Segmentation of the Breast Region in Digital Mammograms and Detection of Masses

Armen Sahakyan

Digital Signal and Image Processing Lab

Institute for Informatics and Automation Problems of NAS of RA

Yerevan, Armenia

Hakop Sarukhanyan

Digital Signal and Image Processing Lab

Institute for Informatics and Automation Problems of NAS of RA

Yerevan, Armenia

Abstract—**The mammography is the most effective procedure for an early diagnosis of the breast cancer. Finding an accurate and efficient breast region segmentation technique still remains a challenging problem in digital mammography. In this paper we explore an automated technique for mammogram segmentation. The proposed algorithm uses morphological preprocessing algorithm in order to: remove digitization noises and separate background region from the breast profile region for further edge detection and regions segmentation.**

Keywords - mammography; image segmentation; ROI; masses detection; breast region

I.   INTRODUCTION

Breast cancer stays in the first place among women malignant neoplasia structures list (about 30%). According to Worldwide Health Corporation it is being #1 of the fundamental reasons of the women's average age mortality. The National Cancer Institute estimates that one out of eight women develops breast cancer at some point during her lifetime [1].

The goal of mammography is to provide early detection of breast cancer through low-dose imaging of the breast. Mammography is considered to be the most efficient technique for identifying lesions when they are not palpable and when there are structural breast modifications [2]. It shows to the physician differences in breast tissue densities and these differences are fundamental to a correct diagnosis. At present, there are no effective ways to prevent breast cancer, because its cause remains unknown [3]. Therefore urgency and importance of mammography image processing is obvious.

Computer-Aided Detection (CADe) and Diagnosis (CADi) systems are continuously being developed aiming to help the physicians in early detection of breast cancer. These tools may call the physician's attention to areas in the mammography that may contain radiological findings. In digital mammography, segmentation is the process of partitioning mammograms into regions, aiming to produce an image that is more meaningful and easier to analyze [4]. After being segmented, the mammogram or the mass lesion region can be further used by physicians, helping them to take decisions that involve their patients' health.

This paper is organized as followed. Section II gives some knowledge about image segmentation. Section III describes an image segmentation technique presented in this paper. In Section IV are shown experimental results of described techniques. In the next section the conclusion and future work are given.

II.   MAMMOGRAM SEGMENTATION

Mammogram segmentation usually involves classifying mammograms into several distinct regions, including the breast border [5], the nipple[6] and the pectoral muscle. The principal feature on a mammogram is the breast border, otherwise known as the skin-air interface, or breast boundary. The breast contour can be obtained by partitioning the mammogram into breast and non-breast regions. The extracted breast contour should adequately model the soft-tissue/air interface and preserve the nipple in profile.

There are a number of problems associated with the accurate segmentation of the breast region. Firstly, owing to the nature of x-ray each pixel in a mammogram represents two or more tissues; indeed all pixels contain a component due to attenuation by skin. Superimposition of different tissue types makes it difficult to differentiate between different regions. As a result of the mammogram acquisition process, there is a region of decreasing contrast near the breast contour where the breast tapers off. This region constitutes the uncompressed region of the breast commonly referred to as the "breast edge", and is caused by a lack of uniform compression of the breast tissue. This tapering effect causes a lack of visibility along the peripheral region of the mammogram, making it difficult to perceive the breast contour and identify the nipple position. The process of digitization may further decrease this visibility through the addition and accentuation of noise.

There have been various approaches proposed to the task of segmenting the breast profile region in mammograms. Some of these have focused on using threshold [7][8], gradients [9], modeling of the non-breast region of a mammogram using a polynomial [10], or active contours [11].

III.   SEGMENTATION ALGORITHM

Digital mammogram images were acquired from the mini-MIAS database [12].

Images acquired consist of left and right breast images of fatty, fatty-glandular and dense-glandular breasts. The acquired mammogram images are classified into three major cases: malignant, benign and normal, all of which are subdivided into five categories as follows: 1) Circumscribed masses 2) Speculated masses 3) Ill-defined masses 4) Architecturally distorted masses 5) Asymmetrical masses. The images are digitized at 200 micron pixel edge and padded in order to obtain all images with a size of 1024×1024 pixels.

A. Digitization Noise Removal

Digitization noises such as straight lines (see Fig. 1(a)) present in the majority of acquired mammogram images are filtered using a two-dimensional (2D) Median Filtering approach in a 3-by-3 neighborhood connection [13]. Each output pixel contains the median value in the 3-by-3 neighborhood around the corresponding pixel in the input images. The edges of the images however, are replaced by zeros. Fig. 1(a) shows the digitization noise present in a mammogram image and Fig. 1(b) shows the same image after noise removal.

B. Image Enhancement

Problems with image acquisition such as scanner-induced artifacts, excessive background noise, scratches and dust artifacts could influence the reliability of this algorithm. The mammogram presented in Fig. 2(a) has a highly non-uniform background and very little contrast in the area above the core breast tissue region. So image enhancement is required before segmentation. As enhancement technique was selected contrast-limited adaptive histogram equalization (CLAHE) [14]. It is a well-known technique of adaptive contrast enhancement. The normal and adaptive histogram equalization may over-enhance the noises and sharp regions in images due to the integration operation. It yields large values in the enhanced image for high peaks in the histogram of the nearly uniform regions in the original image. To solve this problem, the CLAHE uses a clip level to limit the local histogram in order to limit the amount of contrast enhancement for each pixel. This clip level is a maximum value of the local histogram specified by users. An interactive binary search process is used to redistribute the pixels which are beyond the clip level. The CLAHE algorithm has following steps: 1) divide the original image into contextual regions; 2) obtain a local histogram for each pixel; 3) clip this histogram based on the clip level; 4) redistribute the histogram using binary search; 5) obtain the enhanced pixel value by histogram integration. The result of CLAHE technique is shown in Fig. 2(b).

C. Background Separation

Radiopaque artifacts such as wedges and labels in the mammograms images are removed using threshold technique and morphological operations [15][16].

Fig. 3(a) shows a mammogram image with a radiopaque artifact present. After the grayscale mammogram images are converted into binary, as shown in Fig. 3(b) for the image in Fig. 3(a).
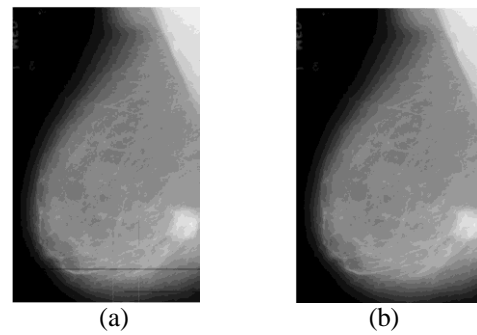

(a)   (b)

Figure 1.   Mammogram digitization noise removal using 2D median filtering. (a) Original image (b) Filtered image after noise removal
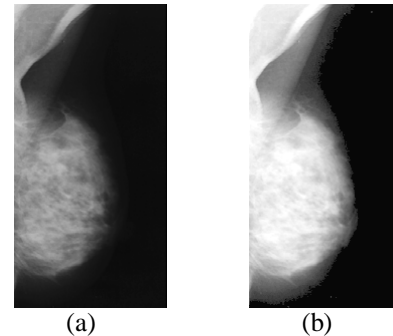

(a)   (b)

Figure 2.   Mammogram enhancement using CLAHE. (a) Original image (b) Enhanced image

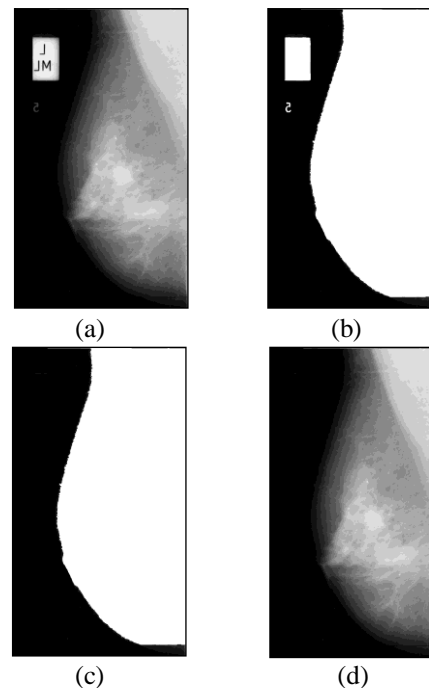(a)   (b)

(c)   (d)

Figure 3.   Suppression of artifacts and labels from a mammogram (a) Original image (b) Threshold image (c) Largest area (object) selected from threshold image 3(b) (d) Mammogram image with radiopaque artifacts suppressed

The algorithm steps to find artifacts and labels and to separate breast profile are as follows:

- Convert grayscale mammogram into binary using threshold technique.

- Binary image objects are labeled and number of pixels in all objects is counted.

- All binary objects are cleaned except the largest one: breast profile (Fig. 3(c)). After which morphological operation to remove isolated pixels is applied.

- After algorithm checks all pixels in a binary image and sets a pixel to 1 if five or more pixels in its 3-by-3 neighborhood are 1's, otherwise, it sets the pixel to 0.

- The resulting binary image is multiplied with the original mammogram image to get the final image without artifacts and etc.

D. Edge Detection

Edge detection is used for getting edge map of the breast region. In the first step horizontal scanning is performed. If any change of pixel intensity is observed it is marked by a black pixel indicating a horizontal edge point. This process is continued for all rows of pixel data to obtain a horizontal edge map. In the next step, image is scanned vertically. Continuing the process for all the columns a vertical edge map image is obtained. Finally, the horizontal edge map is merged with vertical edge map by performing a logical OR operation on the two image files, to obtain the edge map of mammogram image. The algorithm steps for horizontal image map are:

- Scan the image array horizontally from left-most pixel to right-most pixel from first row to last row and take the first pixel intensity value as a reference value.

- Compare intensity of subsequent pixels with the reference value. If the same value go on to next pixel.

- If the value differs, change the value of reference value to the pixel intensity value and mark the pixel black.

- If the last row and column pixel is not reached then go to Step2.

Steps for horizontal image map are the same as implemented for rows.

E. Segmentation of Breast ROI

After previous steps we get the ROI of the Breast. We now consider the edge map that corresponds only to the ROI of the breast. The edge map indicates various closed structures within the breast region that corresponds to the different anatomical regions of the breast. The objective is to identify these regions on the mammogram image. The algorithm starts by identifying the left baseline of the breast image from the edge map. Then a line is drawn vertically from top to bottom identifying the left boundary of the breast. Then the breast boundary is scanned on the right side to locate the rightmost pixel on the breast contour. After the pixel is located another vertical line is drawn from top to bottom passing through the rightmost pixel thus partitioning the image only to the breast ROI.
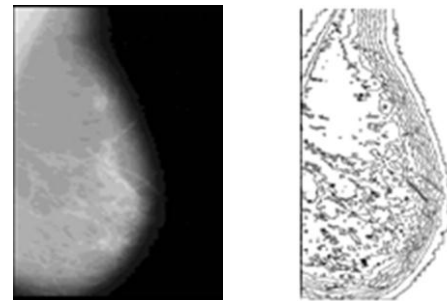

Figure 4. Mammogram after edge detection algorithm

This process optimizes the algorithm and increases the processing efficiency. At this stage is tried to locate all edge paths that are circular or terminate either on the left base line or the bottom of the image, forming a closed structure. Locating all the edge paths that originates from the top margin line namely the first row of the image is started.

The algorithm travels each individual path and stores them on the plotting list. This list is plotted on another image if the edge path is circular or end on the last row of the image or the vertical line representing the left boundary of the breast.

After all pixel paths on the first row are traversed the algorithm repeats similar scanning and traversal of all pixel paths from the row that is indicated by dividing the image vertically into sixteen segments. The first row of each segment is used for locating pixel paths for traversal. All complete paths are then plotted on another image thus providing regions of the breast. The algorithm steps to detect regions are:

- Scan the image from the left side of the image to locate the leftmost pixel of the breast region and draw a vertical line along this pixel from top to bottom representing the left baseline or boundary.

- Scan the image from the right side of the image to locate the rightmost pixel of the breast region and draw a vertical line along this pixel from top to bottom.

- Partition the obtained rectangle horizontally into sixteen segments and start with the first row of the first segment.

- Scan the enclosed rectangle from the right side to left, from the first row of the segment.

- Obtain a pixel that is black indicating an edge path, traverse the pixel path by considering all the surrounding pixels in a clockwise priority and consider the pixel with the highest priority.

- The pixels that surrounded the edge pixel, but are of lower priority are stored in a history stack to be used only if the traversal process reaches a dead end.

- If a dead end is reached, pop out from the history stack a lesser priority pixel and continue with the traversal process.

- Store the pixels traversed in a plotting list for plotting.

- Traversal continues to the next pixel till it reaches the left baseline or the bottom of the image or the start position is reached.

- If the traversal is terminated, the plotting list is erased and continues from Step5. Else plot pixels from the plotting list.

- Continue to Step4 till all black pixels, indicating an edge path, is traversed.

- Move to the first row of the next segment and continue from Step4 to Step9.

IV. EXPERIMENTAL RESULTS

To demonstrate the robustness of the algorithm it has been tested on mammograms with differing breast tissue densities. Overall, for the mammograms evaluated, the mean values for the quality measure for breast region detections were 0.95, signifying that the algorithm seems extremely robust with respect to density types. In Fig. 5 is shown experimental results of detection of breast contours of mammogram. In Fig. 6 is shown results of edge detection and regions segmentation of mammogram.

V. CONCLUSION AND FUTURE WORKS

There are a number of factors which make it difficult to postulate the exact effect digital mammograms on a particular segmentation algorithm. The first of these relates to acquisition parameters, such as exposure time and energy level, which influence the quality of the image registered on film. Secondly, segmentation of the breast region from the background is further hampered by the tapering nature of the breast.

As we can see from the results, mammography segmentation using technique presented in this paper is efficient. Edge detection and regions segmentation algorithms work more efficient when images are preprocessed. For the future work it is planned improvement of the algorithm to derive a smoother breast region contour for image preprocessing. Improvement of edge detection and region segmentation algorithms and abnormalities detection (mass, tumors or calcifications) in segmented images is planned.
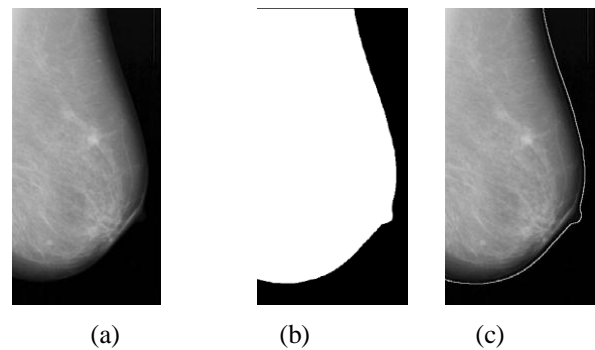

Figure 5. Detection of breast contour of the mammogram (a) Original, (b) Segmented mask, (c) Breast contour superimposed on (a).
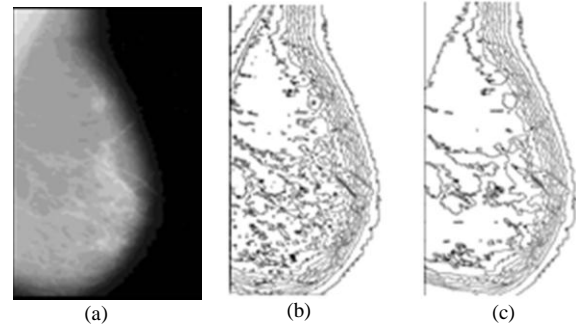

Figure 6. Segmentation of the breast region: (a) original mammogram, (b) mammogram after edge detection, (c) mammogram after regions segmentation.

REFERENCES

[1] L.-M. Wun, R. M. Merrill, and E. J. Feuer, "Estimating Lifetime and Age-Conditional Probabilities of Developing Cancer," Lifetime Data Analysis, vol. 4, pp. 169-186, 1998.

[2] Swanson, G. M. (1992). Breast cancer in the 1990's. Journal of American Medical Women's Association, v. 47, p. 140-148.

[3] D. E. Stewart, et al., "Attributions of cause and recurrence in long-term breast cancer survivors," Psycho-Oncology, vol. 10, pp. 179-183, 2001.

[4] Shapiro, Linda G. & Stockman, George C. (2002). "Computer Vision". Prentice Hall. ISBN 0-13-030796-3

[5] R. Chandrasekhar, and Y. Attikiouzel, "Automatic Breast Border Segmentation by Background Modelling and Subtraction", in 5th International Workshop on Digital Mammography (IWDM), (Yaffe M. ed.), Medical Physics Publishing, Madison, USA, pp. 560–565.

[6] R. Chandrasekhar, and Y. Attikiouzel, "A Simple Method for Automatically Locating the Nipple on Mammograms", IEEE Transactions on Medical Imaging, vol. 16, pp. 483–494, Oct. 1997.

[7] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, and K. Doi, "Automated Segmentation of Digitized Mammograms", Academic Radiology, vol. 2, no. 2, pp. 1–9, 1995.

[8] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3 ed.: Pearson Prentice Hall, 2007.

[9] A.J. Mendez, P.J. Tahoces, M.J. Lado, M. Souto, J.L. Correa, and J.J. Vidal, J.J, "Automatic Detection of Breast Border and Nipple in Digital Mammograms", Computer Methods and Programs in Biomedicine, vol. 49, pp. 253–262, 1996.

[10] R. Chandrasekhar, and Y. Attikiouzel, Y. Automatic, "Breast Border Segmentation by Background Modeling and Subtraction", in: M.J. Yaffe (Ed.), Proceedings of the 5th International Workshop on Digital Mammography (IWDM), Medical Physics Publishing, Toronto, Canada, 2000, pp. 560–565.

[11] M. A. Wirth, A. Stapinski, "Segmentation of the breast region in mammograms using active contours", in Visual Communications and Image Processing, pp.1995–2006.

[12] J. Suckling et al., "The Mammographic Image Analysis Society digital mammogram database", Exerpta Medica., vol. 1069, pp. 375– 378, 1994.

[13] T. Huang, G. Yang, G. Tang, " A fast two-dimensional median filtering algorithm", Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 27, 1979

[14] S. M. Pizer, et al., "Adaptive histogram equalization and its variations," Computer Vision, Graphics, and Image Processing, vol. 39, pp. 355-368, 1987.

[15] A. Sahakyan, "Segmentation of Mammography Images Enhanced by Histogram Equalization", Mathematical Problems of Computer Science 35, Armenia, pp. 109 - 115, 2011.

[16] A. Sahakyan, H. Sarukhanyan, "Automatic Segmentation of the Breast Region in Digital Mammograms", Computer Science and Information Technologies, Proceedings of the Conference, pp. 386 - 389, Yerevan, Armenia, September 26-30, 2011.