ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Breast Cancer Tumour Diagnosis from Mammography Images Using Wavelet Transform and Hidden Markov Model

Sayedeh Somayeh Hosaini[1], Mehran Emadi[2]

Computer Master Student, Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran[1]

Assistant Professor, Department of Electrical Engineering, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Iran[2]

ABSTRACT: Breast cancer is one of the most important reasons of death within women 15 to 54 years old. A woman dies of breast cancer every 13 minutes and 12.6 percent of women are infected during their life. Although mammography is one of the most effective ways of detecting disease; but still it has deficiencies and limitations. Cancerous lesions may not be diagnosed or non-cancerous lesions are detected as cancer in the interpretation of mammography images. Recently, quality of mammography images is increased and image features are extracted using image processing science in order to help radiologists in detection and diagnosis of cancer masses. This will increase detection speed and accuracy rate. In this study, a new method is proposed for detection of suspicious areas of breast cancer tumoursbased on wavelet and hidden Markov model. According to combination of these two methods, the efficiency has been increased compared to method of previous works. In this research, cancer masses are detected as well as percentage of cancer masses becomes clear; this makes to estimate mass growth rate. In this paper, Markov model with tree structure is used in order to extract statistical properties of wavelet transform components. Markov model has special ability in extracting information related to edges and protruding parts of image context due to its features which can accurately detect cancer areas. In this research we try to estimate the appropriate label (clustering) of pixels from a checked image in order to segment cancer areas. Certain joint distribution is assumed for pixels of a region or class; then, the maximum similarity of different areas of an image under review is checked using ML method. Combination of MIAS database and Paden including 150 images is used in order to test the proposed method. The results indicate that proposed method is more accurate compared to methods that use only wavelet transforms method. Detection rate of proposed method is 96% that is improved 24.5% compared to wavelet transform with detection rate of 71.5 percent.

KEYWORDS: Breast Cancer, Wavelet Transform, Hidden Markov Model, Mammography, Distribution Function, Image Processing**.**

I.INTRODUCTION

Cancer is considered as one of the major reasons of death in world. Breast cancer is one of the major reasons of death among women of 15 to 54 years. It can be said that a woman died due to breast cancer every 13 minutes. According to conducted research, the disease probability among patients is 12.6 percent [1].

Early diagnosis of breast cancer plays significant role in improvement and treatment of disease and increases patient's chance of being alive. Sampling through surgery is the most common method of disease diagnose. This method is invasive and time-consuming with very high accuracy method among available methods [2]. Other method of disease diagnosis is mammography. Using mammography we can diagnose breast cancer masses and symptoms ten years earlier than they are tangible [3]. A mammogram is X-ray image of breast tissue, and radiologists can find structural asymmetry between breasts through analysing the abnormal symptoms such as lumps and calcification [4].

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Mammography is now a perfect tool for early detection of breast cancer, but it has limitations. Cancerous lesions may not be diagnosed or non-cancerous lesions are detected as cancer in the interpretation of mammography images. As a result, radiologists will mistake in 10 to 30% of cases [5]. Cases where a lesion is misdiagnosed lead to a biopsy and surgery on patient. Cases where cancer is misdiagnosed lead to patient death.

Image processing techniques can be used in order to improve recognition and increase the sensitivity and accuracy of detecting lesions. Image processing is a method of converting an image digitally and performing some operation on it in order to get an improved image or to extract some useful information.

Mammography image processing objective in this study is image intensification and improvement in order to identify and early diagnosis of breast cancer tumours; it detects cancerous masses in mammography images through combining wavelet transform and hidden Markov model.

MIAS database and Paden images and images gathered from radiology are used in order to provide mammography images for analysis of method presented in this research.

MIAS Database; mammographic images are collected in United Kingdom of England analysis centre. The size of all images in this database is 1024 x 1024 pixels. It includes 322 mammographic images of 161 women left and right breasts with 50 micron pixel edge resolution. In fact, digital mammography images are grey with depth of 8 bits. The images are classified based on lesion type, including normal breast, having masses, having micro-calcification clusters and with asymmetric structure. The database includes 209 normal breasts, 67 ROI with benign lesions and 54 ROI with malignant lesions [6]. Paden database is a database that is collected in Singapore and includes 150 mammographic images from a variety of benign and malignant tumours. There are different types of images at this database. There are images with 800 x 600 pixels size in this database.

The need for accurate diagnosis is arisen, given the importance of complications early diagnosis in diseases treatment and life expectancy of patients. Science help in the field of medical image processing leads to quick and precise detection of breast cancer tumours on mammography images. In this section, an overview of research will be carried.

In 2008, Papadopulou et al examined following methods on MIAS and Nijmegon database images in order to improve the detection of lesions in mammography images: Contrast Limited Adaptive Histogram Equalization (CLAHE), local range modification (LRM), Discrete Wavelet (RDW), traction linear algorithm. The results show that local range modification method is more appropriate in determining lesion compared to other methods [7].SonaliBahadori et al removed normal breast tissue and additional parts using threshold in order to examine and analyse breast tumour growth from mammography images and finally analysed their characteristics by tumour shape [8].Indra Kant Mi Terra et al. (2011), proposed a method in order to detect abnormal masses of mammography images that is divided into two distinct parts of homogeneous blocks formation and determining the colour after processing. This method was applied on MIAS database images. The idea has the potential for further development due to simplicity of method and offering results in shortest time [9].Natarajan et al (2013) removed noises of image using filters, adjusting contrast and traction techniques in order to characterize the tumour in mammographic images; then they used image subtraction of least threshold development and segmentation techniques in order to identify tumour[10].In 2014, Ajo et al offered an algorithm in order to eliminate noise effectively in mammography images through splitting and merging. This method reduces false positives in diagnosis of breast cancer. MIAS base images were used in order to test this idea; 297 images (92.24%) out of 322 images are detected [11].Kanchana et al (2015) have segmented mammographic images using threshold based wavelet in order to diagnose breast cancer; Division of is done. Wavelet transform is used based on four levels of analysis. Mammographic images are analysed in this method using statistical characteristics of wavelet and target vector. MIAS database is used in this search in order to examine research process; the method accuracy is 92.30 percent [12].

The general trend of this paper is shown in Figure 1.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

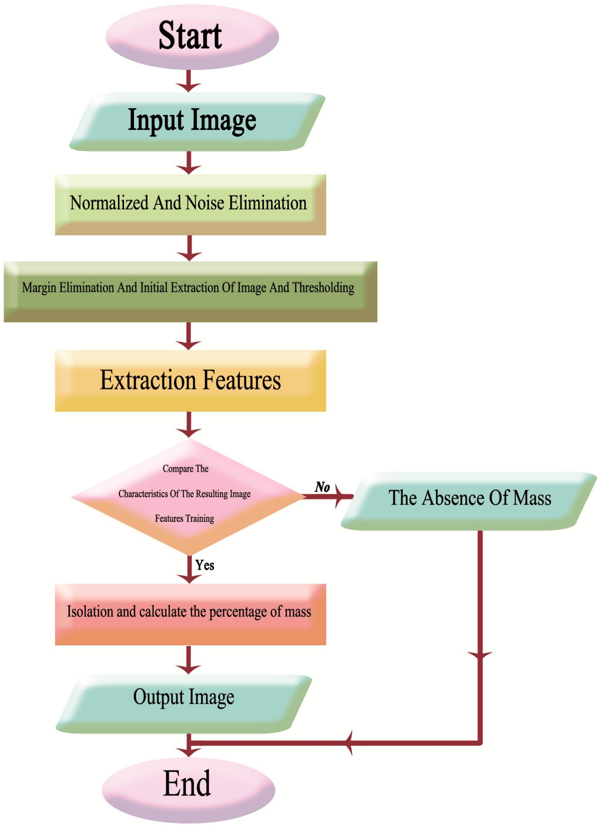

Figure 1: The general trend of this study

II.METHODS

II.I.PREPROCESSING

We use preprocessing techniques before examining mammographic images and extracting images features in order to clarify and improve the quality of mammographic images.

Firstly, we use median filter that averages in neighbourhood of image in order to smooth and remove noises of mammographic images. We use edge detection and vertical and horizontal profiles in order to remove the margin and initial extraction of image. The margins are removed in order to extract the images in horizontal and vertical dimensions after profile edge detection and taking into account the least threshold to zero level. In Figure 2, the results of edge detection, profiling, and the final separation can be seen.

 DOI: 10.15662/ijareeie.2015.0408002

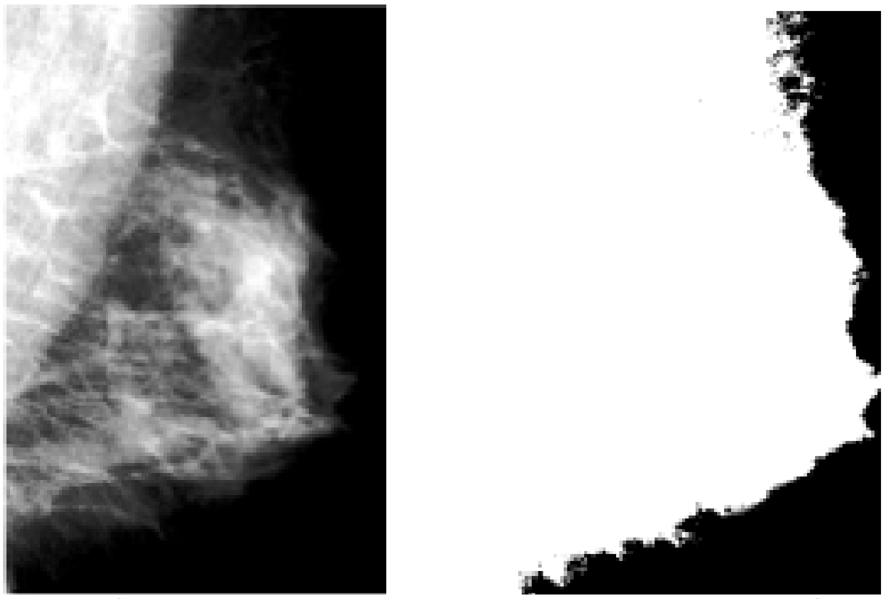

Figure 2: left side is mammography image and right side is isolated margin of image

II.II.PROCESSING

Segmentation can be done step by step and based on a window with specific size in order to segment different areas of an image and detect cancer areas, so that each window pixels belonging to a class is checked. Notably, selecting window size can be very important. Segmentation accuracy will be increased if we use large size windows for segmentation because it increases the number of pixels in other classes studied. Therefore, selecting large windows can segment accurately large and homogenized areas but it makes poor segmentation in different parts of border areas. Also, selecting small windows may reduce different pixels of classes within window, but it has less accurate segmentation due to few statistics. Therefore, using small window within areas and using large windows in other cases seems appropriate. In this study, we used the idea of four blocks (dyadic) in order to implement segmentation windows. To do this, we divide recursively selection window for segmentation into 4 square sub divisions at each step.

With this type of study structure, images are estimated based on class label for each of segmentation windows. This estimation requires PDF model (probability distribution function) for each class pixels (according to windows under review). Notable point of this work is wavelet based transform and wavelet based statistical model. Models of different tissues are very crucial in segmenting cancer suspicious areas because access to pixel-perfect joint distribution is very difficult. The basis for examining areas in images is images range transform using wavelet because the inverse and linear transformation leads to components and modelling their structure is much easier. More images, especially grey images have structures (based on edges and protrusions) and using wavelet transform lead to precise segmentation of areas. In fact, wavelet transform of these cases acts as multi-scale edge detection that expresses images structure based on different scales and three different directions. In fact, wavelet transform leads to large components for edge areas and small components for flat areas.

Many statistical methods have been proposed in order to model the structure of textures, but we use hidden Markov model (HMM) in this paper. Hidden Markov model is used in this researchin order to approximate marginal statistics and wavelet components. Hidden Markov model considers hidden state variable in this application for each of wavelet components in order to control their large or small amount. Then, the marginal distribution of each component is modelled by combining two Gaussian distributions. It must be noted that for Gaussian distribution we use large variance for modelling large areas and small variance for modelling small areas. It is noteworthy that double Gaussian combination is able to model marginal statistical distribution of non-Gaussian wavelet components in actual images. Hidden Markov model extracts the durability of large and small components during scale changes using dependency between Hidden Markov states.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

As it was mentioned, firstly, tree window structure is created in each image under investigation in order to detect suspicious areas of cancer and then windows wavelet transform is calculated in order to extract information and then each of wavelet components are modelled based on double Gaussian distribution combination whose variance is controlled by hidden Markov state variable.

In this step, we collect model parameters including double Gaussian combined variances and the possibility of Markov transient state within a vector M. The hidden Markov model can serve as a model of estimating almost all wavelet components joint distributions W, f (w / M). For example, EM repeated method can be used in order to estimate Hidden Markov model parameters. HMM tree has an interesting structure that matches windows. Each HMM sub-tree is a tree HMM that starts from it node and models statistical behaviour of wavelet components related to di window. Therefore, we consider distribution of each window in image modelling process as below:

$f(d_i|M)$ (1)

Now, we can segment images using this tool. We suppose that we have trained a tree-based HMM for each texture class of $c \in \{1,2,\dots,N_c\}$ based on $M_c$parameters. We consider the wavelet transform mode$\widetilde{w}$of test image$\tilde{x}$that includes these textures. Then, multi-scale distribution computation for different windows based on each tree-based HMM leads to $f\left(\tilde{d}_i \middle| M_c\right), c \in \{1,2,\dots,N_c\}$ for each dyadic sub-images $\tilde{d}_i$(square that is divided at every step based on 4 sub-sections). Therefore, having multi-scale distribution at this stage, test image segmentation is done based on maximum likelihood criterion as follows:

The segmentation process leads to a J member set from various segmentations:

In this study, firstly, we select an image of cancerous tissues and then we use it for training tree-based HMM. We consider an image like below image for segmentation using explained mechanism with tree structure of windows:

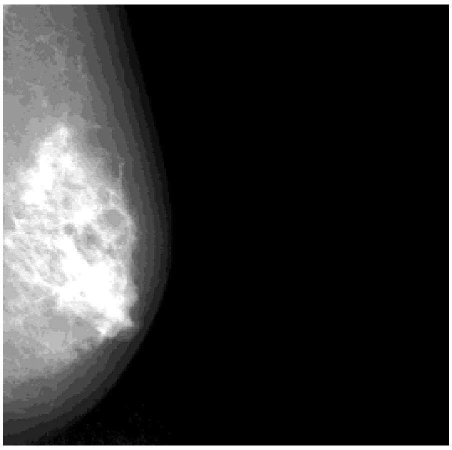

Figure 3: Sample test image

The final segmentation result for intended last scale is seen in Figure 4:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

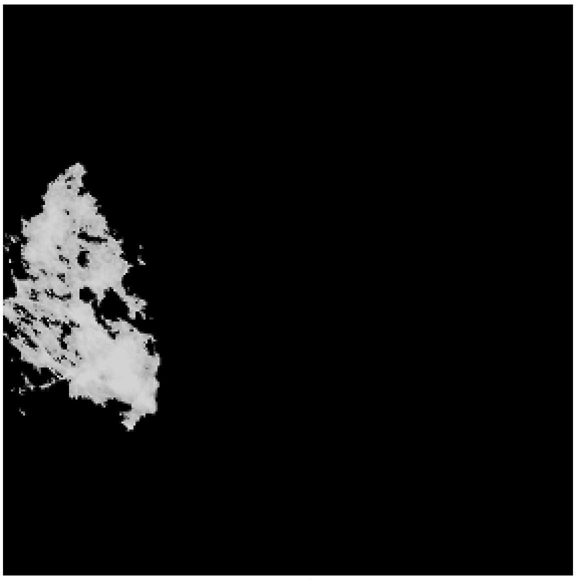

Figure 4: Sample result from study approach

In this research, cancerous tumours will be isolated and detected as well as mass percentage will be calculated and the result will be announced.
Mass growth rate is estimated during mammogram. For example, the mass amount is calculated 12.4127 percent for below image.

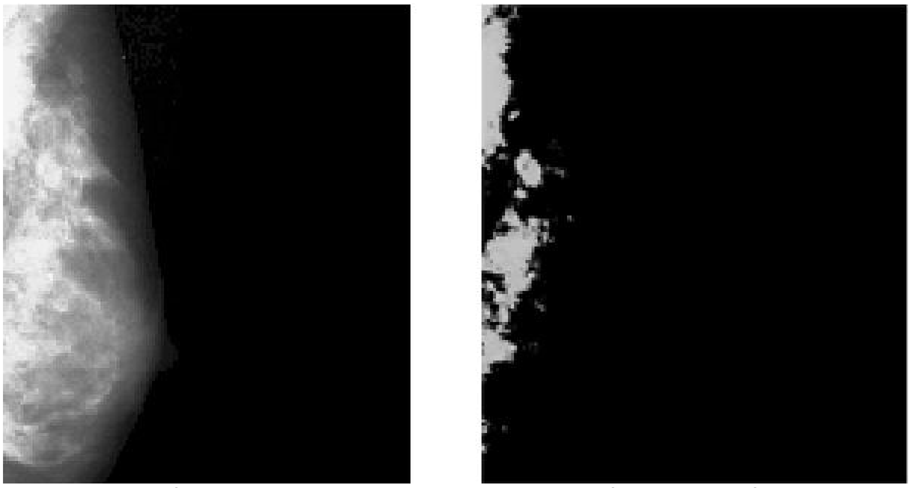

Figure 5: Mammogram image before image processing(a) mass extracted from image after image processing (b) that in mammogram image, the mass volume is 12,4127 percent

III. RESULT AND DISCUSSION

MIAS database and Paden images and images gathered from mammography are used in order to evaluate the proposed approach. The aim is to compare the performance of proposed method and the results of cancer diagnosis from mammographic images by an expert. The results of this method are shown in Table 1.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Sample/number of samples	Result
Simple Sample/98	100 %
Complex Sample/52	92%
Total/150	96%

Table 1: Results of proposed method

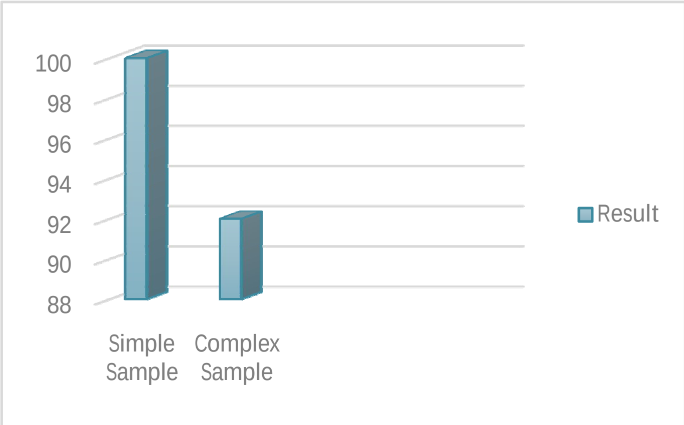

Figure 6:Results diagram of proposed method

As the results show, mammographic images of proposed method have detected all cases; in complex cases, the proposed method has detected 92% of cases where detecting suspicious areas is not simple and overall rate of detection is obtained 96% in proposed method.

The proposed method has benefited from wavelet transform and hidden Markov models in order to present and isolate breast cancer tumours from mammographic images. The results show that above approach is more accurate and reliable compared to approach that uses wavelet transform in order to extract masses from mammographic images. In Table 2, we can see the comparison between two methods.

Sample/number of samples	Result of Wavelet & Hidden Markov Model	Result of Wavelet
Simple Sample/98	100 %	82 %
Complex Sample/52	92%	61%
Total/150	96%	71.5%

Table 2: Results of wavelet transform and proposed method

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

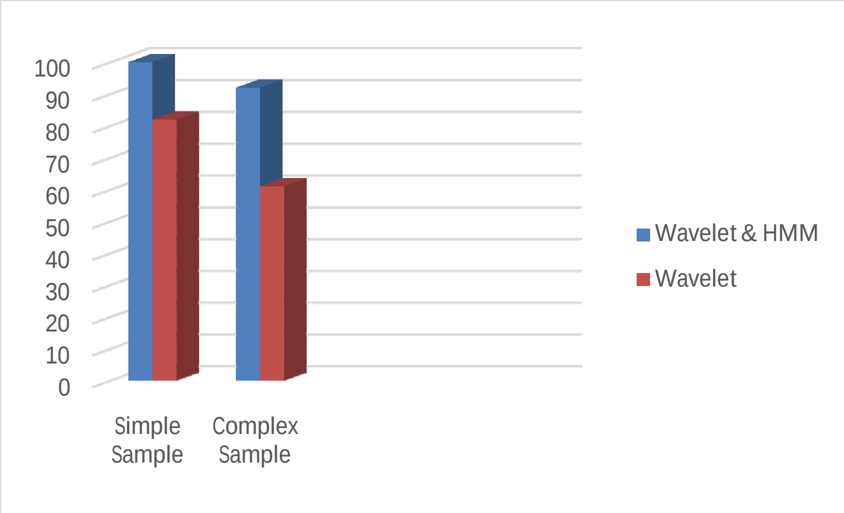

Figure 6: Comparison diagrambetween twoproposed methods andwavelet transform

IV.CONCLUSION

According to recent advances in mammography technology, image processing techniques and using image processing techniques on mammographic images, diagnosis of breast cancer lesions has increased for radiologists and specialists and has increased life expectancy among patients. According to the fact that separation of tumour mass is calculated in this study, we can calculate mass growth or reduction through periodic mammography. This improves medical science in patient diagnosis, treatment and health care with higher confidence. In this paper, Markov model is used with a tree structure in order to extract statistical properties of wavelet transform components which enhance cancerous tumours detection of mammographic images. We use the combination of MIAS database and Paden including 150 images in order to test the proposed method. The results show that proposed method is more accurate compared to methods that use only wavelet transform method. Detection rate of proposed method is 96% that is improved 24.5% compared to wavelet transform with detection rate of 71.5 percent.

REFERENCES

[1] World- Health - Organization. (2013). Cancer, WHO, Fact sheet N°297. Available: http://www.who.int/mediacentre/factsheets/fs297/en/
[2] Fiyuzi M, RezaeiKh, Hdadniya C, providing a new method in identification and extraction of breast cancer tumors from mammography images, the Fourteenth Student Conference on Electric Engineering, Industrial University of Kermanshah, September 2011.
[3] http://www.dralighayour.com.
[4] B. Verma and P. Zhang, "A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms," Applied Soft Computing, vol. 7, pp. 612 -625, 2007.
[5] M. L. Giger, "Computer-aided diagnosis in radiology," Academic Radiology, vol. 9, pp. 1-3, 2002.
[6] http://peipa.essex.ac.uk/info/mias.html.
[7] A. Papadopoulos, D. I. Fotiadis and L. Costaridou, "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques", Computers in biology and medicine, 1045-1055, 2008.
[8] S. Bhadoria, C. Dethe, S. Patra and A. Chaubey, "Breast Tumor Shape and Size Analysis for Growth Estimation Using Mammography",International Journal on Advanced Computer Engineering and Communication Technology.
[9] I. K. Maitra, S. Nag and S. K. Bandyopadhyay, "Identification of abnormal masses in digital mammography images", International Journal of Computer Graphics, 17 -30, 2011.
[10] P. Natarajan, D. Ghosh, K. N. Sandeep and S. Jilani, "Detection of Tumor in Mammogram Images using Extended Local Minima Threshold", International Journal of Engineering & Technology (0975-4024), 2013.
[11] J. Ojo, T. Adepoju, E. Omdiora, O. Olabiyisi and O. Bello, "Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram" 2014.
[12] M. Kanchana, P. Varalakshmi, "Breast Cancer Diagnosis Using Wavelet Based Threshold Method", ISSN 1990-9233,2015.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

BIOGRAPHY

MehranEmadi received the B.S. degree in Electrical Engineering (Electronics) in 1995, the M.S. degree in Electrical Engineering (Electronics) in 1999 from Azad University, NajafAbad branch, Iran, and the Ph.D. degree in Electrical Engineering (Image processing) from University Technology Malaysia (UTM), Malaysia in 2013. His researches include semiconductors in his M.S. degree and image processing, computer vision, biometric applications in his Ph.D. degree period. There have been more than 30 technical papers, published by him and he is an assistant professor in Azad University in Iran and now he is the head of Master students in Azad University Mobarakeh branch. He has been working as the program chair, and a committee member for some international conferences, and has been training several students in Universities by webinar.

SayedeSomayehHosaini received the Associate degree in Computer in 2006 from Azad University,Najafabad, Iran, and received the B.S. degree in Computer Engineering (Software) in 2009 from Applied Science University, Isfahan Iran, studying Computer Engineering (Computer systems architecture)now,Islamic Azad University, Najafabad, Iran.