Computers in Biology and Medicine 42 (2012) 75–82

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Fully automated gradient based breast boundary detection for digitized X-ray mammograms

Pelin Kus*, Irfan Karagoz [1]

Department of Electrical and Electronics Engineering, Gazi University, 06570 Maltepe, Ankara, Turkey

ARTICLE INFO

Article history:
Received 22 February 2011
Accepted 24 October 2011

Keywords:
Image analysis
Breast segmentation
Breast border
Mammogram
MIAS database

ABSTRACT

Accurate segmentation of the breast from digital mammograms is an important pre-processing step for computerized breast cancer detection. In this study, we propose a fully automated segmentation method. Noise on the acquired mammogram is reduced by median filtering; multidirectional scanning is then applied to the resultant image using a moving window 15×1 in size. The border pixels are detected using the intensity value and maximum gradient value of the window. The breast boundary is identified from the detected pixels filtered using an averaging filter. The segmentation accuracy on a dataset of 84 mammograms from the MIAS database is 99%.

1. Introduction

Breast cancer is the most frequent type and the second most dangerous type of cancer in women all over the world [1,2]. Early detection of breast cancer is key in reducing the high death rate. X-ray mammography is the clinical gold standard for the early detection of breast cancer [3]. Therefore, many image analyses and enhancement techniques for digital mammograms have been proposed for computer aided detection (CAD) applications [3–6]. In the pre-processing stage of these image processing algorithms, accurate breast boundary estimation is an important pre-requisite. Inadequate segmentation of breast boundaries could lead to a lesion located near the skin-line being overlooked [7]. Breast boundary estimation can also be used for image size reduction, image registration, and background elimination in automatic breast tissue classification [8,9]. Due to background noise, artifacts, and low density fatty tissue near the border, breast boundary segmentation is a very complicated task. An artifact is defined as any gray level variation in mammographic density not caused by breast tissue attenuations, but caused by the identification labels and type of X-ray view taken in mammograms [9].

There have been considerable efforts to develop breast boundary segmentation algorithms over the past two decades [10–31]. Most of the methodologies for breast boundary segmentation are mainly based on features extracted from the image intensity histogram. The studies started with Kallergi et al.'s [11] manual segmentation method and were followed by morphological treatments and thresholding applications reported by Yin et al. [12] and Bick and Giger [13], while Masek and Attikiouzel [14] used iterative adaptive thresholding. When compared with the other methods, histogram and thresholding based methods are easier and require less computation. Selection of the optimum threshold value, however, is quite difficult and the use of a single threshold value is usually not appropriate. Abdel-Mottaleb et al. [15] used multiple thresholding to obtain different breast masks in order to estimate the boundary. Overall, 98% of the results were found to be "acceptable" when tested on a dataset of 500 mammograms. Ojala et al. [16] used automatic thresholding that consisted of histogram thresholding, morphological filtering, and border fitting. The algorithm was tested over a range of screening mammograms digitized using different scanners. Over the 20 test images, 96% of the results were found to be "acceptable". Sun et al. [17] used a dependency approach, based on the observation that the Euclidean distances from the edge of the stroma to the true breast skin-line are usually uniform in mammographic views. They used adaptive thresholding to estimate an initial skin-line. This dependency approach detected the breast border with a mean error of 0.7 mm. This method may fail if the Euclidean distance between the stroma edge and the breast skin-line in the pectoral muscle area is different from the distance in the breast area. Wu et al. used dynamic multiple thresholding for initial border detection and Sobel filtering for true edge detection. They used three performance metrics on their dataset: the Hausdorff distance, the average minimum Euclidean distance, and the area overlap measure. It was found that 94% of images had a Hausdorff distance less than 6 pixels (4.8 mm), 96% of images had an average minimum Euclidean distance less than 1.5 pixels (1.2 mm), and 99% of images had an area overlap measure

* Corresponding author. Tel.: + 905303337164; fax: + 90 312 418 66 89.
E-mail addresses: pelinkus1@gmail.com (P. Kus), irfankaragoz@gazi.edu.tr (I. Karagoz).
[1] Tel.: +905324943572; fax: +90 312 231 13 40.

0010-4825/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2011.10.011

value larger than 0.9 [18]. Shahedi et al. [19] used local adaptive thresholding. They applied every gray level as a threshold value from the highest gray level to the lowest. They divided the square of the perimeter by the area of the obtained image, and determined the threshold value that made the obtained compactness value the lowest. They compared the method that they put forward to active borders (snake) modeling, which has been frequently applied in recent years, and they stated that local adaptive thresholding yielded better results in terms of accuracy and reliability. Local adaptive thresholding detects the breast border 86% accurately and 14% comparably. The method was able to capture the nipple in 94% of the 36 mammograms [19]. Another technique for breast area segmentation is polynomial modeling. This technique starts with determining the border approximately, and the border is obtained by the estimation of the fitting polynomial. Chandrasekhar and Attikiouzel [20,21] modeled mammogram background in y and z directions with polynomial modeling using approaching theorems. They carried out thresholding by taking out the image that they modeled pixel by pixel from the original image and determined the area. Despite being more reliable than thresholding, this method is quite complicated in terms of application. Apart from the fact that there is no objective evaluation of the output data, it was stated that the borders that Masek [22] enhanced were 29% better and 60% more acceptable. Polynomial modeling needs to be examined in terms of accuracy and applicability to CAD systems because it requires the user to enter parameters such as polynomial degree, and it is not fully automatic.

Raba et al. [23], Ferrari and Rangayyan [24], Wirth and Stapinski [25], and McLoughlin and Bones [26] applied a snake algorithm. In the first step of this method, the breast border was identified by thresholding, and in the second step region extraction was performed using the snake algorithm. The algorithm starts with initial segmentation of the breast by using a threshold determined by the Lloyd-Max quantizer [32] or Rosin's method [33]. In this method, low gray level changes and background noise prevent correct detection of the border. The biggest issue concerning the application of this method is that the breast nipple cannot be extracted. Another disadvantage of the method is that it is complicated.

Padayachee et al. [27] used iso-intensity borders and spatial information to segment the breast area. They achieved a minimum average root mean square difference between the manual and automated iso-intensity borders of 3.0 ± 0.3 mm.

Byung-Woo Hong and Bong-Soo Sohn used iso-level borders and anatomical information to detect and segment the breast boundary and the pectoral muscle. By means of segmentation, they established a breast coordinate system and this provided useful information for identification of masses and registration of the mammograms [28].

Karnan and Thangavel [31] used median filtering, image normalization, and a genetic algorithm to detect the breast border by considering the border pixel intensity values as population strings.

In the present study, firstly an approximate value of the global threshold is estimated. Secondly, the image is enhanced by using histogram stretching, and then, by using a gradient based algorithm, border pixels are detected. Finally, the breast boundary is drawn by interpolation and extrapolation of the detected border pixels. The method is shown in Fig. 1.

2. Materials and methods

2.1. Data set

Eighty-four mammograms randomly chosen from the Mini-Mammographic Image Analysis Society (MIAS) database [34] were used in this study. The mammograms were in medio-lateral oblique (MLO) view with a 200 μm sampling interval and 8-bit gray-level quantization. All images were 1024×1024 pixels in size. In addition to the mentioned database, we used another database containing the results of Ferrari and Rangayyan active borders modeling on 84 mammograms of the mini-MIAS database as well as an exactly extracted boundary drawn by an expert radiologist [24].

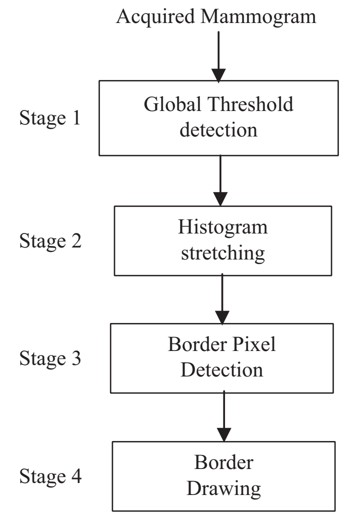

Fig. 1. Flow chart for identification of the breast border estimation.

3. Methods

Herein, we introduce a gradient based approach for breast boundary detection. The method is composed of five stages (see Fig. 1).

Stage 1: An approximate value of the global threshold is estimated by using discontinuity as proposed by Ojala [16].

Stage 2: Histogram stretching is performed to enhance the visibility of the breast border in the digitized mammograms providing greater separation between foreground and background by linearly re-mapping the pixel value from minimum to maximum gray level as given in Eq. (1).

$$I_o(x,y) = 255.\frac{I_i(x,y)-\min}{\max-\min} \tag{1}$$

where $I_o(x, y)$ is the gray level for the output pixel at (x, y) after the stretching process, $I_i(x, y)$ is the intensity value of a pixel at (x, y) in the digitized mammogram, and max and min are the maximum and the minimum intensity value of the digitized mammogram, respectively. A nonlinear median filter 5×5 pixels in size was applied to the image to remove noise from the mammogram without blurring the edges. This filter size was also chosen (as in [8]) as a good compromise between noise reduction and texture preservation of breast tissue.

In Ref. [35], Saha et al. stated that intensity histograms of the mammograms contain a prominent peak at low intensities that corresponds to the background. The intensity corresponding to this peak is considered the background intensity. In the mini-MIAS database we observe that most of the mammograms contain a prominent peak or highly attenuated frequencies at low intensities corresponding to the background. When we threshold the mammogram by using a manually selected threshold value, the estimated breast border either appears erroneously in the top and bottom portions of the mammogram, outside the breast region, or appears through the inside of the breast border. A typical histogram and two different thresholding applications are shown in Figs. 2 and 3, respectively.

An approximate global threshold, assumed to be where the background intensity ends, is determined by the maximum