Computer-Aided Mass Detection on Digitized Mammograms using a Novel Hybrid Segmentation System

N. Riyahi Alam, F. Younesi, M. S. Riyahi Alam

***Abstract*—** A Novel hybrid segmentation method has been developed for detection of masses in digitized mammograms using three parallel approaches: adaptive thresholding method, Gabor filtering and fuzzy entropy feature as a computer-aided detection(CAD) scheme. The algorithm consists of the following steps: a) Preprocessing of the digitized mammograms including identification of region of interest (ROI) as candidate for massive lesion through breast region extraction, b) Image enhancement using linear transformation and subtracting enhanced from the original image, c) Characterization of the ROI by extracting the fuzzy entropy feature, d) Local adaptive thresholding for segmentation of mass areas, e) Filtering the input images using Gabor functions, f) Combine expert of the last three parallel approaches for mass detection. The proposed method was tested on 78 mammograms (30 normal & 48 cancerous) from the BIRADS and local databases. The detected regions validated by comparing them with the radiologists' hand-sketched boundaries of real masses. The current algorithm can achieve a sensitivity of 90.73% and specificity of 89.17%.

This approach showed that the behavior of local adaptive thresholding, Gabor filters and fuzzy entropy technique could be useful for mass detection on digitized mammograms. Our results suggest that the proposed method could help radiologists as a second reader in mammographic screening of masses.

Keywords: BreastMasses,Segmentation,ImageEnhancement,localadaptiveThresholding,FuzzyEntropy

I. INTRODUCTION

Breast cancer is one of the leading causes of early mortality in women [1]. Retrospective studies have shown that in current breast cancer screening between 10% and 25% of the tumors are missed by radiologists [2]. A computer-aided detection (CAD) system that prompts suspicious regions can draw the attention of the radiologist to a tumor he might otherwise have overlooked [3]. The use of CAD methods as a "second opinion" or as a "pre-reader" strategy has been proposed, and when combined with film reading, improves overall performance for detection and classification of breast cancer and may reduce the observer variation, Fig. 1. Masses are in three different sizes: small size (3–15 mm), middle size (15–30 mm) and large size (30–50 mm) [4]. The third type is rare in mammograms. The detection algorithms must therefore be as robust as possible against these variationsFig. 2.

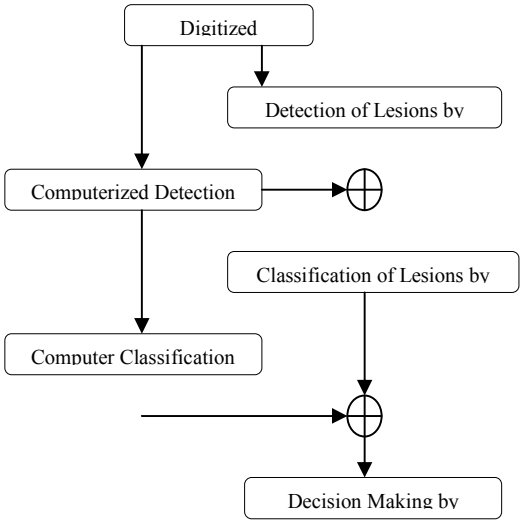


Fig.1. Computer-Aided Decision Making by Radiology

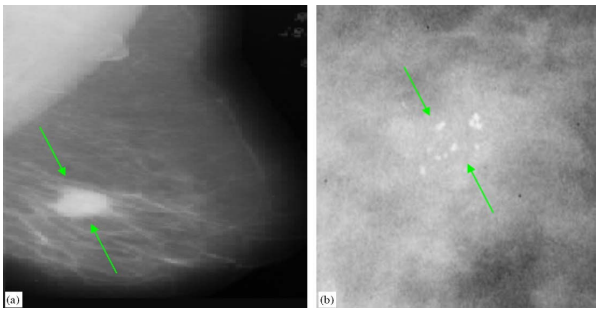


Fig.2. Examples of mammograms: (a) arrows indicate mass area; (b) microcalcifications area

II. Materials and Methods

This section presents the detection procedure used in this work. It is schematically summarized in Fig(3) and the detail of each step in this scheme is explained in the following sections.

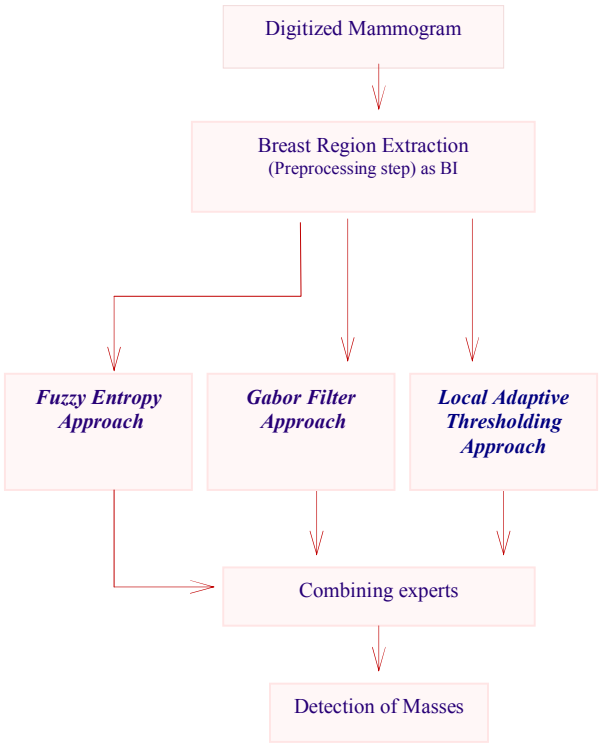


Fig.3. Algorithm for detection of Masses in digitized mammograms

A. Input Data

The data used was obtained from two different databases. A local database (consists of 20 cases: 8 massive and 12 normal) and BIRADS dataset (consists of 58 cases: 40 massive and 18 normal). The original mammogram images from local database were scanned with MicroTek scanner with 42 micrometer spatial resolution and 12 bits per pixel radiometric resolution. Prior to our analysis, each mammogram was resized and down-sampled to 1800 ×1440 pixels and down-quantized to 8 bits per pixel. Another database (from BIRADS) had images with 8-bits pixel depth at the size of 1024 ×1024 pixels [5]. This dataset not required any pre-processing for analysis.

B. Pre-Processing of Local Database Images

The purpose of pre-processing step is to resize the image, remove noise and radiopaque artifact contained within the mammogram and increase region homogeneity, with the objective being to improve in algorithm reliability and robustness.

In mammography screening images, the breast tissue is the interested region for radiologists. As seen in this type of images, more than one third of a mammogram consists of a dark background without any diagnostic information. Considering the computation and system efficiency, extracting the breast tissue as region of interest is the first step of computer automation. So, as the first module, a method to extract this region was developed as follows:

1. The intensity of mammogram was increased by means of multiply the intensity of each pixel by a constant value of k. The proper of k, that analyzed visually, was equal to 2.5
2. Afterwards, by finding a suitable threshold value using Otsu method, each image separated into two sections: foreground and background [6]. Then, 256 gray-level input images are converted into binary images.
3. Since binarized image consists of many holes, in this step all the holes of the image were filled by means of flood-fill operation with 8-connectivity. After filling, all the objects in obtained image were labeled.
4. The area of each object was calculated by counting of all the pixels in any labeled object.
5. Since the biggest object was related to breast tissue, so, by removing all the objects except the biggest one, the breast region could be achieved in binary format. Now, by convolving this image with the original one, we will obtain a breast region image.

Further algorithms work only on this extracted breast image.

C. The approaches to detect masses

Now, all the input data are prepare for analysis. The proposed algorithm consists of two parallel approaches: Local adaptive thresholding method that performs only in spatial domain and Fuzzy entropy approach in the basis of bright object with diffused boundaries, Fig.4.

C.1. First approach: Local adaptive thresholding method

This section presents the detection procedure used in this step. It is schematically summarized in fig.1 and the detail of each step in this scheme is explained in the following section.

C.1.1. Linear transformation filter enhancement

In order to enhance the image, a mass-pattern dependent enhancement approach was designed based on linear transformation of pixel values. This transform enhances the lower gray-level (dark areas) and the inverse of it, enhance the higher gray levels (bright areas). This enhancement step is with all images independently of their initial contrast and background variations. When performed on the original image, linear transformation enhancement filtering was found to useful in visualizing the lesion and increasing the sensitivity of subsequent processing.

C.1.2. Binarization by adaptive local thresholding

Since masses are generally radiographically denser than surrounding tissue, the locally bright spot is binarized using an adaptive thresholding method. the subtracted enhanced image from the original one, a decision is made to classify it

into a potential mass pixel or a normal pixel by the proposed rules. Two windows (large and small in Fig.5) around each pixel is used for choosing empirically the parameters of adaptive thresholding, an example output shown in Fig. 6.

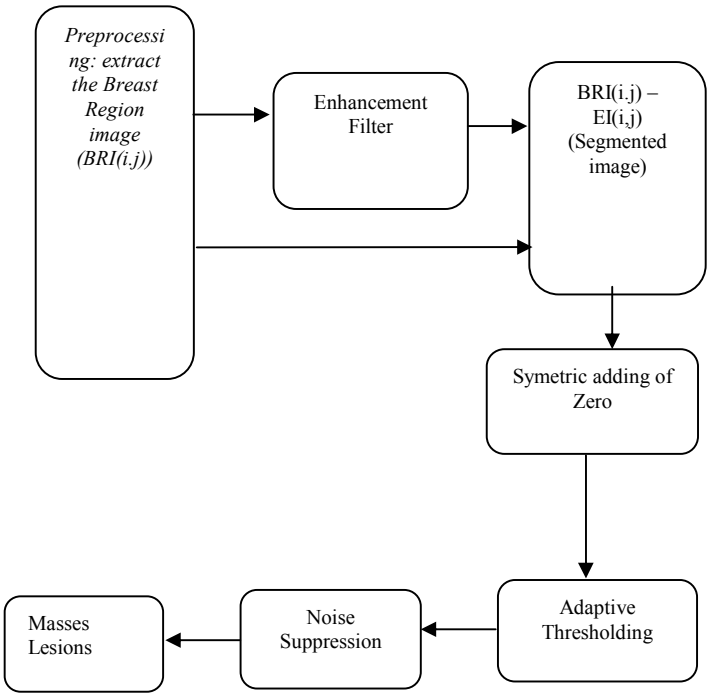


Fig 4. Eenhancement algorithm scheme

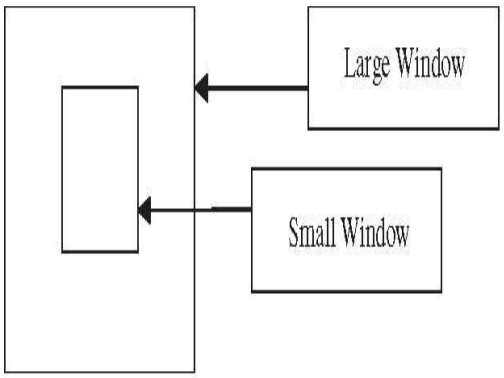


Fig5. Windows for adaptive thresholding

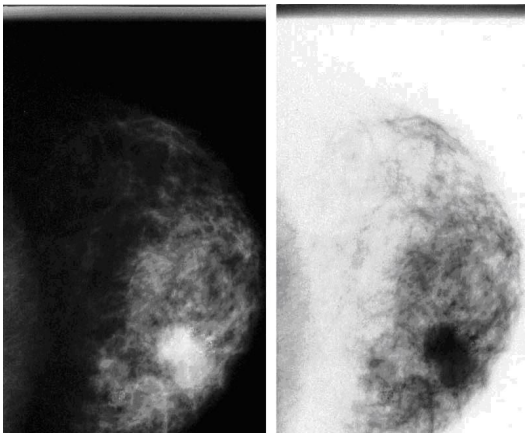


Fig.6. image enhancement output

C.2. Second approach: Fuzzy entropy method

This proposed scheme consists of fuzzy entropy minimizing and mass extraction. Fig.1 shows the block diagram of the method. In order to segment images with diffused and not well-defined areas, the theory of fuzzy sets is used to select a threshold that partitions the image into meaningful regions. The optimal threshold is selected to minimize a given fuzziness measure of the image. In this section, we used the fuzzy entropy defined in [10-14]. The image is viewed as degraded version of an ideal bimodal image which represents light objects on a dark background. A fuzzy set X can be defined for the image in which each image pixel is assigned a membership value is defined by using the relationship between the gray level of the pixel and the average value of its belonging region. By minimizing the fuzzy entropy of the fuzzy set X, the optimal threshold value t, is selected. This optimal threshold value is used to threshold input mammogram image. This step produces a binary image representing the candidate suspicious masses.

Analysis of the mammogram images revealed that small isolated regions are usually irrelevant. Hence, in this scheme such region eliminated from the final segmentation results. This can be achieved by grouping pixels in the segmented image into connected objects. This step produces an object table containing information about the image objects. This information includes a pixel to represent each object and the number of pixels in each object. Finally, the object table is used to eliminate objects with size smaller than a certain threshold value.

C.3.Third approach: Gabor filtering

This approach employs the class of analytical functions kernel as Gabor elementary functions. It is schematically summarized in Fig.7 and the detail of each step in this scheme is explained in the following section.

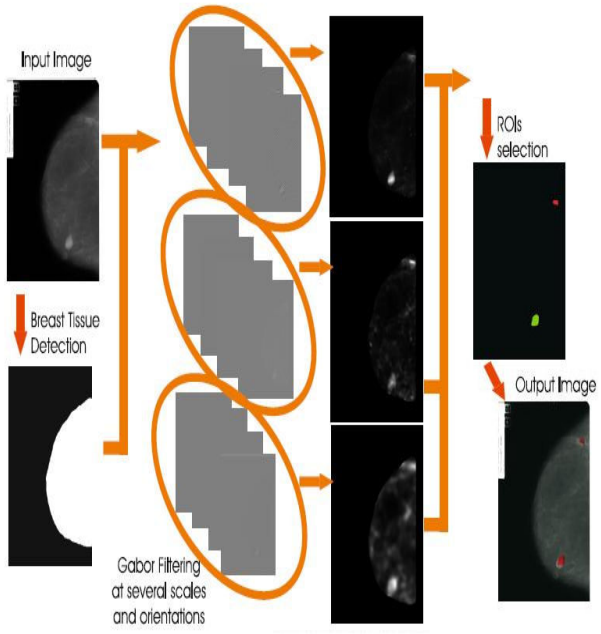


Fig.7. Scheme of Gabor filtering for mass detection

C.3.1. Gabor Functions and Filter Design

A 2D Gabor function g(x,y) and its fourier transform G(u,v) can be written as [11]

$$g(x,y) = \frac{1}{2\pi\delta_x\delta_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\delta_x^2}+\frac{y^2}{\delta_y^2}\right)+2\pi jWx\right]$$

$$G(u,v) = \exp\left[-\frac{1}{2}\left(\frac{(u-v)^2}{\delta_u^2}+\frac{v^2}{\delta_v^2}\right)\right]$$

Where

$$\delta_u = \frac{1}{2\pi\delta_x} \quad , \quad \delta_v = \frac{1}{2\pi\delta_y}$$

Gabor functions form a complete but non-orthogonal basis set. Expanding a signal using this basis provide s a localized frequency description. A class of self-similar functions referred to as Gabor wavelet, is now considered. Let g(x,y) be the mother Gabor wavelet, then this self-similar filter dictionary can be obtained by appropriate dilations and rotations of g(x,y) through the generating function:

$$g_{mn}(x,y) = a^{-m}.G(x',y') \qquad a \succ 1 \qquad , \qquad m,n = \text{int}$$

$$x' = a^{-m}\left(xCos\theta + ySin\theta\right)$$

$$y' = a^{-m}\left(-xSin\theta + yCos\theta\right)$$

$$\theta = \frac{n\pi}{K}$$

And k is the total number of orientations. The scale factor a^{-m} is meant to ensure that the energy is independent of m.

The non-orthogonality of the Gabor wavelets implies that there is redundant information in the filtered images, and the following strategy is used to reduce this redundancy. Let U_l and U_h denote the lower and upper center frequencies of interest. Let K be the number of orientations and S be the number of scales in decomposition. Then the design strategy is to ensure that the half-peak magnitude support of the filter responses in the frequency spectrum touch each other as shown in Fig.8.

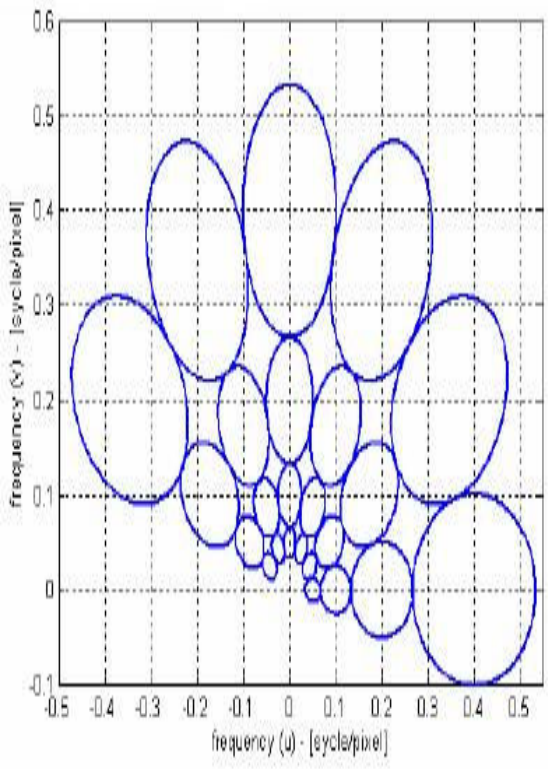


Fig.8. The contours indicate the half-peak magnitude of filter responses in Gabor filter dictionary. The filter parameters are $U_h=0.4$, $U_l=0.05$, $M=4$, $N=6$.

This results in the following formulas for coputing the filter parameters δ_v , δ_u .

$$a = \left(\frac{U_h}{U_l}\right)^{\frac{-1}{S-1}}$$

$$\delta_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ln 2}}$$

$$\delta_v = \tan\left(\frac{\pi}{2k}\right)\left[U_h - 2\ln\left(\frac{\delta_u^2}{U_h}\right)\right]\left[2\ln 2 - \frac{(2\ln 2)^2\delta_u^2}{U_h}\right]^{\frac{1}{2}}$$

In order to eliminate sensitivity of filter response to absolute

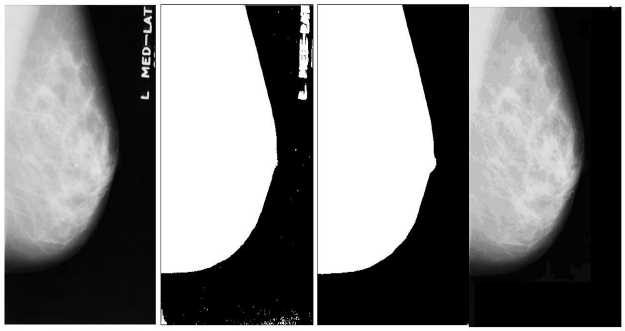
intensity values the real components of 2D Gabor filters are biased by adding a constant to make them zero mean.

In order to extract textural features of an image, the mammograms convolved with a 2D version of Gabor filters. Such filters are linear and local and their convolution kernel is the product of a Gaussian with a plane-wave function. A 2D Gabor filter acts as a local band-pass filter with optimal joint localization properties in the spatial and in the spatial-frequency domain. The Gabor filter decomposes its input image into several sub-images. As mentioned before, the number of orientations and scales defines the number of filters that should affect on input images by multiplying them with each other.

III. Implementation and Results

In this work, the proposed algorithm was tested by using about 78 mammogram image as mentioned in section A. The algorithm implemented in MATLAB 7.0 software packages.

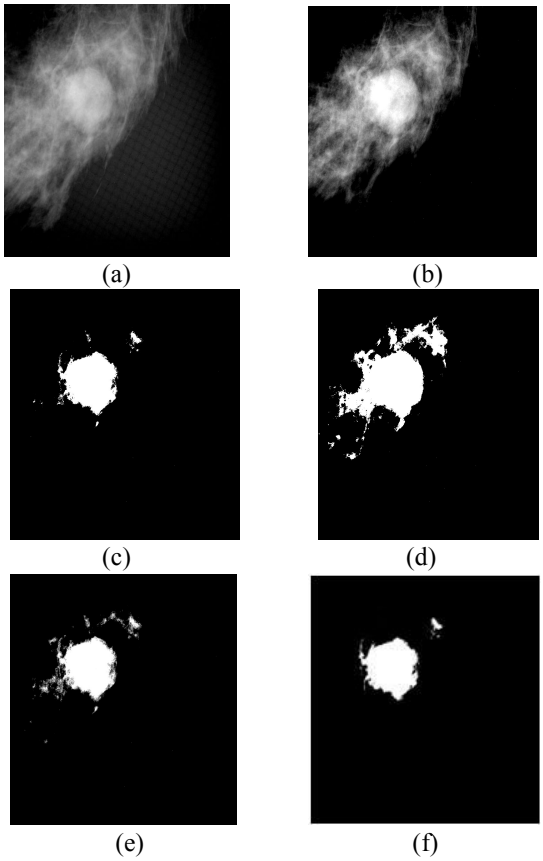
As mentioned before, the smallest and biggest masses are 3-mm and 50-mm respectively. A 3-mm object in preprocessed mammogram occupies about 8–9 pixels and a 50-mm object occupies about 150 pixels. An object with a size of 8-9 pixels is detectable by many computer algorithms. Therefore, the shrinking step is applicable for mass cases and saves computation time. In fig.2 the small and large windows are related to the smallest and largest masses. So, the dimensions of small and large windows are 9×9 and 150×150 pixels respectively.



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

Fig.9. Breast region extraction results: a- Original mammogram, b- Binary image, c- Extracted breast tissue, d- Breast region

Fig.9 (b) shows the segmented region obtained after thresholding in the input image by the optimal threshold value which corresponds to the minimum fuzzy entropy. In the proposed scheme, the object with sizes smaller than 9 pixels removed. The remaining segmented regions represent the suspicious masses in the image.



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

Fig.10. An output of mass detection using the proposed approach. a. Original Image b. Equalized Image c. Fuzzy Output d. Gabor Filter output. e. Adaptive Thresholding Output e. Final Output

For designing the Gabor filter bank, 4 orientations and 3 scales were selected to apply the mammographic images. For orientation, Gabor filters considered at an angle 45 degree apart from each other. A separation among orientations smaller than 30 degrees produces more noise and worse results. As for spatial resolution, for each assigned orientation, the standard deviation assumes the values 1,2,4. these values were selected on the basis of the size of the structures of interests in mammograms. So, 3 scales and 4 orientations were used for implementations. Then 12 sub-images obtained from Gabor filter bank. Each sub-image consists of special textural information. Visually, we assigned some weights for each sub-image before convolved them with each other to improve the convolved results. These weights were selected by experiments. After that, using automatic thresholding approach, the combined images inverted to binary image for other uses. The statistical analysis over 78 images revealed that the proposed scheme achieves an average sensitivity of 90.73% and average specificity of 89.17%.

IV. Conclusion

We have investigated the behavior of local adaptive thresholding and fuzzy entropy technique to detect masses on

digitized mammograms. Once the optimal parameters selected empirically, the processing technique applied to a database of digitized mammograms. Our results suggest that the proposed method could help radiologists as a second reader in mammographic screening of masses.

REFERENCES

[1] A.G. Hauss, M.J. Yaffe, *A categorical course in physics: technical aspects of breast imaging*, in: 78th Scientific Assembly and Annual Meeting of RSNA (Radiological Society of North America), 1994.

[2] J.A.M. van Dijck, L.M. Verbeek, J.H.C.L. Hendriks, *The current delectability of breast cancer in a mammographic screening program*, Cancer 72 (1933) 1938–1993.

[3] I.W. Hutt, S.M. Astley, C.R.M. Boggis, *Prompting as an aid to diagnosis in mammography*, in: A.G. Gale, S.M. Astley, D.R. Dance, A.Y. Cairns (Eds.), Digital Mammography, Elsevier, Amsterdam, 1994, pp.389–398.

[4] H. Li, Y. Wang, K.J. Ray Liu, *Computerized radiographic mass detection. Part II: decision support by featured database visualization and modular neural network*s, IEEE. Trans. Med. Imag. 20 (4) (2001) 302– 313.

[5] Pfarl G. *Breast Imagimg Reporting and Data System*. 2006; Available from: http://www.birads.at, Access at: 2006Oct24.

[6] Otsu N. *A threshold selection method from gray-level histogram*. IEEE Transactions on System Man Cybernetics 1979; 9(1): 62-66.

[7] A. R. Abdel-Dayem, M.R. El-Sakka, *Fuzzy Entropy Based Detection of Suspicious masses in Digital Mammogram Images*. 27th Annual Conference of IEEE, China 2005, September 1-4.

[8] L. Li, W. Qian, L.P. Clarke, *Image feature extraction for mass detection using digital mammography: effects of wavelet analysis*, Proc. SPIE Med. Imag. 3338 (1998) 168–1176 Med. Phys. 26(3) (1999).

[9] L. Li, W. Qian, L.P. Clarke, R.A. *Clark, Improving mass detection by adaptive and multi-scale processing in digitized mammograms*, SPIE (Society of Photographic Scientists and Engineer) Conference on Image Processing, San Diego, CA, vol. 3661, February 1999, pp. 490–498.

[10] Silvano DZ, Luigi C, Stefano L. *Image thresholding using fuzzy entropies*. IEEE Trans. on systems man and cybernetics, Part B 1998; 28(1):15-23.

[11] B. S. Manjunath, W. Y. Ma, *Texture features for browsing and retrieval of image data*, IEEE Trans. Pattern Anal. & Machine Intell., vol.18, no. 8, pp.837–842, 1996.

[12] Three-Dimensional Technique for Automatic Brain Segmentation of the Ventricles Based on Optimal Histogram Thresholds of MRI, *Danmary Sanchez, Malek Adjouadi, Byron Bernal, Nolan Altman,* WSEAS TRANSACTIONS on COMPUTERS, Issue 7, Volume 4, July 2005, ISSN 1109-2750.

[13] Comparison of Three Gaussian Mixture Modeling and Spatial Encoding Methods for Segmenting Human brain MRI, *Mahmood Zeydabadi-Nejad, Reza A. Zoroofi, Hamid Soltanian-Zadeh,* WSEAS TRANSACTIONS on ELECTRONICS, Issue 3, Volume 1, July 2004, ISSN 1109-9445.

[14] Fuzzy Rule Based Classifiers from Support Vector Learning, *Stergios Papadimitriou, Konstantinos Terzidis, Seferina Mavroudi, Lambros Skarlas , Spiridon Likothanasis*, Issue 7, Volume 4, July 2005, ISSN 1109-2750.

This work was supported by the research chancellor of Tehran University of Medical Sciences/University of Tehran, Tehran-Iran.

N. Riyahi Alam is is Associate prof in medical physics and biomedical engineering in Tehran university of medical sciences, school of medicine, dept of medical physics and biomedical engineering, research interest image processing and computer aided diagnostic systems. (corresponding author to provide phone & fax: 00-98-21-66466383; e-mail: riahinad@sina.tums.ac.ir).

F. Younesi is with the Department of Medical Physics & Biomedical Engineering, School of Medicine, Medical Sciences / University of Tehran, Tehran-Iran, He has Msc in Biomedical engineering, with research interest in CAD. (e-mail: fyounesi@razi.tums.ac.ir)

M. S. Riyahi Alam is MSc student in Department of Computer Sciences in Politechnique of Milan, italy, His research interest is in Computer Programming & Applications(email: sadegh.riyahi@gmail.com)