A Breast Tumors Segmentation Algorithm for Digital Mammograms

Sheng-Wen Zheng[#1] ,Yue-Jing He [#2] ,Yung-Tsang Chang[*3] ,Chen-Chung Liu[#4]
[#]*Department of Electronic Engineering, National Chin-Yi University of Technology,*
Taichung, Taiwan
[1]unun1535@hotmail.com
[2]yuejing@ncut.edu.tw
[4]ccl@ncut.edu.tw
[*]*Department of Information Networking Technology, Hsiuping University of Science and Technology,Taiwan*
[3]kenchang@hust.edu.tw

***Abstract*— Breast cancer is frequently occurred in women to cause death. It is a malignant tumor caused by the abnormal division and reproduction of breast duct or acini cells. Many researches have proven that an early detection of the breast cancer can lead to successful treatments to reduce the death rate of breast cancer. For detecting the early-stage breast cancer, mammography is the most popular and effective technique. Extracting breast tumors accurately from a mammogram is a kernel stage for mammography, due to it significantly influences the overall analysis accuracy and processing speed of the whole breast tumor analysis. For this reason, tumors have to be identified and segmented from the breast region in a mammogram before further analysis. This paper aims to presented an accurate and efficient algorithm of breast tumor extraction on medio-lateral oblique (MLO) mammograms. The presented algorithm utilizes modified gradient vector flow (MGVF) snake to determine the breast region from a mammogram image, then uses Otsu thresholding and multiple regression analysis to remove the pectoral muscle from the breast region, and finally applies upper outlier detection and texture complexity analysis to extract the breast tumors. The presented algorithm was employed on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. The experimental results show that the tumor extracted by the presented algorithm approximately follows that extracted by the biopsy of MIAS.**

***Keywords*— mammography, cancer, medio-lateral oblique (MLO), Mammogram Image Analysis Society (MIAS), random walk.**

1. INTRODUCTION

Breast cancer is one of the common cancers to cause death [1- 5] in women. An early detection of the breast cancer really can lead to successful treatment to reduce the death rate of breast cancer. For detecting the early-stage breast cancer, mammography is the most effective and popular technique. Extracting the breast tumors accurately from a mammogram is a kernel stage for mammography. It significantly influences the overall analysis accuracy and processing speed of the whole breast tumor analysis.

A tumor segmenting from mammograms is frequently difficult due to the varying in size, shape, and location of the tumors. Moreover, the low contrast and poor definition of tumors boundaries, and the disturbance from the fatty tissues and veins are also increase the difficulty of segmentation [3, 6]. On the other hand, tumor-tissue regions in a mammogram have different texture patterns and gray levels from the normal tissues. So, those possible tumors may be segmented from a mammogram by utilized conventional image segmentation schemes [2, 7].

Many breast tumor segmentation algorithms were proposed [8-13]. These segmentation approaches can be classified as two major categories: (i) edge-based tumor segmentation schemes and (ii) region-based tumor segmentation schemes. Edge-based segmentation schemes first find object boundaries and then segment areas enclosed by the boundaries without requiring the priori information about the image content. Although edge-based segmentation schemes are computationally fast, but the boundaries frequently do not enclose the object completely. For edge-based tumor segmentation schemes, Yuan et al. [8] applied a radial gradient index (RGI)-based segmentation scheme to generate an initial contour of a tumor. Then they based on the initial contour of a tumor to introduce an automatic background estimation to classify the effective circumstance of the tumor to extract tumors. Song et al. [9] used plane fitting and dynamic programming techniques to detect the optimal contour of a tumor from the candidate edge pixels. Region-based segmentation schemes search pixels that satisfy a given homogeneity criterion, and collect these pixels to form an object (region). For region-based tumor segmentation schemes, Jiang et al. [11] first utilized a Gaussian filter to remove image noise and utilized a gamma correction on a region of interest (ROI) to improve the contrast of ROI. Then, they applied the principle of maximum entropy and selected optimal thresholds to obtain initial segmented tumor regions. Finally, they conduct a morphologic dilation operation on the image with initial tumor regions segmented to obtain the tumor. Xu et al. [12] used the iterative thresholding scheme to extract suspicious regions. Then they used 3-level discrete wavelet transform to decompose the suspicious area into four sub-bands to locate centers of the tumors. Finally, they utilized the region growing scheme to merge adjacent regions to extract tumors.

In this paper, we present an accurate and efficient algorithm of breast tumor segmentation for MLO mammograms. The presented algorithm consists of four phases: (i) adapting breast region extracted by using modified gradient vector flow (MGVF) snake, (ii) utilizing multiple regression analysis to delete the pectoral muscle from the breast region, (iii) applying upper outlier detection and texture complexity analysis to extract the breast tumor candidates, and (iv) employing random walk scheme on the initial breast tumor candidates to segment breast tumors accurately. The presented algorithm is conducted on MIAS database. The remainder of this paper is organized as follows: Section 2 introduces the presented segmentation algorithm of tumor from mammograms. Section 3 presents the experimental results. Finally, the conclusions of this paper are presented in Section 4.

2. PRESENTED BREAST TUMORS EXTRACTION

A robust breast tumor segmentation scheme for digital mammograms must be extremely effective and efficiency. In order to construct an accurate breast tumor segmentation algorithm for digital mammograms, several schemes are used in this paper to achieve the goal. The presented breast tumor segmentation algorithm four phases: (i) breast region extracted using modified gradient vector flow (MGVF) snake, (ii) pectoral muscle deleted using the iterative Otsu thresholding and multiple regression analysis, (iii) breast tumor regions initially segmented utilizing dynamic threholdings and texture complexity analysis, and (iv) breast tumors extracted utilizing random walk scheme. The details of these phases used in the presented breast tumor segmentation algorithm are illustrated in the following subsections.

2. 1. Breast region segmentation [14]

The breast region segmentation by using MGVF snake is employed in this paper, it was introduced in our previous paper; a breast segmentation scheme for digital mammograms using gradient vector flow snake [14]. In the proposed breast region extracting algorithm, median filter was used to filter out the noise, binarization processing and the morphological erosion processing were used to find a rough breast border, the novel gradient adjusting step was used to get a modified edge map, and GVF snake was employed to obtain the accurate breast border. Fig. 1 shows an example of the output of each step in phase 1 of the presented algorithm.

2.2. Pectoral muscle removal [15]

In the breast region of a mammogram, the gray intensity of a pectoral muscle region is similar to that of the breast tumor cells and the pectoral muscle’s texture may also be similar to some abnormalities.

(a) (b) (c) (d) (e)

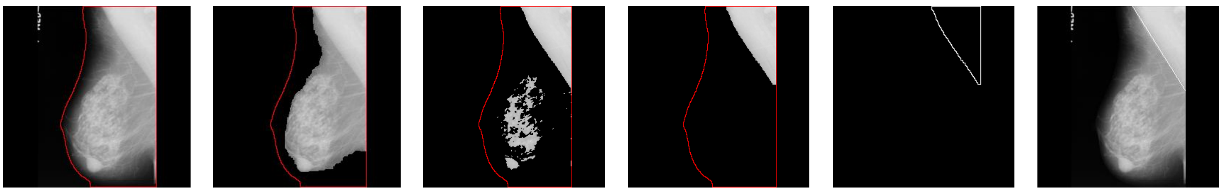
Fig. 1. The output of each step in the presented breast region segmentation algorithm, (a) Original mammogram, (b) after binarization, (c) After non-breast objects removal, (d) Initial breast border segmentation, (e) final breast region extracted by using MGVF snake.

Due to the appearance of pectoral muscle in MLO mammograms will increase the false positive in a breast cancer CAD, pectoral muscle has to be segmented from the breast region in a mammogram before further analysis. The pectoral muscle removal by using multiple regression analysis is employed in this paper, it was introduced in our previous paper; a breast segmentation scheme for digital mammograms using gradient vector flow snake [15].The presented scheme uses an iterative Otsu thresholding on the breast region extracted by MGVF snake, and then utilizes multiple regression analysis to remove the pectoral muscle accurately. Fig. 2 shows an example of the output of each step in the presented pectoral muscle removal algorithm.

2.3. Breast tumors segmentation

In this stage, we want to extract one or more suspicious areas from the pectoral muscle deleted breast region (PMDBR); which is to extract and locate several areas as the suspicious tumor candidates of a PMDBR. Region-based segmentation schemes are more suitable for mammograms since a suspicious region is always brighter than its surrounding tissues and has a fuzzy boundary with almost uniform intensity [16]. In this paper, region segmentation based-thresholding scheme is utilized for extracting suspicious breast tumors from a PMDBR.

For acquiring a better initial segmentation of breast tumors from a PMDBR, the presented algorithm repeatedly utilizes the upper outlier detection scheme with different thresholds to binarize the PMDBR into a black-white image. The binarization procedure shall be terminated while the area ratio of white area to the black area is within in the interval [0.01, 0.1] and the absolute difference ratio of white area between two adjacency iterations is less than 5%. Then the white regions are taken as the breast tumor candidates. And the morphological erosion processing, area value filter, main-axes ratio filter and texture analysis are then used in series on these breast tumor candidates to find the real breast tumors. Some steps are illustrated in the following subsections and the flowchart of the presented breast tumors segmentation algorithm is shown in Fig. 3.



(a) (b) (c) (d) (e) (f)

Fig. 2. The output of each step in the presented pectoral muscle removal algorithm, (a) breast region extracted by using MGVF snake, (b) after Otsu threholding and morphological processing one time, (c) after Otsu threholding and morphological processing two times, (d) after position detecting, (e) extracted pectoral muscle rough border, (f) final extracted pectoral muscle.

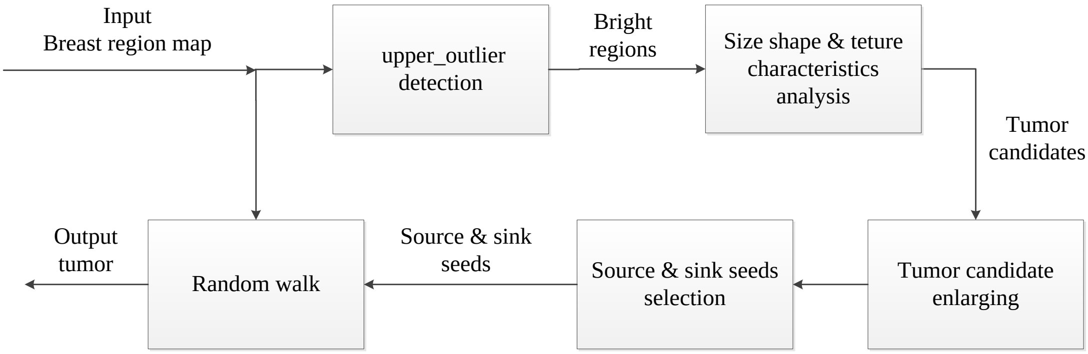


Fig. 3. The flowchart of the presented breast tumors segmentation algorithm.

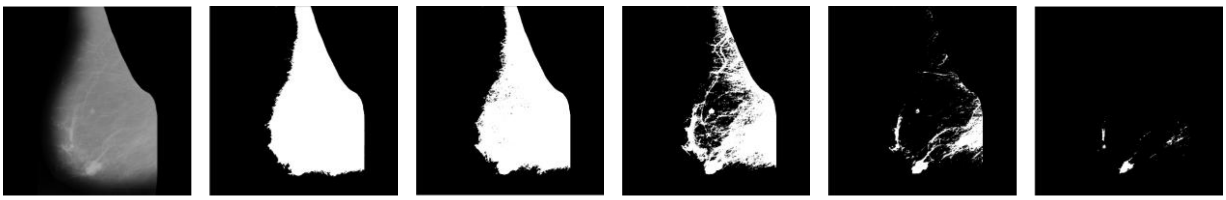
2.3.1. Suspicious areas locating

The grayscale value of the real tumor is always higher than that of the remainder of the tissue within a mammogram. In statistics, this fact means that the breast tumors shall be in the outliers of the gray values distribution of the pectoral muscle deleted breast region. So, the presented algorithm repeats the outlier detection scheme with different threholds to binarize the PMDBR to determine the breast tumor candidates. To define the outliers of a data set, let $\bar{x}$ be the mean and let σ be the standard deviation of the data set. One observation is declared a lower outlier if it is no more than $\bar{x} - k\sigma$, and declared as an upper outlier if it is no less than $\bar{x} - k\sigma$, and the others are declared as inliers, where the value of upper-outlier parameter (UOP) k is usually taken as no more than 3 and no less than 1.

The brightest areas in a grayscale PMDBR are the breast tumor candidates. These bright regions are always located in the upper outliers of the pixel value distribution of a PMDBR. It should be reasonable and effective that the proposed algorithm calculate the mean and standard deviation of a grayscale PMDBR by taking the upper outlier points of the PMDBR as the breast tumor candidates. For obtaining real breast tumors from the PMDBR, the presented algorithm will apply the morphological operation and connected component scheme on the upper outlier regions (suspicious breast tumors) to extract the real tumors in following processes. Fig. 4 shows an example of the outlier detection of a PMDBR with different upper-outlier parameters; the example shows that the detected outlier area is smaller when the upper-outlier parameter is higher.

2.3.2. Tumor characteristics analysis

The presented upper outlier detection scheme although can effectively determine the breast tumor candidates. However, the detected breast tumor candidates may contain other regions of higher grayscale value such as dense tissue, calcifications, and various kinds of noise, etc. On the other hand, a breast tumor is usually a brighter area with special shape and texture characteristics [17, 18]. There are several qualitative and quantitative features for characterizing the shape of tumors in a mammogram. The shape feature of tumors contains the geometric parameters such as area,



(a) (b) (c) (d) (e) (f)

Fig. 4. The detected outliers of a PMDBR for different upper-outlier parameters, (a) Original PMDBR, (b) UOP = 1, (c) UOP = 1.5, (d) UOP = 2, (e) UOP = 2.5, (f) UOP = 3.

area ratio, perimeter, circularity, mean and standard deviation of radial distance, eccentricity, and orientation moment invariants, etc [19]. In this paper, area, area ratio, and ratio of two main axes are taken as criteria to evaluate whether a suspicious breast tumor is a real tumor or not. For obtaining a better and simple binary image of breast tumors to evaluate these shapes' features, the presented algorithm applies the Sobel gradient mask on each suspicious breast tumor to acquire the tumor's shape features. The Sobel gradient mask is one of popular edge detection methods. Fig. 5 shows the Sobel gradient mask; Fig. 5 (a) is a mask type, Fig. 5 (b) and Fig. 5 (c) are the values of masks that are used for detecting the horizontal and vertical edge, respectively.

Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

(a)

-1	-2	-1
0	0	0
1	2	1

(b)

-1	0	1
-2	0	2
-1	0	1

(c)

Fig. 5. The Sobel gradient mask, (a) the mask pixels (b) $f_x(x,y)$ (c) $f_y(x,y)$

2.3.3. Random walk

For image segmentation, random walk scheme is a semi-automated interactive algorithm proposed by Grady [20, 21]. The main steps are: The original image is first presented with its corresponding weighted graph G = (V, W), in which each pixel is the vertex V, and W is the weight between the neighbor vertices. The weight is defined by the Gaussian weight function:

$$w(v_i,v_j) = \exp(-\beta(g(v_i) - g(v_j))^2), \qquad (1)$$

Where $g(v_i)$ is the intensity of the pixel v_i and β is a free parameter.

For the weighted graph, all vertices are divided into a marked vertices set V_M , and an unmarked vertices set V_U , such that $V_M \cup V_U = V$ and $V_M \cap V_U = \phi$. Finally, due to the probability problem for a random walk is the same as a Dirichlet problem, and Dirichlet problem can be evaluate from the graph Laplacian matrix defined as the follows.

So the image segmentation with random walk is transformed to the Dirichlet problem and the image segmentation results are obtained by solving the corresponding Dirichlet problem. In the random walk step, the mass center of a selected candidate breast tumor is taken as the corresponding source seed, and the middle pixel between the candidate breast tumor and the nearest neighbor candidate breast tumor is taken as the sink seed, respectively.

$$L(i,j) = \begin{cases} \sum_{v_i} w(v_i,v_j), & v_i = v_j \\ w(v_i,v_j), & v_i \text{ and } v_j \text{ are adjacent vertices.} \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

3. EXPERIMENTAL RESULTS

For evaluating the effectiveness of the proposed algorithm, the presented tumor segmentation scheme was applied on Mammogram Image Analysis Society (MIAS) database [22]. This MIAS database has 322 mammograms of right and left breast taken from 161 patients. The size of mammogram is 1024*1024 and bit-depth of 8 bits ([0, 255]), and each mammogram is 50um/pixel. In the database, 51 mammograms were diagnosed as malignant, 64 mammograms were diagnosed as benign and 207 mammograms were diagnosed as normal. And different classes of abnormal tissues in mammograms are indicated as ill-defined masses, well-defined circumscribed masses, speculated masses, architectural distortion, calcification, asymmetry and normal. Each of these abnormal tissues has been diagnosed and confirmed as benign or malignant by a biopsy.

The experimental results show that the breast border extrapolated by the presented breast region segmentation algorithm with average of mean error function is 2.2818, average of misclassification error is 0.0144, and average of relative foreground area error is 0.0166 [14].

With reference to the manually demarcated pectoral muscle regions, the segmented regions provided by the presented pectoral muscle removal algorithm are resulted in low average mean error function, misclassification error, and relative foreground area error with 1.7188, 0.0083 and 0.0056, respectively [15].

Successful tumor segmentation is defined as that the overlapping area between the algorithm's extracting area and the area labeled by a biopsy in MIAS is larger than 80%. 15 mammograms with abnormal masses selected from MIAS are tested by the proposed algorithm, and the number of successful tumor segmentation is 14 mammograms. Fig. 5 shows an example of the output of each step of the presented algorithm.

4. CONCLUSION

Breast cancer is a malignant tumor caused by the abnormal division and reproduction of breast duct or acinar cells and is commonly occurred in women to cause death. Successful treatment can reduce the death rate of breast cancer, and an early detection of the breast cancer really leads to successful treatment. For detecting the early-stage breast cancer, mammography is the most effective technique. Extracting breast tumors accurately from a mammogram is a kernel stage for mammography. It significantly influences the overall analysis accuracy and processing speed of the whole breast tumor analysis.

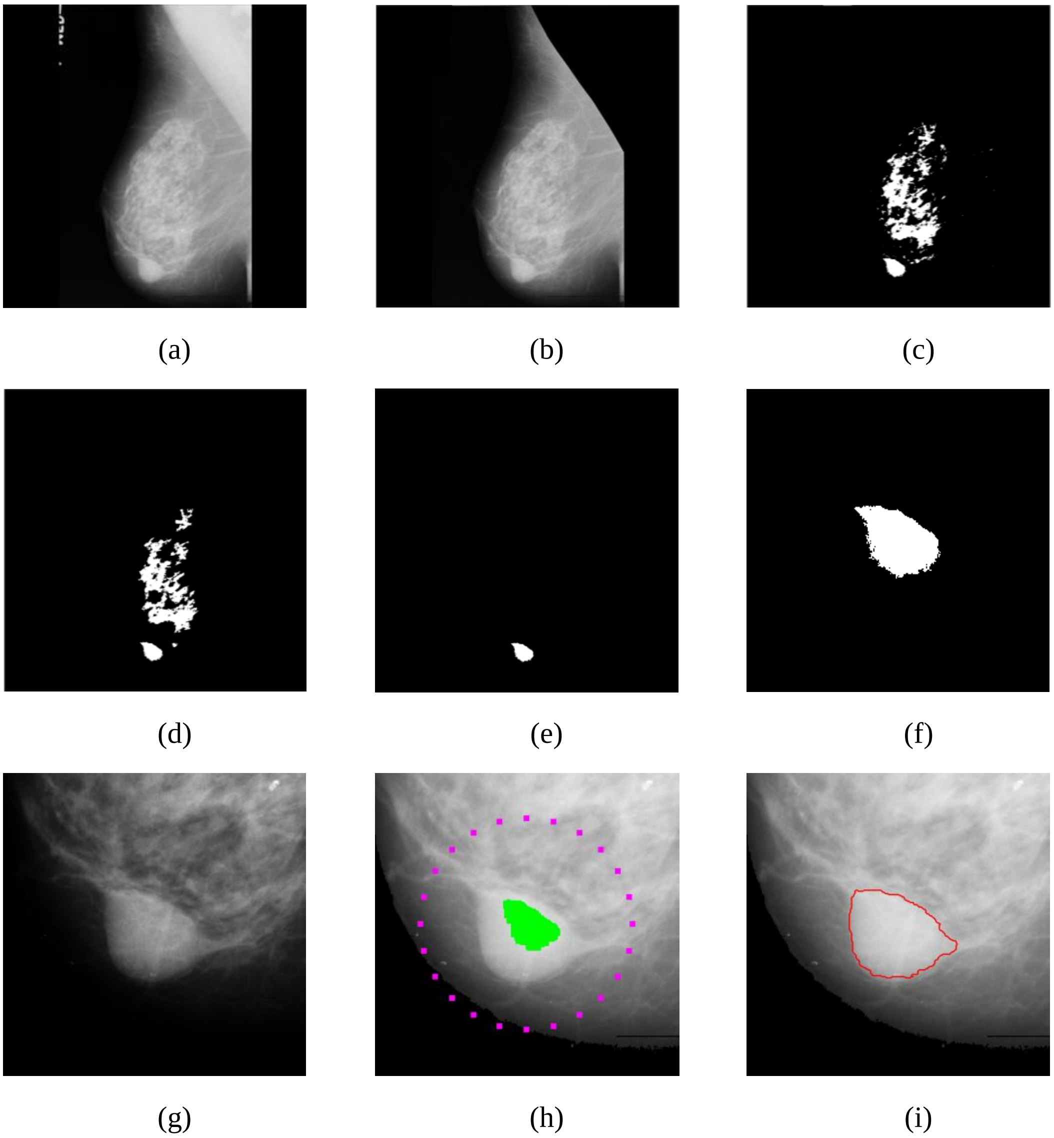


Fig. 6. The output of each step in the presented breast tumors segmentation algorithm, (a) original mammogram Mdb021, (b) corresponding PMDBR, (c) after outlier detection, (d) after breast tumor size and shape characteristics analysis, (e) after breast tumor texture characteristics analysis, (f) after scale up of (e), (g) the part of (f) in the original mammogram, (h) selected source seeks (in green) and sink seeds (pink squares), (i) final extracted tumor by using random walk technology.

For this reason, tumors have to be identified and segmented from the breast region in a mammogram before further analysis.The main goal of this paper is to propose an accurate and efficient algorithm of breast tumor extraction on medio-lateral oblique (MLO) mammograms. The proposed algorithm adapts breast region extracted by using modified gradient vector flow (MGVF) snake to determine the breast region from a mammogram image, uses Otsu thresholding and multiple regression analysis to delete the pectoral muscle from the breast region, utilizes upper outlier detection and texture complexity analysis to segment the initial breast tumor regions, and takes random walk scheme on the initial tumor regions to accurately segment breast tumors. The presented algorithm is tested on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. The experimental results show that the tumor extracted by the presented algorithm approximately follows that extracted by the biopsy of MIAS. In the future we will develop a high performance breast mass analysis based on accurate breast tumor segmentation to power the computer aided detection of breast cancer.

ACKNOWLEDGMENT

This paper has been partially supported by National Science Council of Taiwan with grant number NSC 101-2221-E-167-038.

REFERENCES

[1] R. D. Yapa, K. Harada, “Breast skin-line estimation and breast segmentation in mammograms using fast-marching method,” International Journal of Biomedical Sciences, Vol. 3, No. 1, pp 54-62, 2008.

[2] S. M. Kwok, R. Chandrasekhar, Y, Attikiouzel, M. T. Rickard, “Automatic pectoral muscle segmentation on mediolateral oblique view mammograms,” IEEE Trans. on Medical Imaging, Vol. 23, No. 9, pp.1129-1140, 2004.

[3] T.S. Subashini, V. Ramalingam, S. Palanivel, “Automated assessment of breast tissue density in digital mammograms,” Computer Vision and Image Understanding, Vol. 114, No. 1, pp. 33-43, 2010.

[4] J. Nagi, S. A. Kareem, F. Nagi, S. K. Ahmed, “Automated breast profile segmentation for ROI detection using digital mammograms,” 2010 IEEE EMBS Conference on Biomedical Engineering & Sciences, pp.87-92, 2010.

[5] G. Kom, A. Tiedeu, M. Kom, “Automated detection of masses in mammograms by local adaptive thresholding,” Computers in Biology and Medicine, Vol. 37, pp. 37-48, 2007.

[6] A. R. Domıınguez, A. K. Nandi, “Detection of masses in mammograms via statistically based enhancement, multilevel thresholding segmentation, and region selection,” Computerized Medical Imaging and Graphics, Vol. 32, pp. 304-315, 2008.

[7] Zhang, Y., Tomuro, N., Furst, J. and Raicu, D. “Image Enhancement and Edge-based Mass Segmentation in Mammogram,” In Proceedings of the SPIE Symposium on Medical Imaging (SPIE-2010), pp.13-18, 2010.

[8] Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, “A Dual-stage Method for Lesion Segmentation on Digital Mammograms “, Med. Phys. Vol. 34, No. 11, pp. 4180-4193, 2007.

[9] E. Song, L. Jiang, R. Jin, L. Zhang, Y. Yuan, and Q. Li, “Breast Mass Segmentation in Mammography Using Plane Fitting and Dynamic Programming”, Academic radiology, Vol. 16, No. 7, pp. 826-35, 2009.

[10] A. R. Domínguez1, and A. K. Nandi, “Toward Breast Cancer Diagnosis Based on Automated Segmentation of Masses in Mammograms,” Pattern Recognition, Vol. 42, pp. 1138-1148, 2009.

[11] L. Jiang, E. Song, X. Xu, G. Ma, and B. Zhang, “Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance,” Acad Radiol, pp. 15:1534-1544, 2008.

[12] W. Xu, S. Xia, M. Xiao, and H. Duan, “A Model-based Algorithm for Mass Segmentation in Mammograms,” Engineering in Medicine and Biology 27th Annual Conference, pp. 2543- 2546, 2005.

[13] P. Delogu, M. E. Fantacci, P. Kasae, A. Reticoa, “Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier,” Computers in Biology and Medicine, Vol. 37, pp. 1479-1491, 2007.

[14] C. C. Liu, C. Y. Tsai, T. S. Tsui, S. S. Yu, An improved GVF snake based breast region extrapolation scheme for digital

mammograms, Expert Systems with Applications, Vol.39, No. 4, pp. 4505- 4510, 2012.

[15] C. C. Liu, C. Y. Tsai, J. Liu, C. Y. Yu, S. S. Yu, A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis, Computers & Mathematics with Applications, Vol. 64, pp.1100- 1107, 2012..

[16] V. Raman, P. Sumari, H. H. Then, and S. A. K. Al-Omari, “Review on Mammogram Mass Detection by Machine Learning Techniques,” International Journal of Computer and Electrical Engineering, Vol. 3, No. 6, pp. 873-879, 2011.

[17] B. Nielsen, F. Albregtsen, and H. E. Anielsen, “Low Dimensional Adaptive Texture Feature Vectors From Class Distance and Class Difference Matrices,” IEEE transactions on medical imaging, Vol. 23, No. 1, 73-84, 2004.

[18] L. Alvarez and L. Mazorra, “Signal and Image Restoration Using Shock Filters and Anisotropic iffusion,” SIAM J. Numerical Analysis, Vol. 31, No. 2, pp. 590-605, 1994.

[19] M. Suganthi, M. Madheswaran, “Mammogram tumor classification using multimodal features and Genetic Algorithm,” International Conference on Control, Automation, Communication and Energy Conservation (INCACEC 2009), pp. 1- 6, 2009.

[20] R. Modzelewski, E. Janvresse, T. de la Rue, P. Vera, “Brain perfusion heterogeneity measurement based on Random Walk algorithm: Choice and influence of inner parameters,” Computerized Medical Imaging and Graphics, Vol. 34, pp. 289–297, 2010.

[21] H. K. Hsiao, C. C. Liu, C. Y. Yu, S. W. Kuo, S. S. Yu, “A Novel Optic Disc Detection Scheme on Retinal Images,” Expert System with Applications, Vol. 39, pp. 10600–10606, 2012.

[22] H. Abed, A. Roya, S. Jamshid, “ReliefF-Based Feature Selection for Automatic Tumor Classification of Mammogram Images,” Machine Vision and Image Processing (MVIP), pp. 1 – 5, 2011.