Automatic Breast-Line and Pectoral Muscle Segmentation

Kamila Czaplicka, Justyna Włodarczyk
IPPT PAN – UJ joint Ph.D. studies of Computer Science,
Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland
e-mail: *kamila.czaplicka@.uj.edu.pl*, *justyna.wlodarczyk@.uj.edu.pl*

Abstract. Pre-processing of mammograms is a crucial step in computer-aided analysis systems. The aim of segmentation is to extract a breast region by estimation of a breast skin-line and a pectoral muscle as well as removing radiographic artifacts and the background of the mammogram. Knowledge of the breast contour also allows further analysis of breast abnormalities such as bilateral asymmetry. In this paper we propose a fully automatic algorithm for segmentation of a breast region, based on two types of global image thresholding: the multi-level Otsu and minimizing the measure of fuzziness as well as the gradient estimation and linear regression. The results of our experiments showed that our method can be used to find a breast line and a pectoral muscle accurately.

Keywords: pectoral muscle, breast boundary, mammogram, segmentation.

1. Introduction

1.1. The role of mammography and CAD systems

According to the reports based on data of National Cancer Registry [1] breast cancer is the most frequently diagnosed form of cancer of Polish women, accounting for approximately 20% of all new cancer cases each year. There were 14576 cases diagnosed in 2008 and breast cancer was the cause of 5362 deaths. Since 1999 mortality rates have increased progressively, about 1–3% per year. The main cause of

this situation is late detection of breast cancer. In an attempt to achieve an earlier diagnosis Polish Ministry of Health introduced breast cancer screening, which involves mammographic screening. Mammography is a fundamental method used in detection of breast cancer among asymptomatic patients. The detection of an asymptomatic cancer increases a cure rate significantly and reduces a mortality rate by about 25% [2]. Mammograms can be used to check for breast cancer not only women who have no signs or symptoms of the disease (a screening mammogram) but also to check for breast cancer after a lump has been found (a diagnostic mammogram). The X-ray images make it possible to detect tumors (that often cannot be felt), microcalcifications (small deposits of calcium which sometimes indicate the presence of breast cancer) and architectural disturbances. Sensitivity of mammography in detecting changes in a clinically asymptomatic breast is about 93% (75–80% for a dense breast) [2].

A radiologist investigates mammograms for the identification (detection) of potential abnormalities and classifies them with respect to malignancy (diagnosis). Even high quality mammograms can be misread and misinterpreted by experienced radiologists, whose sensitivity is about 70% [2]. A mammogram misinterpretation can be caused by the great variability in the appearance of normal and abnormal tissue as well as by the so called human factor. The most common cause of wrong interpretation of mammograms is simply overlooking (caused for e.g. by tiredness), then, on the second place, wrong classification of a pathology. It has been shown that an independent second reading of mammograms improves the sensitivity of mammography by as much as 15% [3]. Instead of the second reading done by a physician, computer systems can be used for abnormalities detection and diagnosis (Computer-Aided Detection- CADe and Computer-Aided Diagnosis- CADi), which is often a cheaper and faster option.

In this paper we propose a method that can be used while image pre-processing in the detection process.

1.2. Types of mammograms

The classical mammogram is an analog image: both coordinates of points in the image and gray levels can have any values from fixed continuous ranges, whereas in digital imaging these values are quantized. Except for that and the difference in how the image is recorded and stored, there is no other difference between the digital and analog mammography. However digital information can be enhanced, magnified, or manipulated for further evaluation more easily than information stored on a film. In order to receive a digital mammogram three methods can be used:

I. secondary digitalization of analog mammograms using scanners (with all drawbacks of the original image, like artifacts of the development process or excessive/insufficiently exposed areas),

II. indirect digital imaging (apart from classic analog mammographs it uses digital memory plates),

III. direct digital imaging, where an electronic image of the breast is created on the spot.

In screening mammography proper mammographic examination typically involves taking two views of the breast, from above (a cranial-caudal view, CC) and from an oblique or angled view (a mediolateral-oblique view, MLO). In the CC view, the entire breast parenchyma (glandular tissue) should be depicted. In the lateral view, the pectoral (chest) muscle should be depicted as a narrow light band on at least half of the picture. The nipple should be depicted in a profile and a clear stomach fold should be visible under the breast. A pectoral muscle can be found in images in the MLO view whereas in the CC view sporadically.

1.3. The aim of our work

In our work we perform breast segmentation. This task includes breast line and pectoral muscle extraction as well as removing artifacts and noise. In our work we have chosen pictures in the mediolateral-obligue view, because this view is taken obligatory while screening (whereas the CC view is not) and a pectoral muscle can always be seen there. This kind of segmentation is important before CAD systems start. Their region of interest (ROI) is the breast only; background artifacts (including labels) and the pectoral muscle, which represent the predominant density region, could affect negatively their effectiveness. The knowledge of breast contour helps to find abnormalities in breast line profile as well as it is useful in comparing images of the same person made over certain amounts of time. Accurate breast skin-line detection is especially important for locating malignant lesions near the skin-line.

1.4. An overview of mammogram segmentation methods

Over last three decades scientists have willingly undertaken the research in breast segmentation. Traditional histogram based algorithms provided good results, but nowadays these methods are only the first step in segmentation. They are improved by gradient methods, local histogram analysis etc. Apart from them, there are newer approaches, which have also proved their effectiveness, such as: active contours, classifiers, fuzzy-logic, polynomial modeling, wavelet decomposition, a topographic approach or Markov chains. Tab. 1 shows selected works representing each of the mentioned methods.

Tab. 1. Classification of methods used in breast segmentation

Method	1990's	2000's
Thresholding	Lau, Bischof (1991) [27]; Yin, Giger, Doi et al. (1991) [34]; Abdel-Mottaleb, Carman, Hill et al. (1996) [35]	Raba, Oliver, Marti et al. (2005) [5]; Subashini, Ramalingam, Palanivel (2010) [4]; Abubaker, Aqel, Qahwaji et al. (2005) [29]; Masek, Attikiouzel, deSilva (2000) [32]
Gradient	Mendez, Tahoces (1996) [6]; Karssemeijer, Brake (1998) [9]	Zhou, Chan, Petrick et al. (2001) [7]; Kwok, Chandrasekhar, Attikiouzel (2001) [8]
Active contours	Ojala, Liang (1999) [25]; Bick, Giger, Schmidt et al. (1995) [33]	Liang, McInerney, Terzopoulos (2006) [12]; Marti, Oliver, Raba et al. (2007) [10]; Wirth, Stapinski (2003) [11]
Classifiers	Karssemeijer (1998) [26]	Saha, Udupa (2001) [13]; Tromans, Brady, Warren (2004) [14]
Fuzzy logic		Wirth, Nikitenko, Lyon (2005) [16]
Markov chains		Wang, Zhu, Deng et al. (2010) [15]
Polynomial modeling		Ferrari, Rangayyan (2005) [30]; Mirzaalian, Ahmadzadeh, Kolahdoozan (2006) [31]
Wavelet decomposition		Bozek, Grgic, Mustra (2009) [36]
Topographic approach		Hong, Sohn (2010) [24]

2. A proposed method

2.1. Mammograms' features

Breasts are made up of gland, fat and thickened (fibrous) tissue. The nature of X-ray images has obvious limitation due to presenting 2D information of 3D objects-each pixel on a mammogram represents superposition of two or more overlapping tissues. That makes the differentiation between regions more difficult. The intensity

of each pixel is proportional to the level of absorption X-ray by the tissues in the breast- as the X-ray attenuation increases, the gray-levels in the mammogram also increase. The breast periphery and the skin-line region are less dense compared to the other neighbouring tissues so they are normally very low in gray-level contrast on mammograms and make the process of breast skin-line identification harder. In contrast, the pectoral muscle is more dense and is characterized by pixels of high intensity. The background is the region of a low intensity pixels and low gradient.

2.2. The data set

All testing images come from the Pilot European Image Processing Archive (PEIPA) mini-MIAS database of mammograms [17], which is free to use in scientific research under the license agreement. Every of 322 pgm images in the database was digitized at the 50 micron pixel edge, reduced to the 200 micron pixel edge and clipped/padded so that every image is 1024×1024 pixels. The MIAS database provides appropriate details of each image, such as a character of background tissue, class, severity and image-coordinates of abnormality (if it is present).

2.3. Algorithm details

Our algorithm consists of four steps:

   I. Breast direction recognition, image trimming and reflection.

  II. Morphological opening of an image.

 III. Pectoral muscle extraction.

 IV. Breast skin-line estimation.

The first two steps are general image pre-processing steps, whereas the last two are indeed two separate sub-algorithms for pectoral muscle extraction and breast line estimation.

2.3.1. Breast direction recognition, image trimming and reflection

The MIAS database contains images of left and right breasts, with radiologists' markers, artifacts and empty black regions (Fig. 1). Labels, artifacts and black regions can disturb further analysis so it is better to remove them from the image. To make images uniform and further working with them easier, it is also necessary to recognize the breast direction from the pectoral muscle towards the nipple. In

the next steps we assume that the pectoral muscle is on the top left corner of the image, so now we have to find the side of the image with the pectoral muscle and reflect the image if it is needed. As it was mentioned earlier, as the pectoral muscle is dense, it is characterized by high intensity pixels. It is safe to assume that in the mammogram the pectoral muscle is the biggest object with a regular (triangular) region of similar, only high intensity pixels, that is located on the upper side of the image.
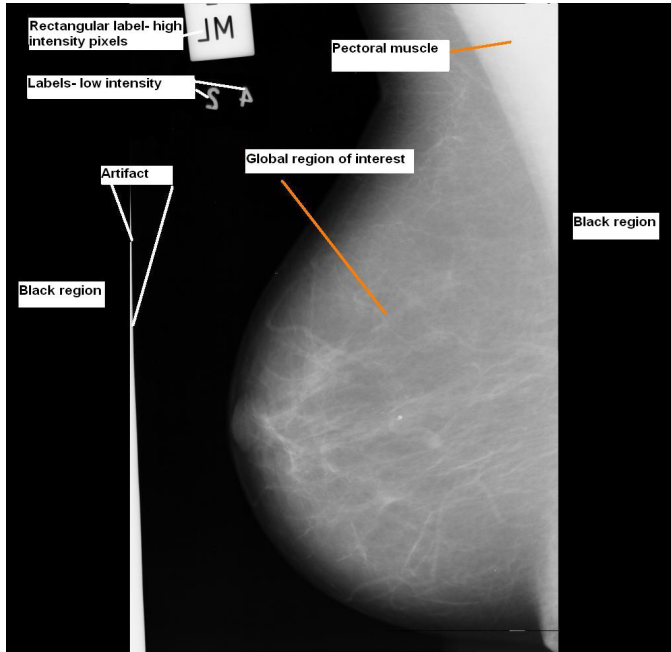

Fig. 1. An example of the image from MIAS database

We also know that unwanted black left and right border regions consist of pixels with zero intensity. This assumption is used while scanning through the image from top to bottom, at regular line intervals (in our tests we scanned about 10% of all image lines). For each scanned line continuous ranges of close-to-zero intensity pixels on both right and left sides are to be detected. The smallest number from all border pixels from the left side and the biggest number from all border pixels from the right side define cut-off points for the black regions.

Assumptions on pectoral muscle features are crucial in the next round of scanning which takes place on the already trimmed image and affects all lines. In this stage for each line we count numbers of pixels with a non-zero intensity from the left border to the center of the image and from the center to the right border. The side with the bigger number of lines with non-zero pixels 'wins' and from now on we consider this side as having the pectoral muscle. If it is the right side, the image is reflected so as to have the pectoral muscle on its top left corner. The effect of trimming and reflecting the image is shown in Fig. 2.
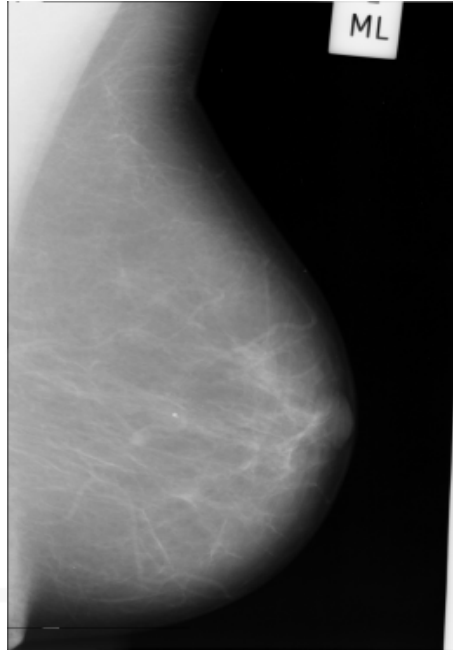
**Fig. 2**. The trimmed image with the pectoral muscle in the top-left corner

Other unnecessary elements such as markers and artifacts are removed as a side-effect of an algorithm of extracting the pectoral muscle.

2.3.2. Morphological opening of the image

The use of a morphological operation of opening is a standard procedure used for smoothing images by removing speckle noise. In our work we applied opening with a structuring element of size 3×3.

2.3.3. Pectoral muscle extraction

In this stage we have smoothed the image with the pectoral muscle on the demanded location and without unnecessary black regions, but still with labels and other artifacts outside the breast. The operation we make on such a prepared image is multilevel thresholding done with the multi-level Otsu (multiotsu) algorithm [18], which classifies pixels in a number of classes based on a number of gray levels (Fig. 3). We know that the pectoral muscle is right-angled triangle shaped, so we do not have to analyze the whole image, but we could cut some part of it which is on the

<center>(a)   (b)   (c)</center>
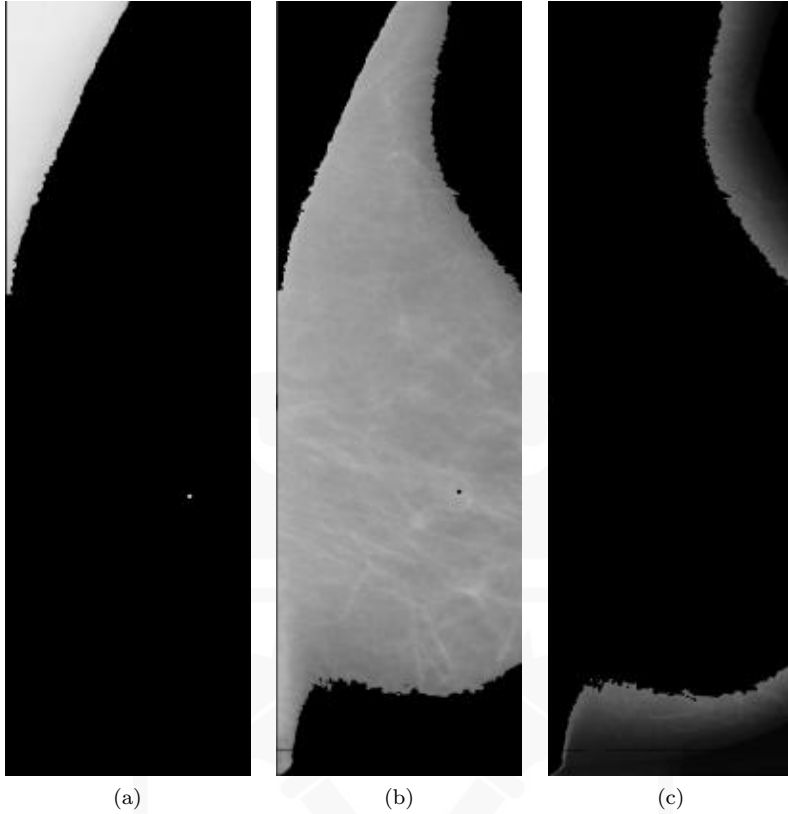
<center>**Fig. 3**. The result of the multiotsu algorithm with two thresholds</center>

right after the right-most high intensity pixels strip on the top. We start with two thresholds and increase this number by one as long as the rest of the operations made in this steps will not bring satisfying results, a definition of which will be explained later. The intention of these proceedings is to have the region with high intensity pixels containing most of pectoral muscle pixels, the other region with low intensity background pixels and at least one intermediate region.

Afterwards we take the region with the pectoral muscle and one of the intermediate regions with pixels of one level below (Fig. 3a and 3b). It should be the region that is adjacent to the first with the pectoral muscle. Next we compute a transitional area between that two regions. If this area contains certain percentage of black pixels it is considered that the second region is useless for further calculations and the multiotsu algorithm is restarted with a greater number of thresholds. On the other case (as in Fig. 3a and 3b) we assume that the chosen area contains the border between the pectoral muscle and the breast.

We calculate hypothetical coordinates of this border by applying the gradient operator put together with some adjustment calculations (e.g. we ignore points which are too far from the sequence of points previously recognized as border points,

Fig. 4a). Having border points we use linear regression to smooth the line formed by the chosen points – the resultant polynomial is acknowledged as a border of the pectoral muscle (Fig. 4b).


(a)　　　　　　　　　　　　　　　(b)
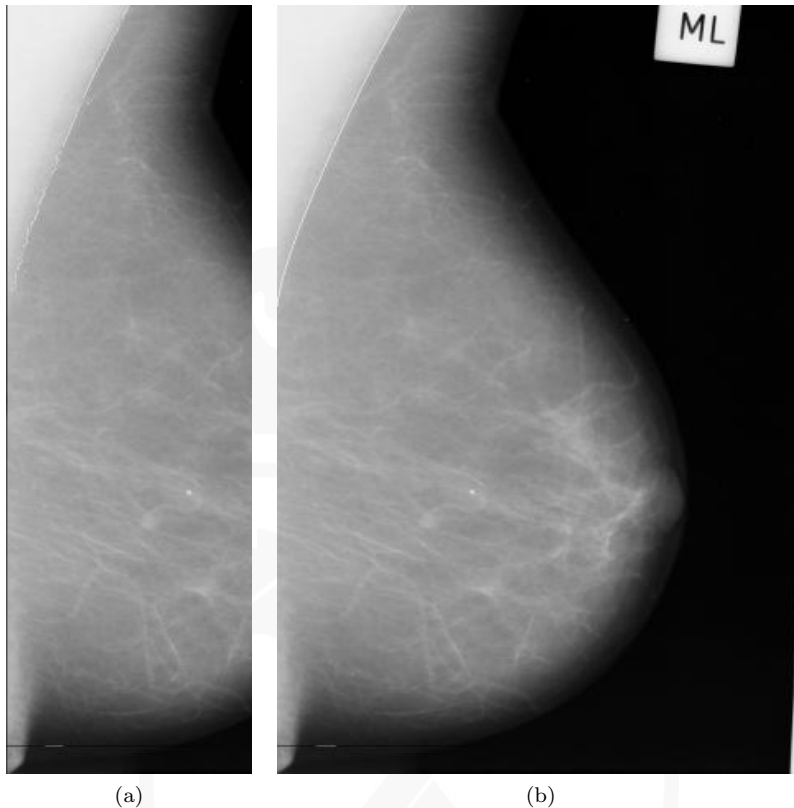
Fig. 4. Border points and the result of linear regression on the trimmed image

As it was mentioned earlier labels and other artifacts are removed on the occasion of previous operations. Indeed, after the multiotsu algorithm we have the image with the region of low intensity pixels containing the background and skin-air interface. Again if we compute the gradient on this image we will estimate the approximate breast border. Then if we subtract everything in the image which is on the right from this rough border we will get the image with labels and artifacts removed (Fig. 5).

2.3.4. Breast skin-line estimation

Detection of the breast border is a two-stage process based on a combination of image binarization and edge detection. In the first stage we use a global thresholding

(a)   (b)
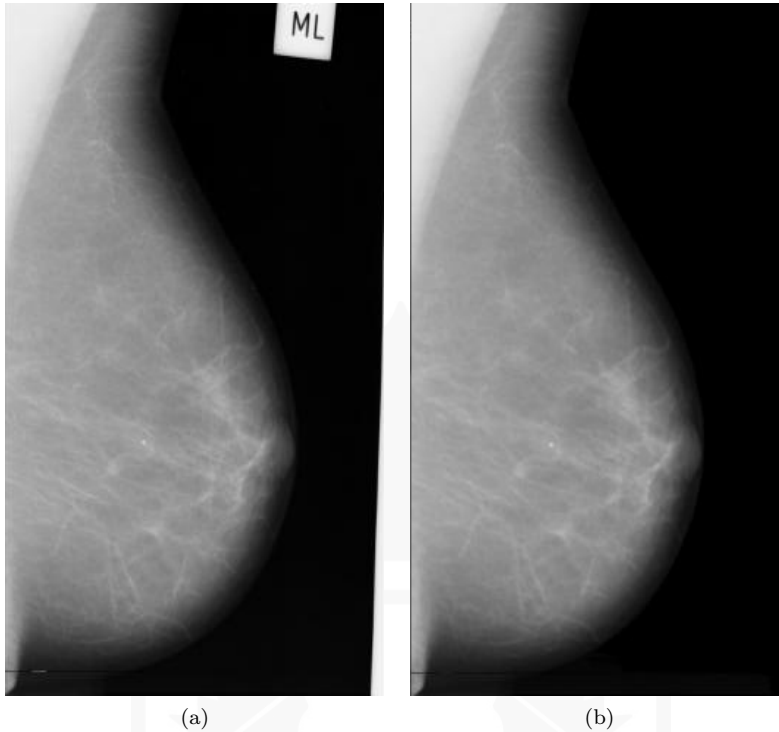
Fig. 5. The image after removing artifacts

method based on minimizing measures of fuzziness of an input mammogram, described by Huang, Wang [19]. After this step we have the image that is a collection of two regions: the background and the union of the breast and the pectoral muscle (Fig. 6a). In the final step we apply the Sobel method to find edges using the Sobel approximation to the derivative. It returns edges at those points where the gradient of horizontal and vertical measure is maximal (Fig. 6b). These edges are assumed as breast skin-line estimation.

3. Results

Our algorithm was tested on 300 selected mammograms from the mini-MIAS database. Each analyzed image was examined by an experienced radiologist who classified it to one of the two categories: correctly segmented or incorrectly segmented. Then the efficiency of our algorithm was evaluated by measuring the total percentage error (TPE) expressed as a percentage quotient of the number of incorrectly segmented images and the total number of images:

(a)   (b)
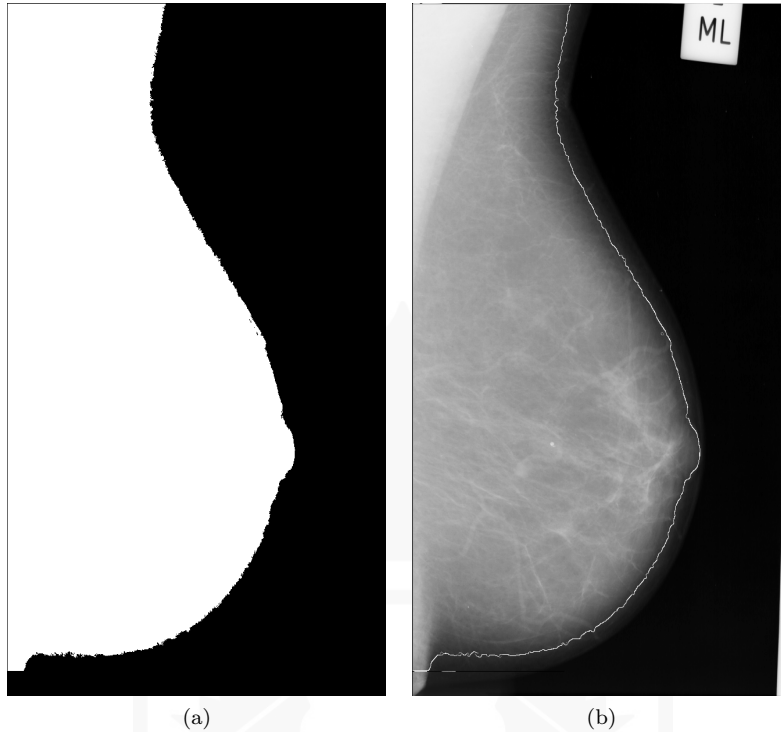
**Fig. 6**. Breast skin-line estimation

$$\text{TPE} = \frac{\text{the number of incorrectly segmented images}}{\text{the number of all images}} 100\%.$$

The algorithm proved its effectiveness achieving 1% level of the TPE rate. The correctness of the estimation separately for the breast line and the pectoral muscle was on the level of 99% and 98%, respectively. Artifacts were properly removed for nearly 100% images.

The results obtained by the proposed method indicate that it can overcome limitations of existing methods such us:

   I. over-segmentation or under-segmentation of the breast region or the pectoral muscle region in the case of the fuzzy boundary between the muscle and the tissue [5] (Fig. 7a, b, c, d, e),

  II. erroneous detection of the pectoral muscle when:

     • the pectoral muscle boundary is a curve (i.e. is not close to a straight line) [20] (Fig. 7a),

     • the intensities of the muscle region change greatly [21] (Fig. 7b),

- the area of the pectoral muscle in comparison with the breast area is larger (Fig. 7c) or smaller (Fig. 7e) than typically [23],

- the pectoral muscle (seems to) consists of several layers [22] (Fig. 7a, b).

(a)   (b)   (c)

(d)   (e)
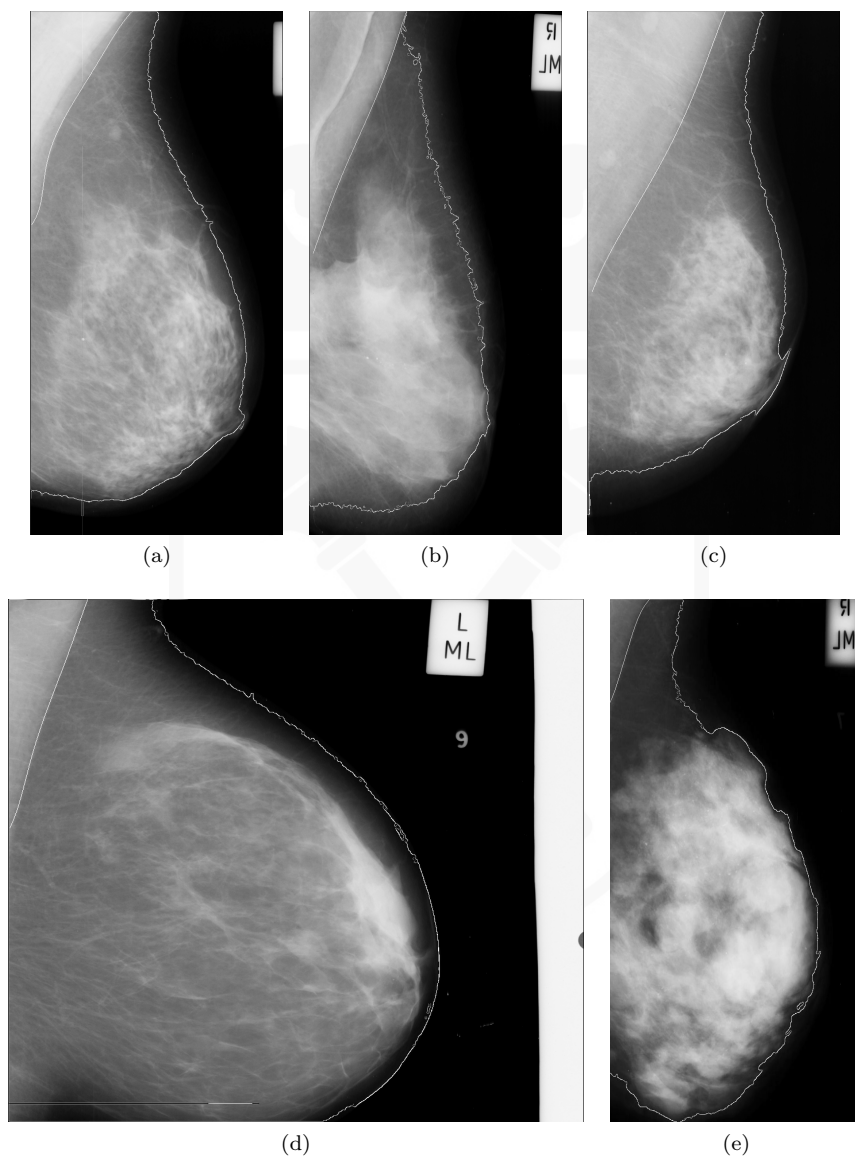
Fig. 7. Results of pectoral muscle extraction on a selection of non-standard muscles

4. Concluding remarks

The results of our experiments showed that the described algorithm is characterized by the high efficiency- the total percentage error was equal to approximately 1%. In comparison with algorithms described in literature [5, 10, 11, 15], where effectiveness ranged from 95% to 97%, the accuracy of our method is higher by about 2%. The approach presented for breast segmentation is among the most effective algorithms and our results are similar to the best algorithms described [16, 28]. The proposed method can almost completely eliminate the possibility of wrong detection of the pectoral muscle region and the breast skin-line. In this context, our fully automatic algorithm for breast segmentation in mammograms can be used as a first processing step in other medical image applications that enable detection of abnormal areas of density, mass or calcification which may indicate the presence of cancer.

5. References

[1] *National Cancer Registry*. Available via `http://epid.coi.waw.pl/krn`.

[2] Dziukowa J., Wesołowska E.; *Mammografia w diagnostyce raka sutka*, MediPage, Warszawa 2006.

[3] Wirth M.A., Stapinski A.; *Segmentation of the breast region in mammograms using active contours*, 1st Canadian Conference on Computer and Robot Vision, 2004, pp. 385–392.

[4] Subashini T.S., Ramalingam V., Palanivel S.; *Pectoral Muscle removal and Detection of masses in Digital Mammogram using CCL*, International Journal of Computer Applications, 1(6), 2010, pp. 66–70.

[5] Raba D., Oliver A., Marti J., Peracaula M. et al.; *Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms*, Pattern Recognition and Image Analysis, Springer, 3523, 2005, pp. 471–478.

[6] Mendez A., Tahoces P., Lado M.J. et al.; *Automatic detection of breast border and nipple in digital mammograms*, Computer methods and programs in biomedicine, 49(3), 1996, pp. 253–262.

[7] Zhou C., Chan H.P., Petrick N. et al.; *Computerized image analysis: estimation of breast density on mammograms*, Medical Physics, 28(6), 2001, pp. 1056–1069.

[8] Kwok S.M., Chandrasekhar R., Attikiouzel Y.; *Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection*, Proceedings of the Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001, pp. 67–72.

[9] Karssemeijer N., Brake G.; *Combining single view features and asymmetry for detection of mass lesions*, 4th International Workshop on Digital Mammography, 1998.

[10] Marti R., Oliver A., Raba D. et al.; *Breast Skin-Line Segmentation Using Contour Growing*, Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, 4478/2007, 2007, pp. 564–571.

[11] Wirth M.A., Stapinski A.; *Segmentation of the breast region in mammograms using active contours*, Visual Communications and Image Processing, 5150, 1995, pp. 1995–2006.

[12] Liang J., McInerney T., Terzopoulos D.; *United Snakes*, Medical Image Analysis, 10, 2006, pp. 215–233.

[13] Saha P., Udupa J.; *Breast tissue density quantification via digitized mammograms*, IEEE Transactions on Medical Imaging, 20, 2001, pp. 792–803.

[14] Tromans C., Brady J., Warren R.; *A high accuracy technique for breast air boundary segmentation and the resulting improvement from its use in breast density estimation*, International Workshop on Digital Mammography, 2004, pp. 17–18.

[15] Wang L., Zhu M., Deng L. et al.; *Automatic pectoral muscle boundary detection in mammograms based on Markov chain and active contour model*, Journal of Zhejiang University-SCIENCE C (Computers & Electronics), 11(2), 2010, pp. 111–118.

[16] Wirth M., Nikitenko D., Lyon J.; *Segmentation of the Breast Region in Mammograms using a Rule-Based Fuzzy Reasoning Algorithm*, ICGST International Journal on Graphics, Vision and Image Processing, 5(2), 2005, pp. 45–54.

[17] *Pilot European Image Processing Archive's database of mammograms.* Available via `http://peipa.essex.ac.uk/info/mias.html`.

[18] Liao P.S., Chen T.S., Chung P.C.; *A fast algorithm for multilevel thresholding*, Journal of Information Science and Engineering, 17(5), 2001, pp. 713–727.

[19] Huang L.-K., Wang M.-J. J.; *Image Thresholding by Minimizing the Measures of Fuzziness*, Pattern Recognition, 28(1), 1995, pp. 41–51.

[20] Kinoshita S.K., Azevedo-Marques P.M., Pereira R.R. et al.; *Radon-Domain Detection of the Nipple and the Pectoral Muscle in Mammograms*, Journal of Digital Imaging, 21(1), 2007, pp. 37–49.

[21] Hong B., Brady M.; *A Topographic Representation for Mammogram Segmentation*, Medical Image Computing and Computer-Assisted Intervention, 2, 2003, pp. 730–737.

[22] Carvalho I.M., Luz L.M.S., Alvarenga A.V. et al.; *An Automatic Method for Delineating the Pectoral Muscle in Mammograms*, IFMBE Proceedings, 18(2), 2008, pp. 271–275.

[23] Xu W.D., Li L.H., Liu W.; *A Novel Pectoral Muscle Segmentation Algorithm Based on Polyline Fitting and Elastic Thread Approaching*, The 1st International Conference on Bioinformatics and Biomedical Engineering, 2007, pp. 837–840.

[24] Hong B.W., Sohn B.S.; *Segmentation of Regions of Interest in Mammograms in a Topographic Approach*, IEEE Transactions on Information Technology in Biomedicine, 14(1), 2010, pp. 129–139.

[25] Ojala T., Liang J.; *Interactive segmentation of the breast region from digitized mammograms with united snakes*, Technical Report 315, Turku Centre for Computer Science, 1999.

[26] Karssemeijer N.; *Automated classification of parenchymal patterns in mammograms*, Physics in Medicine and Biology, 43(2), 1998.

[27] Lau T., Bischof W.F.; *Automated Detection of Breat Tumors Using the Assymetry Approach*, Computers and Biomedical Research, 24, 1991, pp. 273–295.

[28] Yapa R.D., Harada K.; *Breast Skin-Line Estimation and Breast Segmentation in Mammograms using Fast-Marching Method*, International Journal of Biological and Life Sciences, 3(1), 2007.

[29] Abubaker A., Aqel M., Qahwaji R. et al.; *Average Row Thresholding Method for Mammogram Segmentation*, The 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2005, pp. 3288–3291.

[30] Ferrari R., Rangayyan R.; *Automatic identification of the pectoral muscle in mammograms*, IEEE Transactions on Medical Imaging, 23, 2004, pp. 232–245.

[31] Mirzaalian H., Ahmadzadeh M.R., Kolahdoozan F.; *Breast Contour Detection on Digital Mammogram*, Proceedings of 2nd International Conference on Information and Communication Technologies: from Theory to Applications, 2006, pp. 1804–1808.

[32] Masek M., Attikiouzel Y., deSilva C.J.S.; *Skin-air interface extraction from mammograms using an automatic local thresholding algorithm*, 15th Biennial International Conference Biosignal, 2000, pp. 204–206.

[33] Bick U., Giger M.L., Schmidt R.A. et al.; *Automated segmentation of digitized mammograms*, Academic Radiology, 2, 1995, pp. 1–9.

[34] Yin F., Giger M.L., Doi K. et al.; *Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images*, Medical Physics, 18(5), 1991, pp. 955–963.

[35] Abdel-Mottaleb M., Carman C.S., Hill C.R. et al.; *Locating the boundary between the breast skin edge and the background in digitized mammograms*, 3rd International Workshop on Digital Mammography, 98, 1996, pp. 467–470.

[36] Bozek J., Grgic M., Mustra M.; *Breast border extraction and pectoral muscle detection using wavelet decomposition*, Proceedings of the International IEEE Conference EUROCON, 2009, pp. 1426–1433.