

A Novel Approach towards Segmentation of Breast Tumors from Screening Mammograms for Efficient Decision Support System

M.Suganthi and M.Madheswaran

Abstract—This paper presents a novel approach to finding a priori interesting regions in mammograms. In order to delineate those regions of interest (ROI's) in mammograms, which appear to be prominent, a topographic representation called the iso-level contour map consisting of iso-level contours at multiple intensity levels and region segmentation based-thresholding have been proposed. The simulation results indicate that the computed boundary gives the detection rate of 99.5% accuracy.

Keywords—Breast Cancer, Mammogram, and Segmentation.

I. INTRODUCTION

BREAST cancer is the leading cause of death [1-3] among all cancers for middle-aged women. Early diagnosis and treatment can increase the chances of survival. Early detection of breast cancer starts with qualitative analysis of mammographic data by physician. Mammography is currently the only proven and cost-effective method to detect breast cancer in its early state of growth. It has been found that a large number of lesions (approximately 25%) of cancers are missed at screening. For this reason, there has been considerable effort aimed at developing computer-aided diagnosis (CAD) systems that might provide a consistent and reproducible second opinion to a radiologist. Currently, most CAD systems are designed to prompt suspicious regions.

The segmentation of the mammographic image is to extract one or more regions of interest (ROIs) from the background after pre-processing. The principle goal of segmentation is to partition an image into homogenous regions (spatially connected groups of pixels called classes, or subsets) with respect to one or more characteristics or features, such that the union of any two neighboring regions yields a heterogeneous.

Accurate segmentation of the breast region is an important prerequisite in the computer-assisted analysis of digital mammograms. The goal is to exclude the background from the further processing. The segmentation algorithm should identify the breast region independently of the digitization system, the orientation of the breast in the image, image

artifacts and other image characteristics.

To segment the breast region, Bick.U et al. classified the image pixels as either potential breast pixels or other pixels [4]. The classification of a pixel was based on the local intensity range and the analysis of a modified histogram, which only included pixels with a small local intensity range. Region growing and morphological filter were used to identify the final breast region. The method was tested on 740 mammograms digitized with a Konica laser scanner and a Fuji optical drum scanner.

Abdel-Mottaleb et al. applied histogram thresholding to segment the interior of the breast [5]. The resulting binary image was then used to locate the remaining breast region near the breast skin line by an edge detection technique. The final breast region is the union of the two regions. The method was tested on more than 500 mammograms digitized with a Lumisys scanner. Meandez et al. detected the breast skin line by an algorithm that computes the gradient in the smoothed version of the mammogram [6]. The image is divided into three parts, which determine the direction of a tracking process that uses the gradient of nine previous pixels to detect the order pixels. The algorithm was tested on 156 mammograms digitized with a Konica laser scanner.

Rezaee et al. [7] used a multiresolution pyramidal segmentation technique. This research indicated methods in which multiresolution pyramidal segmentation could be combined with fuzzy c-means (FCM) to provide a final segmentation of the image. Starting with the top layer (lowest resolution), a pyramidal segmentation algorithm was applied based on both spatial relationships and similarity of parent-child pixels. This resulted in the lowest image (highest resolution) being classified into regions, usually more than desired. These regions were then merged into the desired number of classes by the FCM algorithm. Strickland and Hahn [8] used a wavelet transformation that acted as a bank of matched filters for the detection of microcalcifications. They thresholded the resulting sub-bands to find the microcalcifications and dilated the microcalcifications to enhance them. They then weighted the sub-bands containing microcalcifications higher for the inverse wavelet transform to create a new output image. Their algorithm was tested with 40 mammogram images obtained from the Nijmegen database located at the University Hospital Nijmegen, The Netherlands. These 40 images represented 21 different patients, each containing at least one verified calcification cluster.

Although previous methods have shown some promise, statistical and local approaches currently generate too many false positives to be useful in clinical practice, and they are relatively slow and inaccurate. Hence in this paper, a novel segmentation methods have been proposed for mammograms and investigate its application to the detection of regions of interest (ROIs). In this work, it has been described and compared two algorithms called a topographic representation called the iso-level contour map consisting of iso-level contours at multiple intensity levels and region segmentation based-thresholding for accurate segmentation of the breast region from the background. Morphological filtering is used to define the boundary of the segmented region after thresholding. To finish the initial segmentation, the segmented bright objects in the background are removed.

II. DATA ACQUISITION AND PREPROCESSING

The mammograms can be acquired with dedicated mammographic systems and digitized with a laser film scanner [Lumisys DIS -1000]. In the development of automated mammographic classification system the analysis of tumor's detection depends on the regions of interests, which are usually of low contrast and noisy nature. Hence an image denoising and enhancement may be required to preserve the image quality, highlighting image features and suppressing the noise. These enhancement techniques make very small tumor structures into more visible, and the features of the calcifications can be made more evident hence the radiologist can be aided in determining whether the calcifications are benign or malignant. Non-linear filters such as median filters have been previously used for the detection of microcalcification by doi et al. [1]. In their work the effects of linear and nonlinear filters and their Region of Support (ROS) on the detection and enhancement of microcalcification are investigated. A median filter with 5x5 supports is found to be most effective on mammogram images with simulated microcalcification. The median filter has been used extensively for image noise suppression and smoothing. The main idea is to estimate the mean of the observation window without the effect of outliers. Such a filter is superior to its linear counterparts in removing impulse noises. It preserves monotonic image features, such as edges and ramps. On the other hand, the application of a median filter can introduce spurious artifacts into the transformed image. The filter attenuates image features that are smaller than the observation window in at least one dimension. In the presence of noise, depending on the shape of the transform window and noise levels, the median filter can generate streaks and amorphous blocks.

In order to overcome the above drawbacks, the shock filter [10] has been used for preprocessing. This filter is applied to remove noisy fluctuations and to enhance edges that contain semantically useful information. However, a major drawback in linear scale-space model is the locations of the region edges can not be well preserved. They are usually shifted from their true locations at the coarser level of multi-scale representation. Because the shape, boundary and intensity variation of

abnormal regions are the most important features in breast cancer detection, a meaningful multi-scale description should meet following requirements so that it could be helpful in improving feature extraction and enhancing their discriminability of malignant from benign tumor.

Consider a continuous image $f: \mathbb{R}^2 \rightarrow \mathbb{R}$. Then a class of filtered images $\{I(x,y,t) \mid t \geq 0\}$ of $f(x,y)$ may be created by evolving f under the process

$$I_t = -\text{sign}(\Delta I) |\nabla I| \quad (1)$$

$$I(x, y, 0) = f(x, y) \quad (2)$$

Where subscript denotes partial derivatives, and $\Delta I = (I_x, I_y)^T$ is the (spatial) gradient of I . The initial condition (2) ensures that the process starts at time $t=0$ with the original image $f(x, y)$.

And the above equation can be rewritten as

$$I_t = -\text{sign}(I_{\eta\eta}) |\nabla I| \quad (3)$$

where η is the direction of the gradient.

The main idea of this filter is to embed the original image with a set of low-pass filter, such as a set of Gaussian kernels with different variance where larger variances correspond to images of coarser resolutions. The common approach to increase robustness is to convolve the signal's second derivatives with a low pass filter, such as a Gaussian is given by

$$I_t = -\text{sign}(G_\sigma * I_{\eta\eta}) |\nabla I| \quad (4)$$

Where G_σ is a Gaussian of standard deviation σ .

The filter eliminates the noise, and dislocates the important edge features in the image. Hence, when moving from the finer to coarser scale, the image gets more and more simplified. This complexity can be overcome by the modified shock filter with diffusion operator. The outcome of the new filter is expected to have denoised image with sharpened edges. This is given by

$$I_t = -\text{sign}(G_\sigma * I_{\eta\eta}) |\nabla I| + cI_{\epsilon\epsilon} \quad (5)$$

where C is a positive constant. This filter is capable of providing high performance compared to existing methods. The approach is based on non-linear diffusion, in which the image gradient was used to weight the diffusion process in order to smooth the mammogram images.

III. SEGMENTATION

A. Iso-level Contour Map

An image is considered as a surface in which the intensity at each pixel represents height in three-dimensional space [11]. An ordered set of connected pixels at the same height (intensity) forms an iso-level contour. That is, an iso-contour C for a given grey level l from the image $I(i,j)$ is given by:

$$C(I) = \{(i, j) | I(i, j) = l\}, \quad \forall (i, j) \in \Omega \quad (6)$$

where Ω is a domain of the image I. An iso-level contour is constrained to be a simply closed curve satisfying the conditions of a Jordan curve. A digital image $I(i,j)$ is generally modeled as a discrete function, the value of which represents intensity where an image domain $\Omega \subset \mathbb{N}^2$. In practice, it is hardly possible to draw a continuous curve of a constant intensity in a discrete domain due to the nature of image acquisition such as intensity quantisation and noise intervention. For this reason, bilinear interpolation is used to obtain a continuous image surface on which isolevel contours are well defined. A topographic representation is obtained by an isolevel contour map consisting of a set of iso-level contours at multiple values over the intensity range of an image. The iso-level contour map $CM(I)$ for an image I is given by:

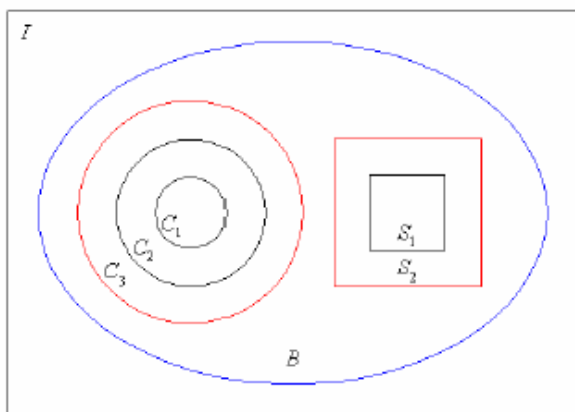
$$CM(I) = \{C(l_i) | l_i \in L, \quad i = 1, 2, \dots, n\}, \quad L = [I_{min}, I_{max}], \quad (7)$$

$$l_i \neq l_j \quad \text{if } i \neq j$$

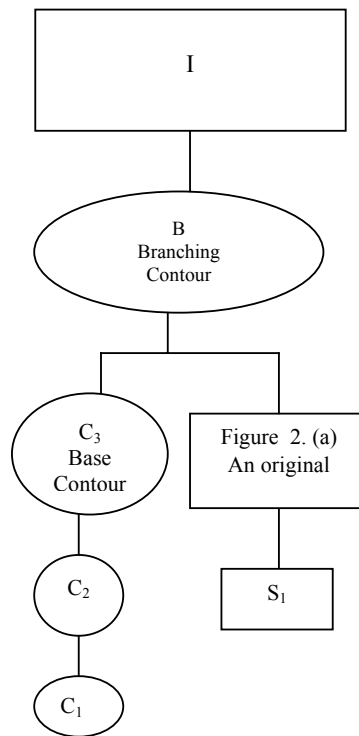
where L is the intensity range of a given image I between minimum intensity I_{min} and maximum intensity I_{max} , and n denotes the quantisation needed to reliably detect features.

B. Inclusion Tree

The iso-level contours that represent distinctive regions with respect to the local background are detected by analysing the iso-level contours generated from the image. The pattern of contours on the iso-level contour map is formed depending on the topological and geometrical properties of objects in the image. A quasi concentric pattern of contours is observed within a dense object, referred to as a salient region. In addition, an abrupt intensity change through a salient region forms a conspicuous dense contour pattern. The topological change of the contour pattern implies a transition from one object to another, or to the background.



(a)



(b)

Fig. 1. (a) An example contour map. (b) Inclusion tree extracted from (a).

The topological and geometrical configuration of the contours in an iso-level contour map can be described by the relationship of enclosure. The enclosure relationship between contours is guaranteed by the Jordan curve theorem. Any iso-level contour C divides an image domain Ω into two disjoint regions, an interior $Int(C)$ and an exterior $Ext(C)$ of the contour C.

The enclosure relationship between two iso-level contours, C_i and C_j , is given by:

$$C_i \subset C_j \quad \text{if } \forall (x, y) \in C_i, \quad (x, y) \in Int(C_j) \quad (8)$$

The enclosure relationship between all the extracted isolevel contours can be represented in the form of a tree structure, called the inclusion tree and shown in Fig 1. The inclusion tree provides a hierarchical representation of the enclosure relationship between contours. In the inclusion tree, each node corresponds to an iso-level contour and the iso-level contour at an ancestral node encloses the iso-level contours at its descendants. In particular, the root node of the inclusion tree represents the boundary of a whole image. A path P in the inclusion tree is an ordered sequence of the nodes defined by:

$$P = (C_1, C_2, \dots, C_n), \quad C_i \in C_j \quad \text{if } i < j \quad (9)$$

where C_i 's are iso-level contours. The path represents a hierarchical sequence of the iso-level contours from the innermost contour to the outermost contour in a nesting

structure. The hierarchical information in the inclusion tree provides an efficient way to examine the topological and geometrical structure of an image.

C. Region based Segmentation

The segmentation by region-based techniques is achieved by identifying all pixels that belong to the object based on the intensity of pixels [12]. They are looking for the regions satisfying a given homogeneity criterion. Since mammographic masses mostly have low contrast and ill-defined edges, it is difficult to determine their boundary with edge-based techniques. Region-based techniques are more suitable for mammograms since suspicious regions are brighter than the surrounding tissues. There are many region-based techniques such as Region growing, Watershed algorithm, and Thresholding. In this paper region segmentation based-thresholding method is also been used for segmentation. Thresholding is based on the image histogram or local statistics such as mean value and standard deviation, or the local gradient. As known, masses appear as bright regions in the mammogram, next step aims to separate those areas from the structured background due to the breast parenchyma. Given an image $I(x,y)$ a single threshold is defined as follow:

$$bin(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{if } f(x, y) < T \end{cases} \quad (10)$$

where $bin(x,y)$ is the resulting binary image and T the threshold value. The finding of the optimal value for T is often one of the most difficult tasks. As already seen, the main purpose of the presented algorithm is to reduce the total area of interest, without losing information about lesions. Here, the most critical issue is to determine the best tradeoff between the reduction of the searching area and the loss of mass information. This threshold algorithm defines a dynamic T value based on mean M and standard deviation σ , estimated within each considered mask. The entire process consist in the translation of a scanning mask over the image, and in the estimation of the statistical values M and σ . At each step, the threshold value T is then calculated as:

$$T = M + \alpha \cdot \sigma + M \cdot k \quad (11)$$

where α and k are two parameters adjustable by the user. In a classic thresholding approach, all the parts inside the breast are considered in the same way (i.e. the values of α and k are unchanged over the entire breast area). However, this could lead to some troubles in the regions close to the borders of the breast. In order to reduce those problems, in the proposed method, it has been considered separately the regions close to the pectoral muscle and those within the external border of the breast. In the first case, it may be noticed that the regions of the breast next to the muscle usually have a brightness distribution quite uniform, with a consequent small σ value. In this situation, a value for T smaller than threshold values of other parts of breast would give better results. A similar

situation is observed on the breast external border. Here, the thickness shrinks very suddenly. As a consequence, a lot of pixels have value over the threshold T . For reducing these problems, a different normalization k value is applied. In particular, the breast is divided into three different parts (external border, border close to the chest and internal breast tissue) processed each one separately. In this way, it is possible to define three different k values to be used inside the breast region, in order to get a more accurate segmentation. The aim of this phase is to reject linear, very small or very large objects, and to reconstruct the correct shape of the objects altered by the local thresholding. In particular, these operations are combinations of erosions and dilations.

IV. RESULT AND DISCUSSIONS

The segmentation algorithm is applied to the detection of regions of interest (ROI's) in mammograms for finding masses. To evaluate the performance of the algorithm, the Digital Database for Screening Mammography (DDSM) was used. For the preprocessing step, an anisotropic diffusion based shock filter was applied to reduce noise and unnecessary small features.

After noise removal and edge enhancement, the isolevel contours were extracted at grey levels that evenly divide the entire range of intensity of an image according to the desired quantisation. This quantisation is related to the size of objects it is desired to segment a finer quantisation is appropriate for more subtle features such as masses. And then a thresholding technique has also been used, in order to identify the bright objects highlighted by the filter. Fig. 2 gives the segmentation results using Iso-level contour map based technique and thresholding. Fig. 2(a) is the original image. Fig. 2(b) is the output of the shockfilter. The boundary selection is shown in the Fig 2(c). The Fig. 2(d) and 2(e) are the output of the Isolevel contour map and thresholding respectively.

The resultant decision on category is validated with the expert and based on the interpretation of the expert, the actual efficiency is estimated. The performance of the proposed method has been evaluated in terms of sensitivity and specificity. Sensitivity (True Positive Fraction) is the probability that a diagnostic test is positive, given that the person has the disease. Specificity (true negative fraction) is the probability that a diagnostic test is negative, given that the person does not have the disease. Overall accuracy is the probability that a diagnostic test is correctly performed. These indices for the various systems have been tabulated in Table.1. The three indices are defined as follows.

$$Sensitivity = \frac{TP}{TP + TN} \quad (12)$$

$$Specificity = \frac{TN}{TP + TN} \quad (13)$$

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN} \quad (14)$$

where TP represents True Positive and TN represent True Negative.

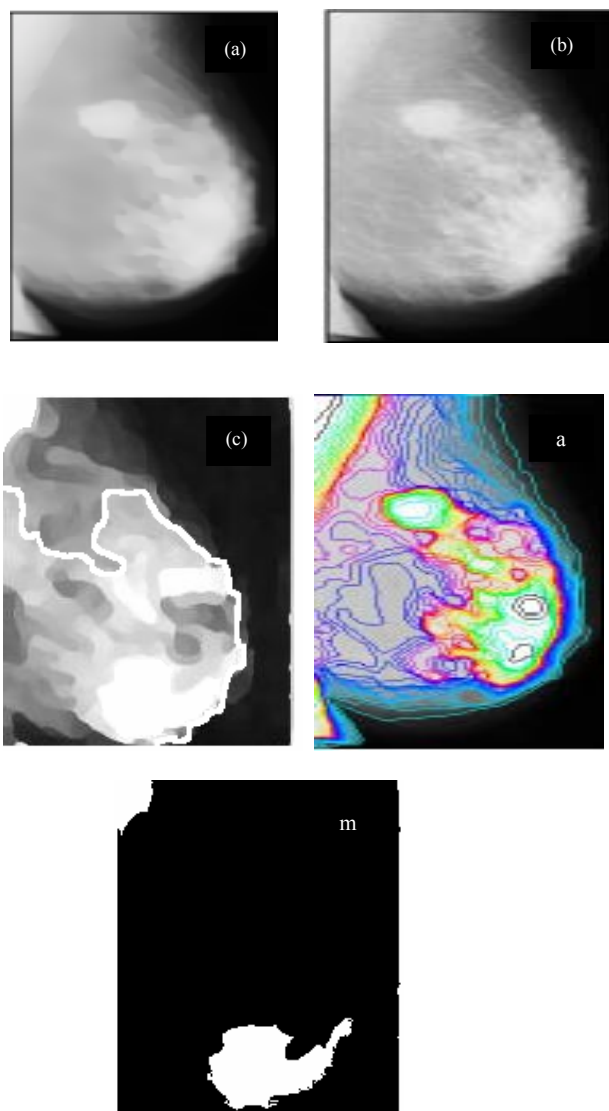


Fig. 2. (a) An original mammogram (b) Denoised Image using shock filter (c) Boundary selection of filtered image (d) Iso-level contour map based on the denoise image (e) Segmented image using thresholding.

Receiver operating characteristic (ROC) curves are frequently used to compare the accuracy of two or more imaging techniques. The effectiveness of the Iso level contours at multiple intensity levels and region segmentation based thresholding algorithm have been compared in ROI detection .It can be seen from the ROC curve presented in Fig .3

The performance of the segmentation algorithms are verified by using area under the receiver operation characteristics (ROC) curve Az and standard error(SE).This is shown in Table 2.From the Table 2.it has been noticed that Region segmentation based thresholding executes with lesser time.

TABLE 1
PERFORMANCE COMPARISON OF VARIOUS METHODS

Segmentation Approach	Sensitivity (%)	Specificity (%)	Overall Accuracy
Manual selection of Regions of Interest	87	66	95.5%
Iso-level contour map consisting of iso-level contours at multiple intensity levels	90	87	96.9%
Region segmentation based-thresholding	96	94	99.5%

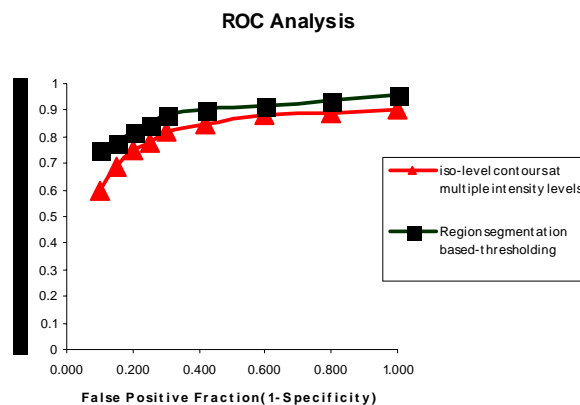


Fig. 3 ROC Analysis of Segmentation Techniques

TABLE II
PERFORMANCE EVALUATION OF SEGMENTATION ALGORITHMS BY MEANS OF
AREA UNDER RECEIVER OPERATING CHARACTERISTICS (ROC) CURVE (AZ)
CORRESPONDING STANDARD ERROR (SE) AND EXECUTION TIME

Segmentation Approach	Az	SE	Time (ms)
Manual selection of Regions of Interest	0.89	0.04	502.02
Iso-level contour map consisting of iso-level contours at multiple intensity levels	0.91	0.03	8.75
Region segmentation based-thresholding	0.99	0.02	0.215

V. CONCLUSION

In this paper automated mass segmentation methods have been proposed. These methods are able to automatically isolate the most suspicious mass regions without any human intervention in the loop. The segmentation algorithm based thresholding has been used to extract features of mammogram images. It is also found that an Automated segmented image can be used to develop Computed Aided Diagnosis (CAD) system for the analysis of breast tumors as malignant or benign.

REFERENCES

- [1] B.Sahiner,N.Petrick,H.P.Chan,L.M.Hadjiishi,C.Paramagul,M.A.Helvie, M.N.Gurcan"Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization," IEEE Tran.Med.Imah.,Vol 20,pp.1275-1284,Dec.2001.
- [2] T. Tweed and S. Miguet. Automatic detection of regions of interest in mammographies based on a combined analysis of texture and histogram", IEEE Computer Society, volume 02, pages 448-452, , 2002..
- [3] Armando Bazzani , Alessandro Bevilacqua, Dante Bollini Rosa Brancaccio and Gianluca Zamboni "Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier" European Symposium on Artificial Neural Networks Brugs, ,pp .195-200 April 2000 .
- [4] Bick U, Giger ML, Schmidt RA, Nishikawa RM, Wolverton DE, Doi K."Automated segmentation of digitized mammograms" Academic Radiology Vol 2 pp.1-9. 1995
- [5] Abdel-Mottaleb M, Carman CS, Hill CR, Vafai S. In: Doi K, Giger ML, "Locating the boundary between the breast skin edge and the background in digitized mammograms Elsevier, . pp. 467-70, 1996.
- [6] Meandez AJ, Tahoces PG, Lado MJ, Souto M, Correa JL, Vidal JJ."Automatic detection of breast border and nipple in digital mammograms"Comput Meth Prog Biomed Vol 49,pp 253-62,1996.
- [7] Rezaee, M.R., van der Zwet, P.M.J., Lelieveldt, B.P.F.,van der Geest, R.J. & Reiber, J.H.C. "A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering" IEEE Transactions on Image Processing, Vol 9, pp. 1238-1248, 2000.
- [8] Strickland, R.N. & Hahn, H.I. "Wavelet transforms for detecting microcalcifications in mammograms". IEEE Transactions on Medical Imaging Vol 15,pp 218-229,1996.
- [9] Karssemeijer N. "Adaptive noise equalization and recognition of microcalcification clusters in mammograms" Int J Pattern Recog. Artificial Intell Vol 7,pp.1357-76, 1993.
- [10] L. Alvarez, L. Mazorra, Signal and image restoration using shock filters and anisotropic diffusion, SIAM Journal on Numerical Analysis 31(2),pp. 590-605,1994
- [11] R. J. Ferrari, A.F. Frere , R.M. Rangayyan ,J.E.L. Desautels ,R. A. Borges ,"Identification of the breast boundary in mammograms using active contour models"Med. Biol. Eng. Comput. Vol 42, pp.201-208, 2004.
- [12] Davies DH, Dance DR. "Automatic computer detection of clustered calcifications in digital mammograms" Phys Med Biol Vol 35:1111-8, 1990



M.Suganthi received Bachelor of Engineering from Government College of Technology Coimbatore in 1998 and received Master of Engineering from K.S.Rangasamy college of Technology, Tiruchengode Tamilnadu. Currently she is working as Assistant Professor, Department of Electronics and Instrumentation Engineering, Muthayammal Engineering, Rasipuram, Tamilnadu. Her area of interest includes Medical image Processing, Neural Network, Fuzzy Logic and Genetic Algorithm.



M.Madheswaran received the BE Degree from Madurai Kamaraj University in 1990, ME Degree from Birla Institute of Technology, Mesra, Ranchi, India in 1992, both in Electronics and Communication Engineering. He obtained his PhD degree in Electronics Engineering from the Institute of Technology, Banaras Hindu University, Varanasi, India, in 1999. At present he is a Principal of Muthayammal Engineering College, Rasipuram, India. He has authored over forty five research publications in international and national journals and conferences. His areas of interest are theoretical modeling and simulation of high-speed semiconductor devices for integrated optoelectronics application, Bio-optics and Bio-signal Processing. He was awarded the Young Scientist Fellowship (YSF) by the State Council for Science and Technology, TamilNadu, in 1994 and Senior Research Fellowship (SRF) by the Council of Scientific and Industrial Research (CSIR), Government of India in 1996. Also he has received YSF from SERC, Department of Science and Technology, Govt. of India. He is named in Marquis Who's Who in Science and engineering in the year 2006. He is a life member of IETE, ISTE and IE (India) and also a senior member of IEEE.