Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Signal Processing

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

Robust automatic breast and pectoral muscle segmentation from scanned mammograms

Mario Mustra *, Mislav Grgic

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

ARTICLE INFO

Article history:
Received 3 February 2012
Received in revised form
28 June 2012
Accepted 16 July 2012

Keywords:
Mammography
Computer aided detection
Breast segmentation
Pectoral muscle extraction
Hough transform
Edge detection

ABSTRACT

Breast skin–air interface and pectoral muscle segmentation are usually first steps in all CAD applications on scanned as well as digital mammograms. Breast skin–air interface segmentation is much more difficult task when performed on scanned mammograms than on digital mammograms. In case of pectoral muscle segmentation, segmentation difficulty of analog and digital mammograms is usually similar. In this paper we present adaptive contrast enhancement method for breast skin–air interface detection which combines usage of adaptive histogram equalization method on small region of interest which contains actual edge and edge detection operators. Pectoral muscle detection method uses combination of contrast enhancement using adaptive histogram equalization and polynomial curvature estimation on selected region of interest. This method makes segmentation of very low contrast pectoral muscle areas possible because of estimation used to segment areas which have lower contrast difference than detection threshold.

1. Introduction

Computers are used more and more in modern detection and diagnostic processes. Breast cancer today is one of the most common diseases and a lot of effort has been put in prevention through different types of screening processes [1]. Because of the huge amount of data collected by various screening programs, usage of computer-assisted or computer-based interpretation of the results became unavoidable. This is mainly the case in digital mammography because there is no need for image digitization as in the case of analog mammograms. Although digital mammography is used for only slightly longer than ten years its advantages are suppressing classic mammography more and more each day. Computer aided detection (CADe) algorithms rely on breast segmentation as a first preprocessing step. To perform breast segmentation as well as other image enhancements or region detections, many image processing methods have been presented. Accurate breast segmentation is important in both scanned film mammograms and digital mammograms but is performed with a different degree of difficulty according to the capturing conditions. Scanned images usually suffer from much more artifacts than digital images which are usually segmented by detector calibration during capturing. Artifacts that are usually found in scanned film mammograms are orientation tags, light leakages caused by non-uniform sensitivity and thickness of film and imperfection of scanning process itself. Another important issue which occurs in segmentation of scanned film mammograms is a non-consistent image orientation and actual breast position in the observed image. That is the reason why scanned images apart from segmentation require accurate registration. Accurate registration provides equal positioning and rotation either for accurate displaying or for post-usage of Computer aided diagnosis (CAD) software. There have been many attempts to develop an optimal breast segmentation algorithm. The main reason was that the two most widely used mammographic databases are

* Corresponding author. Tel.: +385 915 886 283; fax: +385 161 297 17.
E-mail address: mario.mustra@fer.hr (M. Mustra).

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.07.026

consisted of scanned film mammograms and accurate segmentation is important in using scanned film mammograms in CAD applications. The MIAS database [2] contains 322 scanned images of the same size with similar properties. The DDSM database [3] contains 2620 cases which results in total 10,480 images, 5240 in medio-lateral oblique (MLO) view and 5240 in cranio-caudal (CC) view.

In this paper we propose a method for breast skin line estimation in order to perform breast tissue segmentation from the background and a method for pectoral muscle extraction as a further step in mammogram preprocessing. The novelty in breast skin line estimation method is the usage of local contrast enhancement on subdivided breast boundary regions to obtain higher contrast gain and produce larger difference between breast tissue and background. Initial subdivision is achieved by transforming edge of a breast mask obtained by thresholding into strip and further division of the strip into small segments. The breast mask is considered to be of a circular-like shape and therefore a transformation metric for conversion of breast boundary into rectangular strip has been presented. Actual edge is then being detected on those newly created small segments in order to achieve more accurate detection because of similar intensities of the neighboring pixels. Results of the segmentation are compared with hand-drawn segmentation line from a specialized radiologist. In this way we were able to measure segmentation accuracy not just qualitatively but also quantitatively. Removal of the pectoral muscle is also one of the important preprocessing steps in certain CAD applications. The method which we present in this paper uses adaptive local contrast enhancement technique in conjunction with polynomial modeling to provide robust pectoral muscle detection. The main problem with pectoral muscle detection is low visibility of the pectoral muscle in certain types of breasts, usually higher density breasts, where there is almost no intensity variation between the pectoral muscle and the breast tissue. Because of that it is almost impossible to achieve accurate segmentation of that visually almost invisible part of the muscle. In this paper we present a novel polynomial estimation of the pectoral muscle boundary from random points detected by applying a threshold on a contrast enhanced pectoral muscle region. The advantage of our method is its possibility to handle low contrast problems in muscle detection by trying to follow the pectoral muscle shape when it is not possible to perform thresholding because of similar tissue density of the pectoral muscle and surrounding tissue.

This paper is organized as follows. In Section 2 we bring a literature review of the recently proposed methods for both breast segmentation and pectoral muscle extraction. Section 3 presents a method used for the estimation of breast skin line interface. Section 4 explains the pectoral muscle detection. Section 5 brings the experimental results and Section 6 draws the conclusions.

2. Literature review

Most of the developed methods for breast segmentation and pectoral muscle segmentation use either MIAS or DDSM database for benchmarking their accuracy. Some of the early attempts for breast segmentation were global contrast based or intensity based [4]. Raba et al. presented a contrast based method for breast segmentation and pectoral muscle extraction [5]. Contrast based approach proved to be not as accurate when performed on the entire image because of light leakage problem in the edges and inconsistent intensity in different images. For that reason authors decided to use different tools which should provide automatic contrast calculation and threshold estimation. Among the most used methods for mammogram segmentation commonly used are also boundary based and region based approaches. Sun et al. used combination of stroma edge boundary computation and adaptive thresholding with curvature metric calculation for estimation of the breast skin line [6]. Sun et al. also presented the problematics of different contrast enhancement for different pixel neighborhoods for breast segmentation from scanned screen film mammograms [7]. Another approach which is similar to local contrast exploration are active contours. This approach can also be classified among boundary based approaches. Ferrari et al. [8] and Wirth and Stapinski [9] used active contours or snakes approach for the breast boundary detection. Active contours were also used by Thiruvenkadam et al. [10]. Thiruvenkadam et al. used 272 of 322 images from MIAS database and segmentation has been validated by radiologists. Their method achieved 258/272 good or acceptable segmented images while other 14 were unacceptably segmented. Chen and Zwiggelaar also presented a boundary based method which includes polynomial fitting of the edge detection points in order to create a smooth boundary [11]. Tzikopoulos et al. [12] presented a scheme for mammogram segmentation, pectoral muscle detection and breast classification according to density and asymmetry. They used mini MIAS database for benchmarking performance of the proposed algorithm and obtained good results which were compared with other proposed methods and inspected by a radiologist to prove the accuracy. For breast border estimation they used Tannimoto Coefficient (TC) and the Dice Similarity Coefficient (DSC). They obtained mean values of 0.900 and 0.945, for the TC and DSC respectively on the entire mini MIAS database. Tomàs [13] presented a method for breast segmentation and pectoral muscle suppression and achieved 86.65% of good and acceptable segmentation results according to subjective evaluation on the mini MIAS database. For the pectoral muscle removal he used Hough transform to detect straight lines and therefore pectoral muscle detection suffered from accuracy in the case of non-linear pectoral muscle shape. In the literature some other segmentation approaches as morphological segmentation [14] could be found. Morphologically based segmentation lies somewhere in-between boundary based and region based methods. Region growing [15] approach is a commonly used image processing technique. To perform a good region based region growing approach the region of interest needs to have good contrast properties. There are many contrast enhancement techniques among which Contrast Limited Adaptive Histogram Equalization (CLAHE) [16] stands out and is being mostly used for preprocessing of mammograms. Pisano et al. used CLAHE in detection of masses in dense mammograms which have lower contrast to the background tissue [17]. There are many examples of region growing used in mammogram segmentation as well. One of

Please cite this article as: M. Mustra, M. Grgic, Robust automatic breast and pectoral muscle segmentation from scanned mammograms, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.07.026

the recent approaches which combines CLAHE and seeded region growing algorithms for mammogram segmentation was proposed by Maitra et al. [18]. De Carvalho et al. [19] presented a combination of morphological operators and polynomial functions fitting, up to the third degree, according to a least square average error. They tested the proposed method on a set of 100 images from mini MIAS database and achieved 97% of results which are delineating with the radiologist-defined edge. Yapa and Harada [20] proposed a method for breast segmentation and breast skin-line detection using a combination of an improved fast-marching method and mathematical morphological operators such as area morphology, alternating sequential filter, openings and closings. Their method was tested on 100 mini MIAS mammograms achieving 99.1% of accurate segmentations. Some slightly different approaches which have recently been presented are the one from Tzikopoulos et al. which considers both breast skin line segmentation and pectoral muscle extraction [21]. For breast skin line detection they have used row-wise and column-wise interface detection with the choice of a specific threshold. A different approach which uses genetic algorithm for breast skin line estimation was presented by Karnan and Thangavel [22]. In that approach it has also been shown that there is a large correlation between setting the appropriate threshold and obtaining good segmentation. Mustra et al. [23] presented a method for segmentation of digital mammograms and pectoral muscle removal. Using wavelet transformation to select appropriate subbands they achieved 85% of good and acceptable pectoral muscle segmentation results. Among the methods which use some sort of modeling of the muscle edge based on the muscle boundary detection, Hough transform has often been used [8,24]. Most of the referenced segmentation methods give acceptable but not perfectly accurate results and they usually underachieve in pectoral muscle segmentation accuracy which we wanted to improve with the proposed method.

3. Breast segmentation method

Breast skin line detection is the first step in mammogram segmentation. Although this problem is rather trivial in case of digital mammograms, scanned screen film mammograms suffer from obstacles in the way of perfect segmentation. In segmentation process there are couple of problems which should be eliminated. Fig. 1 shows the most common artifacts and scanning imperfection found in the MIAS database. Some of those imperfections are easier to remove than others and usually the largest problem occurs when artifact is present in the breast tissue. An example of artifact in the breast tissue is the presence of a duct tape which significantly reduces intensity of breast tissue and makes segmentation by a fixed threshold impossible. For each of those artifacts we needed to use a standalone procedure for removing them without losing parts of breast tissue. Fig. 2 shows all the steps in the process of removing artifacts and final breast skin line estimation. First step which we have performed was image registration. Mini MIAS images are all of the same size of 1024×1024 pixels with 8 bits per pixel. Even though images are of the same size, the problem is that breast tissue is not positioned equally on all images. Therefore to perform more accurate segmentation it is needed to make some basic image transformation in order for all the images to have breast tissue at similar position.

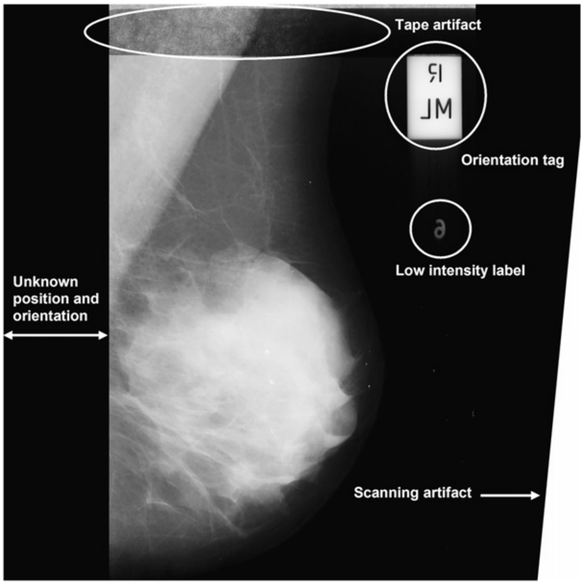

Fig. 1. Common artifacts occurred in MIAS database images.

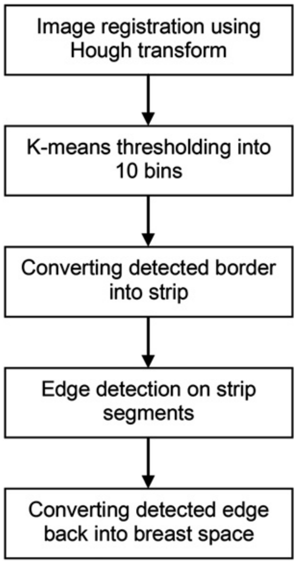

Fig. 2. All steps in the process of removing artifacts and final breast skin line estimation.

3.1. Breast alignment

All images have been oriented so that the pectoral muscle is situated in the top left area of each image. For that purpose images which show left breast medio-lateral oblique (MLO) view needed to be flipped horizontally. Besides flipping, all images should be translated for different number of pixels so that visible breast tissue starts at first left column of the image. For orientation detection and calculation of the distance for translation we have used combination of Hough transform and Sobel edge detection combined with fixed intensity thresholding and mathematical morphology [25]. The first thing we have done was fixed intensity thresholding of the original image. For the threshold 20% of the maximal intensity value in the observed image is used for the lower boundary and 90% of the maximal intensity for the higher boundary. These boundaries should remove dark and bright artifacts from the image without influencing the breast tissue. After this basic thresholding we have got a binary image where breast tissue and almost all artifacts are visible. For coarse removing of orientation tag, low intensity labels and scanning artifacts, we have used morphological opening operator. For the structuring element we have chosen a square of 101×101 pixels. This size of the structuring element experimentally proved to be a good compromise between substantial removal of artifacts and preserving enough detail in image. Of course this size of a structuring element could not be used for images of different size than mini MIAS. If an image is larger so should be a structuring element with linear dependency. After coarse removal of artifacts left after thresholding we have used Sobel edge detection filter to detect straight lines in the image. The longest straight line which should be almost vertical is considered to be the breast tissue edge. By converting pixels to Hough space we can get information about the line orientation and distance as well as line length. Standard Hough transform can be expressed as

$$\rho = x\cos(\theta) + y\sin(\theta) \quad (1)$$

where ρ indicates the distance between the origin of the line and θ is the angle of inclination of the normal line from the x-axis of almost vertical orientation of the breast in the observed image we can easily search for maximums in Hough space which correspond to vertical lines. θ can have values in range $[-\pi/2, \pi/2]$ and we can easily limit our region of interest in Hough space to search for maximum around $\pi/2$. This approach will give us the distance and therefore the position of the longest vertical line which is detected using Sobel filter from the morphologically opened binary image. These steps are enough to create breast image aligned to the top left corner. Results of the mentioned steps for alignment from the original image to final result are shown in Fig. 3 (a)–(d).

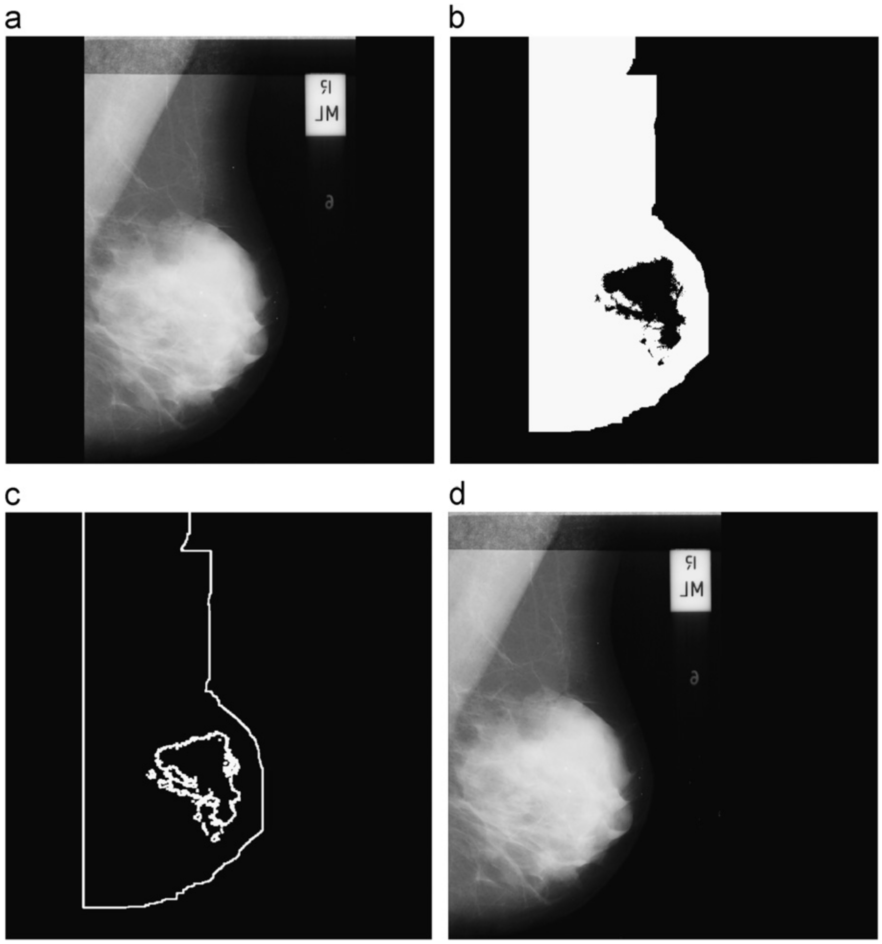

Fig. 3. (a) Original image from the mini MIAS database; (b) binary mask after thresholding; (c) detected edge of the binary mask; and (d) final result of image alignment.

Please cite this article as: M. Mustra, M. Grgic, Robust automatic breast and pectoral muscle segmentation from scanned mammograms, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.07.026

3.2. Breast boundary extraction and segmentation

After aligning images we have proceeded with the next step in the breast tissue segmentation. For better contrast enhancement we have divided our area of detection into smaller regions. The idea behind that is the similarity in intensity between neighboring areas of the breast border. This has provided us with the possibility for setting a better threshold to selected areas instead of trying to set the best overall threshold. Since we are usually performing image processing techniques on matrices representing images and breast skin line is not a straight line we have proposed a transformation between polar and Cartesian space for extracting breast boundary segments. Of course, breast is not perfect circular object, but larger part of the breast in MLO view shows similarity with circular shape objects, therefore we have decided that the polar representation would be most appropriate. For the center of our polar coordinate system we have used 1/3 of overall breast height in vertical dimension and first image pixel in the horizontal dimension. Breast height has been calculated from the binary mask obtained using thresholding. Image has been thresholded using values obtained by k-means clustering of the aligned image into 10 clusters and considering only clusters [3,10]. After that we have proceed with morphological dilation using structuring element of 151×151 pixels. Dilation provides us with the larger image from which we can substitute the original binary mask and get the region in which breast skin line should be contained. Fig. 4 shows the original mask obtained by *k*-means thresholding and region of interest mask obtained by subtracting the original mask from the mask dilated using the above mentioned structuring element. Region of interest creation could be expressed as

$$ROI = I \oplus S - I \tag{2}$$

where *ROI* is region of interest image for detecting the breast tissue border, *I* is the binary image obtained after selecting appropriate clusters after performing *k*-means algorithm and *S* is the structuring element used to perform morphological dilation. Once we have extracted our region of interest it is possible to proceed with conversion from polar to Cartesian coordinate space as explained above. This conversion for arbitrary element inside the *ROI* could be written as

$$x_i = C + (d-i)\sin(\varphi)$$
$$y_i = 1 + (d-i)\cos(\varphi) \tag{3}$$

where x_i and y_i are the coordinates of each pixel, d is the Euclidean distance between the center and edge points and φ is the corresponding angle as shown in Fig. 5. This figure shows the graphical representation of conversion from constructed polar space to Cartesian coordinate space. Coordinate C has been determined as one third of breast height as explained before, W as the width of the segment for boundary detection was chosen to be 100, angle φ takes values in range $[0, \pi]$ and d is the distance between center point and beginning of the *ROI* for certain φ. It is possible to choose different angle resolutions for extraction of segments because if we choose only one degree as angle resolution, we will get not very good resolution in conversion to Cartesian coordinate space. Therefore we have chosen to experiment with larger angle resolution and build our model according to the chosen angle resolution which can arbitrarily vary. The larger the angle resolution is, the smoother the detected border will be. Experimenting with different angle

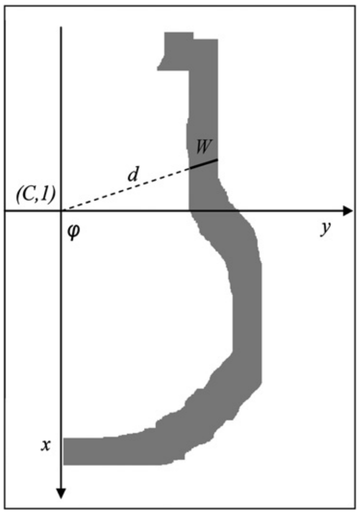

Fig. 5. Graphical representation of conversion of the *ROI* from constructed polar space to Cartesian coordinate space.

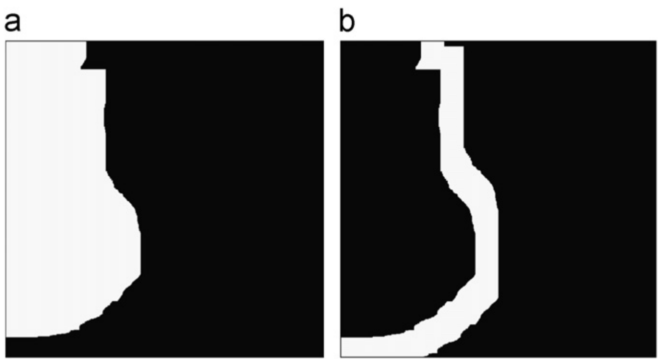

Fig. 4. (a) The original mask obtained by k-means thresholding; (b) *ROI* for breast skin line detection.

Please cite this article as: M. Mustra, M. Grgic, Robust automatic breast and pectoral muscle segmentation from scanned mammograms, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.07.026

Fig. 6. Example of the breast border after conversion to Cartesian coordinate space and before edge detection.

resolutions proved that 1/5 of a degree gives very good results as a compromise between border smoothness and algorithm execution time. Since the entire image size in each dimension is 1024 pixels and angle resolution of 0.2° gives total of 900 points it is understandable that finer resolution will not provide us with the smoother border but mostly increase the execution time of the algorithm. Example of a breast border segment before thresholding with an angle resolution of 0.2° is shown in Fig. 6. After gathering all segments according to the angle resolution we have proceeded with contrast enhancement of each segment in order to get the best possible breast skin line visibility. For contrast enhancement we have used contrast limited adaptive histogram equalization (CLAHE) [16]. This method proved to be efficient for local contrast enhancement and provides good results without large computational effort. After contrast enhancement we can use thresholding on the detected segment and detect edge by using Sobel edge detector on the final binary image. The final binary image has been created by combining two binary images, each with a different threshold. One binary image is created by taking into account all intensities higher than 0.01 of the mean value for the specified region and other by taking all intensities lower than 0.5 of the mean value. These two binary images have been combined using the logical AND operator which provided relatively accurate division into low and high intensity image sections to create the final binary image for the edge detection. Binary image creation process can be expressed as

$$B = B_1 \wedge B_2 \tag{4}$$

where B is the final binary mask and B_1 and B_2 are the two binary masks combined with the AND operator.

4. Pectoral muscle extraction

For pectoral muscle detection we have used a different approach than for breast skin line detection. The reason for that are completely different backgrounds in case of the pectoral muscle towards breast tissue than in the case of breast tissue towards dark surrounding. Usually pectoral muscles are homogenous areas situated in the top left corner of the MLO mammogram, after images have been aligned as described before. This knowledge makes detection of the muscle somewhat easier than in the case of unknown position in the image. In majority of cases pectoral muscles are consisted of brighter pixels than surrounding fat tissue and are usually easy to distinguish from the surrounding tissue. In some cases parts of pectoral muscle or the entire pectoral muscle are not visible. Reasons for that can be low overall contrast of the image or imperfection in mammogram capturing process. These cases present the most problems for automatic detection algorithms. To overcome that problem we have combined standard muscle edge detection with the polynomial estimation of muscle curvature. The polynomial estimation is obtained from visible part of the muscle and helps in detection of non-visible areas of the pectoral muscle. The method for pectoral muscle detection which we propose in this paper is consisted of couple of steps shown in Fig. 7. The first step is to determine the region of interest where pectoral muscle is situated. We have defined our *ROI* for the pectoral muscle detection as 2/3 of the breast height and breast tissue width at the top. Because of this criterion our *ROI* is different for every different breast and should provide a good compromise between relatively small size of the *ROI* and enough size so that the pectoral muscle is not cropped. An example of cropped *ROI* from "mdb002" is shown in Fig. 8. For initial thresholding we have chosen to divide mean intensity value of the *ROI* with 1.5. At this point it is important to notice that our *ROI*, when chosen in this way, should contain very few or no background pixels and because of

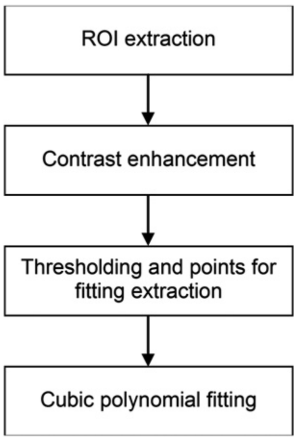

Fig. 7. All steps in the process of the pectoral muscle segmentation.

that we can say that this mean value is actually mean value of the breast tissue including the pectoral muscle in the chosen *ROI*. The next step we have done is contrast enhancement using CLAHE algorithm. The image obtained after contrast enhancement has been adjusted so that it covers the entire possible range of intensities of total 8 bits. After that we have used grayscale morphology opening operator to eliminate background noise and small objects in order to make detection boundary as smooth as possible. For the structuring element we have used a square of 3×3 pixels. This size of a structuring element provides a good compromise of preserving larger details and still is of enough size to remove smaller objects. Now that we obtained more or less clean *ROI* with enhanced contrast, we have used the previously calculated threshold to create a preliminary binary mask. From that mask we have chosen to randomly select 10 points for polynomial fitting of the muscle boundary. Instead of linear function we have used cubic fitting function with 4 coefficients:

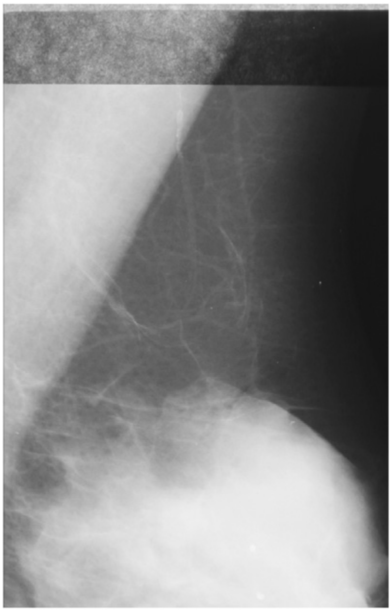

Fig. 8. Example of the extracted region of interest for the pectoral muscle detection.

$$y = p_1x^3 + p_2x^2 + p_3x + p_4 \tag{5}$$

where y is the horizontal coordinate and x is the vertical coordinate as shown in Fig. 5 and *pi* are the coefficients. Cubic polynomial function has been chosen because of the pectoral muscle shape which in majority of cases follows curvature represented by cubic function instead of linear function. The problem of choosing wrong points is solved by linear fit and iterative function which seeks for other points until correct slope of linear function is achieved. Our linear function is defined as

$$y = p_5x + p_6 \tag{6}$$

where y is the horizontal coordinate and x is the vertical coordinate as shown in Fig. 5 and p_5 and p_6 are the coefficients. If p_5 has a negative sign, the iteration stops and the algorithm decides that we have chosen correct points. Fig. 9(a)–(f) shows the entire process of pectoral muscle detection step by step as described in this section for the case of a pectoral muscle which has very low distinction towards the breast tissue.

5. Experimental results and discussion

For testing the methods proposed in this paper we have chosen the mini MIAS database [2]. Mini MIAS database consists of 322 images with size of 1024×1024 pixels with 8 bits per pixel. The difference between MIAS and mini MIAS is in the image size. Mini MIAS mammograms are resized and have spatial resolution of 200 μm per pixel while the original MIAS images have spatial resolution of 50 μm per pixel. All images are MLO view mammograms and this database is relatively outdated. The reason for choosing it as a test set is that it is publicly available and mammograms contain many imperfections. Almost all images in the mini MIAS database contain at least one artifact and many contain two or more artifacts, mainly

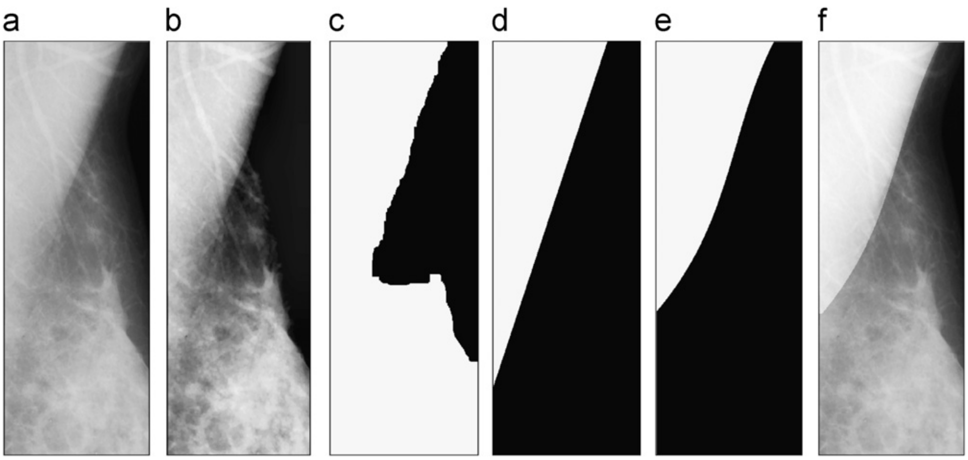

Fig. 9. (a) *ROI* extracted from the image; (b) *ROI* after contrast enhancement and morphological opening; (c) binary mask of the enhanced image; (d) straight line estimation; (e) cubic polynomial modeling of the muscle edge; and (f) overlap of the modeled muscle edge on the *ROI*.

Please cite this article as: M. Mustra, M. Grgic, Robust automatic breast and pectoral muscle segmentation from scanned mammograms, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.07.026

outside of the breast region and this makes breast segmentation somewhat more difficult. Artifacts inside the breast region are mainly caused by duct tape and low intensity tags. Another important reason for choosing the mini MIAS is the number of images which is relatively low and we could gather professionally segmented breast and pectoral muscle mask for this database. Even though the DDSM database [3] contains much more images, usually of higher quality, it was not possible for us to obtain segmentation masks for the entire database and we wanted to avoid any preselection of images which could influence the results. However, this approach for mammogram segmentation can be adopted for using with DDSM images. There can be two possibilities to obtain comparable results from DDSM database: image resizing to make images of comparable dimension with mini MIAS and resizing of the structuring element to fit DDSM database. In both cases resizing should be done with a factor between 4 and 5 because DDSM images have each dimension 4–5 times larger.

The proposed segmentation method does not use training set which means that all images in the database are part of the test set. Segmentation accuracy can be proved by visual inspection by radiologists or quantitatively by comparing segmented areas with hand drawn segmentation masks. In the majority of published papers segmentation accuracy has been proved only by visual inspection and results were classified as good, acceptable or unacceptable. In this paper we present both approaches for both breast and pectoral muscle segmentations.

The proposed method for image alignment proved to be successful in all 322 images with the maximal alignment error in 1 pixel. This gives alignment accuracy of 99.9% according to the entire image size of 1024×1024 pixels. For the segmentation accuracy we have divided successfulness of the proposed algorithm into three categories: successful, acceptable and unacceptable. Comparison has been made by visual inspection that was done with consultation with a radiologist. For the segmentation accuracy to be classified as successful, the image needs to be almost perfectly segmented with no artifacts left in the background and with visible edge of the entire breast. Acceptable category contains images where some very low intensity parts of the breast edge are cropped but without any influence to the later CAD application. Unacceptable category contains images where either some high or low intensity artifacts still exist in the background or some parts of the breast tissue are cropped in a way which could affect the later CAD application. Table 1 shows the division of segmented images using the proposed method into three mentioned categories by visual inspection of segmentation performance.

Table 1
Division of the automatically segmented mini MIAS images into categories according to the segmentation successfulness.

Category	Number of images	Percentage
Successful	295	91.61
Acceptable	24	7.45
Unacceptable	3	0.93

From Table 1 it is clear that the proposed method functions well for 99.07% of cases in the mini MIAS database. Although three images which were classified into unacceptable category do not exhibit severe case of breast tissue cropping, some background has not been removed and that is the reason why they have been classified as unacceptable.

To achieve the quantitative segmentation accuracy test of the proposed method we have compared each automatic segmentation mask with the hand drawn segmentation mask in order to get the percentage difference of these two masks. Segmentation error is defined as

$$E = \frac{\sum (A \; XOR \; A_T)}{\sum A_T} \times 100\% \tag{7}$$

where E is the segmentation error, A is the automatically segmented area, and A_T is the area of the hand drawn mask for the particular breast. To be able to understand the possible detection error and evaluate the impact of its size it is important to know the average breast tissue size in the image. The average breast tissue size when calculated over the entire mini MIAS database is 429,053 pixels which is 40.92% of the entire pixel count of every image. Average breast tissue size should not influence the overall segmentation error because we divide the erroneous pixel count with the pixel count of the hand segmented mask. With the proposed method according to (6) we have achieved average error in detection of 3.71%. Maximal error in segmentation relative to the size of the breast tissue area with the proposed method was 13.97% while minimal error was 1.00%. Both of these cases of maximal and minimal segmentation error are shown in Fig. 10(a) and (b) respectively. From these results it is possible to conclude that the proposed method functions very well on the given test set which includes the entire mini MIAS database without any image preselection. There are however some weaknesses of the proposed method and one of them is the smoothness of the detected boundary. Even though most CAD applications do not concern area around breast skin–air interface as important part of the breast, some detection algorithms can benefit from accurately detected edge. One of those algorithms are automatic nipple detection algorithms. Different authors used different accuracy estimation benchmarks and it is rather difficult to present an exact comparison of the method we have proposed in this paper with other recently proposed methods because of different number of categories authors used for dividing the segmentation successfulness. Although most of validation methods are based on visual inspection of segmentation accuracy, different authors decided to divide segmentation accuracy into different number of categories. We have tried to interpret the results from other authors in a way to get the best comparability among them. Maitra et al. [18] achieved 95.71% of successfully segmented images from the entire mini MIAS database. Their main tool for achieving accurate segmentation was CLAHE which provided more distinct border between breast tissue and background. It is important to stress out that their results of segmentation include both breast and pectoral muscle segmentation. Chen and Zwiggelaar achieved 98.4% of acceptable segmentation on

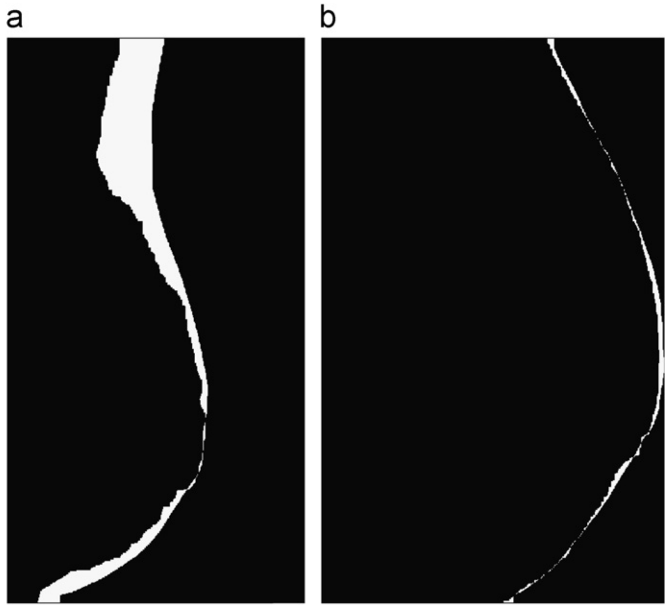

Fig. 10. (a) Case of the maximal segmentation error; and (b) case of the minimal segmentation error.

the test set of 240 mammograms, of which 66.5% were accurate, 25% nearly accurate and 6.9% acceptably segmented [11]. Their method shares some similarity in a way that they have also used coarse masks and further refinement for achieving more accurate edge detection. Raba et al. achieved good segmentation results in 86% of cases [5]. Their method was based on histogram based with iterative search of the appropriate threshold for the entire image. However this method, in our opinion, cannot give as accurate results as our proposed method because it deals with global instead of local threshold determination and therefore is not sensitive to local intensity variations. From this comparison result we can conclude that the method which we propose in this paper gives slightly better overall breast tissue segmentation results than other recent state of the art methods. All of the mentioned methods used only visual inspection to classify results between successful and unsuccessful ones, while we have also tried to quantitatively show the segmentation error in percentage of incorrectly segmented pixels towards the entire breast tissue area in each mammogram. Even though we have not achieved perfect segmentation for all images, improvement in segmentation accuracy as well as relative simplicity of the segmentation process is important in CAD usage. Most of the CAD applications need to be run in real time and therefore there is not much space for very complex calculations and straightforward methods which give satisfactory results are most welcome. On the other hand, the CAD systems need to have the segmented image as good as possible at their input and therefore a good compromise between speed and accuracy of segmentation needs to be achieved.

The proposed method for pectoral muscle segmentation has also been tested on the entire mini MIAS database. As in the case of breast segmentation we have chosen to prove segmentation accuracy by visual inspection with the division into three categories and by quantitative comparison with hand drawn segmentation masks. It is important to state that overall number of images for the pectoral muscle detection was not 322 because on two images pectoral muscle is not visible. Therefore we have calculated the accuracy of our method on 320 images. Results of the visual inspection of the pectoral muscle segmentation accuracy are shown in Table 2.

Table 2
Division of the automatically segmented pectoral muscles from mini MIAS images into categories according to the segmentation successfulness.

Category	Number of images	Percentage
Successful	287	89.69
Acceptable	22	6.88
Unacceptable	11	3.44

To be able to quantitatively check the accuracy of the proposed method we have compared each pectoral muscle mask with the hand drawn mask which is considered to be the ground truth. It is not clear which segmentation accuracy comparison method is the most descriptive so we have chosen to represent detection error by the percentage difference in segmented area. In this case we have used the same equation as for the breast segmentation accuracy testing (7). In Eq. (7) A now stands for the automatically segmented area, and A_T is the area of a hand drawn mask for each pectoral muscle. The average pectoral muscle size when calculated over the entire mini MIAS database is 58,870 pixels which is 5.61% of the entire pixel count of every image. This large difference in size will also affect the average error in segmentation because denominator in (7) will be much smaller in the case of pectoral muscle segmentation. With the proposed method we have achieved the average area detection error of 14.57%. Even though this might seem to be a rather large error, visual inspection of the segmentation proves that segmentation accuracy is actually acceptable for the large majority of cases while the method still functions well in cases when there is no visible muscle–tissue border. We have previously stated that the biggest advantage of the proposed method over the other methods is in segmentation of very low contrast pectoral muscles. Fig. 11 shows the segmentation of one of such cases which is often reported as the problematic one. It is the case of MIAS image "mdb151". In comparison to other recently presented methods for pectoral muscle extraction, our method proves to be slightly more successful. The proposed method performs worse in some cases than on others. There are a couple of scenarios where error can occur and mostly the problem is the intensity variation of the pectoral muscle in certain mammograms. Sometimes pectoral muscles are not homogenous objects in images and point selection from initial mask on which we have done polynomial fitting can introduce some errors. Because of wrong point selection, the resulting fitting algorithm creates slightly different line which denotes pectoral muscle edge. In those cases segmentation line is not perfectly aligned with the actual pectoral muscle edge. To be able to compare our proposed method with other methods in terms of segmentation successfulness we needed to interpret results from other authors in a way to get the best comparability among them, mostly because of different number of categories used to divide results. Besides visual inspection we have also presented numerical estimation of segmentation accuracy obtained by comparing the automatic segmentation results with hand drawn masks. Among the newest and most successful segmentation methods are one of Maitra et al. [18] who achieved 95.71% of correctly segmented mammograms which includes both breast in line and pectoral muscle segmentation. Chen and Zwiggelaar [11] achieved 93.5% of acceptable segmentation on the test set of 240 mammograms of which 62.5% were accurate, 25.4% nearly accurate and 5.6% acceptably segmented pectoral muscles. Raba et al. [5] achieved good segmentation results in 86% of cases. For extraction of the pectoral muscle they have used the region growing method which gave acceptable results but also introduces some problems with determination of appropriate stopping condition. This is generally the biggest problem in region growing algorithms when segmenting low contrast images with not perfectly homogenous regions. We have tried to overcome this problem by creating a polynomial estimation of the pectoral muscle which should prevent cases of completely false pectoral muscle segmentations. In the above mentioned methods for the pectoral muscle segmentation, the authors classified segmentation results into couple of categories according to segmentation performance validated by visual inspection. In the proposed method we have used both visual inspection of segmentation performance and quantitative accuracy measure in order to get results as precise as possible.

6. Conclusions

In this paper we have presented approaches for breast skin line segmentation and pectoral muscle extraction from scanned screen film mammograms. The breast skin line estimation method uses division of estimated boundary in order to get better contrast enhancement and threshold setting for each individual segment. Using this approach we have obtained more accurate threshold and therefore more precise skin line detection. In the case of pectoral muscle detection, we proposed an approach for estimation of the edge using polynomial modeling. For the polynomial function we have used cubic function because it provides best results for majority of cases. The advantage of the proposed algorithm is in the possibility to accurately segment images with part of the pectoral muscle which are not distinguishable from the surrounding tissue. Those areas of the pectoral muscle have been estimated using the knowledge of muscle contour from the visible part of the muscle. Segmentation accuracy for both breast skin line and pectoral muscle has been proven by comparing segmentation masks with the masks hand drawn by a professional radiologist. Hand drawn masks were considered to be ground truth and the methods have been tested on the entire mini MIAS database containing 322 images. Results of the segmentation process prove to be of a high accuracy, not just visually but also quantitatively. The results of comparison with the recently proposed methods show some improvement in segmentation accuracy and reduction in number of falsely segmented images when segmentation objects have no intensity difference than the surrounding tissue. We can conclude that the proposed method can be used

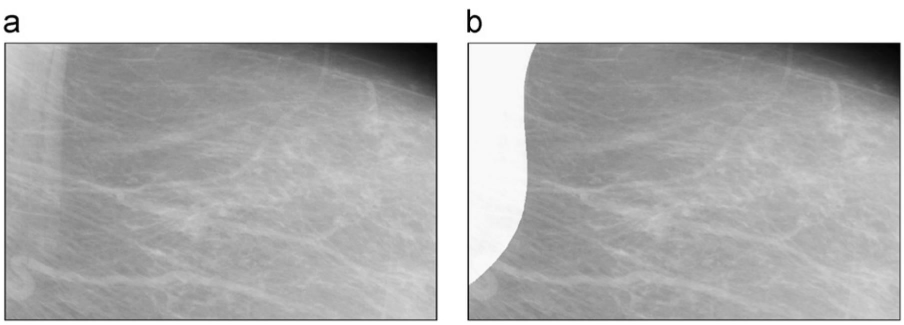

Fig. 11. (a) Original *ROI* for the pectoral muscle detection; and (b) result of the pectoral muscle segmentation for "mdb151".

for automatic segmentation of all scanned screen film mammograms and the pectoral muscle detection method can be used in both scanned screen film and digital mammograms as a preprocessing step in many CAD applications.

References

[1] Breast Cancer Facts and Figures, American Cancer Society, Inc., Atlanta, 2011–2012.

[2] 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, in: Proceedings of the Second International Workshop on Digital Mammography, York, England, 10–12 July 1994, pp. 375–378.

[3] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The digital database for screening mammography, in: Proceedings of the Fifth International Workshop on Digital Mammography, IWDM 2000, Medical Physics Publishing, 2001, pp. 212–218.

[4] P.J. Besl, R.C. Jain, Segmentation through variable-order surface fitting, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-IO (1988) 167–192.

[5] D. Raba, A. Oliver, J. Marti, M. Peracaula, J. Espunya, Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms, vol. LNCS 3523, Springer-Verlag, Berlin Heidelberg, 2005 (pp. 471–478).

[6] Y. Sun, J. Suri, R. Rangayyan, A novel approach for breast skin-line estimation in mammograms, in: Proceedings of the 18th IEEE Symposium on the Computer-Based Medical Systems, 23–24 June 2005, pp. 241–246.

[7] Y. Sun, J. Suri, Y. Zhen, R.M. Rangayyan, R. Janer, Effect of adaptive-neighborhood contrast enhancement on the extraction of the breast skin-line in mammograms, in: Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society, 17–18 January 2006, pp. 3475–3478.

[8] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, A.F. Frère, Identification of the breast boundary in mammograms using active contour models, Medical and Biological Engineering and Computing 42 (2004) 201–208.

[9] M. Wirth, A. Stapinski, Segmentation of the breast region in mammograms using snakes, in: Proceedings of the First Canadian Conference on Computer and Robot Vision (CRV 2004), London, Ontario, Canada, 17–19 May 2004, pp. 385–392.

[10] S. Thiruvenkadam, M. Acharyya, N.V. Neeba, P. Jhunjhunwala, S. Ranjan, A region-based active contour method for extraction of breast skin-line in mammograms, in: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 14–17 April 2010, pp. 189–192.

[11] Z. Chen, R. Zwiggelaar, Segmentation of the Breast Region with Pectoral Muscle Removal in Mammograms, Medical Image Understanding and Analysis (MIUA) 2010, The University of Warwick, Coventry, UK, 2010 (pp. 71–76).

[12] S.D Tzikopoulos, M.E. Mavroforakis, H.V. Georgiou, N. Dimitropoulos, S. Theodoridis, A Fully, Automated scheme for mammographic segmentation and classification based on breast density and asymmetry, Computer Methods and Programs in Biomedicine 102 (1) (2011) 47–63.

[13] J.S. Tomàs, Segmentation of the Breast Region with Pectoral Muscle Suppression and Automatic Breast Density Classification, Université Catholique de Louvain, A Thesis Submitted for the Degree of Master Ingénieur Civil Électricien, 2011.

[14] F. Meyer, S. Beucher, Morphological segmentation, Journal of Visual Communication and Image Representation 1 (1990) 21–46.

[15] R. Adams, L. Bischof, Seeded region growing, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (6) (1994) 641–647.

[16] K. Zuiderveld, Contrast Limited Adaptive Histogram Equalization, Graphic Gems IV, Academic Press Professional, San Diego, 1994 (pp. 474–485).

[17] E.D. Pisano, S. Zong, B.M. Hemminger, M. DeLuca, R.E. Johnston, K. Muller, M.P. Braeuning, S.M. Pizer, Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the Detection of Simulated Spiculations in Dense Mammograms, Journal of Digital Imaging, vol. 11, Springer, New York, 1998 (pp. 193–200).

[18] I.K. Maitra, S. Nag, S.K. Bandyopadhyay, Technique for Preprocessing of Digital Mammogram, Computer Methods and Programs in Biomedicine, Available online from: 〈http://www.sciencedirect.com/science/article/pii/S0169260711001222〉, 12 June 2011.

[19] I.M. de Carvalho, L.M.S. Luz, A.V. Alvarenga, A.F.C. Infantosi, W.C.A. Pereira, C.M. Azevedo, An automatic method for delineating the pectoral muscle in mammograms, IFMBE Proceedings 2 (2007) 271–275.

[20] R.D. Yapa, K. Harada, Breast skin-line estimation and breast segmentation in mammograms using fast-marching method, International Journal of Biological and Life Sciences 3 (1) (2007) 54–62.

[21] S. Tzikopoulos, H. Georgiou, M. Mavroforakis, N. Dimitropoulos, S. Theodoridis, A fully automated complete segmentation scheme for mammograms, in: Proceedings of the 16th International Conference on Digital Signal Processing, 5–7 July 2009, pp. 1–6.

[22] M. Karnan, K. Thangavel, Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications, Computer Methods and Programs in Biomedicine 87 (1) (2007) 12–20.

[23] M. Mustra, J. Bozek, M. Grgic, Breast border extraction and pectoral muscle detection using wavelet decomposition, in: Proceedings of the IEEE Eurocon, 18–23 May 2009, pp. 1426–1433.

[24] A. Sultana, M. Ciuc, R. Strungaru, Detection of pectoral muscle in mammograms using a mean-shift segmentation approach, in: Proceedings of the Eighth International IEEE Conference on Communications (COMM), Bucharest, 2010, pp. 165–168.

[25] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Third edition, Pearson Prentice Hall, New Jersey, 2008.

Please cite this article as: M. Mustra, M. Grgic, Robust automatic breast and pectoral muscle segmentation from scanned mammograms, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.07.026