral region was referred to the radiology department for confirmation before comparison.

The region extracted by the segmentation algorithm (mask), which matches the GT, is denoted as true positive (TP) emphasizing that the algorithm has indeed found a portion of the pectoral muscle. Pixels shown in the GT but not shown in the mask are defined as false negative (FN) classifications. The pixels not in the GT, but in the mask are defined as false positive (FP) pixels. From this we can derive two metrics: completeness (CM) and correctness (CR).

Completeness = \( \frac{TP}{TP + FN} \)  \( (5) \)

Completeness can range from 0 to 1, with 0 indicating that none of the regions with properly partitioned, and one indicating that the regions were all segmented.

Correctness = \( \frac{TP}{TP + FP} \)  \( (6) \)

Like CM, the optimum value for CR is 1. Several authors have concluded that the algorithm can be considered accurate if the percentage of both completeness and correctness is greater than 95% [32].
Fig. 14 – Pectoral muscle suppressed by proposed seeded region growing algorithm.

Fig. 15 – The final output, after suppression of pectoral muscle.

A more general measure of the algorithm can be achieved by combining completeness and correctness into a single measure, with an optimum value of 1.

\[
\text{Quality} = \frac{TP}{FN + FP + TP} \tag{7}
\]

The results obtained are as follows (see Tables 1 and 2).

Fig. 16 – Fatty fibro-glandular tissue mammogram image (014.I).

Fig. 17 – The histogram of Fig. 16.

Fig. 18 – Fatty fibro-glandular tissue mammogram image (014.I) after CLAHE.
Fig. 19 – The histogram of Fig. 18.

Fig. 20 – Cropped rectangle after masking.

Fig. 21 – The histogram of Fig 20.

Fig. 22 – Pectoral muscle suppressed by proposed seeded region growing algorithm.

Fig. 23 – The final output, after suppression of pectoral muscle.
12. Failure assessment

No algorithm can be considered 100% robust, especially considering the heterogeneous nature of mammograms. Problems with image acquisition such as scanner induced artifacts, excessive background noise, scratches and dust artifacts could influence the reliability of this algorithm. Out of the 322 mammograms, only fourteen (4.29%) fell marginally short of the 95% accuracy indicator specified. Take the examples of image mdb098.L, mdb151.R, mdb158.L etc. here either the pectoral muscle is absent or very indistinct. If there is no visible pectoral region then the algorithm will not consider the mammogram for detection and isolation. We cannot consider the same as the failure of the algorithm but may actually point to some inherent inadequacy in the mammogram. But we notice that in very few results show a little bit over or under segmentation. Take the example of mdb179.R, mdb288.L etc. here the intensity level of pectoral region and the breast region is almost same, and there is no visible distinction between two. The results of the segmentation are not acceptable in those few cases. To summarize, the results obtained by the method show that it is a robust approach but it can be improved in terms of accuracy. Even so, we accept this method because it provides encouraging results.

13. Comparative analysis

We have compared the proposed algorithm with Mustra et al. [33], used hybrid method for the pectoral muscle detection. The method uses bit depth reduction and wavelet decomposition for finding pectoral muscle border. They got 85% accurate result. Another paper that can be comparable to our
work by Raba et al. [34], proposed an automated technique for segmenting a digital mammogram into breast region and background with pectoral muscle suppression. Here they used a combination of an adaptive histogram approach to separate the breast from the background and a selective region growing algorithm to obtain pectoral muscle suppression. They have obtained 86% of good extractions. Chen et al. [35], proposed a fully automated breast region segmentation method based on histogram thresholding, edge detection in scale space, contour growing and polynomial fitting, subsequently, pectoral muscle removal using region growing. The method obtained 93.5% success rate for pectoral muscle separation. Sultana et al.

Fig. 28 – Cropped rectangle after masking.

Fig. 29 – The histogram of Fig 28.

Fig. 30 – Pectoral muscle suppressed by proposed seeded region growing algorithm.

Fig. 31 – The final output, after suppression of pectoral muscle.
used mean-shift segmentation approach using the Hough transform obtain an average of 84% detection rate per image and average of 13% false positive rate per image. Our results are comparable with aforesaid algorithms. For the 322 mammograms evaluated, the mean values for CM and CR are 0.976 and 0.980 respectively. Out of the 322 mammograms, fourteen (04.29%) fell marginally short of the 95% accuracy indicator specified.

14. Conclusions

The proposed algorithm of preprocessing has been presented with contrast enhancement, pectoral muscle detection and suppression. The results obtained over MIAS database have shown excellent output. This algorithm is used to reduce noise, edge-shadowing effect, accurately detect pectoral muscle, and suppress the pectoral muscle successfully without losing any information from the rest of the mammogram. The resultant mammogram can be used further for the automated abnormalities detection of human breast like calcification, circumscribed masses, spiculated masses and other ill-defined masses, circumscribed lesions, asymmetry analysis etc. Further work may be conducted to develop and smoothing the pectoral muscle segmentation. This algorithm has the potential for further development because of its simplicity and encouraging results that will motivate real-time breast cancer diagnosis system.

REFERENCES

