

Detection of Breast Masses in Digital Mammograms using Multiple Concentric Layers

Chethan K
4th Sem, M.Tech
Dept of CSE, SJBIT
Bangalore-60

Dr. Krishna A N
Associate Professor
Dept of CSE, SJBIT
Bangalore-60

Abstract: Breast cancer is the leading cancer among the women's. Mammography is one of the most effective method available for the early detection of breast cancer. Mammogram is an x-ray projection of the 3D structure of the breast. A mammogram contains two different regions one is exposed region and another is unexposed region called as background region or non-breast region. A mass can be defined as a space occupying lesion if it is seen in at least two different projections. The implementation of Multiple Concentric Layers (MCL) is applied to mammogram segmentation; this technique focuses on the automatic detection of masses in digital mammograms. Multiple Concentric Layers (MCL) is a promising strategy for mammography to identify the malignant masses. The system involves major two stages: 1) First stage is the segmentation of breast to remove background and labels. 2) Second stage is determining the suspicious region with the presence of concentric layers.

Index Terms: Breast cancer, Mammogram, Segmentation, MCL, Otsu, Granulation.

I. INTRODUCTION

Breast cancer is one of the major health problems in Western European countries, and it constitutes the most common cancer among the women. Breast cancer is the most frequent leading cancer death in women's. Breast cancer is any form of malignant tumor which develops from the breast cells. In Breast cancer the major risk factor is that it cannot be avoided; only it can be detected earlier to increase the survival rate. Detection of breast cancer in early stage will be critical issue for a high survival rate. One of the leading method for diagnosing breast cancer is screening mammography. Screening mammography involves x-ray imaging of breast and performed on asymptomatic women to detect breast cancer earlier. Mammography is high sensitivity to fatty breast and useful for demonstration on microcalcifications, it also suitable for mass screening program. Mammogram is a low energy x-rays which will be passes through compressed breast. Radiologists interpret the mammogram images for detection of cancerous cells such as clustered microcalcifications (MCCs), masses, architectural

distortion, breast edema, and lymphadenopathy. A mammogram mainly contains two regions: the exposed breast region and unexposed non-breast region. There are number of abnormalities that are normally associated with breast cancer:

- Asymmetry between images of left and right breasts.
- Distortion of the normal architecture of the breast tissue.
- Presence of micro-calcifications in the breast.
- Presence of masses in the breast.

The mammogram image usually composes of image background, label, marker, artifact, near-skin tissue, fatty tissue, pectoral muscle and dense glandular tissue. Figure 1 describes about the pectoral muscle, breast composition, image background, subcutaneous fat region in the mammogram image

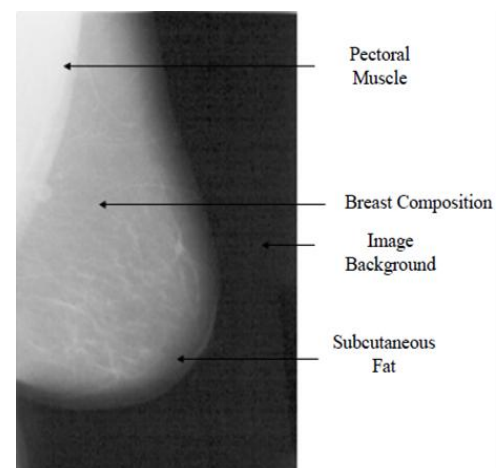


Figure1: A Mammogram Image

Segmentation [1] is a process of partitioning an image into multiple segments formed by the union of connected pixels and having similar properties, such as gray level, color, texture, brightness, contrast, etc. Segmentation is to represent an image into more meaningful and easier to analyze, regions. The aim of breast segmentation is to

separate the breast from the other objects with minimum loss of breast tissue in the mammograms. A mass [2] is defined as a space occupying lesion if it is seen in at least two different projections. A mass with a regular shape has a higher probability of being benign and mass with an irregular shape has a higher probability of being malignant cases. Mammogram images usually with high measures of breast density requires a carefully examination of radiologists. According to the American College of Radiology (ACR) and the Breast Imaging Reporting and Data System (BIRADS), the breast density are classified into four major groups are

- BIRADS 1: the breast is almost entirely fatty
- BIRADS 2: the breast will be having some fibro glandular tissue
- BIRADS 3: the breast will be heterogeneously dense
- BIRADS 4: the breast will be extremely dense.

II. RELATED WORKS

Matthew A. Kupinski and Maryellen L. Giger [3] presented on automated seeded lesion segmentation on digital mammogram. The system has many mass-detection schemes where the segmenting lesions plays main vital role. Hence there are two lesion segmentation techniques 1) single feature called the radial gradient index (RGI) 2) simple probabilistic models to segment mass lesions, or other similar nodular structures from the surrounding background. Both techniques are based on seeded segmentation algorithms which they benign with a point called seed point that defined to be within the suspect lesion.

Hidefumi Kobatake [4] performs new method for the detection of spicules on mammograms. Spicules can be viewed as line patterns radiating from the center of the tumor. The spicules characteristic pattern will be detected by line skeletons and modified Hough transform. Line skeleton processing will be done with the enhancing spinal axes of spicules and reduction in other skeletons. Hough transform is experimented on line skeletons and radiating line structures to obtain the results.

Guido M. Te Brake and Nico Karssemeijer [5] proposed Single and Multiscale Detection of Masses in Digital Mammograms. Whereas scale is an important issue with detection of masses due to its wide range of possible size masses. The proposed method will detect masses in single scale or it may be more appropriate to output at different scales in a multiscale schemes. There are three different pixel-based mass detection methods are 1) convolution of mammograms with the Laplacian of a Gaussian 2) correlation with a model of a mass 3) statistical analysis of gradient-orientation maps.

Naga R. Mudigonda, Rangaraj M. Rangayyan [6] presents a detection of breast masses in mammograms by density slicing and texture flow field analysis. The proposed method introduces an analyzing oriented flow like textural information in mammograms, it also estimates the widths of ribbons that are able to adapt to variations in the size and shape of the mass regions. Features based on flow orientation, the margin of masses are detected as true

mass region or false positive. Texture flow field analysis determines specifically used for computing features that will discriminate masses from false positive.

Nobuyuki Otsu [7] presents a threshold selection method from gray-level histograms. A method to select an optimal threshold from gray level histogram will be a discriminant analysis to maximize the separability of the resultant classes in gray levels. The method which follows were simple based on only zeroth and the first-order cumulative moments of the gray level histogram. Procedure which implements is straight forward extension to multi thresholding problems. Threshold selection is not only standard technique for picture processing but it also applies to an unsupervised decision problems in pattern recognition.

R. Zwiggelaar, T.C. Parr, J.E. Schumm [8] concentrated on the detection of spiculated lesions. A spiculated lesion is characterized by two abnormal patterns one is linear structures and another one is central mass. The statistical model has been developed to describe and detect the abnormal patterns of spiculated lesions. The generic method developed to represent patterns of linear structures, which relies on factor analysis to separate the systematic and random aspects of a class of patterns. For the detection of central mass the local scale orientation signatures are used based on recursive median filtering and principal component analysis (PCA). The techniques that can also detect architectural distortions and abnormal patterns without the central mass or focal regions, it may also does not contain linear structures.

III. DESIGN

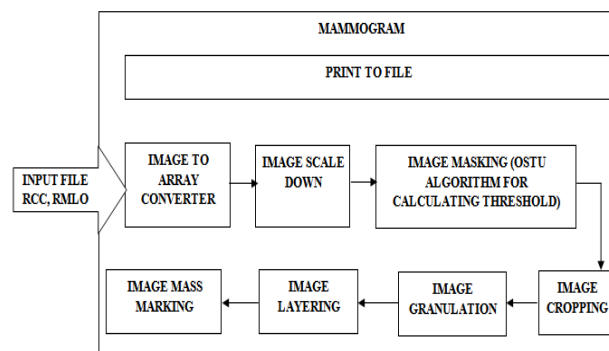


Figure 2: System Architecture

Figure 2 shows an overview of system architecture, where mammogram file contains 8 modules including input file, and each module will describe its own explanation. Then each 8 modules will be called as input file, image to array converter, image scale down, image masking, image cropping, image granulation, image layering, and last image mass marking.

- In input file there are two types of image views called CC and MLO. It also contains right, left mammograms images and both right, left images are passed to input operations.

- Next module is image to array converter this step is whatever the pixels value in the image are taken into consideration and converted into array then it will be stored in a file
- In scale down process image is of 16bit so it will be converted to 8 bit image.
- Binary masking and histogram analysis is done to calculate the threshold value of an image. For calculating threshold value, system uses an algorithm called Otsu method.
- Image cropping is done for removal of background region from breast region.
- Image granulation is done by using neighbor pixels and connectivity rule. Granulation step plays a major role for detecting mass region.
- Image layering is done by taking granulated image as input and different layers of connectivity can be seen in whole mammogram image.
- Final step is Image marking; marking can be done by using rectangular box where the mass region is determined in mammogram image.

IV. MAIN PROGRAM FLOW

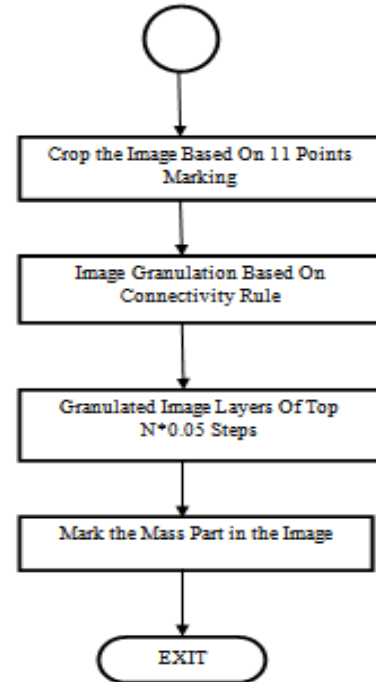
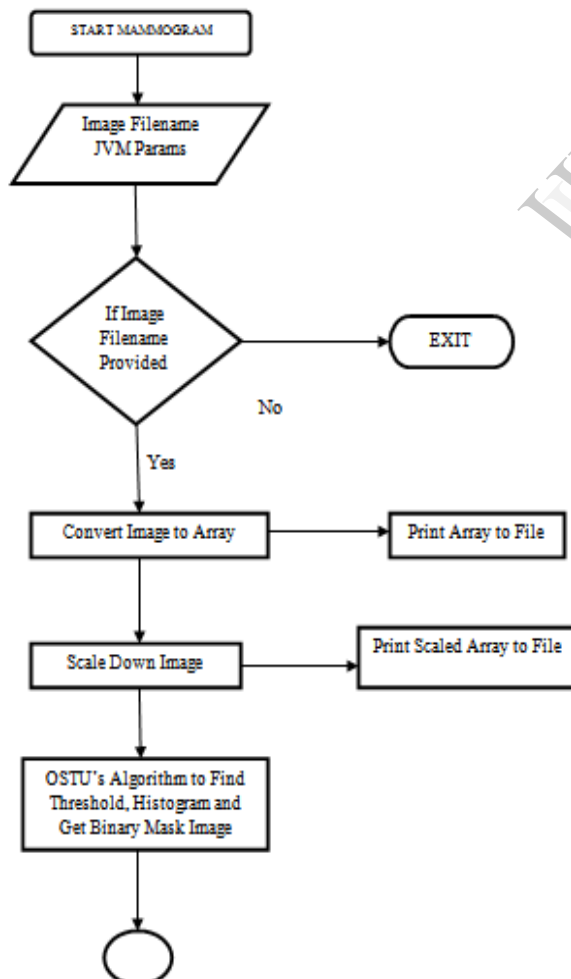


Figure 3: Flowchart of Main Program

Mammogram.java is the main java file which processes the mammogram inputs from the input file such as RCC, RML0, LCC and LMLO, and it invokes various modules such as Image to Array Converter, Print Image to Array, Image Scale Down, Otsu Algorithm, Image Cropping, Image Granulation, Layered Image, and Image Marking.

Initially the input image will be in PNG format and hence it will be huge in size (typically 10 to 25 MB), so the pixel value will be larger. To simplify image in digital image processing the Image should be scaled down by dividing each pixel value by 256, so that each resulting value ranges from 0 to 255.

The scaled down image is given as input to Otsu Algorithm module, that will be able to find out threshold, histogram and to get the binary mask image. Threshold value is used to differentiate the background and the actual breast data area. From the binary mask, the actual data obtained will be merged with new file that contains full of 0's and that will help for the cropping of an image.

After cropping of an image, the image is applied to granulate module and then resultant granulated image is obtained by 8 neighbors, connectivity rule. Then the Granulated image is applied to image layering, image layering will produce layers based on top $N*0.05$ values in a mammogram. Then the conclusive mass part in the image is marked with a bright rectangular box..

V. METHODS

A. Input

In recent new hospitals the screen film mammography is being replaced with digital mammography. With the 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 and for storage purpose. Breast images can be viewed in different angles the most commonly views are Craniocaudal (CC) view and Mediolateral-oblique (MLO) view.

The MLO view is the best opportunity to view an image, and to observe all breast tissue in a single mammogram. It is viewed in superolateral portion of the breast where the most breast cancers are found.

The CC view captures the medial tissue, that the latter view of an image might have been missed. The x-ray beam is directed from the superior portion of the breast to the inferior, or from top to bottom as it implies.

B. Image to Array Converter

Digital mammogram images has thousands of rows and columns, for that the conversion from image to array is needed. First the input image will be taken from the input file, and then input file image will be having rows and columns that cannot be viewed. So initially pixel values of an image will be taken into matrix, that matrix will be equal to size of an image then each row and column will convert to array and stored it in a file.

C. Scale Down

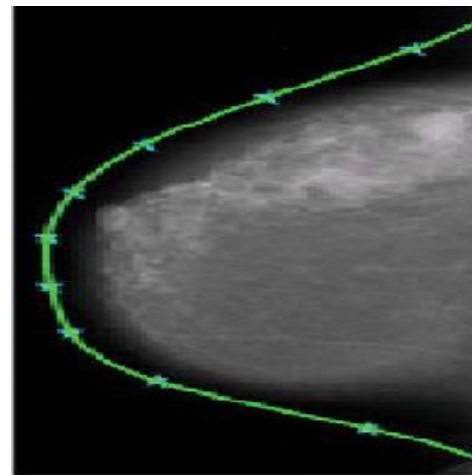
The analysis of image in screening mammograms is challenging for two reasons 1) image size and 2) image complexity. The mammogram images usually range from 20MB to more than 70MB. Mammogram images will be saved in PNG (Portable Network Graphics) extension format. PNG format is a lossless compression image. PNG format was designed to replace the older formats and simple GIF image format. The PNG image format supports up to 48bit true color or 16bit grayscale samples. PNG format will restore or resave an image that will not degrade its quality or information unlike JPEG image. PNG is a raster format; it represents an image as two dimensional arrays (pixels).

In 16bit image the original colors is 65536 and original size is 2728*3920 pixels and image will take around 10MB or more space in the memory. When 16bit image is converted to 8bit image, the 8bit image properties can be viewed as only changing in original colors that from 65536 to 256 and same original size, memory size follows. This transformation will be helpful for further process.

D. Image Masking

Binary image will be having only two possible values for each pixel. A digital image is represented by two dimensional images. Digital image usually may be vector or raster type and most of times digital image refers to raster image or bitmapped images. Binary image also called as bi-level or two-level. Binary mask image are typically obtained by Thresholding from the grayscale image. The pixels in the gray level image are defined as above the threshold are set to 1 (equivalently 255) and rest set to 0.

Before the binary mask image, the mammograms are examined based on the prior knowledge of image intensity, then it determines 11 horizontal sections of equally spaced image. Then each section in the image is visited at fixed intervals of 50 pixels that is from the background to the approaching breast. Segmentation of breast module allocates 11 points from high pixel intensity transition to breast skin region.



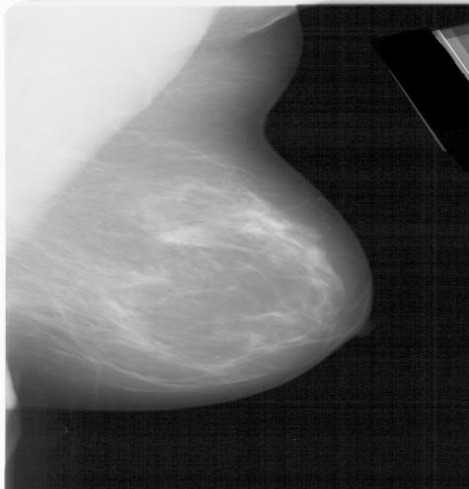
(a)



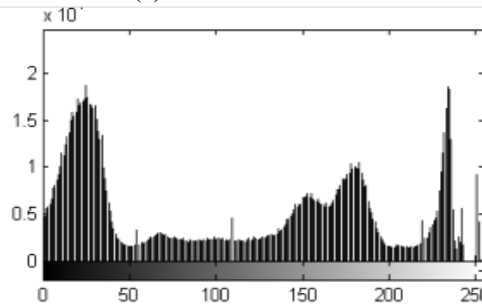
(b)

Figure4: (a) horizontal points in an image and (b) Binary mask image

Threshold calculation of an image can be difficult because some of them considered as 'black art'. Most of the techniques or approaches make use of histogram for the grayscale image.



(a)



(b)

Figure5: (a) The Original Mammogram Image and (b) Its Corresponding Histogram Analysis

Histogram shows probability distribution for each grey level in the image and figure 5 shown is 8-bit image, so the histogram shows up to 256 values. Histogram of an image is computed by recording number of times each pixel intensity is determines then that should be within the valid range of an image.

E. OTSU METHOD

Otsu method [9] is used to automatically perform clustering based image or reduction from grey level image to binary image. The algorithm assumes that threshold containing only bi-modal histogram (that is foreground and background), then the algorithm calculates the optimum threshold separating the bi-modal histogram, so that there should be minimal threshold for the intra-class variance.

The pixels in the image can be represented in gray levels $[1, 2, \dots, L]$. number of pixels at level i is denoted as n_i and total number of pixels is given by $N = n_1 + n_2 + \dots + n_L$. The probability of gray level in an image is:

$$P_i = f_i/N \quad (1)$$

The above equation (1) represents a number of pixels with grey level or total number of pixels. The pixels are divided into 2 classes, C1 pixels with levels $[1, \dots, t]$ and C2 pixels with levels $[t + 1, \dots, L]$. The probability distribution for class occurrence is given by:

$$C1: p_1/w_1(t) \dots p_t/w_1(t) \quad (2)$$

$$C2: p_{t+1} + \frac{1}{w_2(t)} \dots \frac{p_L}{w_2(t)} \quad (3)$$

and also where

$$w_1(t) = \sum p_i \text{ (where } i = 1, 2, 3, \dots, t) \quad (4)$$

$$w_2(t) = \sum p_i \text{ (where } i = t + 1, t + 2, \dots, L) \quad (5)$$

and also the mean for classes are defined as

$$\mu_1 = \sum i p_i/w_1(t) \text{ (where } i = 1, 2, 3, \dots, t) \quad (6)$$

$$\mu_2 = \sum i p_i/w_2(t) \text{ (where } i = t + 1, t + 2, \dots, L) \quad (7)$$

and Otsu method also defines the between-class variance of the threshold image as

$$\sigma_B^2 = w_1(\mu_1 - \mu_T)^2 + w_2(\mu_2 - \mu_T)^2 \quad (8)$$

and finally we calculate within class variance as

$$\sigma_w^2 = W_b \sigma_b^2 + W_f \sigma_f^2 \quad (9)$$

F. Image Cropping

Image cropping is cropping of an image that is removal of unwanted areas. In image cropping the eleven points marked in masking will be helpful for cropping of an image. Breast area part what we get from eleven points in image masking step will be AND operation with newly created background area that is only with value 0. For the image cropping step input file will be scaled down image and output will be cropped image.

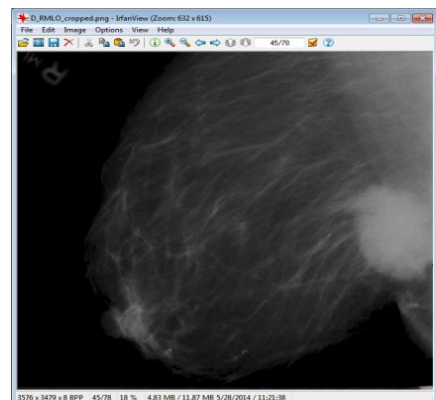


Figure6: Image Cropping

G. Image Granulation

Image granulation step determines that group of pixels that are strongly connected in terms of spatial location and intensity range. In Image granulation step the input file will be the cropped image. Granulation step process involves basic relationship of pixel concepts. In the granulation step system find out maximum value in the image and determines co-ordinates (x,y) positions in the image from that position, then from that position check out pixels are strongly connected and neighboring pixel value is less than 2%. If neighboring pixel value is within 2%, and visited pixel value ranges within that then pixel value is assigned same value to the visited pixel. In granulated image the mass part that is highest group of pixel connected forms cluster that can be viewed brightly in the granulated image.

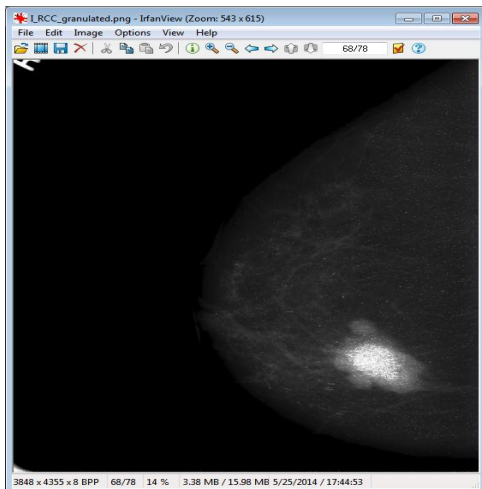


Figure7: image granulation

H. Image Layering

In granulated image the group of pixels that are strongly connected viewed as brighter region like that in image layering step it shows how clusters are formed in each layer. In image layering step top most pixel value will be denoted as N and down to all way will be $0.5*N$, at each layer we can visit at N, $.95*N$, $0.90*N$, $0.85*N$, $0.80*N$, $0.75*N$, $0.70*N$, $0.65*N$, $0.60*N$, $0.55*N$, $0.50*N$ totally there will be 11 levels of layering but actually considering for only top 5 levels only.

I. Image Marking

Granulated image will be major step to detect mass in the mammogram from that granulated image, the image marking is done. Input for marking will be the granulate image and in granulated image the mass part that will be viewed as brighter and that area will be marked. The obtained mass part will be confirmed by some of the radiologists and then it is declared as mass part and marked in the mammograms.

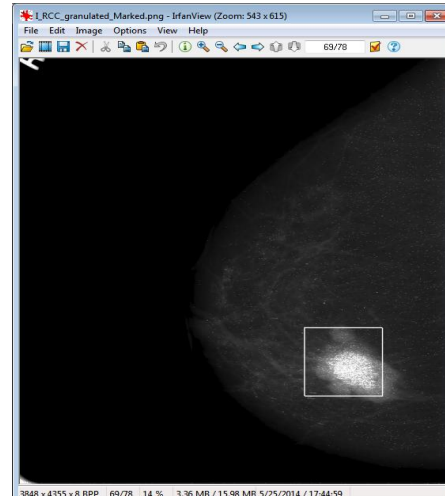


Figure8: Image Marking

VI. CONCLUSION AND FUTURE WORK

The multiple concentric layers analysis for screening mammograms is useful for identifying malignant masses. The DDSM [10] mammograms have two types of images left and right, and also two views CC and MLO. Segmentation is an important step for the mammograms to localize the suspicious regions, providing the objective quantitative assessment and for analysis of anatomical structures. The system is implemented by Java programming and cropping of an image, Granulate image is done automatically. By considering both types of views system can get more accurate results of mammograms. Finally when system evaluated results by performance analysis for 50 mammograms out of those 36 mammograms are achieved, from that the system is getting 72% of satisfied results and that mammograms are confirmed by an expert.

Whenever the mammogram has to be computed it is more sensitive to the light part involved in the mammogram, or some darker part or some pins parts will be seen in mammograms. So sometimes it may not be cropped the image that time some external or manual cropping of an image has to be done. We can improve the accuracy or performance by adopting some new approaches to remove light part and unwanted part automatically.

REFERENCES

- [1] N. H. Eltonsy, G. D. Tourassi, and A. S. Elmaghraby, "A concentric morphology model for the detection of masses in mammography," *IEEE Trans. Med. Imag.*, vol. 26, no. 6, pp. 880–889, Jun. 2007.
- [2] Xinbo Gao, Senior Member, IEEE, Ying Wang, Xuelong Li, Senior Member, IEEE, And Dacheng Tao, Member, IEEE, "On Combining Morphological Component Analysis And Concentric Morphology Model For Mammographic Mass Detection" VOL. 14, NO. 2, MARCH 2010.
- [3] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 510–517, Aug. 1998.

- [4] Hidefumi Kobatake And Yukiyasu Yoshinaga," Detection Of Spicules On Mammogram Based On Skeleton Analysis" VOL ..15, NO..3, JUNE 1996.
- [5] G. M. te Brake and N. Karssemeijer, "Single and multiscale detection of masses in digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 7, pp. 628–639, Jul. 1999.
- [6] N. R. Mudigonda, R. M. Rangayyan, and J. E. Leo Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1215–1227, Dec. 2001
- [7] N. A. Otsu, "Athreshold selection method from gray-level histograms," in *IEEE Trans. Syst., Man, Cybern.*, 1979, vol. SMC-9, pp. 62–66.
- [8] R. Zwiggelaar, T. C. Parr, J. E. Schumm, I. W. Hutt, C. J. Taylor, S. M. Astley, and C. R. M. Boggis, "Model-based detection of spiculated lesions in mammograms," *Med. Imag. Anal.*, vol. 3, pp. 39–62, 1999.
- [9] <http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html>
- [10] <http://marathon.csee.usf.edu/Mammography/Database.html>

IJERT