Mask Extraction from Manually Segmented MIAS Mammograms

Mario Mustra [1], Mislav Grgic [1], Renata Huzjan-Korunic [2]

[1] University of Zagreb, Faculty of EE and Comp, Unska 3 / XII, HR-10000 Zagreb, Croatia
[2] Clinical Hospital Dubrava, Avenija Gojka Suska 6, HR-10000 Zagreb, Croatia
mario.mustra@fer.hr

Abstract - The first step in computer aided detection (CAD) in mammography relies on accurate image segmentation. Testing of CAD algorithms should include comparison with other proposed methods in order to show how a new method compares with ones presented before. Two most popular mammographic databases, which are publicly available, consist of scanned films. This presents a segmentation challenge in order to achieve the best possible results. The major problem which arises in testing automatic segmentation algorithms is defining the ground truth. In this paper we present a method for automatic extraction of manually segmented breast segmentation masks from MIAS database. The manual segmentation process has been performed by the professional radiologist on printed images, which have later been digitized and breast pectoral muscle masks have been extracted. The extraction process differs for breast mask extraction and for pectoral muscle mask extraction because of different manually drawn segmentation line properties.

Keywords - Image Registration; Region Growing; Computer Aided Detection; Image Segmentation

I. INTRODUCTION

Computer-aided detection (CAD) is an important and developing field in modern medical systems. It is developed to help radiologists especially in screening examinations when a large number of patients is examined. CAD allows integration between automatic knowledge extraction and professional diagnosis performed by radiologists, where a good CAD system should prevent possible false negative results while avoiding too large false positive count. In the case of screen-film mammography (SFM), A/D conversion or digitalization of mammogram films, should be performed before a CAD systems can be used. The digitalization process requires additional effort and also introduces artifacts caused by scanning imperfections. Films usually suffer from internal imperfections and different sensitivity which causes non-uniform background and light leakage. All these things present a problem to an automatic segmentation algorithm because breast skin-air interface has similar intensity and therefore is hard to detect the actual breast boundary. Another problem caused by scanning is image alignment. After images have being digitized, their orientation, rotation and size often need to be corrected. For CAD purposes, images need to have known orientation and rotation and we can say that digitization process and image alignment antecede all CAD algorithms. Image segmentation is the first (pre)processing step in development of any CAD system [1]. In the case of mammography CADs, segmentation of the breast tissue from the background is the necessary first step because SFM suffers from uneven background and identification tags as well as film artifacts which need to be removed [2]. In the case of digital mammograms, segmentation of the breast tissue from the background is a much easier task, because precalibration of the device takes care of maintaining background intensity equal to zero. In this paper we propose a method for semi-automatic breast segmentation and mask extraction. Printed mammograms are manually segmented, both breast skin-air interface and pectoral muscle, and corresponding masks are automatically extracted using region growing and flood fill methods. The proposed algorithm consists of printing complemented images, manual segmentation and scanning after which masks are being extracted. This process results in image segmentation masks which can be used to test automatic segmentation algorithms. If algorithms are being tested against the same data set, the obtained results should give a correct algorithm quality evaluation. We have chosen to use the mini-MIAS database [3], which is publicly available and widely used in CAD algorithm development and testing. This database contains 322 scanned SFM mammograms. All images are of the same size, 1024×1024 pixels with 8 bits per pixel. Although this database is rather outdated, besides DDSM [4], it still presents the most common choice for CAD in mammography.

The image registration method considered in this paper can be classified as feature detection method [5]. Features which we have chosen to concentrate on and which should provide correct image transformation coefficients are corner positions. Corner detection has been achieved through line detection using Hough transform [6]. This eliminates the need to compute image transformation matrices and we only have to gather corner coordinates from which scanned images can be arbitrary rotated and resized to match the size of original image. This registration method has similar properties to the rigid image transformation [7].

The mask extraction has been done using the seeded region growing method [8]. Pectoral muscle masks have been detected using a standard seeded region growing method while breast masks were detected using the flood fill algorithm [9] with previous border detection. The reason for

this lies in non-optimal contrast of the scanning process and difficulty to set appropriate threshold for the region growing termination.

This paper is organized as follows. In Section II the entire segmentation and the mask extraction procedure is explained. In Section III the mask extraction results are presented and discussed. Section IV draws the conclusion.

II. IMAGE SEGMENTATION AND MASK EXTRACTION

The proposed mask extraction process consists of steps shown on diagram in Fig. 1.

```
┌─────────────────────┐
│      Printing       │
│  inverted images    │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│      Manual         │
│   segmentation      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   Scanning and      │
│   registration      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   Mask extraction   │
└─────────────────────┘
```
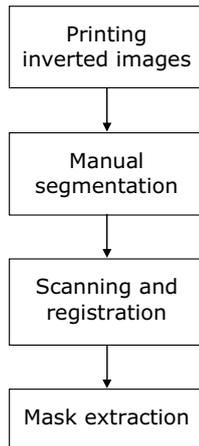
Figure 1.   Diagram of the mask extraction process.

The first step in this process is image inversion which is done for the reason of higher printer efficiency and easier line drawing in manual segmentation step. The second step is manual segmentation performed by the professional radiologist. This segmentation included both breast skin-air interface and pectoral muscle. Segmentation lines for the skin-air interface were drawn by a black marker and the pectoral muscle segmentation lines were drawn using a white marker. This was done to boost the line contrast as much as possible against the background. The third step includes both scanning and registration, because its output should provide an image on which the region growing algorithm can be used. The registration process has two steps, image alignment followed by translation and rotation of the breast area if necessary. The goal is to make the pectoral muscle area situated in top left corner of the image. The final step is mask extraction. This step can also be divided into two different steps. One is pectoral muscle mask extraction and other is breast mask extraction.

The result obtained with printing of inverted images results in 16×16 cm large grayscale image. The resulting image quality is not very important, because it is only used to detect pectoral muscle and breast boundaries. An example of a scanned image printed on a monochrome laser printer is shown in Fig. 2.
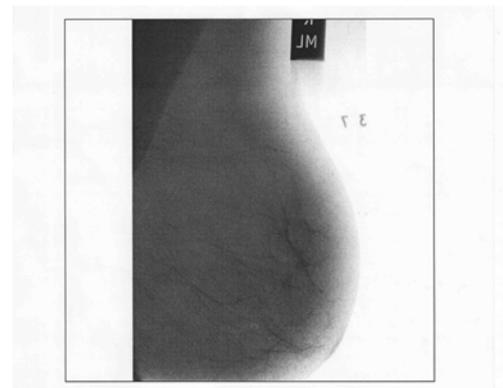
Figure 2.   Scanned image which was printed on a monochrome laser printer.

Fig. 3 shows the same printed image after manual segmentation and scanning.
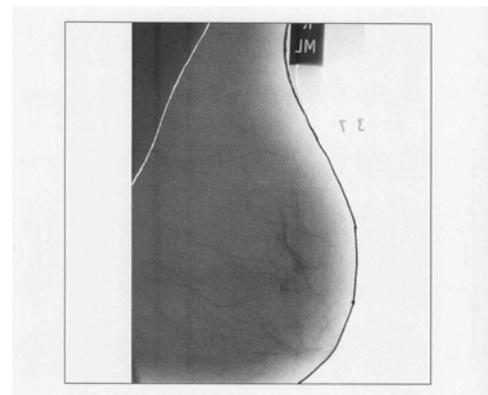
Figure 3.   Scanned manually segmented image.

From Fig. 3 it is visible that the scanning process resulted in a slight rotation and a bit altered overall contrast with some visible artifacts. This could be avoided by using a better quality scanner, but the proposed mask extraction method works well with lower quality scanning process as well. For the registration of the entire scanned image with the original image we have used Hough transform (1) with image corners as markers:

$$r = x\cos(\theta) + y\sin(\theta), \qquad (1)$$

where r represents the distance between the line and the origin, while θ is the angle of the vector from the origin to the observed point. After detecting the actual corner coordinates, it is possible to perform image rotation and resizing in both horizontal and vertical dimensions if those coefficients differ one from another. The rotation angle has been calculated as the mean value of differences between corner coordinates on each of four edges (2):

$$\varphi = \arctan\left(\frac{n1 - n2 + n4 - n3 + m3 - m2 + m4 - m1}{m1 - m2 + m4 - m3 + n2 - n3 + n1 - n4}\right), \quad (2)$$

where *mi* and *nj* are *x* and *y* coordinates of the detected corners, as shown in Fig. 4.
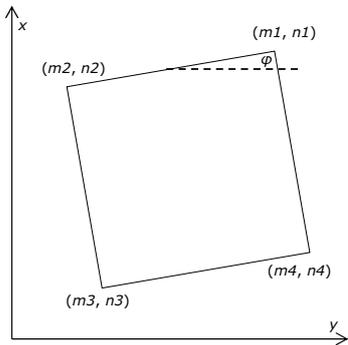
Figure 4.   Corner coordinates and rotation angle φ of the scanned image.

The final step in registration process is positioning image in the top left corner and rotation of mirrored images. This is performed on the original images using vertical line detection with Hough transform. Vertical lines are created using a simple Sobel edge detector on a breast mask created by simple thresholding. The final result of the proposed method on the same image is shown in Fig. 5.
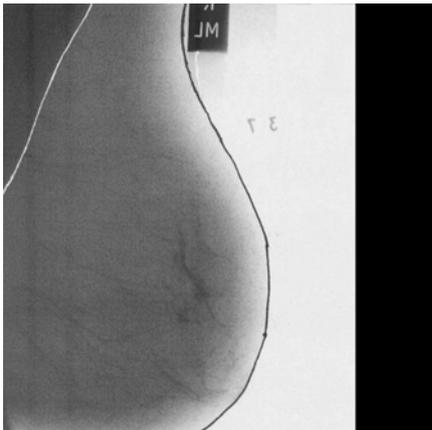
Figure 5.   Registered scanned image positioned in the top left corner.

After the registration and alignment processes are finished it is possible to extract the pectoral muscle and breast masks. The pectoral muscle masks are extracted using the seeded region growing algorithm. For this purpose we have used the 4-connected neighborhood region growing method and the result of pectoral muscle extraction for the same image is shown in Fig. 6 where only a cutout of the entire image only containing pectoral muscle mask is shown.
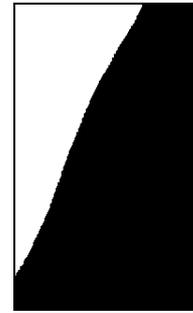
Figure 6.   Cutout of the image showing the extracted pectoral muscle mask.

Region growing would give the same result for the automatic breast mask extraction, but because of a low overall contrast, this method should be implemented in a different manner. First thing which needs to be done is to make sure that the image contrast is as high as possible. After stretching image intensities to obtain the maximal contrast, we have proceeded with thresholding at 25% of the overall intensity range. The obtained binary image was first morphologically closed with structuring element larger than the thickness of the segmentation line we wish to detect, in our case the structuring element was a square 11 pixels wide. Closed image was then opened with the same structuring element and dilated with a larger structuring element, a square 15 pixels wide. Those two images were then subtracted. The subtraction process results in a segmentation line with some "noise" inside and outside of the breast area caused by morphological operations. After line detection, the breast area is filled using flood fill algorithm. When the breast area was filled we have used the islanding method which finds all groups of pixels that are interconnected to remove smaller objects that do not belong to the breast mask and are outside of the breast area. This process results in a breast mask shown in Fig. 7.
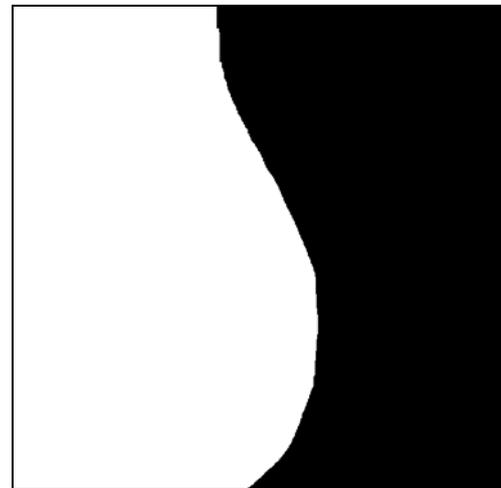
Figure 7.   Breast segmentation mask for image "mdb006".

III. Experimental Results

The experimental test set which we have used contains 322 8-bit grayscale images with size of 1024×1024 pixels belonging to the mini-MIAS database. All images have been prepared for printing and printed on a standard monochrome laser printer resulting in the size of 16×16 cm on an A4 paper format. After manual segmentation by the professional radiologist, images have been scanned and stored in grayscale JPEG format. Automatic registration algorithm has been applied to all images in the test set. Registration was followed by the pectoral mask extraction using the region growing method and the breast mask extraction using a segmentation line detection and the flood fill algorithm. The proposed method works well with all images in the MIAS database except images "mdb151" and "mdb152", because those images do not show pectoral muscle and are not captured within the MLO projection standard. Segmentation masks that are obtained from this manual segmentation can be used in testing of the performance of automatic segmentation algorithms in mammography, because they can be considered as the ground truth. Fig. 8 shows the result of segmentation of original MIAS mammogram "mdb006" with the mask extracted using the method described in this paper.
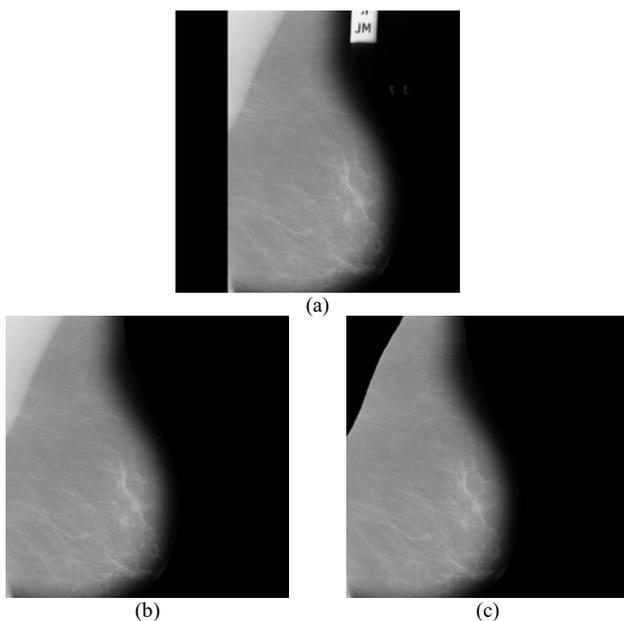
Figure 8. (a) Original image "mdb006". (b) Application of the breast mask to the original image. (c) Application of the pectoral muscle mask to the segmented breast image.

IV. Conclusion

In this paper we have presented a method for automatic extraction of the manually segmented pectoral muscle and breast masks from the mini-MIAS mammograms. The proposed method includes image registration using markers and Hough transform and region growing and flood fill algorithms to create binary masks from manually segmented printed mammograms. As markers, we have used image corners which were detected using Hough transform. The proposed method works well on the entire mini-MIAS database containing 322 images. The only two exceptions are images "mdb151" and "mdb152" on which pectoral muscle is not visible and images do not correspond to the MLO projection standard. Because of non optimal scanning process, the overall contrast was not the best possible and therefore we needed to use line detection method with the flood fill algorithm instead of the region growing algorithm. The results obtained using these methods are accurate pectoral muscle and breast masks which can be used as ground truth for testing of automatic segmentation algorithms. Our future work in this field will be development of automatic segmentation algorithms and testing of currently available algorithms in order to achieve results which are as close as possible to the masks obtained with the method described in this paper.

References

[1] T. Ojala, J. Nappi, O. Nevalainen, "Accurate segmentation of the breast region from digitized mammograms", Computerized Medical Imaging and Graphics, Vol. 25, Issue 1, pp. 47-59, January 2001.

[2] S. Tzikopoulos, H. Georgiou, M. Mavroforakis, N. Dimitropoulos, S. Theodoridis, "A fully automated complete segmentation scheme for mammograms", 16th International Conference on Digital Signal Processing, pp. 1-6, 5-7 July 2009.

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

[4] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, "The Digital Database for Screening Mammography", Fifth International Workshop on Digital Mammography, IWDM 2000, Medical Physics Publishing, 2001, pp. 212-218.

[5] B. Zitova, J. Flusser, "Image registration methods: a survey", Image and Vision Computing, Vol. 21, Issue 11, pp. 977-1000, 2003.

[6] D.H. Ballard, "Generalizing the Hough transform to detect arbotrary shapes", Pattern Recognition, Vol. 13, No. 2, pp. 111-122, 1981.

[7] J.B.A. Maintz, M.A. Viergever, "A survey of medical image registration", Medical Image Analysis, Vol. 2, Issue 1, pp. 1-36, March 1998.

[8] R. Adams, L. Bischof," Seeded region growing", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 16, Issue 6, pp. 641-647, June 1994.

[9] P. Soille, "Morphological Image Analysis: Principles and Applications", Springer-Verlag, pp. 173-174 , 1999.