Pre-processing of Mammogram Images

Prof. Samir Kumar Bandyopadhyay
Dept. of Computer Sc. & Engg,
University of Calcutta
92 A.P.C. Road, Kolkata – 700009, India

Abstract

Digital mammogram has become the most effective technique for early breast cancer detection modality. Digital mammogram takes an electronic image of the breast and stores it directly in a computer. High quality mammogram images are high resolution and large size images. Processing these images require high computational capabilities. The transmission of these images over the net is sometimes critical especially if the diagnosis of remote radiologists is required. The aim of this study is to develop an automated system for assisting the analysis of digital mammograms. Computer image processing techniques will be applied to enhance images and this is followed by segmentation of the region of interest (ROI). Subsequently, the textural features will be extracted from the ROI. The texture features will be used to classify the ROIs as either masses or non-masses.

Key Words Breast Cancer, Mammogram, Masses, Homogeneous Blocks, Colour Quantization

Introduction

Early detection is the best way to improve breast cancer prognosis since the causes of the disease are still unknown to us. Breast cancer has become a significant health problem in the world. Early detection is the primary solution for improving breast cancer prognosis. Screening can be done through digital mammogram, ultrasound, magnetic resonance imaging (MRI) or breast Ultrasound produces a good contrast image but it does not contain enough detailed information which can be found in digital mammogram.

Ultrasound produces a good contrast image but it does not contain enough detailed information which can be found in digital mammogram. Although MRI is more sensitive than digital mammogram, its results can also lead to false positive diagnosis which then leads to unnecessary additional tests, biopsies and increased patient anxiety.

In addition, the American Cancer Society recommends MRI for women with approximately 20-25% or greater lifetime risk of breast cancer, including women with strong family history of breast or ovarian cancer [1, 2]. The benefit of digital mammogram in helping to detect breast cancer early, obviously outweigh the other methods discussed previously. This support the fact that many studies have found that digital mammogram is better at detecting early stage breast cancer [3-14, 16-21].

Although digital mammogram has been proven to be an effective method for detecting breast cancer, interpretation of such mammograms requires skill and experience by a trained radiologist. It is noted that about 10-30% of breast lesions are missed during routine screening [3]. Independent double reading by two radiologists has been shown to improve the sensitivity, but it also increased the cost of the screening process [4]. Thus, computer-aided detection (CAD) can act as a second reader where the final decision will be made by the radiologist.

Imaging techniques play an important role in helping perform digital mammogram, especially of abnormal areas that cannot be felt but can be seen on a conventional mammogram. Before any image-processing algorithm of mammogram pre-processing steps are very important in order to limit the search for abnormalities without undue influence from background of the mammogram. These steps are needed only on digitized screen film mammography (SFM) images because digital mammography devices perform this step automatically during the image storing process. 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 easier.

Most mammogram images are large size and high resolution images that require specialized computing facilities to enables efficient processing. To facilitate the transmission of these images over computer networks image compression techniques are usually applied. In this paper, we present a size reduction algorithm that can be implemented on most mammogram images as a pre-processing step to reduce their size without affecting their quality.

Previous work from many authors used mammography image databases including this paper, especially MiniMIAS [12] and DDSM [13, 14], both comprised of scanned and digitized SFM images. In this paper we have proposed three steps of pre-processing of raw digital mammogram for future automatic elaborate investigation of abnormalities using some image processing technique.

Review Works

Mammograms are medical images that are difficult to interpret, thus a pre-processing phase is needed in order to improve the image quality and make the segmentation results more accurate. The first step involves the removal of artefact and unwanted parts in the background of the mammogram. Then, an enhancement process is applied to the digital mammogram.

Image enhancement operations can be used to improve the appearance of images, to eliminate noise or error, or to accentuate certain features in an image.

The contrast limited adaptive histogram equalization (CLAHE) method seeks to reduce the noise produced in homogeneous areas and was originally developed for medical imaging [15]. This method has been used for enhancement to remove the noise in the pre-processing of digital mammogram [16].

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 that might be present in the image. The experimental results of enhancement on digital mammogram using CLAHE have been reported [17].

Figure 1 shows a mammogram image with many fine details. Figure 2 shows the image in Figure 1 after 20x20 averaging filter has been applied to reduce the detail.

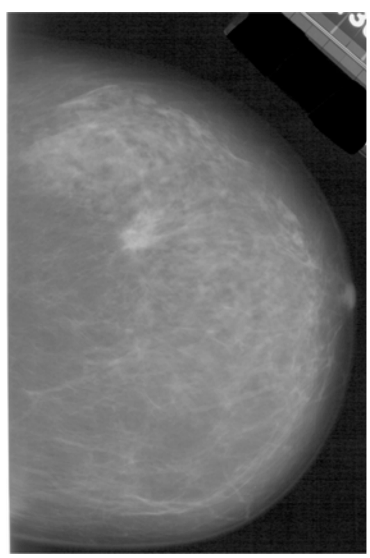

Figure 1 Mammogram image with fine details

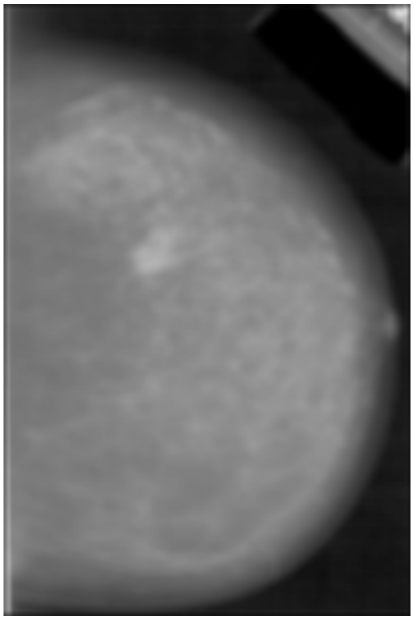

Figure 2 Image from Figure 3.9 after 20x20 average filter

In analysing mammogram image, it is important to distinguish the suspicious region from its surroundings. The methods used to separate the region of interest from the background are usually referred as the segmentation process. In other words, the process of dividing an image into distinct, meaningful regions is called image segmentation. This is useful in the analysis of digital mammogram images because the image is divided into various components and areas of interest.

There are numerous methods for image segmentation; among these are two basic types of image segmentation algorithms. This first group includes methods which attempt to differentiate regions of the image by locating the boundary or edges between different regions. These methods are referred to as edge based techniques. The second group of algorithms is those which attempt to divide the image into segments by grouping individual pixels into various regions. These are called region based.

Edge based segmentation methods attempt to define the regions of an image by locating the edges which form the boundaries for the regions. Figure 3 shows an image with simple regions. Using the Laplacian kernel method, the edges in Figure 3 are detected and shown in Figure 4.

Figure 3 Example image with simple region

Figure 4 Edges outlining regions in Figure 3

This operation works very well for segmenting the simple image. The benefits of this are that the method requires no training data or prior knowledge of the image contents. However, this method alone is not suitable for images with more complex regions.

Region based segmentation methods typically evaluate each pixel in the image and compares it to the other pixels to determine distinct groups. This often requires a training set of data or a classification scheme to cluster the pixels into distinct groups. The training area which yields the highest probably for a given pixel is considered the matching region of that pixel. Thus the pixel is classified into a distinct region.

Figure 5 shows the results of performing this segmentation method on the example image in Figure 3.

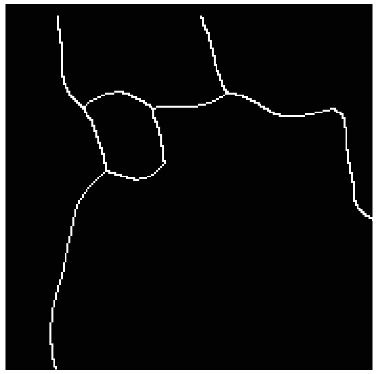

Figure 5 Poisson based region segmentation

Digital images are typically compressed to conserve space. These compressions can be either lossless or lossy. Lossless compression preserves the information in the original image exactly, while lossy compression stores an approximation of the original image data. Medical imaging applications require lossless compressions. The images used in this research were lossless JPEG images, which are converted to PNG images for compatibility reasons to other software packages. Both are lossless.

Database Resources

Most image processing systems applies a pre-processing stage as a first stage. The system we introduce here could aid radiologists by highlighting the suspicious regions in mammograms.

Proposed method is implemented and performed on 382 mammographic images from USF (university of South Florida) and MIAS databases (i.e., 64 from USF and the remaining from MIAS). The USF database is a Digital Database for Screening Mammography (DDSM) and it is publicly available. These images are available with the same specification (3000×4500 pixels with 16-bit pixel depth). This database is classified to four volumes to represent different types of diagnosis: normal, cancer, benign and benign without call back. Normal cases are formed for patients with normal exam results that have had previous normal exams in the last four years. A normal screening exam is one in which no further "work-up" is required. Cancer cases are formed from screening exams in which at least one pathology proven cancer is found. Benign cases are formed from screening exams in which something suspicious is found, but it turned out not to be malignant (by pathology, ultrasound or some other means). The term benign without call back is used to identify benign cases in which no additional films or biopsy is done to make the benign finding.

Proposed Method

Mammograms show a projection of the breast that can be made from different angles. The two most common projections are medio-lateral oblique and cranio-caudal. 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 will be shown in upper part of the image, which is superimposed over a portion of the breast. 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 earlier one for its advantage but pectoral muscle detection is one more difficult task in the breast segmentation process. Reason for detecting pectoral muscle is to remove. Suppression can help in some auto detect procedures such as finding bilateral asymmetry etc.

It is important to detecting the pectoral muscle and defines the region of interest (ROI) for further analysis. This operation is important in medio-lateral oblique (MLO), where the pectoral muscle, slightly brighter compared to the rest of the breast tissue, can appear in the mammogram. To detect the same it is very important to determine whether the mammogram of the left or the right breast is viewed. We consider right breast first. At first, a straight line (AB) is plotted between the left background of the mammogram and starting of actual part of breast image. In second step, is to determine middle point (C) at the top margin of the mammogram and plot a straight line (CD) from the middle point (C) to lower left corner point of the mammogram. The line CD cross the line AB at point E resulting an inverted right angle triangle (ACE) that is our region of interest (ROI) to detect the pectoral tissues from mammogram, the process is depicted in figure 6 and figure 7. Except the ROI rest part of mammogram is converted to background color to make said ROI prominent. Finally to reduce the computational complexity inverted right angle triangle (ACE) is cropped from the original mammogram shown in figure 8 for further processing.

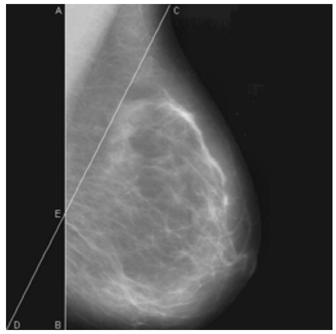

Figure 6 The Inverted Right Angle Triangle (ACE)

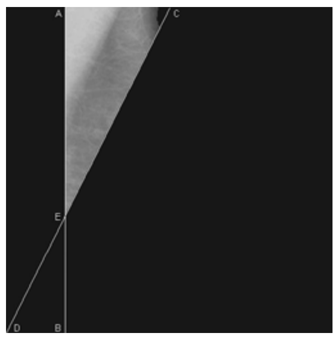

Figure 7 Detected ROI from Mammogram


Figure 8 Isolated Part of ROI from Mammogram for Further Processing

Now the question will arise that in all the cases the said triangle will cover the entire region of pectoral muscle. It gives absolute success ratio on 80 percent on different pairs of mammogram of different shapes and size. In case of left breast we will consider the mirror image to run the said process.

Conclusions

A set process of preprocessing has been presented with contrast enhancement, pectoral muscle detection and suppression. The results obtained over MiniMIAS database have shown a general good behavior. Using this preprocessing segmentation processes reduce noise and edge-shadowing effect, accurately detect region of interest (ROI) for pectoral muscle, suppress the pectoral muscle successfully. So, the processed mammogram can be used for the automated abnormalities detection of human breast like calcification, circumscribed masses, speculated masses and other ill-defined masses speculated, circumscribed lesions, asymmetry analysis etc.

References

[1] Breast Cancer Facts & Figures, 2009-2010, American Cancer Society, Inc.

[2] Norum J. Breast cancer screening by mammography in Norway. Is it cost-effective? Ann Oncol 1999, 10: 197-203.

[3] Michaelson J, Satija S, Moore R, et al. The pattern of breast cancer screening utilization and its consequences. Cancer. Jan 1 2002; 94(1): 37-43.

[4] Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist: Estimates of interobserver agreement and potential delay in cancer detection in the national screening study. Invest Radiol 1990, 25: 971-6.

[5] Wallis MG, Walsh MT, Lee JR. A review of false negative mammography in a symptomatic population. Clin Radiol 1991, 44: 13-5.

[6] Sterns EE, “Relation between clinical and mammographic diagnosis of breast problems and the cancer/ biopsy rate,” Can. J. Surg., vol. 39, n°. 2, p 128-132, 1996.

[7] Highnam R and Brady M, Mammographic Image Analysis , Kluwer Academic Publishers, 1999. ISBN: 0-7923- 5620-9.

[8] Kekre HB, Sarode Tanuja K and Gharge Saylee M, "Tumor Detection in Mammography Images using Vector Quantization Technique", International Journal of Intelligent Information Technology Application, 2009, 2(5):237-242

[9] FDA Web site, http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMQSA/mqsa.cfm

[10] RadiologyInfo.org developed jointly by Radiological Society of North America and American College of Radiology

[11] National Cancer Institute (NCI) Web site, http://www.cancernet.gov

[12] Suckling J., Parker J., Dance D.R., Astley S., Hutt I., Boggis C.R.M., Ricketts I., Stamatakis E., Cernaez N., Kok S.L., Taylor P., Betal D., Savage J., "The Mammographic Image Analysis Society Digital Mammogram Database", Proceedings of the 2nd International Workshop on Digital Mammography, York, England, 10–12 July 1994, Elsevier Science, Amsterdam, The Netherlands, pp. 375-378.

[13] Heath M., Bowyer K., Kopans D., Moore R., Kegelmeyer P. Jr., "The Digital Database for Screening Mammography", Proceedings of the 5th International Workshop on Digital Mammography, Toronto, Canada, 11-14 June 2000, Medical Physics Publishing, 2001, pp. 212-218.

[14] Sanmeet Bawa, A thesis on "Edge Based Region Growing", Department of Electronics and communication Engimeering, Thapar Institute of Engineering & Technology (Deemed University), India, June – 2006

[15] P. J. Besl and R. C. Jain, “Segmentation through variable-order surface fitting,” IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-IO, pp.167-192, 1988

[16] R. M. Haralick and L. G. Shapiro, “Image segmentation techniques,” Comput. Vis. Graph. Image Process., vol. 29, pp. 100-132, 1985.

[17] P. K. Sahoo, S. Soltani, and A. K. C. Wong, “A survey of thresholding techniques,” Comput. Vis.. Graph. Image Process., vol. 41, pp. 233-260, 1988.

[18] L. S. Davis, “A survey of edge detection techniques,” Compur. Graph. Image Process., vol. 4, pp. 248-270, 1975.

[19] S. W. Zucker, “Region growing: Childhood and adolescence,” Comput. Graph. Image Process., vol. 5, pp. 382-399, 1976.

[20] S. L. Horowitz and T. Pavlidis, “Picture segmentation by a directed split-and-merge procedure,” Proc. 2nd Int. Joint Conf Pattern Recognit., 1974, pp, 424-433.

[21] F. Meyer and S. Beucher, “Morphological segmentation,” J. Vis. Commun. Image Represent., vol. 1, pp. 2146, 1990