A global approach for detecting mass in Digital Mammograms

K.Vaidehi, T.S.Subashini

1Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar – 608 002, India.
2Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar – 608 002, India.
Email: 1vainakrishna@gmail.com; 2tramsuba@gmail.com

ABSTRACT

This paper proposes some image processing methods for detecting breast masses in mammograms. This approach is global in the sense the breast region along with the background and pectoral muscle is used as input for detecting mass. First, the mammogram is preprocessed for noise reduction and image enhancement. We introduce a thresholding method to obtain rough region of interest. Connected component labeling is then applied to the binarized image to get discrete regions or objects. The objects are subjected to various yes or no criteria which narrows down the region that are most likely to be the mass. The proposed work is done using the Mammography Image Analysis Society (MIAS) database. This work shows that connected component labeling can be used to detect masses in digital mammograms effectively.

Keywords: Digital mammogram, connected component labeling, segmentation, mass detection.

1 INTRODUCTION

Breast cancer is the second most common malignancy in India. A women succumbs to breast cancer every ten minutes. According to Indian council of medical research statistics, 10,000 breast cancers are being diagnosed every year in India and more than 70% of them are diagnosed in advanced stage. By 2020, the incidence of breast cancer in India is expected to double. The cancer registries’ data shows that urban women are at almost having double the risk of breast cancer than rural women [1,2].

The aim of mammography is the benign detection of breast cancer. Digital mammograms take an electronic image of the breast and store it directly in a computer. Breast abnormalities that can indicate cancer are masses, calcifications, architectural distortions and bilateral asymmetry. Segmentation is one of the most important step in computer aided detection/diagnosis especially for masses, because breast masses can have unclear borders and are sometimes obscured by glandular tissue in mammograms. During the search for suspicious areas, it is possible that masses of this type are overlooked by radiologists. In this work, mass segmentation is carried out using connected component labeling.

2 PREVIOUS WORK

Various image processing methods have been applied for mass detection. [3] employs a difference of gaussians and derivative based features. [4] presents a segmentation method based on the fuzzy sets theory to divide a mammogram into different regions and produces regions of mass candidates. [5] applies the watershed algorithm for detecting masses. [6] characterizes the suspected regions with features based on the iris filter output and gray level, texture, contour-related and morphological features extracted from the image. [7] proposes the use of Local Binary Patterns (LBP) for representing the textural properties of the masses. [8] applies least squares curve fitting procedure to obtain the best-fit ellipse for each of the manually segmented regions. The set of best fit ellipses detects the mass region. [9] uses the K-Means algorithm for mass segmentation and co-occurrence matrix to describe the texture of segmented structures. [10] implements Local thresholding technique and Otsu method for segmenting the tumor regions. Our previous work, [11] detects masses after eliminating background and pectoral muscle that is, it restricts the processing to the breast region alone. However the proposed work detects mass without eliminating the background and pectoral muscle.

3 DATA SOURCE

The proposed work is done using the digital mammogram images obtained from the Mammography Image Analysis Society (MIAS) database. MIAS is an organization of research groups interested in the understanding of mammograms, and has produced a digital mammography database. The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Program and digitized with a Joyce-Loebel scanning micro densitometer to a resolution of 50µm x 50µm and 8 bits represent each pixel. It has been reduced to 200 micron pixel edge and clipped/padded so that every image is 1024 x 1024 pixels[12].

4 PROPOSED WORK

In the first stage, The original image is cropped to remove the unwanted background region. Then the noise is removed and the image is smoothed using median filter and average filter respectively. Gray scale transformation is applied to improve the contrast of suspected masses. Connected component labeling is applied to detect masses. The block diagram of the proposed method is shown in Fig. 1. The modules involved in this section are explained below:

The appearance of an image, as well as possibilities of recognizing the diagnostic features, depends crucially on the image contrast. Though primarily determined by the properties of the imaging modality and adjustment of the imaging system, the displayed contrast can be substantially influenced by applying simple contrast transforms. Contrast transforms is performed with the purpose of increasing the difference between the dark and bright pixels. In this paper, the contrast of the mammogram image is improved using the cubic function given by equation (1).

\[
T(i, j) = \text{round}\left(\frac{P(i, j)^3}{f * M^2}\right)
\]  

Where \(M = \max(P(i, j))\)

\(M\) is the maximum gray value of the pixels of the original image. The gray scale is modified based on the cubic function given in equation (1). \(P(i,j)\) corresponds to the pixels of the original image. \(T(i, j)\) is the transformed image. \(f\) is the brightener factor that modifies the function. The brightener factor \(f\) can have values 0.4, 0.6 and 1 [15]. The factor value 1 is chosen in this paper since it enhances mass areas much better than 0.4 and 0.6.

### 4.3 Mass Detection

A threshold value \(t\) is chosen that retains bright pixels in the image. Pixels with values greater than \(t\) are set to white (1) and values less than \(t\) are set to black (0). Connected component labeling is applied to the binary image using eight pixel connectivity to indicate each discrete region in the binary segmented image. These discrete regions are subjected to three acceptance/rejection criteria which select the most important candidate regions that strongly resemble a circumscribed mass in terms of their area and their statistical characteristics such as their pixel’s intensity, higher order moments, etc.

(a) **Criterion 1:** From the ground truth given in the database, it is found that area of the mass ranges between 900 to 5000 pixels. So the region whose area lies between 900 pixels and 5000 pixels is considered to be suspicious. This rule is applied to each segmented region and this reduces the number of the candidate regions to \(R_i, i = 1, ..,M\). Regions that don’t meet this requirement are rejected.

(b) **Criterion 2:** Each remaining region is considered a suspicious region if its third order moment (skewness) is negative,
otherwise they are rejected. Third order moment is the measure of the asymmetry of the pixel values around the image mean.
(c) Criterion 3: Each remaining region is still considered suspicious if its mean intensity is higher than a threshold value $T_m$. The regions that do not satisfy this criterion are rejected. The threshold value is chosen according to the character of the background tissue as can be seen in Table 1. These threshold values were chosen after experimenting with the images in the database.

Table 1: Threshold values for different types of background tissue

<table>
<thead>
<tr>
<th>Background</th>
<th>Threshold Value $T_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty</td>
<td>$160 &lt; T_m &lt; 170$</td>
</tr>
<tr>
<td>Glandular</td>
<td>$171 &lt; T_m &lt; 180$</td>
</tr>
<tr>
<td>Dense</td>
<td>$T_m &gt; 181$</td>
</tr>
</tbody>
</table>

![Fig. 2: Mass detection using CCL - Global processing](image)

(a) Cropped original image. (b) Median filtered image. (c) Smoothed image. (d) Transformed image. (e) Thresholded image at $(t = 79)$. (f) Connected component labeled image. (g) Image after applying first criterion. (h) Image after applying second criterion. (i) Image shows final mass after applying third criterion.

5 EXPERIMENTAL RESULTS

In this proposed work, Mini-MIAS database is used. This work has been done using Microsoft Visual C++ 6.0 and OPENCV Image Processing Library. The final segmented region is shown in Fig. 2.

6 CONCLUSION

Computer aided detection and diagnosis is not only the kind of second readers of medical images but also useful for patients to avoid unnecessary biopsies, stress and cost. This paper presented several methods of image processing techniques that can be implemented for detection of masses in digital mammography. In this paper, global thresholding, connected component labeling are applied and the three criterions are satisfied. This work shows that connected component labeling can be used in conjunction with the image processing technique to detect masses in digital mammograms.

ACKNOWLEDGMENT

The work has been done under University Grants Commission (UGC) Major Research Project. The financial support of UGC is greatly acknowledged with appreciation.

REFERENCES

[1] www.healthindia.com

