Detection of Cancer Cells in Gabour Filtered Mammogram Using Gray Level Co-Occurrence Matrices

S. Meenalosini and Dr. J. Janet

***Abstract*---**This paper presents a hybrid technique which aims to assist radiologist in identifying breast cancer at its earlier stages through mammograms. It is difficult to identify masses in raw mammogram. Hence, in this paper an intelligent system is designed to diagnose breast cancer in mammogram using intelligent techniques such as Gabor filter and gray level co-occurrence matrices. Preprocessing, Segmentation and mass extraction are the three major steps involved in the proposed method. In preprocessing, down sampling and quantization is applied on input mammogram, following it noise removal is efficiently performed using median filter and finally Region of Interest is extracted using histogram matching. . In segmentation, a band-pass filter is formed by rotating a 1-D Gaussian filter(off center) in frequency space, termed as —Circular Gaussian Filter (CGF). A CGF can be uniquely characterized by specifying a central frequency and a frequency band. Usually mass appears as a brighter region on a mammogram. Mass region can be segmented out using a threshold that is adaptively decided upon the histogram analysis of the CGF-filtered mammogram. Finally extraction of masses is performed using gray level co-occurrence matrices (GLCM) features. GLCM Features like entropy, contrast and homogeneity is analyzed in order to detect whether extracted region contains masses or normal tissue. Efficiency of the proposed method is calculated by analyzing true positive, true negative and false positive, false negative results. Receiver Operating Characteristics curve method is used to analyze efficiency of the proposed method. Thus, the proposed approach would be helpful for automated real time breast cancer diagnosis.

***Keywords*---**Gabour Filter, Gaussian Filter, Gray Level Co-Occurrence Matrices, Histogram Matching, Mammogram, Masses, Segmentation

I. Introduction

BREAST cancer is the second most common cause of cancer death in women and the main cause of death in women ages 45 to 55. Mammography is at present the best available technique for early detection of breast cancer. The most common breast cancer abnormalities which are indicated as breast cancer tissue are masses and microcalcification. In the mammogram mass areas usually appear brighter than healthy tissues. Calcification is nothing but small calcium deposits. But all calcification will not be the sign of breast cancer. Normal calcification tends to be numerous, clustered, small, varying in size and shape. But the calcification associated with benign diseases is generally larger, more rounded, smaller in number and more homogenous in size and shape.

At present mammogram readings are performed by radiologists and mammographers who will examine the mammograms visually for the presence of deformalities that can be interpreted as cancer. But reading mammogram image for cancer detection is challenging task as the image occasionally show low contrast difference especially on dense breast. It will be time consuming to train an expert person in this area. Manual reading may result in misdiagnosis due to human errors. Hence performance of radiologist varies from 65 to 85% effectiveness. In many cases there are two types of errors occur while reading mammograms, False Positive(FP)which consists of the detection of a sign that appears to be malign which usually lead to unnecessary intervention and False Negative(FN) which are detection of malign signs that are classified as benign by radiologist.

To overcome the above problem in diagnosing breast cancer and to improve the diagnostic accuracy and efficiency of screening mammography, computer aided diagnosis techniques are introduced. Over the past two decades many attempts have been made by computer scientists to assist radiologists in detection and diagnosis of masses by developing computer aided tools for mammography interpretation. Image processing and intelligent synthesis are two main streams of computer technologies that have been constantly explored in the development of computer aided mammography systems.

In this paper Gaussian filter, gabour filter is used to segment the mass region in mammogram then mass region is verified by using gray level co-occurrence matrices.

II. Literature Review

Microcalcifications are small calcified structures that appear as clear points on a mammogram [3,4].For mammograms manifesting masses this corresponds to the detection of suspicious mass regions. A number of image processing methods have been proposed to perform this task. S. M. Lai et al [5] and W. Qian et al [6] have proposed using modified and weighted median filtering, respectively, to enhance the digitized image prior to object identification. Researchers have concentrated on identifying areas in mammograms that may contain cancerous changes. Steps have been taken to fully automate mammogram analysis. Various technologies such as wavelet based image denoising multiresolution based image processing[7]and Markov random field(MRF)[8],have been used Even though many algorithms are available for tumor

Manuscript received on January 31, 2011, review completed on March 10, 2011 and revised on March 19, 2011.

Meenalosini, Research Scholar, Veltech Dr.RR. & Dr.SR Technical University, Chennai, India. E-Mail: losini_2005@yahoo.co.in

Dr. J. Janet, Dean of computing, Veltech Dr.RR. & Dr.SR Technical University, Chennai, India. E-Mail: janetjude1@rediffmail.com

Digital Object Identifier No: DIP042011010.

detection the detection rate is still not high. To optimize this MRF based segmentation, Particle

Swarm Optimization (PSO) algorithm is implemented to compute the adaptive optimal threshold value. [9]. Wavelet transforms were used for the detection of microcalcifications by Strickland in [10]. An adaptive method for enhancing the contrast of mammographic features of varying size and shape is presented in [11]. Ref. [12] presents a neural network based method to classify mammograms as malignant or benign. Thus, mammographic images with high breast density value should be examined more carefully by the radiologists, creating a need for auto-matic breast parenchymal density estimation algorithms. In [13], such algorithms in the literature are presented and a new technique, introducing a histogram distance metric achieves good results. Some existing algorithms, e.g., [14,15] use the texture information of mammograms, in order to extract more features for the breast density estimation

III. Methodology

A. Mammogram Preprocessing

Goal of preprocessing is to prepare the image for next two steps segmentation and mass identification. Preprocessing is performed by some basic image processing techniques such as down sampling, quantization, noise removal and normalization to improve the quality of mammogram.

To reduce the computational load without losing the information the original mammogram should be down sampled by factor 4. So that the resulting mammogram will be reduced to 1/16th of it's original size. Sampled image is quantized down to 8 bits per pixels. So that sampled image will be spreaded to 256 gray levels. Due to sampling and quantization mammogram will be digitized with high resolution and fidelity.

Median filter is applied to remove the noise from mammogram. In this process intensity value of every pixel is replaced with median value of neighborhood pixel with 3x3 windows. In order to reduce the variation in brightness among different mammomgrams which occurs during image capturing and to achieve computational consistency images are normalized by mapping all mammograms into fixed intensity range.

Result of all the preprocessing methods mentioned above are shown in figure.1

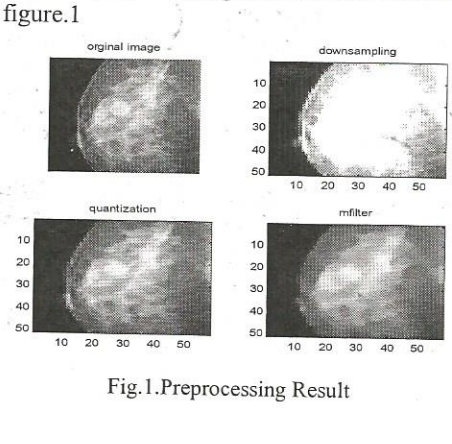

Fig.1.Preprocessing Result

B. Segmentation

The goal of segmentation is to find all suspicious regions that should contain as many cancers (masses or calcifications) as possible. The region of interest on a mammographic image (breast ROI) is extracted to reduce the processing time (by ignoring the dark areas). Region of interest is extracted by manual segmentation. As a result of this process darker region is ignored and only the suspected brighter region will be considered as region of interest.

A set of band-pass filters formed by rotating a 1-D Gaussian filter (off center) in frequency space, termed as —*Circular Gaussian Filter (CGF)* . CGF can be uniquely characterized by specifying a central frequency (f) and a frequency band (σ). A mass or calcification is a space-occupying lesion and usually appears as a bright region on a mammogram. On the filtered mammograms with a set of CGFs, the highlighted regions correspond to mass or calcification segments. Consequently, the suspicious mass or calcification segments can be extracted out using a threshold adaptively decided upon histogram analysis. Typically, the CGF parameters of (f = [12 24 48], σ = [6 12 24]) produce promising segmentation results. In our case f=12 and σ=*24* produced better result. Hence it was taken in to process.

In Fourier frequency domain, a *Circular Gaussian Filter* (CGF) is defined as follows:

$$CGF(u,v) = \frac{1}{2\pi\sigma} e^{-\frac{u_1^2+v_1^2}{2\sigma^2}} \quad (1)$$

Where $u_1 = (u - f\cos\theta)\cos\theta - (v - f\sin\theta)\sin\theta$ (2a)

$v_1 = (u - f\cos\theta)\sin\theta + (v - f\sin\theta)\cos\theta$
where $(\theta = 0 \sim 2\pi)$ (2b)

Where f specifies a central frequency and σ defines a frequency band.

Alarm pixels and *alarm segments* are generated with the following procedures:

(1) *Alarm pixels* are produced by thresholding three CGF-filtered images (I_Fm) pixel-by-pixel. The *alarm threshold* (T_Am) is determined by histogram analyses. For each of three CGF-filtered images (I_Fm, m = 2, 3,4), initialize a corresponding alarm image, I_Am, with zero pixel values, and then:

(a)Compute the histogram and *accumulated histogram*: H_Fm and AH_Fm.

(b)Find the locations of peaks in H_Fm by using histogram gradient changes (of sign pattern [+ + -]): {LP_1, LP_2, LP_q}; and assumed this set is in the order from the lowest (LP_1) to the highest (LP_q) gray level.

(c)Choose the candidates of *alarm threshold*: Tk = {LPi | when (the *selected alarm area*) < (10% entire breast ROI area); $i = 1 \sim q$}, k = p, p+1, …, q ($2 \le p \le q$). Use AH_Fm to calculate the selected alarm area.

(d) Let the *alarm threshold* be one of $\{T_k;\ k = p \sim q\}$, *i.e.*, $T_{Am} = T_l$, $p \le l \le q$, such that $|LP_l - LP_{l-1}|$ is the maximum among $\{|LP_k - LP_{k-1}|;\ k = p \sim q\}$.

(e) Mark a pixel at (x, y) as a candidate of alarm pixel if $I_{Fm}(x, y) > T_{Am}$ by assigning $I_{Am}(x, y) = 4 - m$, where $m = 2,3,4$

(f) A pixel at (x, y) is considered as an *alarm pixel* if $\sum_{m=1}^{3} I_{Am}(x, y) \ge 4$.

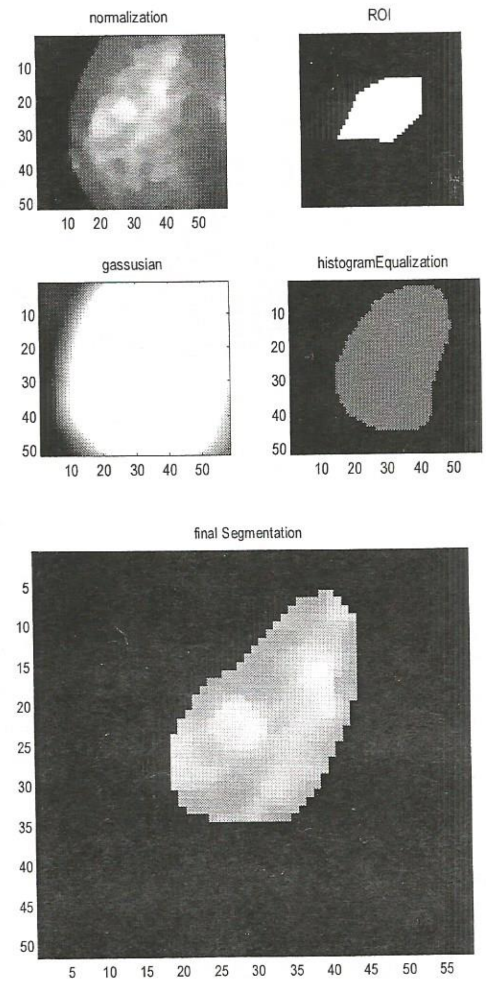

Fig.2. Segmentation with Gaussian Filter and Histogram Equalization

Gabor filters have been used in many applications, such as texture segmentation, target detection, edge detection, retina identification, image coding and image representation. A *Gabor filter* can be viewed as a *sinusoidal* plane of particular frequency and orientation, modulated by a *Gaussian* envelope. It can be written as:

$$g(x, y) = e^{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right]} e^{-j2\pi(u_0 x + v_0 y)} \quad (3)$$

where in fourier frequency domain ,filter's response consists of two 2D Gaussian function

$$G(u, v) = G_1 + G_2 = e^{-\frac{1}{2}\left[\frac{u_1^2}{\sigma_u^2}+\frac{v_1^2}{\sigma_v^2}\right]} + e^{-\frac{1}{2}\left[\frac{u_2^2}{\sigma_u^2}+\frac{v_2^2}{\sigma_v^2}\right]} \quad (4)$$

where $\sigma u = 1/(2\pi\sigma x)$ and $\sigma v = 1/(2\pi\sigma y)$ are the standard deviation along two orthogonal directions (which determines the width of the Gaussian envelope along the *x*- and *y*-axes in spatial domain), and assume that the origin of the Fourier transform has been centered. The intermediate variables are defined as following:

$$u_1 = (u - f\cos\theta)\cos\theta + (v - f\sin\theta)\sin\theta \quad (5a)$$

$$v_1 = -(u - f\cos\theta)\cos\theta + (v - f\sin\theta)\sin\theta \quad (5b)$$

$$U_2 = (u + f\cos\theta)\cos\theta + (v + f\sin\theta)\sin\theta \quad (5c)$$

$$v2 = -(u + f\cos\theta)\cos\theta + (v + f\sin\theta)\sin\theta \quad (5d)$$

From each mammogram, a total of 20 Gabor filtered images (I_{Gmn}, $m = 1\sim5$, $n = 1\sim4$, in spatial domain) are produced with 20 Gabor filters distributed along five bands ,located from low to high frequencies(f=6,12,24,48,80) and by four orientations (vertical, 45°, horizontal, and 135°).

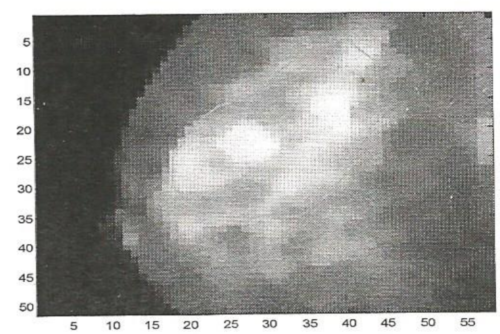

Fig.3. Segmented Area

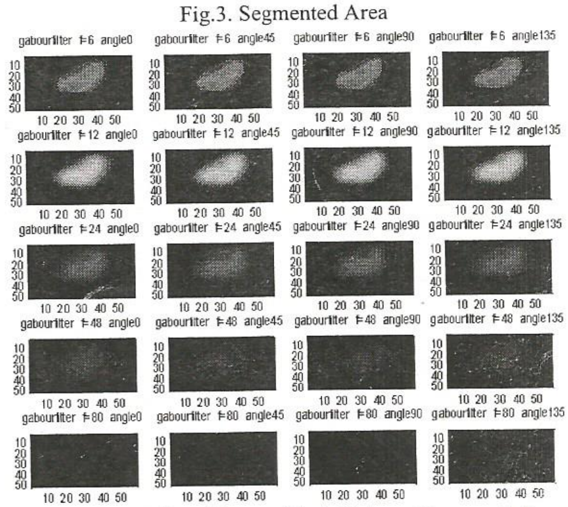

Fig.4.Gabour Filtered Image (Each Column Represents Four Orientations in 0°,45°,90° and 135° from Left To Right)

0974-9691/ICST-046/06/$20/$100 © 2011 CiiT Published by the Coimbatore Institute of Information Technology

C. Features Extraction and Selection

Texture features have been proven to be useful in differentiating masses and normal breast tissues. The texture features are extracted using gray level co-occurrence matrices (GLCM). The matrices are constructed for direction of θ given as 0°,45°,90° and 135°.A single direction might not give enough and reliable texture information. For this reason, four directions are used to extract the texture information for each masses and non-masses tiles area.

The texture descriptor derived from GLCM are contrast, energy and homogeneity. Table 1 provides the equations for the four features. The contrast measures the amount of local variations present in an image, while energy is the sum of squared elements in GLCM. Energy may also be referred as uniformity or the angular second moment. The homogeneity descriptor refers to the closeness of the distribution of elements in GLCM to the GLCM diagonal. Lastly, correlation will show how correlated a pixel is to its neighbour over the whole image.

All the 20 gabour filtered images are splitted into 8x8 windows. Then the homogeneity value, energy value and contrast value is calculated for all 8x8 tiles in every gabour filtered images. From the data base analysis it is noted that the region which has the homogeneity value greater than 0.9, energy value greater than 0.6 and contrast value less than 0.2 contains masses. Hence the above condition is checked in all 20 gabour filtered images for all the four orientation. The tile region which satisfies the above three conditions is concluded as mass region. The region which shows mass effect in the maximum number of 20 gabour filtered image can be confirmed as mass region which is shown in the table2.

TABLE 1
FEATURES OFGLCM

Feature	Formula
Contrast	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
Energy	$\sum_{i,j=0}^{N-1} (P_{i,j})^2$
Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
Correlation	$\sum_{i,j=0}^{N-1} P_{i,j}\frac{(i-\mu)(j-\mu)}{\sigma^2}$

$P_{i,j}$ = Element i, j of the normalized symmetrical GLCM

N = Number of gray levels in the image as specified by number of levels in under quantization on the GLCM

μ = The GLCM mean, calculated as: $\mu=\sum_{i,j=0}^{N-1} iP_{i,j}$

σ^2 = The variance of the intensities of all reference pixel in the relationships that

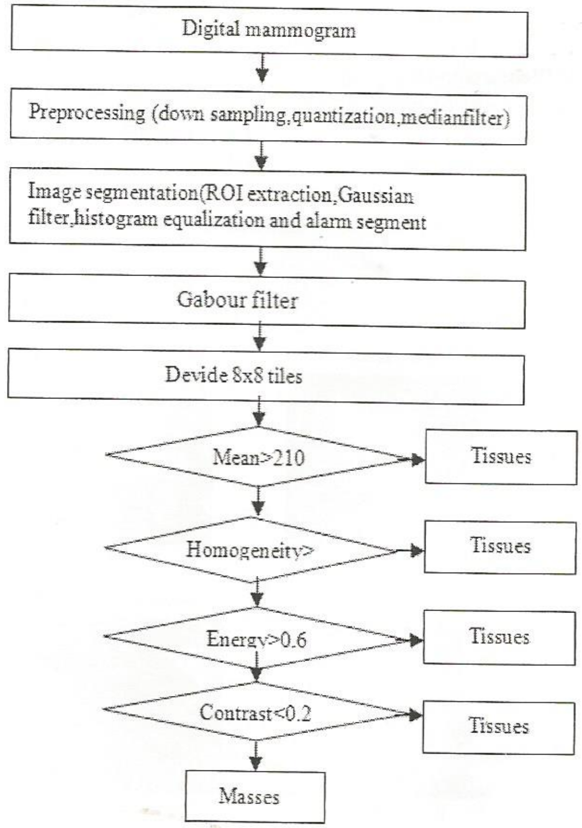

Fig.5. Flow Chart for The Entire Process

IV. RESULTS AND DISCUSSIONS

All the 20 gabour filtered images are analyzed for the GLCM feature condition which is shown in the flow chart. In each image If the 8x8 tile region satisfies all the three conditions then it is confirmed as mass area else it can be considered as tissue area. It is found that the best features for discriminating masses are the features of GLCM constructed at frequency 48 for the selected image. But it differs from mammogram to mammogram. For masses area, the contrast, homogeneity and energy ranges are 0.00-0.07, 0.96-1.00 and 0.70 -1.00, respectively. Similarly for non-masses area, the contrast, homogeneity and energy ranges are 0.27-0.73, 0.63-0.80 and 0.16-0.36, respectively. Some samples of range of values homogeneity, energy and contrast of masses and normal tissues of 8x8 tile area is shown in table.2. (m=1,2,3,4 denotes frequency 6,12,24,48 respectively and n=1,2,3,4 denotes direction 0°,45°,90°,135° respectively).

Table2 shows the samples of row and column number of some tile regions in which mass is detected according to homogeneity, energy and contrast value which is given in first three columns of the table. The algorithm is analyzed for different mammograms. The algorithm is analyzed both for cancer detected and cancer less mammograms, in which mass and normal tissues were correctly classified by the algorithm for all the given inputs. Efficiency and sensitivity of the algorithm is analyzed. It is found that the algorithm shows 98% efficiency and 90% sensitivity, while comparing with the previous methods it is quite high. Efficiency of the algorithm is plotted in Fig.6.

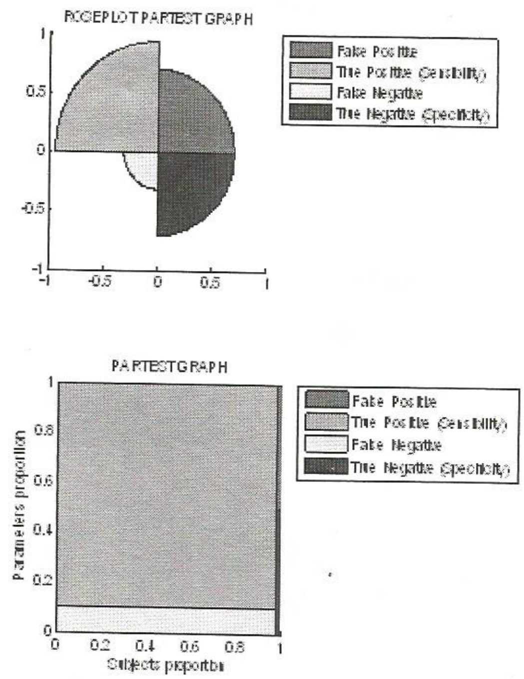

Fig.6 Experimental Results in the Detection of Masses

V. Conclusion

In this paper the suspected mass region is segmented using gabour filter and segmented region is verified by gray level co-occurrence matrices such as homogeneity, energy and contrast value. The algorithm is checked with different mammogram sets which contain both cancer detected and cancer less samples. False positive (FP) and false negative (FN) cases are considered as errors in these experiments because they will degrade the overall performance of the detection techniques. The proposed method shows greater true positive and true negative values and nil false positive and false negative values for the given samples. Efficiency of the algorithm is checked with performance measurement tool of MATLAB which shows 98% efficiency and 90% sensitivity. Efficiency of the proposed method is compared with the existing methods by using Receiver Operating Characteristics curve method which shows better results for various mammograms than the existing methods. Thus the Efficiency of the proposed method is quite high while comparing with the existing methods.

TABLE 2
SAMPLES OF DETECTED REGION BASED ON THE RANGE OF HOMOGENEITY, ENERGY AND CONTRAST VALUE

Homogeneity	Energy	Contrast	Frequency	Angle	Row	Column
1.000	1.000	0.000	m 4.000	n 1.000	j 3.000	i 8.000
1.000	1.000	0.000	m 4.000	n 1.000	j 4.000	i 8.000
0.993	0.973	0.014	m 4.000	n 1.000	j 5.000	i 8.000
0.979	0.894	0.042	m 4.000	n 1.000	j 6.000	i 8.000
0.979	0.820	0.042	m 4.000	n 1.000	j 7.000	i 8.000
0.972	0.730	0.056	m 4.000	n 1.000	j 8.000	i 8.000
1.000	1.000	0.000	m 4.000	n 2.000	j 6.000	i 1.000
0.993	0.973	0.014	m 4.000	n 4.000	j 8.000	i 5.000
0.993	0.973	0.014	m 4.000	n 4.000	j 6.000	i 6.000
0.993	0.946	0.014	m 4.000	n 4.000	j 7.000	i 6.000
0.986	0.894	0.028	m 4.000	n 4.000	j 8.000	i 6.000
0.986	0.946	0.028	m 4.000	n 4.000	j 6.000	i 7.000
0.986	0.894	0.028	m 4.000	n 4.000	j 7.000	i 7.000
0.979	0.820	0.042	m 4.000	n 4.000	j 8.000	i 7.000
0.993	0.973	0.014	m 4.000	n 4.000	j 5.000	i 8.000
0.979	0.894	0.042	m 4.000	n 4.000	j 6.000	i 8.000
0.979	0.820	0.042	m 4.000	n 4.000	j 7.000	i 8.000
0.972	0.730	0.056	m 4.000	n 4.000	j 8.000	i 8.000

References

[1] S.Meenalosini.;Dr.J.Janet "Detection of masses in mammograms using gabor features and gray level co-occurrence matrices"

[2] yufeng zheng,' Breast Cancer Detection with Gabor Features from Digital Mammograms,algorithms,1999-4893,January2010

[3] ACR BI-RADS — Mammography, Ultrasound & Magnetic Resonance Imaging, 4th ed.; American College of Radiology: Reston, VA, USA, 2003.

[4] Chang, S.F.; Sikora, T.; Puri, A. Overview of the MPEG-7 standard. IEEE Trans. Circ. Syst. Video Technol. 2001, 11, 688–695.

[5] Gonzalez, R.C.; Woods, R.E. Digital Image Processing, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.

[6] Bhangale, T.; Desai, U.B.; Sharma, U. An unsupervised scheme for detection of microcalcifications on mammograms. In Proceedings of IEEE International Conference on Image Processing, Vancouver, BC, Canada, September 10–13, 2000; pp. 184–187.

[7] A Mohd. Khuzi, R Besar, WMD Wan Zaki, NN Ahmad., Identification of masses in digital mammogram using gray level co-occurrence matrices, Biomedical imaging and intervention journal, May2009

[8] Mencattini A, Caselli F, Salmeri et al. Wavelet based adaptive algorithm for mammographic images enhancement and denoising. IEEE International Conference on Image Processing. 2005: I-1141-4.

[9] Dominguez AR, Nandi AK. Enhanced multi-level thresholding segmentation and rank based region selection for detection of masses in mammograms. IEEE International Conference on Acoustics, Speech and Signal Processing. 2007: I-449-I-452.

[10] Xu W, Li L, Xu P. A new ANN-based detection algorithm of the masses in digital mammograms. IEEE International Conference on Integration Technology. 2007: 26-30.

[11] Ball JE. Digital mammogram spiculated mass detection and spicule segmentation using level sets. Proceedings of the 29th Annual International Conference of the IEEE EMBS. 2007: 4979-84.
[12] Mohd Khuzi A, Besar R, Wan Zaki WMD. Texture features selection for masses detection in digital mammogram. International Conference on Biomedical Engineering. 2008: 629-32.
[13] Dr.H.B.Kekre,Ms.Saylee M.Gharge, Image segmentation using extended edge operator for mammographic images ,International journal on computer science and Engineering,Vol 02,No 04,2010,1086-1091
[14] Dr.H.B.Kekre, Ms.Saylee M.Gharge,Tanuja K.Sarode Tumor detection in mammography images using vector quantization technique, ,International journal on intelligent information technology Application,2009 2(5):237-242
[15] Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers. Palo Alto, USA: HP Laboratories, 2004.
[16] Liu, S.L.; Babbs, C.F.; Delp, E.J. MultiResolution Detection of spiculated lesions in digital mammograms. IEEE Trans. Image Process. 2001, 10, 874–884.

Dr.J. Janet completed her B.E. and M.E. degrees from University Of Madras and obtained her doctorate in 2006. She has published 15 papers in international journals and 20 papers in international conferences. Currently, she is working on a chapter to be published for a book "Advances In Telemedicine", for INTECH, Austria. Her interests include Knowledge Based Systems, Telemedicine Systems and Medical Image Processing.

S.Meenalosini received M.Sc degree from Madurai Kamaraj University and M.Tech degree from Manonmaian Sundaranar University. She has been awarded with gold medal for securing first rank in M.Tech. She is currently working as Assistant professor in the department of computer science at Veltech Dr.RR & Dr.SR Technical University, Chennai.She is currently pursuing her research in Medical image processing under the guidance of Dr.J.Janet.