Identification of Abnormal Masses in Digital Mammography Images

Indra Kanta Maitra  
Research Fellow,  
Dept. of Computer Science & Engineering,  
University of Calcutta, India  
E-mail: ikm.1975@ieee.org

Sanjay Nag  
Research Fellow,  
Dept. of Computer Science & Engineering,  
University of Calcutta, India  
E-mail: sanjaynag75@gmail.com

Prof. Samir Kumar Bandyopadhyay  
Professor and Senior Member IEEE,  
Dept. of Computer Science & Engineering,  
University of Calcutta, India  
E-mail: skb1@vsnl.com

Abstract

Mammography is at present one of the available method for early detection of masses or abnormalities which is related to breast cancer. The most common abnormalities that may indicate breast cancer are masses and calcifications. The challenge lies in early and accurate detection to overcome the development of breast cancer that affects more and more women throughout the world. Masses appear in a mammogram as fine, granular clusters, which are often difficult to identify in a raw mammogram. Digital mammogram is one of the best technologies currently being used for diagnosing breast cancer. Breast cancer is diagnosed at advanced stages with the help of the digital mammogram image. In this paper, a method has been developed to make a supporting tool. This will make it easy and less time consuming for identification of abnormal masses in digital mammography images. The identification technique is divided into two distinct parts i.e. Formation of Homogeneous Blocks and Color Quantization after preprocessing. The type of masses, orientation of masses, shape and distribution of masses, size of masses, position of masses, density of masses, symmetry between two pair etc. are clearly sited after proposed method is executed on raw mammogram, for easy and early detection of abnormalities.

Keywords: Breast Cancer, Mammogram, Masses, Homogeneous Blocks, Color Quantization, GLCM, Contrast, Homogeneity, Energy.

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 [5].

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 [13]. In India, a death rate of one in eight women has been reported due to breast cancer.

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. Testing is somewhat more accurate in postmenopausal than in premenopausal women [11]. An increasing number of countries have started mass screening programmes that have resulted in a large increase in the number of mammograms requiring interpretation [1]. 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 per cent of breast radiologists miss cancers during routine screening [1][20]. In order to improve the accuracy of interpretation, a variety of screening techniques have been developed.

Breast image analysis can be performed using mammography, magnetic resonance, nuclear medicine or ultrasound. So far the most effective and economical breast imaging modality has been mammography due to its simplicity, portability and cost effectiveness. Segmentation is the fundamental process which partitions a data space into meaningful salient regions. Image segmentation essentially affects the overall performance of any automated image analysis system thus its quality is of the utmost importance.

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 [6] system received U.S. Food and Drug Administration (FDA) approval in 2000. An example of a digital mammography system is the Senographe 2000D. The images are displayed on a computer monitor and can be enhanced (lightened or darkened) before they are printed on film. Images can also be manipulated; the radiologist can magnify or zoom in on an area. From the patient’s perspective, the procedure for a mammogram with a digital system is the same as for conventional mammography [12].

Digital mammography may have some advantages over conventional mammography. The images can be stored and retrieved electronically. Despite these benefits, studies have not yet shown that digital mammography is more effective in finding cancer than conventional mammography [14].

Initial mammographic or MRI images themselves are not usually enough to determine the existence of a benign or malignant disease with certainty. If a finding or spot seems suspicious, your radiologist may recommend further diagnostic studies. Interpretations of mammograms can be difficult because a normal breast can appear differently for each woman. Also, the appearance of an image may be compromised if there is powder or salve on the breasts or if the patient has undergone breast surgery.

Recent studies showed that the interpretation of the mammogram by the radiologists give high rates of false positive cases. 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. They could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis [8][9][18].

 Imaging techniques play an important role on mammogram images, especially of abnormal areas that cannot be physically felt but can be seen or processed on a conventional mammogram or with ultrasound [12]. In this paper we have proposed a new technique, and we have developed a supporting tool for easy identification of abnormal masses in mammography images. This will reduce false positive (FP), false negative (FN) detection.

2. Previous Works

Numerous promising approaches are being proposed by various authors. Few of these recent works have been extensively studied by us and discussed in this paper.

H.S. Sheshadri, et al, 2005, proposed an important approach for describing a region is to quantify its structure content. In their paper the use of functions for computing texture based on statistical measures is prescribed. MPM (Maximizer of the posterior margins) algorithm is employed. The segmentation based on texture feature would classify the breast tissue under various categories. The algorithm evaluates the region properties of the mammogram image and thereby would classify the image into important segments. Images from mini-MIAS database have been considered to conduct the experiments. The segmentation thus obtained is comparatively better than the other normal methods. The validation of the work has been done by visual inspection of the segmented image by an expert radiologist [17].

H.S. Sheshadri, et al, 2006, proposed method employs simple thresholding the region of interest and the use of filters for clear identification of microcalcifications. The method suggested for the detection of microcalcifications from mammogram image segmentation and analysis was tested over several images taken from mini-MIAS (Mammogram Image Analysis Society, UK) database. The algorithm was implemented using MATLAB codes and hence can work effectively on a simple personal computer with digital mammogram as stored data for analysis [16].

S. Saheb Basha et al, 2009, presents a research on mammography images using Morphological operators and Fuzzy c – means clustering for cancer tumour mass segmentation. The first step of the cancer signs detection should be a segmentation procedure able to distinguish masses and micro calcifications from background tissue using Morphological operators and finally fuzzy c- means clustering (FCM) algorithm has been implemented for intensity – based segmentation. The proposed technique shows better results [15].

A. Mohd. Khuzi, et al, 2009, studied MIAS database, masses are grouped into either speculated, circumscribed or ill-defined. Additional information includes location of masses centres and radius of masses. The extraction of the textural features of ROIs is done by using gray level co-occurrence matrices (GLCM) which is constructed at four different directions for each ROI. Analysis of GLCM properties i.e. contrast, energy and homogeneity resulted in receiver operating characteristics (ROC) curve area of $Az = 0.84$ for Otsu’s method, $0.82$ for thresholding method and $Az = 0.7$ for K-mean clustering. ROC curve area of 0.8-0.9 is rated as good results. The authors’ proposed method contains no complicated algorithm. The detection is based on a decision tree with five criterions to be analyzed. This simplicity leads to less computational time. Thus, this approach is suitable for automated real-time breast cancer diagnosis system [10].

Dr.H.B.Kekre, et al, 2009, in their paper proposed segmentation using vector quantization technique. Here they used Linde Buzo and Gray (LBG) for segmentation of mammographic images. Initially a codebook of size 128 was generated for mammographic images. These
code vectors were further clustered in 8 clusters using same LBG algorithm. These 8 images were displayed as a result. This approach does not lead to over segmentation or under segmentation. For the comparison purpose displayed results of GLCM and watershed segmentation along with this method [9].

D.M.Garge, et al, 2009, reported a low cost wavelet based image processing technique using MATLAB for detecting the calcification on a relatively less sophisticated computing platform. The paper analyzed sample mammograms and reports, the success of the algorithms is the unit cost of patient treatment is much lower than existing one [7].

J. Subhash Chandra Bose, et al, 2010, presented an intelligent system that was designed to diagnose breast cancer through mammograms, using image processing techniques along with intelligent optimization tools such as GA and PSO. The suspicious region is extracted or segmented using two different approaches such as asymmetry approach and Markov Random Field (MRF) hybrid with Particle Swarm Optimization (PSO) algorithm [3].

3. Proposed Technique

This paper is basically concentrated to develop a technique to identify the abnormal growth of masses in breast using very simple algorithms. The tool will only identify the masses with some distinguishing features to ease further investigation. In this new process we have selected the digital mammogram, which has become the most effective technique for early breast cancer detection. The mammogram images used in this experiment are taken from the mini mammography database of MIAS [19]. The original MIAS Database (digitized at 50 micron pixel edge) has been reduced to 200-micron pixel edge and clipped/padded so that every image is 1024 pixels x 1024 pixels. All images are held as 8-bit gray level scale images with 256 different gray levels (0-255) and physically in portable gray map (.pgm) format. 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 three types of breast tissues i.e. Fatty, Fatty-glandular, Dense-glandular. Different types of abnormalities like calcification, well-defined or circumscribed masses, speculated masses and other ill-defined masses have also been considered by our study. We have considered more than hundred samples as a test case. The identification technique is divided into two parts. They follow after preprocessing of digital mammogram.

3.1. Preprocessing

Mammograms are medical images that are difficult to interpret, thus a preprocessing phase is needed in order to improve the image quality and make the segmentation results more accurate. The first step involves the removal of unwanted parts in the background of the mammogram. The main objective of this process is to improve the quality of the image, to make it ready for further processing. Removing the irrelevant parts of the image is done by increasing contrast of the mammogram using threshold value.
3.2. Formation of Homogeneous Blocks

By analyzing mammogram image, we segmented the mammogram into very small blocks. To reduce the complexity of the algorithm, we first degenerate the image into 2X2 pixel blocks. Check the intensity or the pixel value of the blocks and obtain the pixel value which presents maximum occurrence within the block. Propagate the value in the adjacent pixels of the block as demonstrated in figure 2. Now the entire block contains the same pixel value. So, the whole mammogram image now consists of 2X2 homogeneous blocks.

Now, the 2X2 homogeneous blocks are arranged to obtain a 4X4 pixel block containing 4 arranged 2x2 blocks. The similar process is continued with this 4X4 block as done with the initial 2X2 blocks. The process is demonstrated in figure 3. Now, the whole mammogram image consists of 4X4 homogeneous blocks. We repeat the same process to produce mammogram image of 8X8 homogeneous blocks by arranging 4 individual 4X4 blocks likewise as stated above.

At first we do not produce the 8X8 block. This is not done to reduce the complexity of the algorithm and to preserve the granular block pixel value property. The said process is depicted in Algorithm1.
3.2.1. Algorithm 01

Input: Preprocessed Mammogram Image (PMI)
Isiz = Size of the Image
B = Block Size i.e. 2, 4, 8 etc
P [255] = Calculate the Maximum Occurrence Pixel Value
Iseg [B][1024] = Segment of Image Read Each Time

Output: Homogeneous Image (HI)

Begin
Step1. Open PMI file.
Step2. Open HI file.
Step3. Loop J=0, Isize/(1024*B)
   Read 1024*B from PMI
   Loop I=0, 1024/B
      Loop K=0, 255
         P[K]=0
         K = K + 1
      End Loop
      Loop R=0, B-1
         Loop C = I*B, (I*B) + (B-1)
            P[Iseg[R][C]]= P[Iseg[R][C]]+1
            C=C+1
         End Loop
         R=R+1
   End Loop
   Hvalue=0
   Maxvalue=0
   Loop K=0, 255
      IF Maxvalue <= P[K]
         Hvalue = K
         Maxvalue = P[K]
      End IF
      K=K+1
   End Loop
   Loop R=0, B-1
      Loop C = I*B, (I*B) + (B-1)
         Iseg[R][C]= Hvalue
         C=C+1
      End Loop
      R=R+1
   End Loop
   I=I+1
End Loop
Step4. Close PMI and HI
End
3.3. Color Quantization

Uniform color quantization technique is used to break the color space of mammogram image into eight equal sized regions. Although color quantization is a lossy process but we utilize this. The idea is sampling the original mammogram image for color statistics, choosing a color map based on those statistics and mapping the colors to their representative properties in the color map. The 8X8 homogeneous mammogram image is now clearly segmented into different color regions and each region representing specific part and properties. The algorithm 2 describes the process logic.

3.3.1. Algorithm 02

Input: Homogeneous Image (HI)
Isize = Size of the Image
Byte = one byte of data offset
Pbyte = to hold the previous value of byte

Output: Color Quantized Image (CQI)

Begin
Step1. Open PMI file
Step2. Open HI file
Step3. Loop J=0, Isize
    Read one Byte from HI
    Loop I=0, 256
        IF Byte < I
            Byte = I - 32
            Break
        End IF
        I = I + 32
    End Loop
    IF J=0
        Pbyte = Byte
    ELSE
        IF Pbyte-Byte>32 OR Byte-Pbyte>32
            Pbyte = Byte
        ELSE
            Byte = Pbyte
        End IF
    End IF
    Write one byte to CQI
    J = J + 1
End Loop
Step4. Close HI, CQI

End

4. Test Result

The success of the proposed technique is determined by the extent to which potential abnormalities can be extracted from analogous mammograms based on analysis of their image. The MIAS Database is used to evaluate the proposed technique. More than hundred
bilateral image pairs were used for testing. A randomly selected set of bilateral pairs drawn from the database, with calcification, circumscribed masses, speculated masses and other ill-defined masses speculated and circumscribed lesions was used for the same to obtain results.

Major objective of the algorithms is to eliminate the non-masses area from the mammogram to identify the presence of abnormality clearly. The stage, intensity, type, feature and treatment can only be detected on the basis of type of masses, orientation of masses, shape and distribution, size, position of masses, density of masses, symmetry between two pair etc. The outputs of aforesaid algorithms are depicted in the following figures for masses and non-masses mammograms along with the histogram and colormap of the images.

Figure 4. Normal Mammogram with Histogram and Colormap after Preprocessing

Figure 5. Mammogram containing masses with Histogram and Colormap after Preprocessing

Figure 6. Normal Mammogram with Histogram and Colormap after Formation of Homogeneous Blocks
After processing, type of masses, orientation of masses, shape and distribution of masses, size of masses, position of masses, and density of masses are clearly visible. Histogram observations are very interesting. Presence of color blocks in mammogram with mass and normal mammogram is depicted in the table 1.

### Table 1. Ratio of Color Block in Mass and Normal Mammogram

<table>
<thead>
<tr>
<th>Color Palette</th>
<th>0</th>
<th>32</th>
<th>64</th>
<th>96</th>
<th>128</th>
<th>160</th>
<th>192</th>
<th>224</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammogram with Mass</td>
<td>9221</td>
<td>428</td>
<td>449</td>
<td>556</td>
<td>2143</td>
<td>2767</td>
<td>621</td>
<td>199</td>
</tr>
<tr>
<td>Normal Mammogram</td>
<td>9293</td>
<td>676</td>
<td>673</td>
<td>818</td>
<td>2817</td>
<td>1727</td>
<td>326</td>
<td>54</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.99</td>
<td>0.63</td>
<td>0.67</td>
<td>0.68</td>
<td>0.76</td>
<td>1.6</td>
<td>1.9</td>
<td>3.69</td>
</tr>
</tbody>
</table>

The ratio of color blocks present in mammogram with mass compared to normal mammogram, are substantially high in the higher part of the palette. Where the average ratio
of color region is between two mammograms is 1.365 but in color region 160, 192 and 224 the ratio is 1.6, 1.9 and 3.69. So, it is indicating presence of comparatively very large amount of opaque area in the mammogram with mass, than the normal one, as displayed in the figure 10.

![Figure 10. Ratio of Color Block vs. Color Palette of Mass and Normal Mammogram](image)

Using different colormap in figure 11 abnormality or the presence of mass is clearly visible along with other areas in distinction with normal mammogram. Determination of type of mass, calculation of orientation, shape and distribution of masses, estimation of size, position and density of masses and symmetry between two pair is absolutely clear. The said information is adequate for decision-making and future course of action.

![Figure 11. Normal Mammogram and Mammogram with Masses using Different Colormap.](image)

5. Result Analysis

Texture features have been proven to be useful in differentiating masses and normal breast tissues \[4][2][21]. Texture features are able to isolate normal and abnormal lesion with masses and micro-calculifications. In the experimental work, the texture features are extracted using gray level co-occurrence matrices (GLCM). The matrices are constructed at a distance of \(d = 1\) and for direction of \(0^\circ\) given as \(0^\circ, 45^\circ, 90^\circ\) and \(135^\circ\). A single direction might not give enough and reliable texture information. For this reason, four directions are used to extract the texture information for each masses and non-masses tiles area \[4][2][21][10].

The texture descriptors derived from GLCM are contrast, energy, homogeneity and correlation of gray level values. The contrast measures the amount of local variations present in an image, while energy is the sum of squared elements in GLCM. Energy may also be
referred as uniformity or the angular second moment. The homogeneity descriptor refers to the closeness of the distribution of elements in GLCM to the GLCM diagonal. Based on the database and resultant images derived from algorithms are tested with the range of values of contrast, homogeneity, energy of masses and non-masses tissues of 8x8 tile area are shown in Table II, Table III and Table IV, respectively. The graphical presentations of the contrast, homogeneity and energy values are also sited in Figure 12, Figure 13 and Figure 14, respectively. It is observed that the values of contrast, homogeneity and energy for image, containing masses and image containing non-masses are highly as per expectation and significantly discriminated. This has proven the usefulness of the proposed method using three texture descriptors in differentiating the masses and non-masses tissues.

Table 2. Contrast Value of Masses and Non Masses

<table>
<thead>
<tr>
<th>Contrast at direction θ</th>
<th>0°</th>
<th>45°</th>
<th>90°</th>
<th>135°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>0.0241</td>
<td>0.0285</td>
<td>0.0047</td>
<td>0.0282</td>
</tr>
<tr>
<td>Non Mass</td>
<td>0.0282</td>
<td>0.0354</td>
<td>0.0082</td>
<td>0.0356</td>
</tr>
</tbody>
</table>

Table 3. Homogeneity Value of Masses and Non Masses

<table>
<thead>
<tr>
<th>Homogeneity at direction θ</th>
<th>0°</th>
<th>45°</th>
<th>90°</th>
<th>135°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>0.9964</td>
<td>0.9942</td>
<td>0.9977</td>
<td>0.9943</td>
</tr>
<tr>
<td>Non Mass</td>
<td>0.9959</td>
<td>0.9936</td>
<td>0.9975</td>
<td>0.9935</td>
</tr>
</tbody>
</table>

Table 4. Energy Value of Masses and Non Masses

<table>
<thead>
<tr>
<th>Energy at direction θ</th>
<th>0°</th>
<th>45°</th>
<th>90°</th>
<th>135°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>0.4083</td>
<td>0.4073</td>
<td>0.4096</td>
<td>0.4073</td>
</tr>
<tr>
<td>Non Mass</td>
<td>0.4021</td>
<td>0.4009</td>
<td>0.4036</td>
<td>0.4009</td>
</tr>
</tbody>
</table>

Figure 12. Contrast Value at θ = 0°
6. Conclusion

It can be very difficult to decide who may have a breast cancer and who may have a non-cancerous breast condition. Advances in computing and telecommunications have resulted in the availability of a range of tools for use in mammography quality assurance and support system. The majority focuses on either enabling mammography to examine and diagnose cases, or providing image archives that serve as reference material. Limited emphasis has been placed on analyzing the diagnostic process used by mammography to reach a diagnosis and using this as a resource for improving diagnostic performance. This method has potential for further development because of its simplicity. This will motivate online or real-time breast cancer diagnosis in providing diagnostic opinion.
References


Authors

Indra Kanta Maitra (First Author)

The Author is Research Fellow of University of Calcutta, India and Senior System Analyst, B P Poddar Institute of Management and Technology. The author received Master in Computer Application in the year 2002 from IGNOU. He received Best Poster Award in 96th Indian Science Congress, 2009 at Shillong for in section of Information and Communication Science and Technology. He is Associate Member of Computer Society of India Kolkata Chapter, currently Managing Committee Member, Ex Student Coordinator (East) 2005-06 and Associate Member IEEE. He is Author of more than fifteen Research publications in National and International Journal and Conference including IEEE etc. Author published two text books for engineering students. Field of Specialization is Biomedical Image Analysis, Image Processing, Network Security, Data Structure, Programming Language etc.

Sanjay Nag

The Author is Research Fellow of University of Calcutta, India. The author received Master in Computer Application from IGNOU. He is Author of more than five Research publications in National and International Journal and Conference. Field of Specialization is Biomedical Image Analysis, Image Processing, Network Security, Data Structure, Programming Language etc.

Prof. Samir Kumar Bandyopadhyay (Correspondence Author)