Automatic Mammographic Mass Segmentation based on Region Growing Technique

K. Yuvaraj, and U.S. Ragupathy

Abstract—Breast cancer is one of the leading cancers in women worldwide both in developed and developing nations as per the records from World Health Organization (WHO). American Society identified that by the end of 2012, about 2,26,000 cases were diagnosed and 40,000 resulted in death. Physician uses mammography as one method for breast cancer detection and interpretation. Mass segmentation plays an important step for the cancer detection. Notable researches were done and still moving towards the effective detection of masses in mammograms. In most of the segmentation techniques, the region of interest is chosen manually. To overcome this, a fully automatic mass segmentation scheme is proposed. The proposed method includes automatic seed selection by extracting the statistical features and the region growing technique is employed. The difference in the mean of the manual markup by an expert and the proposed segmentation obtained is 0.356.

Keywords—Mammography, Mass, Region Growing, Segmentation.

I. INTRODUCTION

A mammogram is an X-ray of the breast. It is the best screening tool used today to find breast cancer. Mammographic images can be stored on film (standard) or on a computer (digital). A mammogram can find breast cancer before it can be felt. When mammography is combined with clinical breast exam, the chances for finding cancer are even greater. For women with dense breast tissue, digital mammography may be more accurate than standard mammography. Other imaging tests can provide valuable information; however, these tests are not always able to tell the difference between dense benign (non-cancerous) lumps and cancer. And, Sometimes they miss tiny calcium deposits that are the earliest sign of a tumor. Thus mammogram plays as a useful tool in breast cancer detection and identification. Masses are defined as space occupying lesions that are described by their shape and marginal properties. A benign is smoothly marginated, whereas a malignancy is characterized by an indistinct border that becomes more spiculated with time.

Masses in mammograms possess various scales, complex shapes, and indistinct margins. Some of them are embedded and hidden in surrounding tissues. Thus, mass segmentation is a challenging task [10]. In recent years, there have been numerous studies on mass segmentation [4], [6], [7], [8]. In the Level Set based segmentation [9], [11], [12] and the Active Contour region based segmentation [13], the region of interest is chosen manually which is overcome by the proposed method of segmentation.

The region-growing algorithm is a typical image segmentation approach. Many extensions of region growing have been proposed to adapt to the mass segmentation problem [2], [3], [5]. The problem lies in the manual initial seed selection which is of time consuming and it can be affected by several factors. The proposed algorithm chooses the seed automatically and consumes less time. Also the proposed work is compared with the manual segmentation done by an expert.

The rest of the paper is organized as follows. Section II discuss about the proposed method of mass segmentation. Results and discussion are drawn in Section III. The conclusion is made in section IV.

II. PROPOSED AUTOMATIC REGION GROWING BASED MASS SEGMENTATION

Many researches were done for automatic segmentation of the mass in the mammogram [1]. The proposed method aims at the automatic seed selection method based on the statistical features to increase the segmentation accuracy. Fig. 1 illustrates the block diagram of the proposed automatic region growing based segmentation. Mammographic image with mass is taken as an input.

The image considered is not subjected to region of interest (no cropping is done) selection. A mask is initialized as a part of seed selection. For the regions within the mask the statistical features are extracted. If the extracted features matches with the predefined features of mass then the seed is fixed. The input image is subjected to thresholding. Then finally gray level segmented image is obtained.

The mask is moved in case of mismatching with the extracted features. These steps are detailed in the following subsections.

K. Yuvaraj is with Department of Electrical and Electronics Engineering, Kongu Engineering College, Perundurai, Erode, 638052 India. (Phone: 09944882450; e-mail: yuvarajkumarasamy@gmail.com).

U.S. Ragupathy is with Department of Electronics and Instrumentation Engineering, Kongu Engineering College, Perundurai, Erode, 638052 India. (Phone: 09842851315; e-mail: ragupathy.us@gmail.com).
A. Mammogram Images

Mammographic images are collected from the Clarity Imaging Center, Coimbatore and from the Mammographic Image Analysis Society (MIAS) database. Sample mammographic images collected are shown in Fig. 2.

B. Feature Extraction

Prior information about the mass region is studied by defining the mask of size 20x20. This is the optimal size to hold the maximum area within the mass. Twenty two images with mass affected region and the normal region are studied.

Gray Level Co occurrences Matrices (GLCM) for 0º, 45 º, 90 º, 135 º are formed and the average of these matrices is taken as resultant mean GLCM. With the defined mask size, the features are extracted from the resultant GLCM. Initially 17 Haralick features are extracted. Among which only five features are able to clearly distinguish between the normal and the mass regions and it is evident from table I.

The window size beyond 20x20 lacks accuracy. The five features includes: Mean, Dissimilarity, Sum average, Sum variance and Correlation.

Hence in the proposed method mask of size 20x20 is defined and the above mentioned five statistical features are extracted.

B.1 Mean

It gives the average pixel value of a given image. The range defined for mass finding is 190 to 200.

B.2 Dissimilarity

Dissimilarity gives the difference in the values of the adjacent pixels. The range for mass finding is 0.10 to 0.3.

B.3 Sum average

It gives the summed value of the element and the values of the neighboring pixel elements. 7 to 9.8 is the range specified.

B.4 Sum variance

This feature puts relatively high weights on the elements that differ from the average value. The masked region should satisfy the range between 30 to 90.

B.5 Correlation

Correlation is the measure of the gray level linear dependence between the pixels at the specified position relative to each other. The specified range is 20 to 35. For the image collected from the imaging center, the feature values that satisfy the mass region are listed in table II.
TABLE I
FEATURE SELECTION FROM GLCM

<table>
<thead>
<tr>
<th>Feature Extracted from GLCM</th>
<th>Normal Mdb123 (x)</th>
<th>Mass Mdb69 (y)</th>
<th>Mean M=(x+y)/2</th>
<th>Variance ((M-x)+(M-y))/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>190</td>
<td>240</td>
<td>215</td>
<td>25</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.3806</td>
<td>0.4526</td>
<td>0.4166</td>
<td>0.0363</td>
</tr>
<tr>
<td>Correlation</td>
<td>-1.3026</td>
<td>-1.2566</td>
<td>1.2796</td>
<td>0.0230</td>
</tr>
<tr>
<td>Dissimilarity</td>
<td>0.0568</td>
<td>1.0459</td>
<td>0.5513</td>
<td>0.4945</td>
</tr>
<tr>
<td>Energy</td>
<td>0.5347</td>
<td>0.4869</td>
<td>0.5108</td>
<td>0.0239</td>
</tr>
<tr>
<td>Entropy</td>
<td>-0.4232</td>
<td>-0.3958</td>
<td>0.4095</td>
<td>0.0137</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.9922</td>
<td>0.9894</td>
<td>0.9908</td>
<td>0.0014</td>
</tr>
<tr>
<td>Maximum Probability</td>
<td>0.6438</td>
<td>0.4977</td>
<td>0.5707</td>
<td>0.0731</td>
</tr>
<tr>
<td>Sum average</td>
<td>9.0698</td>
<td>6.9194</td>
<td>7.946</td>
<td>1.0752</td>
</tr>
<tr>
<td>Sum variance</td>
<td>120.6634</td>
<td>90.7037</td>
<td>105.6835</td>
<td>14.9799</td>
</tr>
<tr>
<td>Sum entropy</td>
<td>0.3298</td>
<td>0.4143</td>
<td>0.3720</td>
<td>0.0422</td>
</tr>
<tr>
<td>Difference variance</td>
<td>0.3806</td>
<td>0.4526</td>
<td>0.4166</td>
<td>0.0360</td>
</tr>
<tr>
<td>Difference entropy</td>
<td>0.0618</td>
<td>0.0866</td>
<td>0.0742</td>
<td>0.0124</td>
</tr>
<tr>
<td>Information measure of correlation</td>
<td>0.9844</td>
<td>0.9868</td>
<td>0.9856</td>
<td>0.0012</td>
</tr>
<tr>
<td>Inverse difference normalized</td>
<td>0.9961</td>
<td>0.9950</td>
<td>0.9955</td>
<td>0.0006</td>
</tr>
<tr>
<td>Inverse difference moment normalized</td>
<td>0.9966</td>
<td>0.9959</td>
<td>0.9962</td>
<td>0.0004</td>
</tr>
<tr>
<td>Correlation</td>
<td>41.6196</td>
<td>31.8959</td>
<td>36.7577</td>
<td>4.8619</td>
</tr>
</tbody>
</table>

TABLE II
STATISTICAL FEATURES EXTRACTED FOR IMAGE SE003

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>192</td>
</tr>
<tr>
<td>Dissimilarity</td>
<td>0.2762</td>
</tr>
<tr>
<td>Sum average</td>
<td>7.7810</td>
</tr>
<tr>
<td>Sum variance</td>
<td>40.2691</td>
</tr>
<tr>
<td>Auto correlation</td>
<td>24.2833</td>
</tr>
</tbody>
</table>

C. Thresholding

If the extracted features match with the mass region, then the original image is subjected to thresholding. The aim of this thresholding is to make the mass region clear. The threshold value chosen is 70. Mass are of with different intensities. To discriminate the mass with the back ground tissue thresholding is performed.

The threshold value beyond 70 will not define the mass clearly as shown in Fig.3 (a).

![Fig. 3 Thresholded Image with (a) with Threshold Value of 75 and (b) with Threshold Value of 70](image)

D. Region Growing

For the mask defined, since the features match with the mass features, it is evident that the initial pixels of the mask can be taken as the seed. With this automatic initial seed selection, the region growing technique is employed. The algorithmic steps are as follows:

- Calculate the mean of the window selected to act as mask
- Collect the number of pixels in the mask region
- Assign free memory to store neighbors of the (mask) region
- Initialize the distance of the region’s new pixel to the region mean
- Initialize neighbor locations (footprint)
- Start region growing until distance between region and possible new pixels become higher than 0.2. The value higher than 0.2 will results in leakage of the region growing
- Add new neighboring pixels
- Calculate the neighbor coordinate
- Check if the neighboring pixel is inside or outside the region
- Add neighbor if inside and not already part of the segmented area
- Add pixel to the region having intensity nearest to the mean of the region
- Calculate the new mean of the region
- Save the coordinates of the pixel
- Remove the pixel from the neighbor list
- Return the segmented area as logical matrix

At the end of region growing technique, binary segmented image is obtained as shown in Fig. 4.

![Fig. 4 (a) Original Image (b) Thresholded Image (c) Binary Segmented Image](image)
The resulting binary image is checked pixel by pixel. If the pixel value is 1, then the corresponding pixel value of the preprocessed image is assigned. If the pixel value is zero, then it is marked as zero. Resulting image will have the gray content of mass and the remaining part as zero.

![Fig. 5 (a) Original Image, (b) Mass Segmented Result and (c) Gray level Mass Segmented Result](image)

Fig. 5 (a) shows the original image. Mass segmented binary image is illustrated in Fig. 5 (b), and Fig. 5 (c) shows the gray level mass segmented image.

III. RESULTS AND DISCUSSION

Fig. 6 shows the cropped version of Fig. 5 to enhance the visualization. Fig. 7 (a) provides the manual segmentation by an expert and Fig. 7 (b) illustrates the proposed automatic seed selected region growing based segmented image.

![Fig. 6 Cropped Images: (a) Original Image, (b) Mass Segmented Result and (c) Gray level Mass Segmented Result](image)

![Fig. 7 (a) Manual Segmentation by an Expert (b) Proposed Mass Segmentation](image)

Table III shows the difference in the mean value of the manual markup by an expert and the proposed segmentation is 0.356 which indicates the closeness of the segmentation.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Process</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>Manual</td>
<td>141.462</td>
</tr>
<tr>
<td>Proposed</td>
<td>Automated</td>
<td>141.690</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>0.356</td>
</tr>
</tbody>
</table>

If the extracted features are not matched with the mass features then the mask position is moved until it satisfies the mass features. Thus the mask is scanned for the entire image. Twenty two images from MIAS database are examined and it ensures the segmentation with high accuracy.

IV. CONCLUSION

In this paper a new computer aided fully automatic mass segmentation scheme is proposed. The results of the segmentation are compared with the segmentation done by an expert. From the result it is inferred that the proposed method holds good for mass segmentation. In future, more number of samples can be incorporated and classification can be carried out by any one of the classifier to classify the nature of the mass.

REFERENCES


**K.Yuvaraj** received Bachelor's Degree by the Anna University, Chennai for Electrical and Electronics Engineering, studied at Kongu Engineering College, Erode, Tamil Nadu, India in 2009. He is currently pursuing Masters Degree for Applied Electronics in Kongu Engineering College under Anna University, Chennai. He had presented a paper titled ‘Mammographic Mass Segmentation based on Iterative Active Contours’ at an international conference on Biomedical Signals, Systems and Imaging in Indian Institute of Technology (IIT) Madras. His areas of interest include Digital Image Processing and Soft Computing techniques.

**U.S.Ragupathy** received his Bachelor's degree in Electrical and Electronics Engineering in 1999 from University of Madras, Master’s degree in Applied Electronics from Anna University Chennai in 2004 and Ph.D in the area of Medical Image Processing from Anna University Chennai in 2011. He is currently working as Professor and Heading the department of Electronics and Instrumentation Engineering at Kongu Engineering College. His areas of interest include Digital Image Processing, Digital Signal Processing, Wavelets, VLSI Signal Processing and Soft Computing Techniques. He has published 23 papers in journals, national and international conferences.