CHAPTER 3

PREPROCESSING AND ENHANCEMENT

3.1 Introduction

Image preprocessing and enhancement methods inquire about how to improve the visual appearance of mammograms. This chapter deals with preprocessing and enhancement activities such as removal of film artifacts and labels, filtering the image, normalization and removal of pectoral muscle region. Figure 3.1 shows the methods involved in the preprocessing and enhancement steps.

Figure 3.1: Flow diagram for preprocessing

A gradient-based image enhancement method for mammography images is based on the first derivative and the local statistics. The proposed enhancement method consists of four processing steps. In the
first step, the given images are identified as left or right breast image and the film artifacts such as labels and X-ray marks are removed from, the mammogram. In the second step, the high frequency components are removed using median filtering. In the third step, the mammogram images are normalized to avoid the difference in brightness and contrast between the mammograms caused by the recording procedure. In the fourth step, the pectoral muscle region, is removed from the breast region to increase the reliability of the segmentation and classification of microcalcifications. The performance of the proposed method is also evaluated by means of Signal-to Noise-Ratio (SNR) using the mammographic images from the Mammography Image Analysis Society database.

3.2 MIAS Data Base

Obtaining real medical images for carrying out research is highly difficult due to privacy issues, legal issues and technical hurdles. Hence, the MIAS database is used in this thesis to study the efficiency of the proposed intelligent system since it is a benchmark database available online for research.

The MIAS, which is an organization of UK-based research groups interested in the understanding of mammograms, has produced a digital mammography database (ftp://peipa.essex.ac.uk). The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Programme and digitized with a Joyce-Lobel scanning microdensitometer to a resolution of 50 μm x 50 μm, 8 bits representing
each pixel. Table 3.1 and Figure 3.2 summaries the different categories of mammograms in MIAS database.

Table 3.1 **Summary of MIAS mammogram categories of 161 pairs**

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>MIAS Category</th>
<th>Image Labels</th>
<th>No. of Image Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal</td>
<td>002, 004, 017-028, 031, 033-034, 037, 039, 043-044, 047, 057, 068-070, 077, 081, 087, 115, 129-31, 138-144, 146-155, 159-161</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>Circumscribed masses</td>
<td>001, 003, 005-006, 008-014, 030, 035, 040, 046,066, 071, 122, 135,145,158</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>Calcifications</td>
<td>105-114, 116-121, 123-128</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Spiculated masses</td>
<td>145-148, 175-188, 189-208</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>Architectural Distortions</td>
<td>58-65, 75-76, 78-80, 82-86</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Asymmetries</td>
<td>036, 038, 041-042, 045, 048-056</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>Miscellaneous</td>
<td>007, 015-016, 029, 032, 067, 072, 132-134, 136-137, 156-167</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>161</strong></td>
</tr>
</tbody>
</table>

**Figure 3.2: MIAS mammogram categories of 161 pairs**
The different categories of mammograms are shown hereunder:

Figure 3.3: Normal mammogram (mdb246)

Figure 3.4: Mammogram image with circumscribed masses (mdb091)
Figure 3.5: Mammogram image with calcifications (mdb209)

Figure 3.6: Mammogram image with spiculated masses (mdb202)
Figure 3.7: Mammogram image with architectural distortion (mdb0125)

Figure 3.8: Mammogram image with asymmetries (mdb081)

The MIAS Digital Mammogram Database contains 322 images Mediolateral Oblique representing 161 bilateral mammogram pairs. The
database is divided into seven categories such as microcalcifications, circumscribed masses, spiculated lesions, ill-defined masses, architectural distortion and asymmetric densities of normal and abnormal mammograms.

3.3 Gradient-Based Mammogram Image Enhancement

The images in the MIAS database are of the size of 1024 x 1024 pixels. A tracking algorithm-! is used to check whether the image is left or right image. Starting from the middle row of the mammogram image, each pixel in the first column is checked to see if it is greater than zero and the average intensity value of the column is calculated. If the average value is greater than 75, the image is considered the right image otherwise the given image is the left image. Whenever a left image occurs, it is flipped horizontally and stored.

3.3.1 Removal of Film Artifacts

The film artifacts are removed using tracking algorithm II. Starting from the first row and first column of the mammogram image, the Intensity value of the pixel is copied into a new two-dimensional array (1024x1024) when the following conditions are satisfied. Initially a flag value is assigned as zero. If the intensity value is zero then it is copied to the new image. If the intensity value is greater than zero, it is copied to a new image and the flag value is set as one. Again the next value is compared; if it is a non-zero value then that value is stored in a new array; if it is zero, the flag value is set as zero and the remaining pixels in that row are skipped. The next iteration is started with the next row. In
this manner, only the breast region is copied into the new array; thus the film artifacts are removed from the mammogram image.

Figure 3.9 shows the original mammogram image and the image with no artifacts. The image also contains a black region before the breast region. The black region is removed and the breast region is moved to the first column in order to remove the pectoral region. The column in which the breast region starts is found by using the tracking algorithm I. Once the column is identified, the pixels between the last columns are transformed. The intensity value of the non-breast region in adjusted to zero in order to maintain the size of the image as 1024x1024. Hence the breast region starts from the first column.

Figure 3.9: (a) Original mammogram, (mdb010.pgm), (b) Removal of film artifacts and (c) Replacement
3.3.2 Median Filtering

The high frequency components in the mammogram image are removed with median filtering. The merit of using median filter is that it can remove the noise without disturbing the edges. For each pixel 11x11 window of neighborhood pixels are extracted, and the median value is calculated for that window. The intensity value of the center pixel value is replaced with the median value.

Figure 3.10: Median filtering

The values are arranged in ascending order such as 20, 20, 20, 20, 21, 21, 22, and 50. The median is equal to 21. The center pixel value is replaced with 21. This procedure is followed for all the pixels in the image to smoothen the entire mammogram image [116]. Figure 3.11 shows the Median filtered mammogram image.

Figure 3.11: (a) Original mammogram and (b) Median filtered image
3.3.3 Normalization

A closer inspection of the mammograms reveals several difficulties in the diagnosis. The global appearance (brightness, contrast, etc.) of the breasts may differ, usually due to variations in the recording procedure and this can be avoided using the normalization method. In order to reduce the variation and achieve computational consistency, the images are normalized by mapping all mammograms into a fixed intensities range between $r_1$ and $r_2$ ($0 < r_1 < r_2 < 255$). Assume an image $g_s(x, y)$ whose maximum gray level value is $\max G_j$ and minimum gray level value is $\min G_j$, transform $g_j(x, y)$ into $g_k(x, y)$

$$g_k(x, y) = \frac{r_1 + (g_i(x, y) - \min G_j \times (r_2 - r_1))}{\max G_j - \min G_j}$$

In this thesis, $r_1$ and $r_2$ are assigned the values 60 and 210 respectively [25] due to the fact that maximum intensities and minimum intensities of the microcalcifications are not beyond $r_2$ and below $r_1$ with certainty after investigating a large number of mammograms. Figure 3.12 shows the mammogram, image after applying median filter and normalization.

![Figure 3.12: (a) Median filtering and (b) Normalized image (mdb010.pgm)](mdb010.pgm)
3.3.4 Removal of Pectoral Muscle Region

The pectoral muscle appears as a bright triangular region in the image corner towards the chest wall and the top of the breast region. The gray level mammogram image is converted to a binary image to segment the pectoral muscle region. A histogram based thresholding technique is used to separate the pectoral muscle region. Figure 3.13 shows the image histogram. The global optimum in the histogram is selected as the threshold value. The gray values smaller than the threshold values are changed to black (zero), and the gray values greater than the threshold are changed to white (one). Erosion and dilation operators are used to preserve the pectoral muscle region [147].

![Figure 3.13: Mammogram image histogram](image)

The white pixels in the upper left corner of the binary mammogram image indicate the pectoral muscle region. The corresponding spatial coordinate of the white pixels in the gray level image is changed to black.
(zero) to remove the pectoral muscle region from the breast region. Figure 3.14 shows the breast region of left and right mammograms after the removal of pectoral muscle region.

Figure 3.14: (a) Pectoral region and (b) After removal of pectoral region.

The steps involved in the removal of pectoral muscle region is summarized in the following algorithms

Step 1. \( S \) \{Read the mammogram image\}
Step 2. \([m \ n]\) \{Size of the image\}
Step 3. \( th \) \{Peak threshold value from the histogram\}
Step 4. for each pixel in \( S_{ij} \) \{i=1..m; j=1..n\}
   a. if \( S_{ij} < th \)
      \( B_{ij} = 0 \);
   b. else
      \( B_{ij} = 1 \); end
      \( B_{ij} \leftarrow \) erosion \( (B_{ij}) \)
Step 5. \( B_{ij} \leftarrow \) dilation \( (B_{ij}) \)
Step 6. for each pixel in \( B_{ij} \)
   a. if \( B_{ij} = 1 \)
      \( S_{ij} = 0 \)
   b. end
end
Step 7. \( S_{ij} \leftarrow \) image after the removal of pectoral muscle region

Figure 3.15: Algorithm of removal of pectoral muscle region
3.4 Experiments and Results

It is hard to measure the improvement of the enhancement objectively. If the enhanced image can make the observer perceive the region of interest better, then we can say that the original image has been improved. In order to compare different enhancement algorithms, it is better to design some methods for the evaluation of enhancement objectively. The statistical measurements such as variance or entropy can always measure the local contrast enhancement. However, that shows no consistency for the mammograms.

The Peak Signal-to-Noise Ratio (PSNR) and the Average Signal-to-Noise Ratio (ASNR) are used to evaluate the enhancement performance [104; 105]. The noise level is measured by the standard derivation \( \sigma \) of the original mammogram image:

\[
\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (b_i - b)^2},
\]

where \( b \) is the mean gray level value of the original image and \( b_i \) is the gray level value of a surrounding region, and \( N \) is the total number of pixels in the surrounding region. The PSNR and the ASNR are defined as follows:

\[
\text{PSNR} = \frac{p - b}{\sigma}
\]

\[
\text{ASNR} = \frac{f - b}{\sigma}
\]

where \( p \) is the maximum gray level value and \( f \) is the average gray level value of a enhanced image. If the values of the two indices are larger, the proposed preprocessing and enhancement method performance is better.

The proposed method produces 0.94 and 0.92 for the PSNR and the
ASNR respectively. Table 3.2 compares the performance of the proposed algorithm with the existing methods.

In this thesis, a gradient-based image enhancement method using first derivative and local statistics has been proposed. The final enhanced mammogram image is validated using SNR values. The merit of the gradient-based image enhancement is that it is very simple because enhancing the mammogram image only by local statistics. So the proposed method produces the output faster than any other existing method. Also it produces the same output every time.

The Signal-to-Noise Ratio can be used as an optimal parameter to analyze the performance of the enhancement techniques. The statistical result shows that the proposed method performs better than the existing methods. The work is published in [172].

Table 3.2 Performance analysis of enhancement techniques

<table>
<thead>
<tr>
<th>Authors</th>
<th>Methods</th>
<th>PSNR</th>
<th>ASNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.R. Mudigonda et al., (2001) [121]</td>
<td>Gaussian smoothening</td>
<td>0.82</td>
<td>0.8</td>
</tr>
<tr>
<td>K.J. McLaughlin et al., (2004) [114]</td>
<td>Noise equalization</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>L. Bocchi et al., (2004) [21]</td>
<td>Matched filtering</td>
<td>0.85</td>
<td>0.81</td>
</tr>
<tr>
<td>H.D. Cheng et al., (2004) [25]</td>
<td>LOG filtering</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>Proposed approach</td>
<td>Tracking algorithms</td>
<td>0.94</td>
<td>0.92</td>
</tr>
</tbody>
</table>
3.5 Summary

This chapter proposed a gradient-based image enhancement method using first derivative and local statistics and showed the validity of detection of microcalcifications. Initially the film artifacts are removed and the breast region alone is copied into a new Image. With this new image median filtering is applied to remove the high frequency components, and the mammogram images are normalized to avoid the difference in brightness and contrast. Then the pectoral muscle region is removed from the breast region. The final enhanced mammogram image is validated using SNR values. The statistical result shows that the proposed method can perform better than other existing works.

Figure 3.16: Performance analysis of enhancement techniques

...