[17] Computer diagnosis of tumors in mammograms

Sadykov S.S., Bulanova Yu.A.
Murom Institute (branch) of the Federal State Budget Institution of Higher Professional Education "Vladimir State University n.a. Alexander G. and Nicholas G. Stoletovs"
Zakharova E.A.
Non-state health care facility "Departmental hospital at the station Murom Open Joint Stock Company "Russian Railways"

Abstract

This paper proposes a technology of computer diagnosis of tumors in mammary gland of three main types: cyst, fibroadenoma, and breast cancer. It is noted that the use of such technology in the health care setting reduces the chance of missing a tumor when viewing mammograms, at an early stage of development, and also minimizes the subjectivity of diagnosis. Work on a set of technologies studied real mammograms from MIAS database with known diagnoses. The results of experiments show the possibility of using the developed integrated cabinet in mammography based on this technology as well for general research as for screening.

Keywords: *image processing, preprocessing, texture analysis, diagnosis, mammograms, technology, tumor, cyst, fibroadenoma, breast cancer.*

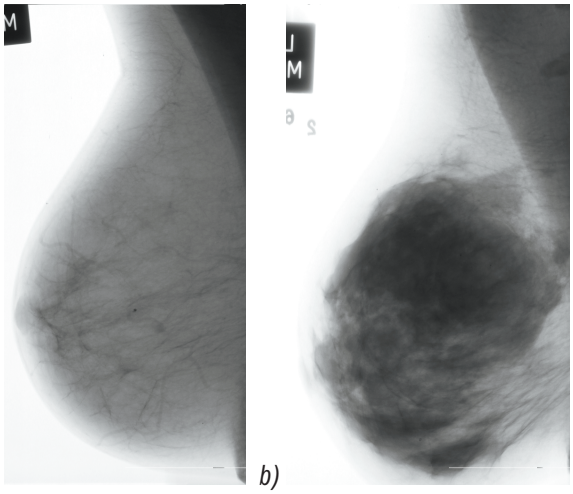
Citation: *Sadykov S.S. Computer diagnosis of tumors in mammograms / Sadykov S.S., Bulanova Yu.A., Zakharova E.A. // Computer Optics. – 2014. – Vol. 38(1). – P. 131-138.*

Introduction

Mammography is the X-ray method of noninvasive examination of mammary gland to identify its pathology. This technique is now considered to be the only method for diagnosis of benign and malignant tumors, i.e. breast cancer (malignant tumor of breast glandular tissue). Oncologic tumors don't cause any discomfort or pain feelings at early stages when medical treatment is still possible, therefore it is important to make an accurate diagnosis at an early stage when a tumor size is approximately equal to 7 mm [1].

Computer Aided Detection systems (CAD) in mammography are mainly focused on expensive digital mammography, which unfortunately are now very slightly used in Russian hospitals. Analogue mammography systems don't involve computer aided systems and, therefore, a common problem in diagnosis of breast pathologies is a human factor, and a lack of automation technology in mammography examination [2].

It is difficult to identify pathologies due to the structure of mammary gland. Mammary gland consists of three types of tissue which can be observed in mammograms: fibroic, glandular and adipoid. Fibroic and glandular tissues have some similar radiographic density and they can't be adequately differentiated in mammograms. Adipoid tissue can better pass X-rays that result in increase of the image contrast. Since soft tissues slightly differ in their X-ray absorption coefficients, the image has a flat contrast, therefore, detection of minor changes in tissues at early stages of disease and identification of small-size tumors is difficult. When projecting mammary gland images, different tissue parts are superimposed on one another in mammograms that also distort overall changes in tissues. It is practically impossible for a doctor-radiologist to detect mammary gland pathologies with various types of mastopathy (mammary tissue accrementition). For example, Fig. 1*a* shows a healthy breast mammogram, and Fig. 1*b* shows how densely mammary tissue accrementition looks, and that breast cancer can't be diagnosed herein [2].



a) *b)*

Fig. 1. Mammograms: of a healthy woman (a), of breast cancer on mastopathy (b)

The foregoing analysis shows the applicability of the technology of computer diagnosis of tumors based on analogue mammograms available for both subspecialists and general practitioners. We can also say that the technology and the system based herewith will be much cheaper than the use of special digital mammography systems for mammary gland research and can be widely applied throughout Russian medical institutions.

1. Test images

Development of the technology of computer diagnosis of tumors is performed in mammograms from MIAS database [3] with known types of diseases and their localization sites.

The original mammogram is represented as some certain function I(x,y). The mammogram is a monochrome image, i.e. there is only one brightness channel. Gradations of image point brightness are distributed within the range of [0, 255]. The mammogram region includes the background, the pectoral muscle, and mammary gland.

2. Mammary gland detection

At the first stage it is necessary to detect mammary gland regions in mammograms, since the processing of the whole mammogram provides distorted results. For this purpose we use the Markov random field as a hierarchical mammogram model [4]. Suppose the brightness distribution is normal in the image, and the image is split into blocks of 7×7 pixels. So, Markov model components shall be formed for a set of blocks $B=\{b_1, b_2,...,b_k\}$, where B – is the set of blocks, k – is the number of blocks into which the image has been split up.

$$W_i = \bigcup_{b_j \bigcap b_i \neq \varnothing} b_j \quad (1)$$

where Wi – is the neighborhood of the block i, $b \in B$, i, j = 1, 2, ...,k.

Two indicators shall be calculated for every neighborhood, i.e. the mean brightness and the root-mean-square deviation (RMSD). Accurate intervals shall be identified for each area, and then similar blocks are to be connected. Studies have shown that the mean brightness of the mammary gland region may vary from 120 to 150, and RMSD – from 12 to 30. Further processing shall be carried out only within a boundary that outlines mammary gland. Using this model in mammary gland detection enables to automatically limit the region for further processing and, moreover, it takes into consideration structural features of mammary glands individually for each patient.

Fig. 2 shows a mammogram model, the background is herein highlighted with black, the pectoral muscle region – with white, and the mammary gland region – with grey.

3. Histogram transformations of mammary gland regions

Specialists perform preprocessing with view of the density of patient's mammary gland using algorithms of image histogram transformation and filtration [1].

A cyst is a benign tumor representing a cavity filled with some stuff. The cyst represents a round– or oval-shaped homogeneous shade in mammography which is similar in its density to the breast glandular tissue [1].

Fig. 2. Mammogram model

In order to detect the cyst on the breast background, the image histogram transformation is used taking into consideration the mean mammogram contrast [1, 5]. Fig. 3 shows an example of the cyst mammogram processing.

Breast cancer is a malignant tumor disease which represents a fuzzy stellate structure with a low density [6].

Improved parameters of breast cancer mammograms requires a radiologist's intervention, i.e. PC-and-radiologist automated work, since for the purpose of diagnosis of this type of tumors it is necessary to consider characteristics of the patient's breast tissue (adipose involution, fibrocystic disease, mammary gland adenosis). The authors have developed three algorithms of preprocessing of mammograms [5, 7, 8], which take into account properties of breast tissues and enable to detect tumors being under mastopathy which can't be observed by the radiologist. Fig. 4 displays the original (Fig. 4a) and processed mammograms (Fig. 4b) with breast cancer, as well as their histograms.

Fibroadenoma is the benign tumor, one of the forms of nodal mastopathy. On mammograms fibroadenoma represents a round– or oval-shaped tumor with a sharp smooth contour. Its density is higher or comparable to the density of the mammary gland tissue [9].

Histogram transformations of the mammary gland region containing fibroadenoma are analytically recorded by formulas (2) and (4).

$$Q_1(x,y) = k_{\max} \times p_1[k(x,y)] - \frac{p_1[k(x,y)] - p_1[0] + k_{mb}}{m \times n - p_1[0]} \quad (2)$$

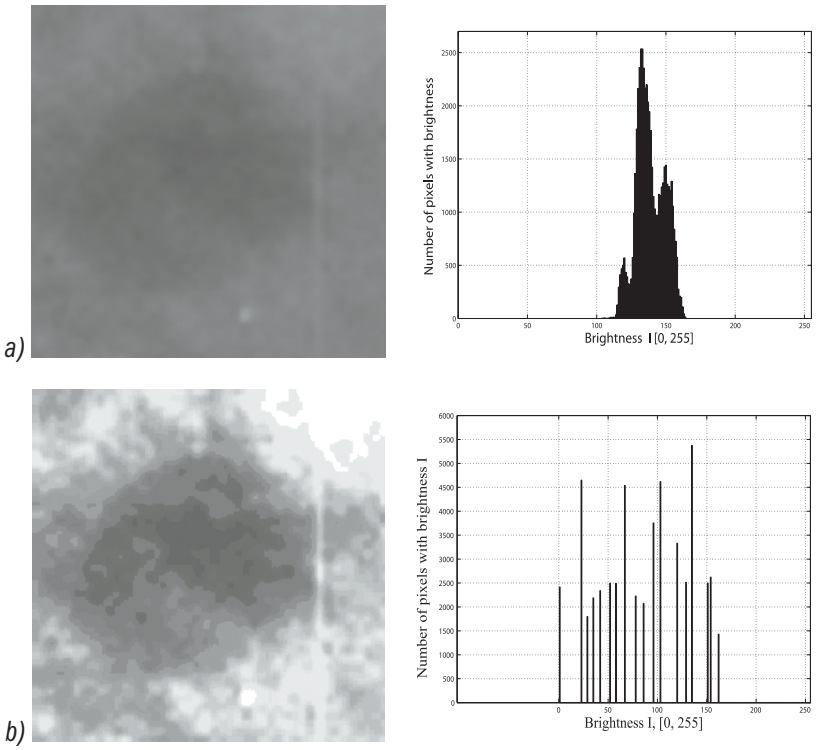


Fig. 3. Cyst mammogram processing: the original mammogram mdb104 and its histogram (a), the processed mammogram and its histogram (b)

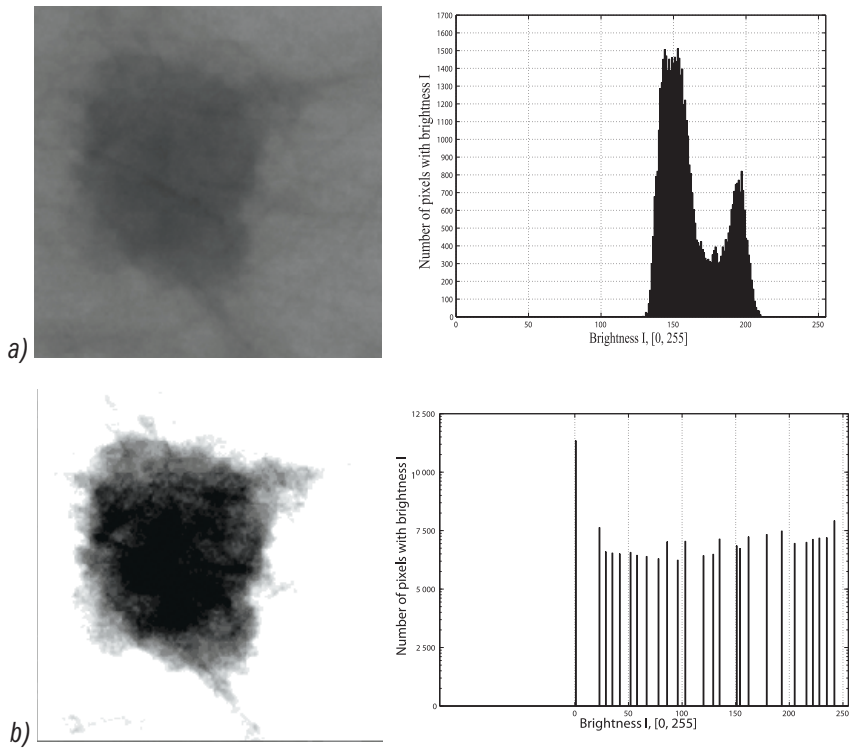


Fig. 4. Breast cancer mammogram processing: the original mammogram mdb134 and its histogram (a), the processed mammogram and its histogram (b)

where $Q_I(x,y)$ – is the output image, $Y(x,y)$ – is the current brightness value of the original mammogram, k_{max}, k_{mb} – are maximum and mean brightness values of the original mammogram, respectively,

$$p_1[i] = \sum_{j=0}^{i} h_1[j], \quad (3)$$

where i – is the value of point brightness gradation of the output image (i=0,1,2,…,255), j – is the value of point brightness gradation of the original mammogram (j=0,1,2, …, 255), H_I – is the distribution area of histogram elements of the original image, $h_I[j]$ – is the value of histogram elements of the original image with the brightness j ($h_I[j] \in H_I$), P_I – is the range of histogram element changes of the output image $Q_I(x,y)$, $p_I[i]$ – is the value of histogram element of the output image with the brightness i ($p_I[i] \in P_I$), $p_I[0] = p_I[Y(x,y)=0]$ – is the value of histogram element of the processed image with the point brightness equal to 0, m×n – is the image size.

$$Q_2(x,y) = \frac{}{n \times m} \times (u[Q_1(x,y)] - u[0]) + \frac{Q_{2\_max} \times (Q_1(x,y) - Q_{1\_mb})}{Q_{1\_mb} \quad Q_{1\_min}} \quad (4)$$

where $Q_2(x,y)$ – is the output image, Q_{2_max} – is the maximum brightness value of the output image, Q_{I_mb}, Q_{I_min} – are the mean and minimum values of the mammogram brightness, respectively, r – is the experimentally selected coefficient (r=0.09).

$$p_2[i] = \sum_{j=0}^{i} h_2[j], \quad (5)$$

where i – is the value of point brightness gradation of the output image (i= 0,1,2, …,255), j – is the value of point brightness gradation of the original mammogram (j=0,1,2, …, 255), H_2 – is the distribution area of histogram elements of the original image, $h_2[j]$ – is the value of histogram element of the original image with the brightness j ($h_2[j] \in H_2$), P_2 – is the range of histogram element changes of the output image $Q_I(x,y)$, $p_2[i]$ – is the value of histogram element of the output image with the brightness i ($p_2[i] \in P_2$), $p_2[0] = p_2[Q_I(x,y)=0]$ – is the value of histogram element of the processed image with the point brightness equal to 0, m×n – is the image size.

4. Filtration

Histogram transformations used in the mammary region usually cause some additional noise; therefore, in order to eliminate them, we use algorithms of median [10] and sigma [11] filtration.

5. Texture segmentation, binarization, boundary detection

At the next step depending on selected preprocessing algorithms, the image is subjected to texture segmentation.

Studies have found that in order to detect cyst, fibroadenoma, and breast cancer tumors we used, as the best suitable, the watershed [12] and Fuzzy C-mean [13] algorithms.

Before we select image boundaries, the image texture map [7] is subjected to binarization. Binarization algorithms are different for each tumor type, for example, breast cancer detection is performed based on the following equation (6).

$$b(x,y) = w \times e(x,y)^5 - u[0] \quad (6)$$

where $b(x,y)$ – is the output image, $e(x,y)$ – is the image texture map,

$$u[i] = \sum_{j=0}^{i} z[j] \quad (7)$$

where i – is the value of point brightness gradation of the output mammogram (i=0,1,2, …,255), j – is the value of point brightness gradation of the original mammogram (j=0,1,2, …,255, j<=i,), z[j] – is the value of histogram element of the original mammogram with the brightness j ($z[j] \in Z$), u[i] – is the value of histogram element of the output image with the brightness i ($u[i] \in U$), $u[0] = u[e(x,y)=0]$ – is the value of histogram element of the processed image with the point brightness equal to 0, w – a coefficient that affects a display image of different types of tissues (if w=0.01, the mammogram only contains pectoral muscle and breast cancer regions) (0 < w<= 10), m×n is the image size.

Boundaries of the detected region are extracted using the algorithm from [14].

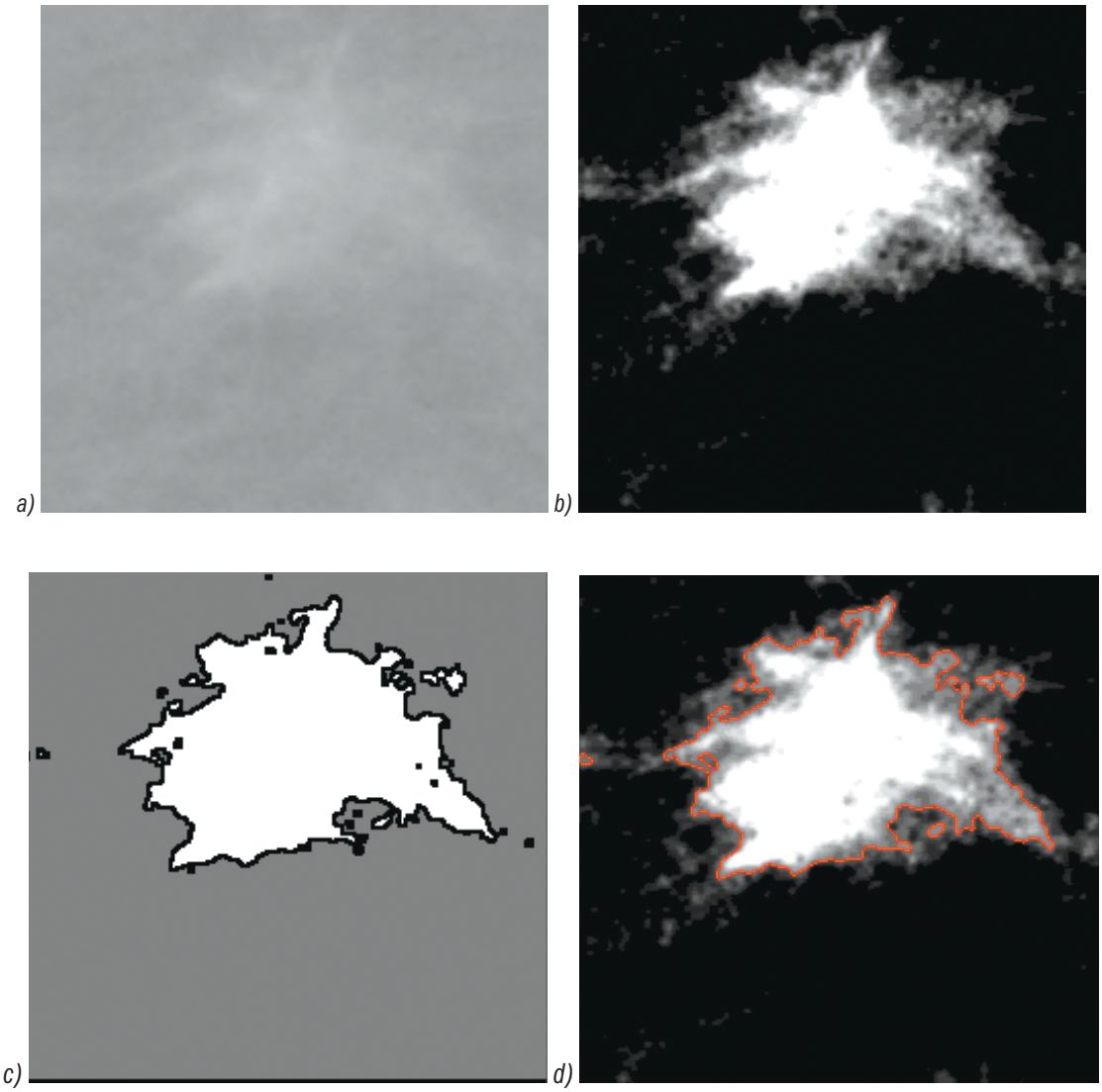


Fig. 5. Breast cancer mammogram processing: the original mammogram mdb 144 (a), preprocessing of the original mammogram (b), texture segmentation (c) superimposed boundaries of the dedicated tumor (c)

The images in Fig. 5 show the process of breast cancer detection in mammography: the original image (Fig. 5*a*) involves the developed algorithm [7] of histogram transformations (Fig. 5*b*) to be applied, then this image is subjected to texture segmentation (Fig. 5*c*), and the boundaries of the detected tumor are superimposed on the processed mammogram (Fig. 5*d*).

6. Textural features

Textural features shall be calculated for each detected region. Full description of tumors is performed using Haralick textural features of the second order. Thus, in order to determine them, we calculate the adjacency matrix [15, 16], which contains relative frequencies p_{ij} of available adjacent elements located at a distance d from each other, with the brightness i, j∈G, G∈[0. 255], where G – is the set of brightness values of a halftone image. For each image we calculate four adjacency matrices in directions of 0°, 45°, 90°, 135°, then we therethrough identify 14 textural features of the second order such as the second angular momentum, a contrast, the difference momentum, the reverse difference momentum, a mean accumulated value, total entropy, difference variation, correlation, a coefficient of variation, the diagonal moment, the second diagonal moment, the multiplication moment.

In order to form the adjacency matrix in this research we used the distance equal to 1 in 4 directions (0°, 45°, 90°, 135°); thus, the total number of features that characterize each detected region equals to 64.

Further, we forme a feature vector of the detected region $V[v_1, v_2, ..., v_{64}]$, and at the next step the obtained feature vector is compared with single reference vectors $V_э[v_{э,1}, v_{э,2}, ..., v_{э,64}]$, i.e. with features form a knowledge database; in this case we calculate minimum values of average amounts of squared deviations [17]. We herewith determine a proper type of tumor, i.e. the computer-assisted diagnosis of breast disease is performed.

7. Experimental studies

Studies were performed in a mammography office of the Non-state health care facility "Departmental hospital at the station Murom Open Joint Stock Company "Russian Railways" (Murom, the Vladimir Region). The office in this hospital is equipped with Mammodiagnost MD-RA analogue mammography device (Philips). It was used to obtain mammograms (about 500 pieces) in the form of X-ray images of the mammary gland on a film 18×24 cm in size. In this type of X-ray studies we used a Kodak MIN-R 2000 film with KODAK MIN-R 2000 and KODAK MIN-R 2190 screens. Characteristics of this screen-to-film combination are as follows: the mean gradient is 3.75, the optical density is over 4D, a supertransparency blue substrate (GF 0.19) and low graininess [12]. To digitize images a scanner Epson Perfection V750 Photo (Epson) was used, which has the following main characteristics: optical density – 4D, resolution – 6400×9600 dpi. [18]. We have filtered a subset involving 222 mammograms with the confirmed diagnosis from MIAS database and 100 mammograms selected from 500 images recorded during the patients examination whose diagnoses were confirmed by biopsy. The size of digitized images is 1024×1024 pixels.

The image subset was processed under close radiologist supervision using the aforementioned tumor detection technology. A match vector was formed for each region of the detected tumor and based to the knowledge database.

Results of tumor detection in mammograms of the used subset via an integrated system constructed on the basis of the developed technology with the aid of a practitioner radiologist have been integrated in Table 1.

Table 1. Results of the tumor detection diagnosis

Parameters		Number of mammograms	Radiologist		System	
"Standard" screenshots	Adipose involution	50	50	100%	50	100%
	Fibrocystic disease	105	105	100%	90	85.7%
	Mammary gland adenosis	28	28	100%	22	78.5%
Cysts screenshots	Adipose involution	39	39	100%	35	89.7%
	Fibrocystic disease	24	11	45.8%	20	83%
	Mammary gland adenosis	6	3	50%	4	67%
Fibroadenoma screenshots	Adipose involution	7	7	100%	6	85.7%
	Fibrocystic disease	9	6	66.7%	8	88.9%
	Mammary gland adenosis	6	3	50%	5	83.3%
Breast cancer screenshots	Adipose involution	23	23	100%	22	95.6%
	Fibrocystic disease	20	8	40%	17	85%
	Mammary gland adenosis	5	2	40%	4	80%

Conclusion

The authors of this paper have proposed the technology of computer diagnosis of tumors in mammary gland of three main types, i.e. cyst, fibroadenoma, and breast cancer, using mammograms

The developed technology of computer diagnosis of tumors has proper advantages over any other similar technologies and CAD systems, i.e. a low cost, a high speed of initial diagnosis, detection of tumor regions that are not visible for radiologists, the possibility of wider use due to prevalence of analogue mammography facilities in Russia, recording of types of breast tissues, and the possibility to process both analogue and digital mammograms.

The experimental data results have shown that a system, as well as a physician can accurately detect all tumors on the background of adipose involution, however the physician have experienced some problems with diagnosis in terms of fibrocystic disease and mammary gland adenosis. Thus the physician has detected maximum 50% of both benign (cyst, fibroadenoma) and malignant tumors in terms of fibrocystic disease. But the system can properly detect maximum 80% of tumors. Therefore, the

presence of any type of mastopathy reduces accuracy of correct diagnosis, since it shadows practically the whole breast region which increases the probability of missing tumors by the radiologist at early stages when screening. This disadvantage is eliminated by a complex system based on the proposed technology, thanks to which, when processing mammograms with complicated mastopathy, a top layer of the tissue is removed, and the tumor becomes visible; moreover, with respect to possible boundary detection of suspicious areas, we can easily identify the type which they relate to (benign or malignant). Using this system, we have succeeded to detect tumors with a diameter less than 7 mm, while the existing CAD systems can detect tumors at later stages when they are insusceptible to medical treatment.

The work has been performed with support of the RFBR (the Russian Foundation for Basic Research) within the framework of the research project No. 13-07-00825.

References

1. Sadykov, S.S. Technology selection cysts on mammogram / S.S. Sadykov, E.A. Zakharova, Y.A. Bulanava // Herald RGRTU, Ryazan. – 2013. – Vol. 1. – P. 7-11. (In Russian).

2. Sadykov, S.S. Interactive system for the analysis of mammographic images / S.S. Sadykov, Y.A. Bulanova, E.A. Zakharova // Algorithms, Methods, and Data Processing Systems. – 2012. – Vol. 19. – P. 155-187. (In Russian).

3. Database of Mammograms MIAS // URL: http://peipa.essec.ac.uk/pix/mias (date of access 18.09.2013).

4. Akhmetshin, A.M. Segmentation of low contrast images using a hierarchical model of Markov random fields / A.M. Akhmetshin, A.E. Fedorenko // Artificial Intelligence. – 2002. – Vol. 4. – P. 414-421. (In Russian).

5. Sadykov, S.S. Use of information technology for the detection of cysts of the breast on mammography / S.S. Sadykov, E.A. Zakharova, Y.A. Bulanova // Journal of Radiology. – 2013. – Vol. 3. – P. 15-20. (In Russian).

6. Sadykov, S.S. Preprocessing of mammographic images / S.S. Sadykov, Y.A. Bulanava, V.S. Yashkov // Proceedings of the International Symposium "Reliability and Quality". – 2013. – Vol. 1. – P. 340. (In Russian).

7. Bulanova, Y.A. The use of information technology for the localization of breast cancer on mammograms with a predominance of glandular component / Y.A. Bulanova // Caspian Magazine: Management and High Technology of 2013. – Vol. 3(23). – P. 100-111. (In Russian).

8. Sadykov, S.S. Algorithm of localization of breast cancer in the background of mastopathy / S.S. Sadykov, Y.A. Bulanova. – 11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA– 11-2013). – 2013. – Vol. 2. – P. 717-721.

9. Korzhenkova, G.P. Integrated X-ray sonographic diagnosis of breast disease, 1st Edition / G.P. Korzhenkova. – Moscow: "Firm STROM" Publisher. – 2004 – P. 128. (In Russian).

10. Sadykov, S.S. Allocation algorithm of cysts on the lowcontrast mammography / S.S. Sadykov, Y.A. Bulanova, E.A. Zakharova // Information Technology. – 2013. – Vol. 8. – P. 53-57. (In Russian).

11. Lukac, R.B. Plataniotis et al. Generalized adaptive vector sigma filers / R.B. Lukac, K.N. Smolka // International Conference on Multimedia Computing and Systems / International Conference on Multimedia and Expo – ICME (ICMCS). – 2003. – Vol. 1. – P. 537-540.

12. Sadykov, S.S. Investigation of the watershed for the isolation of breast cancer / S.S. Sadykov, Y.A. Bulanova, E.A. Zakharova, V.S. Yashkov // Algorithms, Methods, and Data Processing System. – 2013. – Vol. 23. – P. 56-64. (In Russian).

13. Sadykov, S.S. Texture segmentation algorithm to identify areas of cysts on mammography / S.S. Sadykov, Y.A. Bulanova, A.G. Romanov // Algorithms, Methods, and Data Processing Systems. – 2013. – Vol. 19. – P. 50-55. (In Russian).

14. Sankar, R.P. Parallel Architecture for Implementation of Contrast Limited Adaptive Histogram Equalization / R.P. Sankar, Rao Srinivasa // International Journal of Advanced Engineering Sciences and Technologies (IJAEST). – Vol. 10(1). – P. 047-051.

15. Haralick, R.M. Textural Features for Image Classification / R.M. Haralick, K. Shanmugan, I. Dinstein // IEEE Transactions on Systems, Man and Cybernetics. – 1973. – Vol. SMC-3(6). – P. 610-621.

16. Gaydel, A.V. Investigation of textural features for the diagnosis of bone diseases by X-ray images / A.V. Gaydel, S.S. Pervushkin // Computer Optics. – 2013. – Vol. 37(1). – P. 113-119. (In Russian).

17. Sadykov, S.S. Automating the detection of tumors on mammographic images / S.S. Sadykov, Y.A. Bulanova, E.A. Zakharova, V.V. Kariaev. – Proceedings of the International Symposium "Safety and quality". – 2012. – Vol. 2. – P. 376– 379. (In Russian).

18. Shcherbakov, S. At the junction of two eras: Flatbed Scanner Epson Perfection V750 Pro / S. Shcherbakov // Foto & Video. – 2006. – Vol. 7. – P. 60-65. (In Russian).