Breast Region Segmentation and Pectoral Muscle Removal in Mammograms

Marijeta Slavković-Ilić, *Student Member, IEEE*, Ana Gavrovska, *Member, IEEE*, Milan Milivojević,
Irini Reljin, *Senior Member, IEEE*, and Branimir Reljin, *Senior Member, IEEE*

Abstract — **The first step in most computer aided diagnosis systems is an accurate segmentation of breast region, which affects not only the accuracy but also the speed of the analysis because it significantly reduces the area of the image to be examined. The second step usually includes removal of pectoral muscle region, which is seen in mediolateral oblique view mammograms. This is primarily done to reduce the number of false positive breast cancer detections. In this paper, a method for the segmentation of breast region based on contrast enhancement and k-means algorithm is proposed. To extract pectoral muscle, a region of interest is found, its contrast is enhanced and the pectoral muscle is identified using k-means algorithm. Cubic polynomial fitting is used for the estimation of muscle's boundary. The method is validated with mammograms from miniMIAS database.**

Keywords — **Breast segmentation, computer aided detection, MIAS database, pectoral muscle.**

I. Introduction

THE most frequent form of cancer in women is breast cancer [1]. Mammography is the best way for non-invasive detection of malignant breast disease at an early stage. Mammography is an X-ray breast examination method, which utilizes a specially designed X-ray device that uses a beam of low energy (about 20 kV) to minimize the side effects of radiation on the patient. Older generation (analogue) devices use film to capture X-ray beams and doctor-specialists visually analyze the image. In digital mammography, a digital detector catches X-ray beams and converts them to electronic signals thus creating an electronic image. Use of digital mammography devices, or digitized films obtained from analogue devices, provide all the advantages of digital image processing, viewing, archiving and transfer of files, which facilitates and improves analysis of the image and enables more accurate diagnosis. Digital mammography devices are better because they require fewer doses of radiation, while directly generating an electronic image which is processed and analyzed by computer. The use of mammography screening can significantly reduce mortality from breast cancer [1]. Therefore, many techniques for the processing and improvement of digital mammograms are being used in CAD (Computer Aided Detection) systems [2]–[9]. A large number of such systems have two important preprocessing steps: detection of breast and pectoral muscle regions. Assessment of the actual edge of the breast is important in order to accurately distinguish it from the background and artifacts. Segmentation of breast object in digital mammograms enables reduction of an image region to be tested for anomalies, elimination of background and artifacts that are often found in the image (for example, a label that defines whether it is a left or right breast, which mammography view was the image taken at - CC (Cranio-Caudal) or MLO (Mediolateral-Oblique). Segmentation of breast is a difficult task due to the presence of artifacts, noise, but mostly because of the low density of fat tissue close to its border. The intensity of pixels near the border of the breast tissue differs little from the intensity of the background which can lead to inadequate segmentation of breast region. This way, anomalies that are located close to the edge could be overlooked.

A number of algorithms for the segmentation of breast and pectoral muscle have been developed. The authors in [10] found a rough edge of the breast using a simple binarization and morphological erosion, then using active contours based on the gradient the actual edge of the breast is obtained. In [11] a global threshold from original mammogram is found, then the image is enhanced by histogram stretching. Algorithm based on the gradient is used to detect border pixels, and the border itself is drawn using interpolation and extrapolation of detected pixels. Extracting pectoral muscle based on the region growing is proposed in [12]. In the first step, authors improve image contrast with CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm, then define the region of interest in which the pectoral muscle is located and eventually propose an algorithm for pectoral muscle elimination based on seeded region growing. In [13], the rough boundary of muscle is found with the iterations of

Marijeta Slavković-Ilić is with the Innovation Center of the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Serbia (phone: 381-11-3370; e mail: mariyeta667@yahoo.com).
Ana Gavrovska is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Serbia (e mail: anaga777@gmail.com).
Milan Milivojević is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Serbia (e mail: msmilance@etf.rs).
Irini Reljin is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Serbia (e mail: irini@etf.rs).
Branimir Reljin is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Serbia (e mail: reljinb@etf.rs).

Otsu's thresholding method and mathematical morphology. The actual boundary of the pectoral muscle is obtained by using a multiple regression algorithm.

This paper proposes a method for the segmentation of breast and pectoral muscle regions in a digital mammogram using methods of digital image processing. It is an extended version of the work published in [14]. The first step in the proposed method is the segmentation of breast region. Contrast of the image is improved using the AGCWD (Adaptive Gamma Correction with Weighting Distribution) method [15], followed by binarization of the image using the k-means algorithm. The application of morphological operations to a binary image gives a finer border of the breast. After segmenting the pectoral muscle, the breast orientation is found and all images are positioned so that the muscle is in the upper left corner. A region of interest is allocated and its contrast is enhanced in order to increase the difference of intensities between pectoral muscle and the surrounding tissue. Based on the k-means method, a binary mask representing the pectoral muscle is obtained. The estimation of the muscle's edge is carried out by using a third order polynomial fitting [16]. The efficiency of the method was tested on mammograms from the reference miniMIAS (Mammographic Image Analysis Society Database) database [17], [18].

II. BREAST REGION SEGMENTATION

In order to extract the breast region, it is necessary to find the breast border. Breast segmentation may be a complicated task because the intensity of pixels near the border is similar to the background. This often leads to so-called sub-segmentation of the breast in which parts of the tissue near the border are characterized as background. In this way, anomalies that are located near the border may not be registered. To find the actual breast border, AGCWD algorithm proposed in [15] is applied for contrast enhancement. The authors propose a new algorithm for modifying the histogram combining gamma correction and histogram equalization. Gamma correction can be represented as:

$$T(l) = l_{\max}\left(\frac{l}{l_{\max}}\right)^{\gamma} \qquad (1)$$

where $l_{\max}$ is the maximum intensity of the input image, and γ is an adaptive parameter. The intensity value l of the input image becomes $T(l)$ after gamma correction.

Histogram equalization is defined as:
$$T(l) = F(l) \cdot l_{\max} \qquad (2)$$
where $F(l)$ is the cumulative distribution function which is calculated based on the probability density function, f. The probability density function f is defined as:

$$f(l) = \frac{n_l}{n_{tot}} \qquad (3)$$

where n_l is the number of pixels with intensity value l and n_{tot} is the total number of pixels in the image. Cumulative distribution function F is calculated by:

$$F(l) = \sum_{k=0}^{l} f(k) \qquad (4)$$

The authors in [15] propose an adaptive gamma correction (AGC) based on the modified cumulative distribution function as:

$$T(l) = l_{\max}\left(\frac{l}{l_{\max}}\right)^{\gamma(l)} = l_{\max}\left(\frac{l}{l_{\max}}\right)^{1-F_w(l)} \qquad (5)$$

This method increases the value of pixels of low intensity while avoiding significant degradation of high intensity value pixels. The introduction of weighting distribution (WD) somewhat modifies the statistical histogram and reduces negative effects. A weighting distribution function is defined as:

$$f_w(l) = f_{\max}\left(\frac{f(l) - f_{\min}}{f_{\max} - f_{\min}}\right)^{\alpha} \qquad (6)$$

where α is the adjusted parameter, $f_{\max}$ and $f_{\min}$ are the maximum and minimum values of the probability density function.

The modified cumulative distribution function is calculated as:

$$F_w(l) = \sum_{k=0}^{l} \frac{f_w(k)}{\sum f_w} \qquad (7)$$

where the sum of modified probability density function is defined as:

$$\sum f_w = \sum_{k=0}^{l_{\max}} f_w(k) \qquad (8)$$

The gama parameter that depends on the modified cumulative distribution function is:
$$\gamma(l) = 1 - F_w(l) \qquad (9)$$

The resulting gamma parameter is inserted into equation (5) and new pixel values are calculated.

In order to provide the same size, 1024×1024, for all images from miniMIAS database, black columns (columns in which all pixels have a zero intensity value) are added from the left and right sides of the scanned mammograms (Fig. 1). We leave out those columns and continue with further processing on the newly obtained images (Fig. 2a). The intensities of background pixels around the breast are not equal to zero but have some small value (from 5 to 15 depending on the image). In order to avoid the increase of the intensity of these pixels, we find the first non-zero pixel from the left and right side of the image. A pixel with a lower intensity is stored as a new variable t. If the value of this variable is small enough ($t < 10$), the intensity of all pixels in the image is reduced by the value t and AGCWD method is applied. Figure 2 shows the original image and the image after subtraction of the value t, and the results of the AGCWD algorithm on these images. Results obtained after AGCWD on two images are different on the external part of the breast. Subtracting the value t from the original image only slightly suppresses the background, while important information is retained.

In order to further suppress unwanted background around

the breast, we subtract images obtained with top-hat and bottom-hat transformations from AGWCD image, where a disk of radius 5 pixels was used as a structural element (Fig. 3).
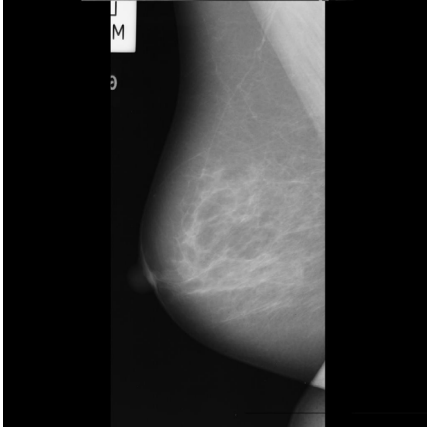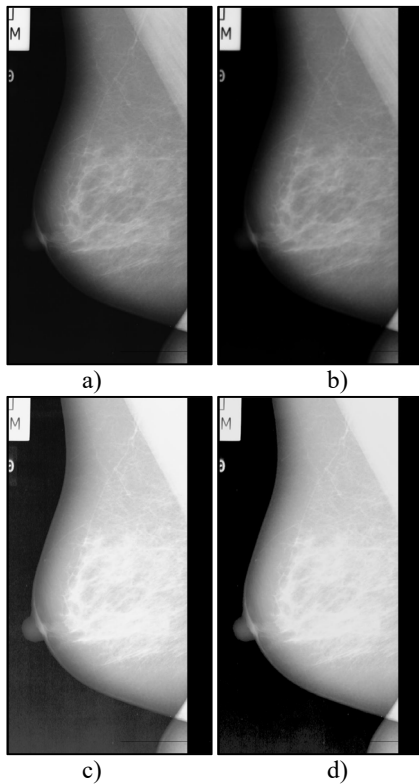

Fig. 1. Original image (mdb237).

a)   b)

c)   d)

Fig. 2. a) Input image b) Image subtracted by value t
c) AGCWD applied to a)   d) AGCWD applied to b).

The k-means algorithm is applied to the resulting image from the previous step in order to obtain a binary image. This method is a method of cluster analysis which aims to divide n objects into k clusters so that each object belongs to a cluster with the nearest mean. Each cluster is represented by its mean. The observed object joins a cluster with the nearest mean value. The initial selection of mean values is random. The mean value is calculated as the arithmetic mean of the objects in the cluster. In our algorithm, the image is divided into 10 clusters (Fig. 4a) and the cluster with the lowest mean is rejected as it

represents the background. Morphological opening with a disk of radius 10 is applied to the image in order to remove any small objects and retain only the largest one that represents the breast. Filling of small holes is done and averaging filter is applied to get a smoother border of the breast. The filter is of size 20×20 and it is assumed that pixels with intensity values above 0.5 form the breast (Fig. 4.b). The obtained image is multiplied by the original mammogram and only the desired object, which represents the breast region, is kept (Fig. 4c). A breast border marked with a green color on the initial mammogram can be seen in Fig. 4.d.
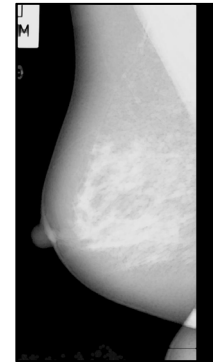

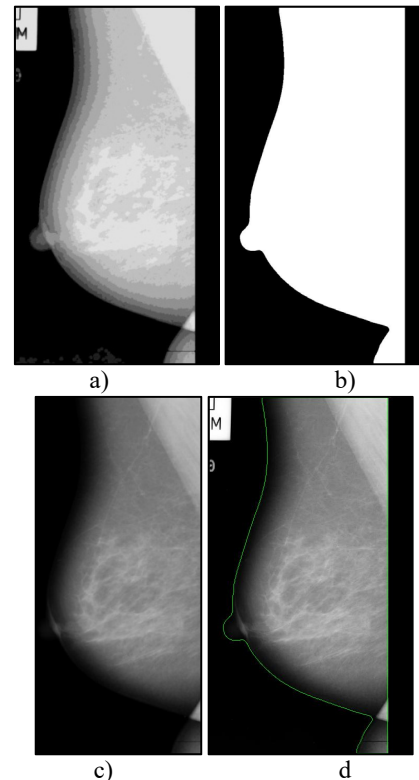Fig. 3. Suppression of unwanted background using mathematical morphology.

a)   b)

c)   d

Fig. 4. a) Clusterized image ($k=10$) b) Mask of breast region c) Final image of breast region d) Input image with breast border marked.

III. PECTORAL MUSCLE REMOVAL

Image obtained in the previous step (Fig. 4.c) is used as the input image for the segmentation of pectoral muscle. Our goal is to have the pectoral muscle located in the upper left corner in every image. To achieve this, we must first determine the orientation of the breast. An image mask that covers only the region of the breast is divided into four equal quadrants, and we compare the amount of white pixels in the upper two quadrants. If there are more white pixels in the right quadrant, we conclude that the pectoral muscle in that image is on the right side and the image needs to be flipped horizontally. A region of interest starts in the top left corner and its width is equal to the width of the breast at the top. The height of the region is equal to 70% of the breast height. The resulting image of the ROI (Fig. 5.b) is normalized and image contrast is enhanced with:

$$I_1 = 255 \cdot \left(\left(\frac{I}{255} - a \right) \cdot b + a \right) \qquad (10)$$

where I and I_1 are input and output images, a and b are constants which in our case take the values of 0.7 and 2 respectively. The obtained image is shown in Fig. 5.c.

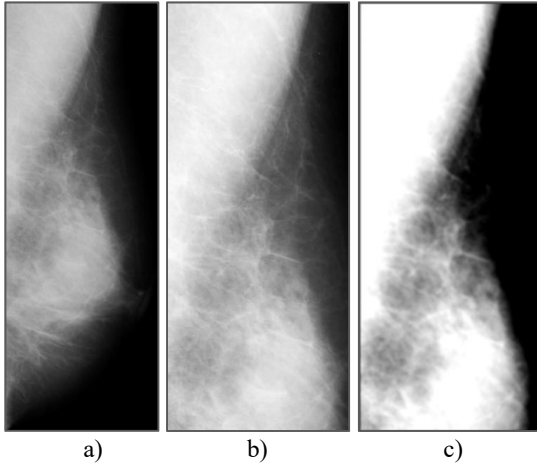a)   b)   c)
Fig. 5. a) Input image (mdb035)   b) ROI
c) ROI with enhanced contrast.

The next step is clustering which is done using the k-means algorithm. Clustering is performed with three different numbers of clusters simultaneously: in the first case the image is divided into two, in the second into three, and in the third into four clusters (Fig. 6). In every image we retain only the cluster with the highest mean value. To remove small objects and to break the weak links, morphological opening of the image, with a disk of radius 5 as a structuring element, is performed. Of the three largest objects in the image, we keep the one that is closest to the top left corner and assume that it represents the pectoral muscle. After that, filling of small holes is performed (Fig. 7).

Selection of the appropriate clustering is done based on the percentage of the image that represents the object. Since sometimes a part of the breast is segmented along with the pectoral muscle (Fig. 6a), the percentage of segmented object would not represent the actual percentage of the pectoral muscle. Also, the difference between the two

adjacent clusterizations would be much greater because of the segmented breast part that does not belong to a muscle. Therefore, we decided to compare only the upper half of the images based on the following parameters:

$$p_i = \frac{p_{i,white}}{p_{total}} \cdot 100, \quad i = 1, 2, 3 \qquad (11)$$

where $p_{i,white}$ is the number of pixels of the upper half of the image belonging to the object, p_{total} is half of the image area, and p_i is the percentage of pixels belonging to an object in the observed part. The variable $i = 1, 2, 3$ indicates whether it is the first ($k = 2$), second ($k = 3$) or third ($k = 4$) case. The percentage difference of objects between adjacent clusterizations is defined as:

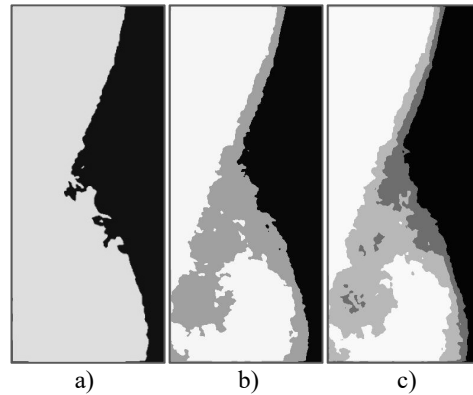$$r_{i,i+1} = \frac{p_{i,white} - p_{i+1,white}}{p_{i,white}} \cdot 100, \quad i = 1, 2 \qquad (12)$$

a)   b)   c)
Fig. 6. Clusterized images: a) $k=2$; b) $k=3$; c) $k=4$.
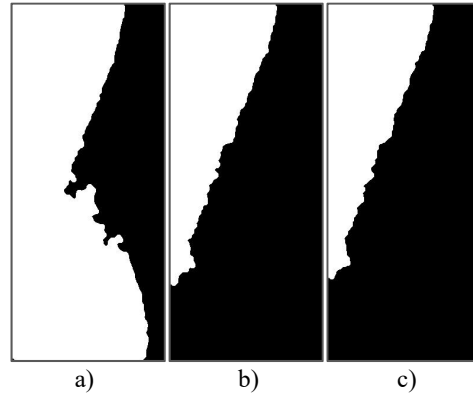

a)   b)   c)
Fig. .7. Pectoral muscle region for a) $k=2$; b) $k=3$; c) $k=4$.

Based on the visual analysis of images, we set the criteria for selecting the best clusterization. If the object in the first clusterization occupies less than 40% of the upper half of the image ($p_1 < 40$), the first clusterization is chosen. If an object in the second clusterization takes up more than 65% of the upper half of the image ($p_2 > 65$), we choose the third clusterization. For cases that do not fall into these two ranges, the selection is made based on the percentage difference. If $r_{1,2} < 15$, that is, if the object from the second clusterization covers an area larger than 85% of the object from first clusterization, we choose the first clusterization.

If $r_{1,2} > 15$ and $r_{2,3} < 15$, the second clusterization is chosen. In all other cases, the third clusterization is the final one. The selection criteria are shown in detail in Fig. 8.

The final region of the pectoral muscle, for example from Fig. 6, is shown in Fig. 9.a, where the first clusterization ($k = 2$) is chosen as the final one.

$$if \ p_1 < 40$$
$$k = 2;$$
$$else$$
$$if \ p_2 > 65$$
$$k = 4;$$
$$else$$
$$if \ r_{1,2} < 15$$
$$k = 2;$$
$$else$$
$$if \ r_{2,3} < 15$$
$$k = 3;$$
$$else$$
$$k = 4;$$
$$end$$
$$end$$
$$end$$
$$end$$

Fig. 8. Best clusterization decision criteria.

The obtained region of pectoral muscle represents a binary mask which is used for the estimation of the muscle's boundary using a polynomial function of the third degree:

$$y = c_1 \cdot x^3 + c_2 \cdot x^2 + c_3 \cdot x + c_4 \qquad (13)$$

where y is the horizontal coordinate, x is the vertical coordinate, and c_i are the coefficients of polynomial.

For curve estimation, we take into account the points that are located on the border between the pectoral muscle and the background, so long as the value of the horizontal coordinate decreases (Figure 9.b). The final region of the pectoral muscle is shown in Figure 9.c.

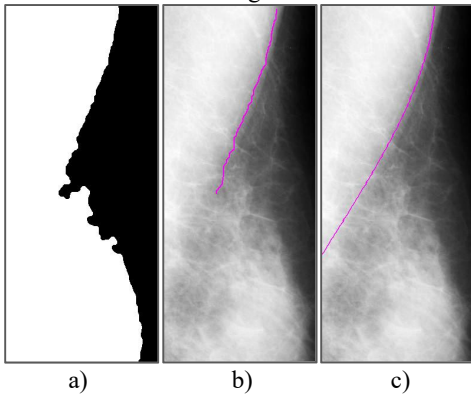a)   b)   c)
Fig. 9. a) Mask image; b) Points for curvature estimation; c) Marked pectoral muscle.

IV.   EXPERIMENTAL RESULTS AND DISCUSSION

To test the proposed method for breast region segmentation and pectoral muscle removal, we used reference miniMIAS database [17], [18]. The database contains 322 images of the same size 1024×1024 with 8 bits per pixel. The breast segmentation results obtained by visual comparison of the actual and regions segmented by the proposed method are shown in Fig. 10. From the figure it can be concluded that the proposed method gives promising results in 97.51% cases. An example of good segmentation is shown in Fig. 11.a, while acceptable segmentation can be seen in Fig. 11.b. An example of bad segmentation is shown in Fig. 11.c where a label was not removed due to its nearness to the breast border.
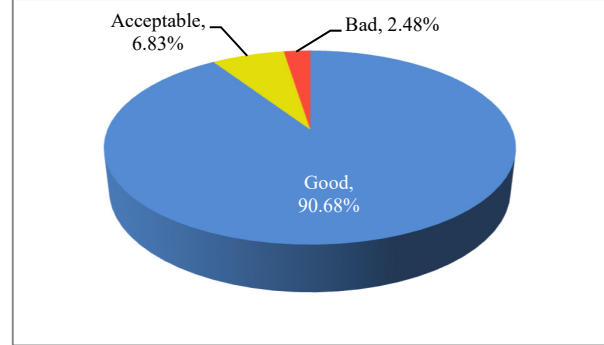

Fig. 10 Results of breast region segmentation.

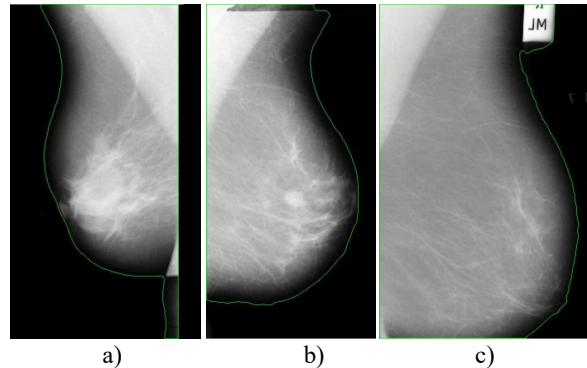a)   b)   c)
Fig. 11. Breast region segmentation: a) Good (mdb207); b) Acceptable (mdb010); c) Bad (mdb006).

The results of pectoral muscle removal are shown in Fig. 12. The success rate is lower due to large differences in the intensity and texture of pectoral muscles in different images. Example of good segmentation is presented in Fig. 13, while an example of acceptable segmentation can be seen in Fig. 14. Bad segmentation can be seen in Fig. 15 where, due to big differences of contrast in the pectoral muscle, just one part of it is detected.
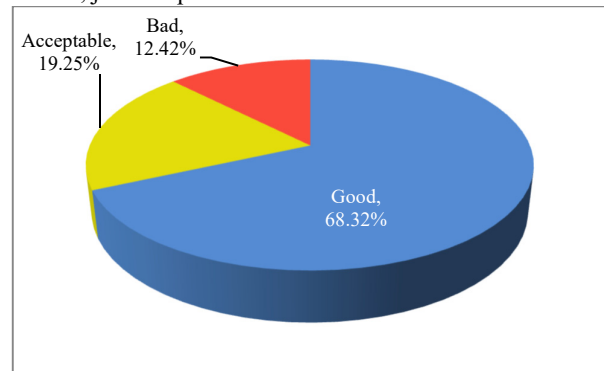

Fig. 12. Results of pectoral muscle removal.

V. CONCLUSION

A new method for breast region segmentation and pectoral muscle removal is proposed. Breast segmentation is done using adaptive gamma correction with a weighting distribution to improve the contrast of the image, after which the k-means method has been applied. Morphological operations on a binary image give a finer breast border.
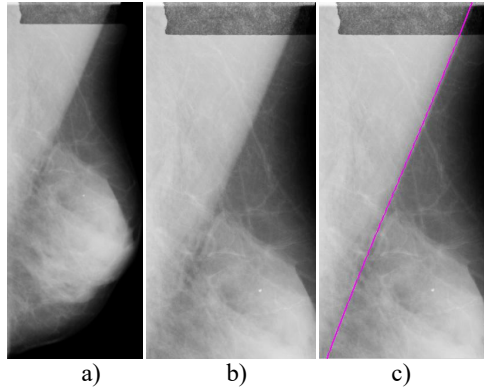

Fig. 13 Good pectoral muscle removal
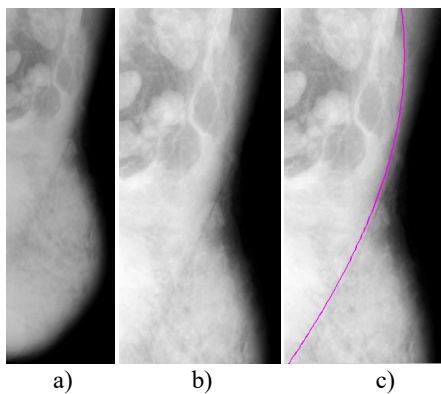a) Input image (mdb289); b) ROI; c) Pectoral muscle.

Fig. 14. Acceptable pectoral muscle removal
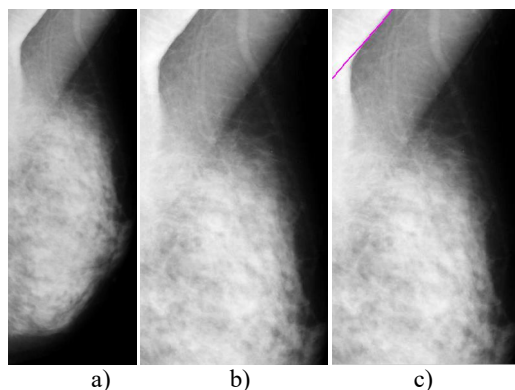a) Input image (mdb061); b) ROI; c) Pectoral muscle.

Fig. 15. Bad pectoral muscle removal
a) Input image (mdb039); b) ROI; c) Pectoral muscle.

To remove the pectoral muscle, it is necessary to determine the orientation of the breast and position all the images so that the pectoral muscle is in the upper left corner. Contrast enhancement of the region of interest is preformed and the binary mask representing the pectoral muscle is found using the k-means algorithm. Estimation of curve representing the border of the pectoral muscle is done by using the third order polynomial fitting. The efficiency of the method was tested on 322 images from reference miniMIAS database. The breast region is well segmented in 90.68% cases, acceptably in 6.83% and bad in 2.48% cases. The success rate of pectoral muscle removal is lower because of the large difference in the intensity and texture of muscles in images. Pectoral muscle is well segmented in 68.32% cases, acceptably in 19.25% , and bad in 12.42% cases.

REFERENCES

[1] "WHO | Breast cancer: prevention and control," *WHO*, 2016.
[2] T. Berber, A. Alpkocak, P. Balci, and O. Dicle, "Breast mass contour segmentation algorithm in digital mammograms," *Comput. Methods Programs Biomed.*, vol. 110, no. 2, pp. 150–159, May 2013.
[3] B. Reljin, Z. Milosevic, T. Stojic, and I. Reljin, "Computer aided system for segmentation and visualization of microcalcifications in digital mammograms," *Folia Histochem. Cytobiol.*, vol. 47, no. 3, pp. 525–532, 2009.
[4] M. M. Eltoukhy, I. Faye, and B. B. Samir, "Breast cancer diagnosis in digital mammogram using multiscale curvelet transform," *Comput. Med. Imaging Graph.*, vol. 34, no. 4, pp. 269–276, 2010.
[5] J. Dheeba, N. Albert Singh, and S. Tamil Selvi, "Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach," *J. Biomed. Inform.*, vol. 49, pp. 45–52, Jun. 2014.
[6] C. Varela, P. G. Tahoces, A. J. Méndez, M. Souto, and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms.," *Comput. Biol. Med.*, vol. 37, no. 2, pp. 214–26, Feb. 2007.
[7] X. Lladó, A. Oliver, J. Freixenet, R. Martí, and J. Martí, "A textural approach for mass false positive reduction in mammography.," *Comput. Med. Imaging Graph.*, vol. 33, no. 6, pp. 415–22, Sep. 2009.
[8] Z. Wang, G. Yu, Y. Kang, Y. Zhao, and Q. Qu, "Breast tumor detection in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 128, pp. 175–184, Mar. 2014.
[9] P. Agrawal, M. Vatsa, and R. Singh, "Saliency based mass detection from screening mammograms," *Signal Processing*, vol. 99, pp. 29–47, Jun. 2014.
[10] C.-C. Liu, C.-Y. Tsai, T.-S. Tsui, and S.-S. Yu, "An improved GVF snake based breast region extrapolation scheme for digital mammograms," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4505–4510, 2012.
[11] P. Kus and I. Karagoz, "Fully automated gradient based breast boundary detection for digitized X-ray mammograms.," *Comput. Biol. Med.*, vol. 42, no. 1, pp. 75–82, Jan. 2012.
[12] I. K. Maitra, S. Nag, and S. K. Bandyopadhyay, "Technique for preprocessing of digital mammogram.," *Comput. Methods Programs Biomed.*, vol. 107, no. 2, pp. 175–88, Aug. 2012.
[13] C.-C. Liu, C.-Y. Tsai, J. Liu, C.-Y. Yu, and S.-S. Yu, "A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis," *Comput. Math. with Appl.*, vol. 64, no. 5, pp. 1100–1107, 2012.
[14] M. Slavkovic-Ilic, A. Gavrovska, M. Milivojevic, I. Reljin, and B. Reljin, "The algorithm for automatic segmentation of breast and pectoral muscle regions from mammograms," in *2015 23rd Telecommunications Forum Telfor (TELFOR)*, 2015, pp. 803–806. (in Serbian)
[15] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
[16] M. Mustra and M. Grgic, "Robust automatic breast and pectoral muscle segmentation from scanned mammograms," *Signal Processing*, vol. 93, no. 10, pp. 2817–2827, Oct. 2013.
[17] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The Mammographic Image Analysis Society Digital Mammogram Database," *Expert. Medica, Int. Congr. Ser.*, vol. 1069, pp. 375–378, Jan. 1994.
[18] "The mini-MIAS database of mammograms." [Online]. Available: http://peipa.essex.ac.uk/info/mias.html. [Accessed: 05-Oct-2015].