International Journal of Image Processing and Vision Science

Volume 1 | Issue 2 Article 4

October 2012

ANALYSIS OF MAMMOGRAM FOR DETECTION OF BREAST CANCER USING WAVELET STATISTICAL FEATURES

S.K. NOWSHAD
QUBA College of Engineering & Technology Venkatachalam Nellore, nowshad719@gmail.com

UMAR FAROOQ
QUBA College of Engineering & Technology Venkatachalam Nellore, ufq@gmail.com

Recommended Citation

NOWSHAD, S.K. and FAROOQ, UMAR (2012) "ANALYSIS OF MAMMOGRAM FOR DETECTION OF BREAST CANCER USING WAVELET STATISTICAL FEATURES," *International Journal of Image Processing and Vision Science*: Vol. 1 : Iss. 2 , Article 4.
DOI: 10.47893/IJIPVS.2012.1016
Available at: https://www.interscience.in/ijipvs/vol1/iss2/4

ANALYSIS OF MAMMOGRAM FOR DETECTION OF BREAST CANCER USING WAVELET STATISTICAL FEATURES

[1]SK. NOWSHAD & [2]UMAR FAROOQ
[1,2]QUBA College of Engineering & Technology Venkatachalam Nellore-320
Email: nowshad719@gmail.com

Abstract -Early detection of breast cancer increases the survival rate and increases the treatment options. One of the most powerful techniques for early detection of breast cancer is based on digital mammogram. A system can be developed for assisting the analysis of digital mammograms using log-Gabor wavelet statistical features. The proposed system involves three major steps called Pre-processing, Processing, and Feature extraction. In pre-processing, the digital mammogram can be de-noised using efficient decision-based algorithm. In processing stage, the suspicious Region of Interest (ROI) can be cropped and convolved with log-Gabor filter for four different orientations. Then gray level co-occurrence matrix (GLCM)can be constructed for log-Gabor filter output at four different orientations and from that first order statistical features and second order statistical features can be extracted to analyze whether the mammogram as normal or benign or malignant. The proposed method can allow the radiologist to focus rapidly on the relevant parts of the mammogram and it can increase the effectiveness and efficiency of radiology clinics.

Keywords: *Digital mammogram; log-Gabor filter; GLCM; textural features*

I. INTRODUCTION

Breast cancer is among the most common and deadly of all cancers, occurring in nearly one in ten women. Mammography is a uniquely important type of medical imaging used to screen for breast cancer. All women at risk go through mammography screening procedures for early detection and diagnosis of tumors. Special x-ray machines developed exclusively for breast imaging are used to produce mammography films. These machines use very low doses of radiation and produce high-quality
x-rays. A typical mammogram is an intensity x-ray image with gray levels showing levels of contrast inside the breast which characterize normal tissue, different calcification and masses. The contrast level of a typical mammogram image is proportional to the difference in x-ray attenuation between different tissues. In general, a clear separation between normal functioning tissue and abnormal cancerous tissues is difficult to identify since their attenuation is very similar. Radiologists visually search mammograms for specific abnormalities. Some of the important signs of breast cancer that radiologists looking for clusters of micro calcifications, masses and architectural distortions. A mass is defined as a space-occupying lesion seen in at least two different projections.

Masses are described by their shape and margin characteristics. Most mass detection algorithms consists of two stages (a)Detection of suspicious regions on the mammogram and (b)classification of suspicious regions as mass or normal tissue. Computer-aided detection (CAD) systems have been developed to aid radiologists in detecting mammographic lesions that may indicate the presence of breast cancer. These systems act only as a second reader and final decision is made by the radiologists. Recent studies have also shown that CAD detection systems, when used as an aid, have improved radiologists accuracy of detection of breast cancer[30]. It is an x-ray examination of the breasts in a woman who is asymptomatic. The diagnostic mammography examination is performed for symptomatic women who have an abnormality found during screening mammography. Nowadays, in most hospitals the screen film mammography is being replaced with digital mammography. With digital mammography the breast image is captured using a special electronic x-ray detector which converts the image into a digital mammogram for viewing on a computer monitor or storing.

Computer-aided detection (CAD) and computer-aided diagnosis (CADx) systems can improve the results of mammography screening programs and decrease number of false positive cases. Most image processing algorithms consist of a few typical steps depicted in Fig.

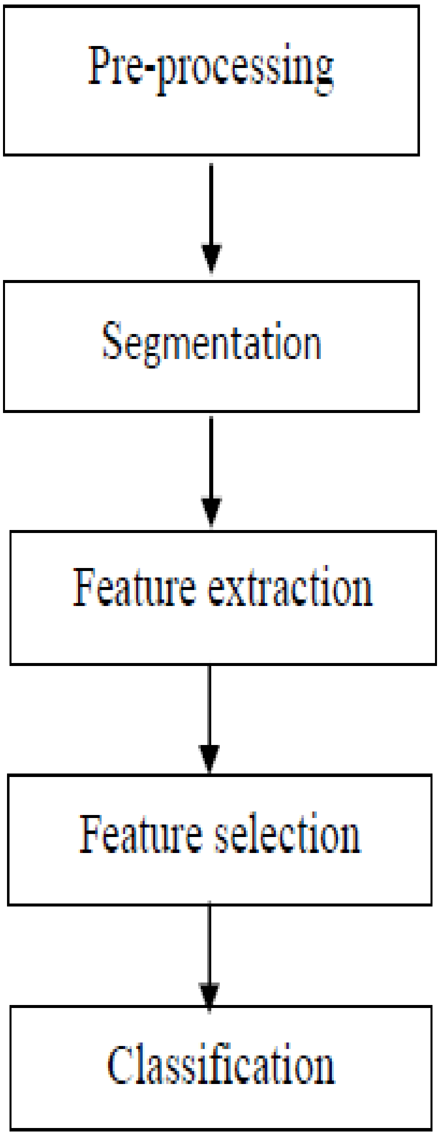
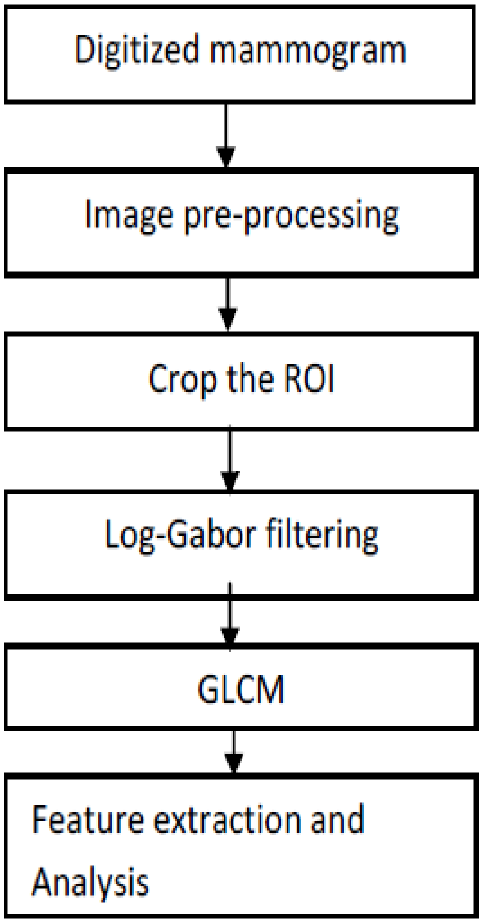


Fig:Main steps involved in the computer aided detection.

The screen film mammographic images need to be digitized prior the image processing. This is one of the advances of digital mammography where the image can be directly processed. The first step in image processing is the pre-processing step. It has to be done on digitized images to reduce the noise and improve the quality of the image. Most digital mammographic images are high quality images. The segmentation step aims to find suspicious regions of interest (ROIs) containing abnormalities. In the feature extraction step the features are calculated from the characteristics of the region of interest.

The ultimate aim of the CADx task is to help the radiologist in making recommendations for patient management. If a mass is suspected to be malignant, a biopsy must be performed.If not, the patient is either scheduled for a short-term followup or is returned to the normal screening population. If the information is insufficient for the radiologist to make a decision, special radiographic views are taken and complementary modalities like ultrasound or mangnetic resonance imaging(MRI) are used to obtain additional information.

II. SYSTEM FOR MAMMOGRAM ANALYSIS



The block diagram of system for mammogram analysis is as ahown.The proposed system involves three major steps called Pre-processing, Processing and feature extraction.

Digitized Mammogram:

The mammogram images used in this experiment were taken from the mini mammography database of mammographic image analysis society (MIAS) (http://peipa.essex.ac.uk/ipa/pix/mias/). All images are held as 8-bit gray level scale images with 256 different gray levels (0-255) and physically in portable gray map (pgm) format with size 1024 pixels x 1024 pixels. In this project mammogram is analysed using Log-Gabor filter statistical features and coded in MATLAB.

Image Pre-Processing:

Images are often corrupted by impulse noise, also known as salt and pepper noise. Linear filtering technique fails to remove the impulse noise. Non-linear digital filters are used to remove the impulse noise i.e Standard Median Filters are used to remove impulse noise as well as to preserve edges.The main drawback of a standard median filter (SMF) is that it is effective only for low noise densities.In SMF, each and every pixel is processed and is replaced by the median of its neighborhood values.New decision-based algorithm processes the corrupted image by first detecting the impulse noise.

The New decision-based algorithm is as follows:

Step 1) A 2-D window "S" of size 3×3 is selected. Assume the pixel to be Processed is P(X, Y).

Step 2) the pixel values inside the window are sorted, and $Pmin$,Pmax $and\ Pmed$, and are Determined as follows.

a) The rows of the window are arranged in ascending order.

b) The columns of the window are arranged in ascending order.

c) The right diagonal of the window is now arranged in ascending order. Now the first element of the window is the minimum value $Pmin$, the last element of the window is the maximum value Pmax and the middle element of the window is the median value $Pmed$,

Step 3)

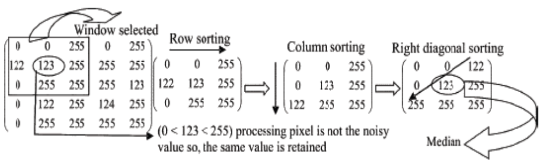
Case 1) The P(X, Y) is an uncorrupted pixel if $Pmin$<P X,Y <Pmax and $Pmin$> 0 and Pmax <255, the pixel being processed is left unchanged. Otherwise, P(X, Y) is a corrupted pixel.

Case 2) If P(X, Y) is a corrupted pixel, it is replaced by its median value if $Pmin$< $Pmed$<Pmax .

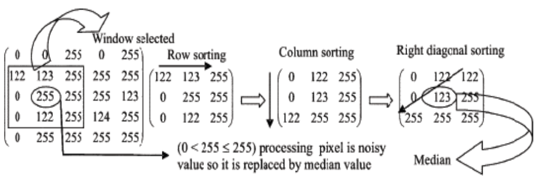
Case 3) $Pmin$ < $Pmed$ < Pmax If is not satisfied, then is a noisy pixel. In this Case, the (X, Y) is replaced by the value of left neighborhood pixel value.

Step 4) Steps 1 to 3 are repeated until the processing is completed.

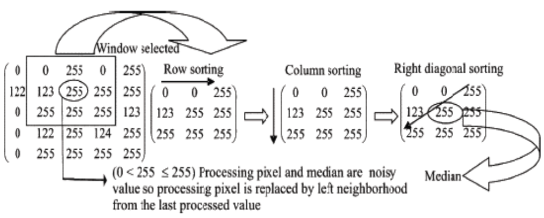
Case 1)



Case 2)



Case 3)



Crop the ROI(Region Of Interest):

In processing stage the suspicious region of interest (ROI) is cropped. For that first order statistical features such as mean and standard deviation are calculated by using the equations:

$$mean = \frac{1}{N^2}\sum_{i,j=1}^{N} R(i,j)$$

$$standard\ deviation = \sqrt{\frac{1}{N^2}\sum_{i,j=1}^{N}[R(i,j) - mean]^2}$$

Where $R\ i, j$ is ROI matrix.

d)LOG-GABOR Filters:

Log Gabor Filter is linear filter used for edge detection. Frequency and orientation representation of Gabor filters are similar to those of the human visual system. The cropped ROI is convolved with Log-Gabor filter for four different orientations. Gabor filters are a traditional choice for obtaining localized frequency information.log-Gabor filter has a transfer function.

$$G(W) = \exp\left(-\log(W/W_o)^2\right)/\left(2(\log(k/W_o)^2)\right)$$

Where wo is the filter's centre frequency.

The Log-Gabor filters are defined in the log-polar coordinates of the Fourier domain as Gaussians shifted from the origin.

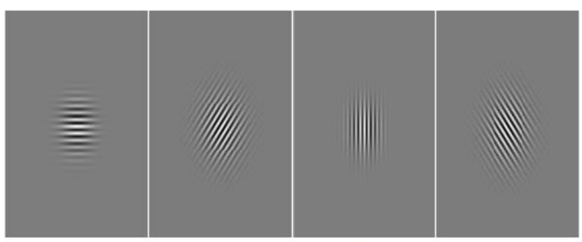
$$G_{(s,t)}(\rho,\theta) = \exp\left(-\frac{1}{2}\left(\frac{\rho-\rho_s}{\sigma_\rho}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{\theta-\theta_{(s,t)}}{\sigma_\theta}\right)^2\right)$$

$$\text{With}\begin{cases}\rho_s = \log_2 n - s \\ \theta_{(s,t)} = \begin{cases}\frac{\pi}{n_t}t & if\ s\ is\ odd \\ \frac{\pi}{n_t}\left(t+\frac{1}{2}\right) & if\ s\ is\ even\end{cases}\end{cases}$$

$$(\sigma_\rho, \sigma_\theta) = 0.996\left(\sqrt{2/3}, \frac{1}{\sqrt{2}}\frac{\pi}{n_t}\right)$$

Where (ρ, θ) are the log-polar coordinates (in log2 scale, indicating the filters are organized in octave scales). ns is the number of scales of the multiresolution schemes and nt is the number of orientations.

Log-Gabor filter is applied on the cropped ROI for single scale (ns =1) and four different orientations (n =4) those are at (0deg, 45deg, 90deg and 135deg) is given as,

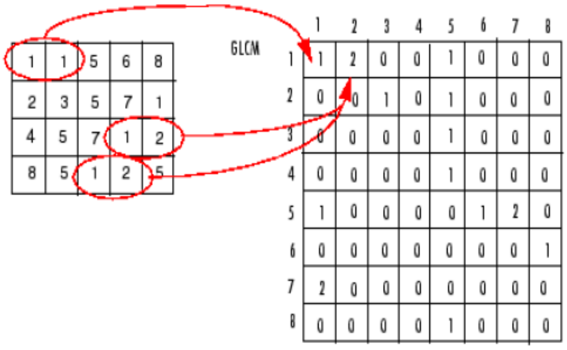


for θ=0° for θ=45° for θ=90° for θ=135°

e) **Gray Level Co-Occurrence Matrices(GLCM):**

The GLCM is a tabulation of how often different combination of pixel brightness values (grey levels) occur in an image. Co-occurrence matrices are very rich representations of an image.

In particular, ratios of the co-occurrence matrix have been shown to be good texture descriptors. This is because they capture the relative abundance of certain image characteristics. The following figure shows how to calculate several values in the GLCM of the 4-by-5 image. Element (1, 1) in the GLCM contains the value 1 because there is only one instance in the image where two, horizontally adjacent pixels have the values 1 and 1. Element (1, 2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2.



Pictorial calculation of GLCM matrix

After applying Log-Gabor filter on the cropped ROI, GLCM matrix is formed. From that second order statistical features, i.e.Co-occurrence matrix features such as contrast, energy, homogeneity, correlation, entropy, cluster shade, cluster prominence are

$$\text{Contrast} = \sum_{i,j=1}^{N} (i-j)^2 C(i,j)$$

$$\text{Correlation} = \sum_{i,j=1}^{N} \frac{(i-\mu)(j-\mu)}{\sigma^2} C(i,j)$$

$$\text{Energy} = \sum_{i,j=1}^{N} C^2(i,j)$$

$$\text{Entropy} = \sum_{i,j=1}^{N} C(i,j)\log_2 C(i,j)$$

$$\text{Homogeneity} = \sum_{i,j=1}^{N} \frac{1}{1+(i-j)^2} C(i,j)$$

$$\text{Clustershade} = \sum_{i,j=1}^{N} (i-M_x+j-M_y)^3 C(i,j)$$

$$\text{Cluster prominence} = \sum_{i,j=1}^{N} (i-M_x+j-M_y)^4 C(i,j)$$

calculated using the equations:

where

$$M_x = \sum_{i,j=1}^{N} iC(i,j) \quad and\ M_y = \sum_{i,j=1}^{N} jC(i,j)$$

μ –mean, σ-variance of co-occurrence matrix.

The physical meanings of the above features are explained as follows.

Contrast: The intensity contrast between a pixel and its neighbor over the whole image.

Correlation: Statistical measure of how correlated a pixel is to its neighbour over the whole image; Range = [-1 1].Correlation is 1 or -1 for a perfecty positively or negatively correlated image. Correlation is NaN for a constant image.

Energy: summation of squared elements in the GLCM; Range = [0 1].Energy is 1 for a constant image.

Homogeneity: Closeness of the distribution of elements in the GLCM to the GLCM diagonal; Range = [0 1]. Homogeneity is 1 for a diagonal GLCM.

Entropy: Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image.

Cluster shade and **cluster prominence**: are measures of the skewness of the matrix, in other words the lack of symmetry.

f) **Feature Extraction:**

The list of first order statistical features extracted from cropped ROI and Co-occurrence matrix features extracted from the Log-Gabor filtered output at four different orientations. They are given as follows.

f1 = Mean of the ROI .(mean)

f2 = Standard deviation of ROI .(std)

f3 = Standard deviation of Log-Gabor filter output orientation at θ=0^0 .(std1)

f4 = Contrast of Log-Gabor filter output orientation at θ=0^0.(cont1)

f5=Correlation of Log-Gabor filter output orientation at θ=0^0 .(corr1)

f6 = Energy of Log-Gabor filter output orientation at θ=0^0.(eng1)

f7=Homogeneity of Log-Gabor filter output orientation at θ=0^0.(homo1)

f8=Entropy of Log-Gabor filter output orientation at θ=0^0.(ent1)

f9=Cluster shade of Log-Gabor filter output orientation at θ=0^0.(clushd1)

f10= Cluster prominence of Log-Gabor filter output orientation at θ=0^0.(clupro1)

f11 = Standard deviation of Log-Gabor filter output orientation at θ=45^0.(std2)

f12= Contrast of Log-Gabor filter output orientation at θ=45^0.(cont2)

f13=Correlation of Log-Gabor filter output orientation at θ=45^0.(corr2)

f14 = Energy of Log-Gabor filter output orientation at θ=45^0.(eng2)

f15=Homogeneity of Log-Gabor filter output orientation at $\theta=45^0$.(homo2)
f16 = Entropy of Log-Gabor filter output orientation at $\theta=45^0$.(ent2)
f17=Cluster shade of Log-Gabor filter output orientation at $\theta=45^0$.(clushd2)
f18= Cluster prominence of Log-Gabor filter output orientation at $\theta=45^0$.(clupro2)
f19 = Standard deviation of Log-Gabor filter output orientation at $\theta=90^0$.(std3)
f20= Contrast of Log-Gabor filter output orientation at $\theta=90^0$.(cont3)
f21=Correlation of Log-Gabor filter output orientation at $\theta=90^0$.(corr3)
f22 = Energy of Log-Gabor filter output orientation at $\theta=90^0$.(eng3)
f23=Homogeneity of Log-Gabor filter output orientation at $\theta=90^0$.(homo3)
f24 = Entropy of Log-Gabor filter output orientation at $\theta=90^0$.(ent3)
f25=Cluster shade of Log-Gabor filter output orientation at $\theta=90^0$.(clushd3)
f26= Cluster prominence of Log-Gabor filter output orientation at $\theta=90^0$.(clupro3)
f27 = Standard deviation of Log-Gabor filter output orientation at $\theta=135^0$.(std4)
f28= Contrast of Log-Gabor filter output orientation at $\theta=135^0$.(cont4)
f29=Correlation of Log-Gabor filter output orientation at $\theta=135^0$.(corr4)
f30 = Energy of Log-Gabor filter output orientation at $\theta=135^0$.(eng4)
f31 = Homogeneity of Log-Gabor filter output orientation at $\theta=135^0$.(homo4)
f32 = Entropy of Log-Gabor filter output orientation at $\theta=135^0$.(ent4)
f33 = Cluster shade of Log-Gabor filter output orientation at $\theta=135^0$.(clushd4)
f34= Cluster prominence of Log-Gabor filter output orientation at $\theta=135^0$.(clupro4).

RESULT:

Experiment is conducted with 20 malignant images, 20 benign images and 20 normal images were taken from mammographic image analysis society (MIAS) database. In pre-processing, the digital mammogram is denoised using efficient decision-based algorithm. In processing stage, the suspicious Region of Interest (ROI) is cropped and convolved with log-Gabor filter for four different orientations. For each mammographic image first order statistical features such as mean and Standard deviation, and co-occurrence matrix features such as contrast, energy,homogeneity,correlation,entropy, Cluster shade , cluster prominence are extracted to analyze whether the mammogram is normal or benign or malignant.

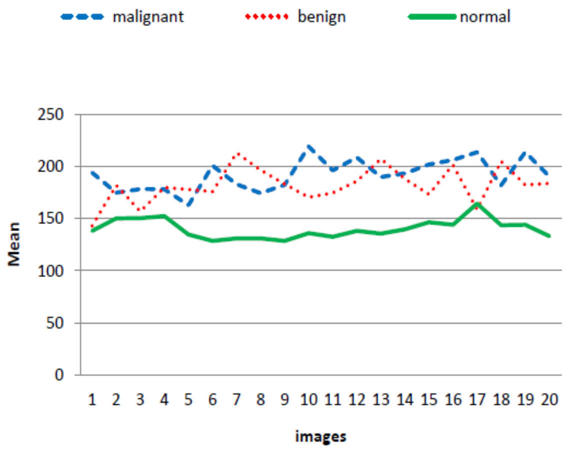


Fig:Mean of ROI for three cases

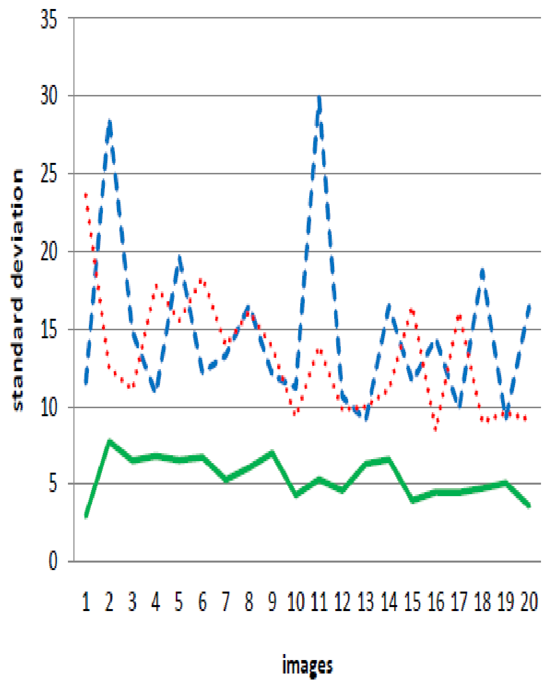


Fig:Standard deviation of ROI for three cases.

	Entropy(0^0)	Cluster shade(0^0)	Entropy(90^0)
Malignant	1.94×10^5 to 2.23×10^5	-1.76×10^18 to -6.68×10^17	1.91×10^5 to 2.24×10^5
Benign	1.78×10^4 to 1.72×10^5	-7.38×10^17 to -4.94×10^14	1.74×10^4 to 1.62×10^5
Normal	1.74×10^5 to 1.93×10^5	-9.22×10^18 to -1.7×10^18	1.69×10^5 to 1.91×10^5

Table: Range of values for significant feature

CONCLUSION:

Based on the results , one concludes that the system proposed for mammogram analysis is based on Log-Gabor wavelet statistical features is accurate for distinguishing normal, benign and malignant. Since the Log-Gabor function has the advantage of the symmetry on the log frequency axis and it spread information equally across the channels, this filter produces an uncorrelated and less redundant representation for mammogram texture compared with ordinary wavelet filters. Because of this, it is observed that the feature, distinguishing only normal mammogram from malignant mammogram but failed in distinguishing benign from normal and malignant. Note that feature vector is a combination of first order and co-occurrence statistical features.

REFERENCES

[1] Christoyianni, I.,et al., (2002) ―Computer aided diagnosis of breast cancer in digitized mammograms‖. *Computerized Medical Imaging and Graphics*, 26(5), 309-319.

[2] Cheng, H.D.,et al., (2003)―Computer-aided detection and classification of microcalcifications in mammograms: a survey‖. *Pattern Recognition*, 36(12), 2967-2991.

[3] Skalansky, J. (1978) ―Image segmentation and feature extraction‖, *IEEE Trans. System Man Cybernat.*, 8(4), 237-247.

[4] J.G.Daugman, (1993) ―High confidence visual recognition of persons by a test of statistical independence‖,*IEEE Trans. Pattern Anal. Machine Intell*, 15, 1148-1160.

[5] J.G.Daugman, (1988) ―Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression‖, *IEEE Trans. Acoustic Speech Signal Processing*, 36, 1169-1179.

[6] Cheng,H.D., et al., (2006) ―Approaches for automated detection and classification of masses in mammograms‖. *Pattern Recognition*, 39, 646-668.

[7] Field,D.J. (1987) ―Relation between the statistics of natural images and the response properties of cortical cells‖ *J. Opt. Soc. Am. A*, 4(12),2379–2394.

[8] Raghu, P.P., Yegnanarayana, B., (1996) ―Segmentation of Gabor filtered textures using deterministic relaxation‖, *IEEE Trans. Image Process*, 5(12), 1625-1636.

[9] Chellappa,R., Chatterjee,S., (1985)―Classification of texture using Gaussian Markov Random Fileds‖, *IEEE Trans.Acoustic Speech Signal Process, ASSP* – 33(4), 959 – 963.

[10] Derin, H., Elliot, H.,(1987) ―Modeling and segmentation of noisy and textured images using Gibbs random fields‖, *IEEE Trans. Pattern Anal. Machine Intell., PAMI* – 9(1), 39-55.

[11] Arivazhagan, S., Ganesan, L., (2003) ―Texture classification using wavelet transform*", Pattern Recognition Letters*, 24, 1513-1521.

[12] Chang,T., Jay Kuo,C.C., (1993) ―Texture analysis and classification with tree-structured wavelet transform‖,*IEEE Trans. Image Process*, 2(4), 429-441.

[13] Randen,T., Husoy,J.H.,(1999) ―Filtering for texture classification: A Comparative Study‖, *IEEE Trans. Pattern Anal. Machine Intell.*, 21(4), 291-310.

[14] Gross,M.H. and Koch, R. (1995) ―Visualization of multidimensional shape and texture features in laser range data using complex valued Gabor wavelets,‖ *IEEE Trans. Visual. and Comput. Graphics*, 1(1),44–59.

[15] A.K.Jain and F.Farrokhnia, (1991) ―Unsupervised texture segmentation using Gabor filters,‖ *Pattern Recognition*, 23, 1167-1186.

[16] C.A.Pomalaza-Racz and C.D.Macgillem, (1984) ―An adaptive non linear edge preserving filter,‖ *IEEE Trans. Acoustic Speech Signal Processing* , 571–576.

[17] K.S.Srinivasan and D.Ebenezer, (2007) ―A New Fast and Efficient Decision-Based Algorithm for Removal of High-Density Impulse Noises‖ *IEEE Signal Processing Letters,* 14(3).

[18] E.J.Feuer, L.wun, C.C.Boring,et al., (1993) ―The lifetime risk of developing breast cancer,‖*J Natl Cancer Inst*,85,892-897.

[19] Jain, A.K., Duin, R.P.W., Mao, J, (2000) ―Statistical Pattern Recognition‖, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37.

[20] Wei, C.H., C.T. Li, and R. Wilson (2005) ― A General Framework for Content-Based Medical Image Retrieval with its Application to Mammograms*", SPIE Medical Imaging*.

[21] Cohen, F.S., Fan, Z., Patel, M.A (1991) ―Classification of rotated and scaled textured images using Gaussian Markov random field models‖, *IEEE Trans. Pattern Anal. Machine Intell.*, 13(2), 192-202.

[22] Deivalakshmi,S.,Arivazhagan,S., Ganesan,L (2005) ―Texture classification using wavelet statistical features,‖ *IE(I) Journal-ET*,85,51-55.

[23] Weszka, J.S., Dyer, C.R., Rosenfeld, A (1976) ―A comparative study of texture measures for terrain classification‖, *IEEE Trans. System Man Cybernat, SMC*-6(4), 269-285.

[24] R.Hralick, M.Shanmugam, I.Dinstein (1973) ―Texture features for image classification,‖*IEEE Trans. System Man Cybernat*.3,610-621.

[25] D. Dunn, W. Higgins, and J. Wakeley (1994) ―Texture segmentation using 2-D Gabor elementary functions‖,*IEEE Trans. Pattern Anal. Machine Intell*, 16, 130-149.

[26] R . Mehrotra, K. R. Namuduri, and N. Ranganathan (1992) ―Gabor filter-based edge detection‖, *Pattern Recognition*,25, 1479-1493.

[27] A. C. Bovik, M. Clark, and W. S. Geisler (1990) ―Multichannel texture analysis using localized spatial filters‖,*IEEE Trans. Pattern Anal. Machine Intel*, 12, 55-73.

[28] Liapis, S., E. Sifakis, and G. Tziritas (2004)―Colour and texture segmentation using wavelet frame analysis, deterministic relaxation, and fast marching algorithms‖. *J. Vis. Commun. Image*, 15 ,1-26.

[29] A Mohd.Khuzi, R Besar, WMD Wan Zaki, NN Ahmad (2009)―Identification of masses in digital mammogram using gray level co-occurrence matrices,‖ *Biomedical Imaging and Intervention Journal.*

[30] Li, L.H., Qian, W., Clarke, L.P., Clark, R.A., Thomas, J (1999) ―Improving Mass Detection by Adaptive and Multi-Scale Processing in Digitized Mammograms‖. *Proceedings of SPIE—The International Society for Optical Engineering* ,3661, 490–498.

[31] Sylvain Fischer, Filip Sroubek, Laurent Perrinet, Rafael Redondo and Gabriel Cristobal, (2007) ―Self-Invertible 2D Log-Gabor Wavelets,‖ *International Journal of Computer Vision* ,75(2),231-246.

[32] Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, H.N. (2006) ―Approaches for Automated Detection and Classification of Masses in Mammograms‖, *Pattern Recognition* 39(4), 646–668.