

# Detection of Cancerous Masses for Screening Mammography using DWT based Multiresolution Markov Random Field

Lei Zheng, Andrew K. Chan, Gary McCord\*, Steve Wu<sup>+</sup>, J. Steve Liu<sup>+</sup>

Department of Electrical Engineering

\*Department of Anatomy

<sup>+</sup> Department of Computer Science

Texas A&M University

College Station, Texas 77845

Tel: (409) 845-7458

Email: akchan@wavelet3.tamu.edu

## Abstract

When cancerous masses are embedded in and camouflaged by varying densities of parenchymal tissue structures, they are very difficult to be visually detected from mammograms. This paper presents a novel algorithm based on the *Discrete Wavelet Transform* (DWT) and *Multiresolution Markov Random Field* (MMRF) that will single out the suspicious masses to assist the attending radiologist in making his decisions.

## 1. INTRODUCTION

Breast cancer is one of the leading causes of cancer related mortality among American women. Early diagnosis and treatment play the most important role to increase the chance of survival. Currently, screening mammography is recommended as the most reliable method for early detection of breast cancer. Women over forty years old are recommended to obtain regular mammogram for screening. However, this program will produce a huge number of mammograms to be examined by a limited number of radiologists and mammographers, resulting in misdiagnoses due to human errors caused by visual fatigue. To improve the diagnostic accuracy and efficiency of screening mammography, computer-aided-diagnosis (CAD) has been introduced into screening mammography.

The objective of this research is not to construct a high-tech radiologist. The algorithm is intended to serve as a second opinion for the radiologist. We focus our effort on the detection of cancerous masses from mammograms. Our goal is to locate suspicious regions in the mammogram for more detail examination by the attending physicians. Currently, there is a number of image processing methods proposed for the detection of tumors in mammograms. In [1], Brzakovic and Neskovic described an algorithm that is based on fuzzy pyramid linking and multiresolution segmentation. In [2], Petrick et al. reported a two-stage adaptive density-weighted contrast enhancement (DWCE) algorithm for tumor detection. In this paper, we present a novel algorithm based on the Discrete Wavelet Transform (DWT) and Multiresolution Markov Random Field (MMRF) that can locate the suspicious regions and mark them on the mammogram.

Our algorithm consists of two steps: segmentation and classification. In the first step, a mammogram is segmented using the DWT and MMRF. The procedure is applied only to the low-frequency subband (LL-band) in every resolution. In the initialization of the MMRF segmentation, fractal analysis is first employed to the LL subband image of the lowest resolution to remove most areas that do not contain tumors. The *Dogs & Rabbits Cluster* algorithm is then applied to the preprocessed image to initialize the segmentation. The segmentation procedure at the next resolution is initialized by the result of the previous resolution. The background clutters including image of fatty tissues, veins, etc. are smoothed out by segmentation. The classification step identifies the regions obtained from step one based on six criteria: edge distance variation, mean gradient of region boundaries, mean intensity difference, area size, compactness, and intensity variation. Segments that satisfy parameters set for these

criteria are considered suspicious regions. Our results using data from Mammographic Image Analysis Society (MIAS) indicate the algorithm is effective in detecting the masses with minimal false positives.

In section 2 of this paper, we introduce the segmentation algorithm that is based on wavelet decomposition, Fractal Analysis and Markov Random Field. In section 3, we explain the classification algorithm that uses the *Binary Decision Tree*. Section 4 shows the detection results and section 5 contains our conclusions.

## 2. MAMMOGRAM SEGMENTATION

Because tumor tissues in a mammogram have different textures and gray levels other than the normal ones, it is possible to use image segmentation method to differentiate them from the background. However, image segmentation of cancerous tumors is reported to be very difficult not only due to their high variance of size and texture, but also due to disturbances from the fatty tissues, veins and glands in the background. In order to avoid these disturbances, we introduce a new algorithm for segmentation of the mammograms.

### 2.1 Wavelet Decomposition

*Discrete Wavelet Transform* (DWT) has been proven to be a very effective tool for various signal-processing applications. It is based on multi-resolution analysis (MRA). Let us define  $L = (g(n))$  and  $H = (h(n))$ , for all  $n \in Z$  ( $Z$  is the integer set), where  $g(n)$  and  $h(n)$  are discrete sequences defined by equation (1) and (2), and  $\psi(x)$  is the scale function for wavelet  $\psi(t)$ . The DWT can be implemented using the perfect reconstruction scheme outlined in [7] where  $L$  denotes a lowpass filter and  $H$  denotes a highpass filter.

$$g(n) = \frac{1}{2} [\psi(x/2) + \psi(x-n)], \tag{1}$$

$$h(n) = (1)^n g(1/n). \quad (2)$$

We use the two-dimensional DWT (2D-DWT) to analyze digital images. The easiest way to obtain the two-dimensional wavelets and scaling functions is to use the tensor product of the 1-D wavelet and the 1-D scaling function. Using the ordered tensor product, we obtain three 2-D wavelets and one 2-D scaling function. In 2-D wavelet decomposition, this means the L and H filters are applied to the image in both horizontal and vertical directions. The outputs are subsampled by 2 to produce three highpass subbands HH, LH and HL, and one lowpass subband LL. This process can be repeated on the LL-band to generate the next level of decomposition. More detailed discussion on two-dimensional wavelet analysis can be found in [7].

After a given mammogram is decomposed by DWT, the LL subband image contains only low frequency information of the original image. The HL and LH parts reflect the horizontal and vertical high frequency components in the original image while HH subband image contains the high frequency components in diagonal directions. Thus, the high frequency disturbances from the tissue fibers, glands and veins in the original mammogram will be smoothed out in the LL subband images.

## 2.2 Multiresolution Markov Random Field (MMRF) Segmentation

### 2.2.1 Markov Random Field

*Markov Random Field* (MRF) is used to find the *Maximum a Posteriori* (MAP) estimation for the original image in image segmentations. Let us denote the observed image data as  $Y$  and use a one-dimensional vector to represent  $Y = \{y_i\}$ , where  $i = \{1, 2, \dots, W \times H\}$ ,  $W$  is the width,  $H$  is the height and  $W \times H$  is the size of image  $Y$ . After segmentation, the result is labeled as  $X = \{x_i\}$ ,  $i = \{1, 2, \dots, W \times H\}$ .  $x_i = k$  means the pixel at position  $i$  in the original image  $Y$  belongs to label- $k$  in the segmented image  $X$ .

The set  $k$  equals  $\{1, 2, \dots, K\}$  where  $K$  is the total number of distinct labels. According to the Hammersley-Clifford theorem [8], the probability density of  $X$  is given by a Gibbs density (3):

$$P(X) = \frac{1}{Z} e^{-U(X)} \quad (3)$$

$$Z = \sum_{X \in \mathcal{X}} e^{-U(X)}$$

$$U(X) = \sum_{c \in C} V_c(X)$$

where  $U(X)$  is the energy function that is in turn the summation of all clique potentials  $V_i(X)$  over all possible cliques in  $C$ . In MRF, a clique consists of a set of pixels close to each other in a neighborhood. The potential function  $V_i(X)$  depends on the local configuration of the clique. For the two-pixel clique used in our application, it is defined as:

$$V_2(x_i, x_j) = \begin{cases} \psi_j, & \text{if } x_i = x_j \text{ and } i, j \in C \\ \psi_j, & \text{if } x_i \neq x_j \text{ and } i, j \in C \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where  $\psi_j > 0$  is a constant associated with  $x_j$ .

The potential function for one-pixel cliques is defined as:

$$V_1(x_i) = \psi_k \quad \text{if } x_i = k \text{ and } i \in C, \quad (5)$$

where  $\psi_k$  is a parameter associated with label  $k$ . The energy function of MRF is represented by a summation of cliques grouped by their sizes (equation (6)).

$$U(X) = \sum_{i \in C_1} V_1(x_i) + \sum_{i, j \in C_2} V_2(x_i, x_j) + \sum_{i, j, k \in C_3} V_3(x_i, x_j, x_k) + \dots \quad (6)$$

According to Bayes' theorem

$$p(X | Y) = p(Y | X)p(X), \quad (7)$$

where  $p(X)$  is the *a priori* density of the region, and  $p(Y | X)$  is the conditional density of the observed image given the distribution of the region. For a given image  $Y$  of size  $W \times H$ , the *a posteriori* probability mass function for pixel labels  $X$  is expressed as

$$p(X | Y) = \frac{1}{Z} e^{-U(X|Y)}, \quad (8)$$

where  $Z$  is a normalizing constant. According to equation (3) and (7), the corresponding energy function is

$$U(X|Y) = U(X) + U(Y|X), \quad (9)$$

where  $U(X)$  is given by (6). Assuming that the observed image  $Y$  is the combination of a original image  $X$  plus an independent Gaussian noise, we have  $Y = X + E$ , where  $E = \{e_i, 0 \leq i < W \times L\}$ , and  $e_i \sim N(x_i, \sigma^2)$ .

We have

$$U(Y|X) = \sum_{i=1}^{W \times L} \left[ \frac{1}{2} \ln(\sigma^2) + \frac{(y_i - x_i)^2}{2\sigma^2} \right], \quad (10)$$

$X = \{x_1, x_2, \dots, x_i, \dots, x_{W \times L}\}$  and  $Y = \{y_1, y_2, \dots, y_i, \dots, y_{W \times L}\}$ ,

The objective of the MRF segmentation scheme is to assign a label to each pixel in the original image  $Y$ , such that the energy function  $U(X|Y)$  is minimized and  $X$  is the *MAP* estimation of image  $Y$ . For more details on MRF theories, the reader is referred to [8].

### 2.2.2 Initialization of MMRF in the coarsest level

Initialization is the first step for the MMRF segmentation. It determines how many sets should be there in the final segmentation result and the exact value of each set. It is easy to understand that if the tumor has the same set of values as its background glands, it will be impossible to distinguish the tumor from the background. However, this is a common occurrence because during the initialization procedure, if we consider the whole mammogram as the data input of the cluster algorithm, the data

values from glands, muscles and other normal tissues will disturb the segmentation. In our approach, we first discard all those areas that certainly do not contain tumors using fractal analysis.

In fractal analysis, the model measures the roughness of a certain  $n \times n$  block in an image. The image texture of a block containing tumor is within certain scale. Thus, the roughness can be used to coarsely locate the tumors in a given mammogram. The area of fractal surface can be expressed:

$$A_r \approx K r^{2D}, \quad (11)$$

where  $A_r$  is the area,  $r$  is the elementary ruled area,  $K$  is the scaling constant and  $D$  is the fractal dimension which is used to calculate the roughness of a given region.

To calculate the parameter  $D$ , the *Blanket Method* is chosen. Let's assume the volume of a graph region is  $u$ , and the gray value at  $(i, j)$  is  $p(i, j)$ , then:

$$u_0(i, j) = p(i, j) \quad (12)$$

$$u_r(i, j) = \max_{(k,j) \in [i, i+1] \times [j, j+1]} \{u_{r+1}(i, j) - 1, \max_{(k,j) \in [i, i+1] \times [j, j+1]} \{u_{r+1}(i, j) - 1\}, r \geq 1,$$

The blanket volume is then obtained by

$$V_r = \sum_{(i,j)} [u_r(i, j) - p(i, j)], \quad (13)$$

The blanket surface area at  $r$  is:

$$A_r = V_r + V_{r+1} \quad (14)$$

Finally, from (15),  $D$  can be calculated.

$$\log(A_r) = (2 + D) \log(r) + K', \quad (15)$$

For two dimensional images,  $D$  is within 2~3. The larger the  $D$  is, the rougher the surface.

To apply the fractal algorithm in the MMRF initialization procedure, the input mammogram is first divided into  $16 \times 16$  blocks. For each block, its fractal dimension  $D$  is calculated. The blocks that

have a very smooth surface or a too rough surface are discarded. Only the data within remaining blocks are sent in the initialization procedure for the calculation of data sets.

Usually, the k-means algorithm is chosen to be the initialization procedure for MRF segmentation. However, as reported in [6], the k-means algorithm can not give a consistent segmentation result for the same data set. Since the classification procedure uses parameters extracted from the segmentation result, this will greatly affect the MRF classification result and finally affect the tumor detection procedure. However, the *Dogs & Rabbits Cluster* algorithm does not have this disadvantage.

The *Dogs & Rabbits Cluster* algorithm is an iterative procedure that uses a dynamic process to move cluster centers and positions to the centers of clusters in a data set. It is more stable than k-means algorithms. The algorithm can be described as following:

1. *Initialize K dogs (cluster centers) to random positions.*
2. *Select a random data point (rabbit) in the data set;*
3. *Calculate the distance of each dog from the rabbit and find the closest dog;*
4. *Move the dogs towards the rabbit according to the dynamic and retarding the movement of the dogs.*
5. *If the dog is closest to the rabbit, increase the fatigue of the dog;*
6. *Repeat step 2~5 until a convergence criterion on the dogs has been reached.*

In step 4, the movement of the closest dog is determined by:

$$\dot{d}^i = \frac{2D}{(1+D)^f} (x^i - d^i), \quad (16)$$

and for the other dogs, they are determined by:

$$\dot{d}^i = \frac{D}{A+D} \frac{2D}{(1+D)^f} (x^i - d^i), \quad (17)$$

Here,  $\hat{d}$  and  $\hat{d}'$  are the previous and current positions of the dog respectively.  $D = \|\hat{x} - \hat{d}\|$  is the Euclidean distance between the current dog and rabbit, and  $f?1$  is the fatigue of the dog which determines how much the current dog should move to the appeared data(rabbit).  $A > 0$  in (17) determines the inhibition of the movement for the other dogs except the one closest to the rabbit. After the *Dogs & Rabbits Cluster* initialization, all data will be divided into K sets. Because these sets are extracted from the data corresponding to the tumor and those around the tumor, they simplify the identification of tumor in the segmentation step.

### 2.2.3 Multiresolution Markov Random Field

MMRF is an extension of the normal MRF and has been used for image segmentation applications in [3], [4]. In [3], Bouman and Liu reported that the MMRF segmentation can help the *MAP* estimation achieve the optimum resolution better because the segmentation procedure in finer subbands can be guided by the solutions for coarser ones. However, most of the MMRF techniques are based on down-sampling techniques. Comparing with the normal MMRF techniques and because the LL subband images are smoothed by wavelet filters, high frequency noises are easier to be removed during the segmentation procedure in our algorithm

In our application, we choose the *Iterated Conditional Modes (ICM)* to find the *MAP* estimation for the MRF because of its high efficient performance as reported in [5]. Fig.3 describes the segmentation procedure. An image is first decomposed using DWT and the segmentation starts from the LL subband of the coarsest resolution level. To avoid over-smoothing at this level, we restrict the segment size to be at least 128x128. For the coarsest resolution level, we use the *Dogs & Rabbit Cluster* algorithm and Fractal Analysis for initialization. Then, based on the self-similarity mapping

relations between different levels, we use results from the previous to initialize the current. Assuming that  $X_{k-1}[i][j]$  is the segmentation label value of pixel  $(i, j)$  in the  $(k-1)^{\text{th}}$  level, after initialization, the label value of pixel  $(m, n)$  in level  $k$  can be expressed as  $X_k[m][n] = X_{k-1}[i][j]$  (if  $|m/2 - i| < 1$  and  $|n/2 - j| < 1$ ). After initialization, ICM is used to get the new segmentation result. This is repeated until the finest level.

### 3. TUMOR CLASSIFICATION

#### 3.1 Parameters chosen for classification:

After MMRF segmentation, the mammogram will be segmented into regions with different gray level and features. To locate the regions that are suspicious for tumors, we must select certain features that can be used as criteria. The features selected in our approach are given as follows:

?? *Area*: It equals to the number of pixels within a certain extracted region;

?? *Compactness*: It reflects the shape of the given region and equals to:

$$cmp = \frac{\text{Area of current region}}{\text{Area of rectangle covering current region}}, \quad (18)$$

?? *Mean gradient within current region (mwg)*: This will measure the average gradient of each pixel in the current region.

$$mwg = \frac{1}{N} \sum_{k=1}^N g_k, \quad (19)$$

where  $N$  equals the total number of within pixels within the region and  $g_k$  is the gradient at each pixel  $k$ ;

?? *Mean gradient of region boundary (mg)*: This will measure the sharpness of each region boundary.

$$mg = \frac{1}{N} \sum_{k=1}^N g_k, \quad (20)$$

where  $N$  equals the total number of pixels on the edge of current region;

?? *Gray value variance (var)*: It measures the smoothness of the extracted region.

$$\begin{aligned} \text{var} &= \frac{1}{N} \sum_{(i,j) \in A} (A(i,j) - \bar{A})^2, \\ \bar{A} &= \frac{1}{N} \sum_{(i,j) \in A} A(i,j), \end{aligned} \quad (21)$$

Where  $A(i,j)$  is the gray value of each pixel within region  $A$  and  $N$  is the total number of pixels in the current region.

?? *Edge Distance Variance (edv)*: This measures the shape of the region and its rotational symmetry:

$$\text{edv} = \frac{\frac{1}{N} \sum_{k=1}^N d_k^2}{\bar{d}^2}, \quad (22)$$

where  $d_k$  represents the distance from pixel  $k$  on the edge to the center of the region and  $\bar{d}$  is the mean value of all edge distances.

?? *Mean Intensity Difference (diff)*: This measures the gray value difference between the values inside the region and those outside the region but inside the smallest rectangle cover the region.

$$\text{diff} = \frac{1}{N_a} \sum_{(i,j) \in A} I(i,j) - \frac{1}{N_{c'}} \sum_{(i,j) \in C'} I(i,j) \quad (23)$$

where  $N_a$  is the total number of pixels in region  $A$ ,  $N_{c'}$  is the total number of pixels in region  $C'$ , which represents covered by the rectangular but not inside the region.

All feature values of a given extracted region are input into the binary decision tree described in figure 1. Each node in this tree contains a threshold for the feature it represents and the input data will flow along the arrows according to the result at each node. When the input reaches the end, the

classification result will be made. Comparing with the neural networks employed in many pattern recognition cases, the decision tree is much simpler and computes faster. Due to its high efficiency in computation and implementation, the binary decision tree is becoming increasingly important.

#### 4. EXPERIMENTAL RESULTS

Our experiment is based on the database of Mammographic Image Analysis Society (MIAS) where there are total of 322 mammograms. The total number of well-defined masses and other ill-defined masses in the database are 36.

During the experiment, the fractal dimension used to process the mammograms is between 2.6 to 2.7. The block size during this step is 16? 16. Figure 2(a) gives the example of one of the mammograms with a suspicious tumor in it. Figure 2(b) is the output of the fractal model preprocessing procedure. Figure 2(c) shows the segmentation result of DWT based MMRF scheme. Here, we choose a 3-level DWT decomposition. And figure 2(d) shows the final output of the binary decision tree where the suspicious regions are marked out.

After applying the algorithm on all images in the database, we compute the following statistics:

?? *Average number suspicious regions per image:*

$$\frac{\text{Total number of suspicious regions}}{\text{Total images}}$$

$$? \frac{1298}{322} ? 4.03$$

?? *Number of False Negative per image:*

$$\frac{\text{Total number of undetected positive regions}}{\text{Total number of images}}$$

$$? \frac{1}{322} ? 0.00315$$

?? *Number of false negative per image:*

$$\frac{\text{Total number of false alarms}}{\text{Total number of images}}$$

$$? \frac{(1298 - 36)}{322} ? 3.92$$

?? *Sensitivity:*

$$\frac{\text{Detected True Positives}}{\text{Real number of Positives}}$$

$$? \frac{36}{37} ? 97.30\%$$

## 5. CONCLUSIONS

In this paper, we describe an algorithm that acts as a preprocessor for marking out suspicious regions for tumor in a mammogram. To increase the segmentation accuracy, fractal analysis and the Dogs & Rabbits Cluster algorithm are used in the segmentation initialization procedure while DWT based MMRF segmentation is used for segmentation. During the classification procedure, a binary decision tree is employed. The results proved that the proposed algorithm is very successful and has high detection accuracy with low false alarm rate.

### Reference:

- [1] D. Brzakovic and M. Neskovic: *Mammogram screening using multiresolution-based image segmentation*. Int. J. Pattern Recog. Artificial Intelligence; 7-6, 1437-1460, 1993
- [2] N. Petrick, H. P. Chan, and D. Wei: *An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection*. IEEE Trans. Med. Imaging 15-1; 59-67, 1996
- [3] C. Bouman and B. Liu: *Multiple resolution segmentation of textured images*. IEEE Trans. Pattern Anal. Machine Intel. Vol. 13. No. 2. 99-113, 1991

- [4] H.D. Li, M. Kallergi, L.P. Clarke, V.K. Jain and R. A. Clark: *Markov Random Field for Tumor Detection in Digital Mammography*. IEEE Tran. on Medical Imaging, VOL. 14, NO. 3, Sept. 1995
- [5] R.C. Dubes, A.K. Jain, S.G. Nadabar, and C.C. Chen: *MRF model based algorithms for image segmentation*. Proc. ICCAS, pp. 808-814, 1990
- [6] Patricia McKenzie and Michael Alder: *Initializing the EM Algorithm for use in Gaussian Mixture Modeling*. Pattern Recognition in Practice IV, Vlieland, 1994
- [7] C.K. Chui, *Wavelets: A tutorial in theory and applications*. Academic Press, Inc., 1992
- [8] J. Besag, *Spatial interaction and the statistical analysis of lattice systems*. Journal of the Royal Statistical Society, Series B 36; 192-236, 1974

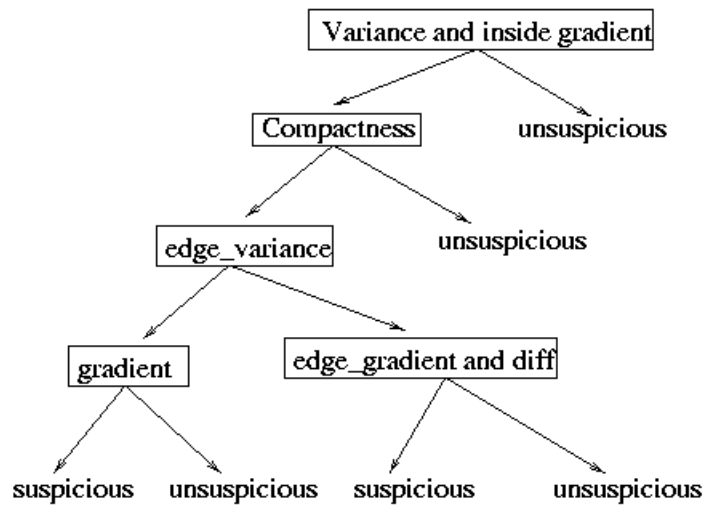


Figure 1. The structure of binary decision tree and the threshold in each node.

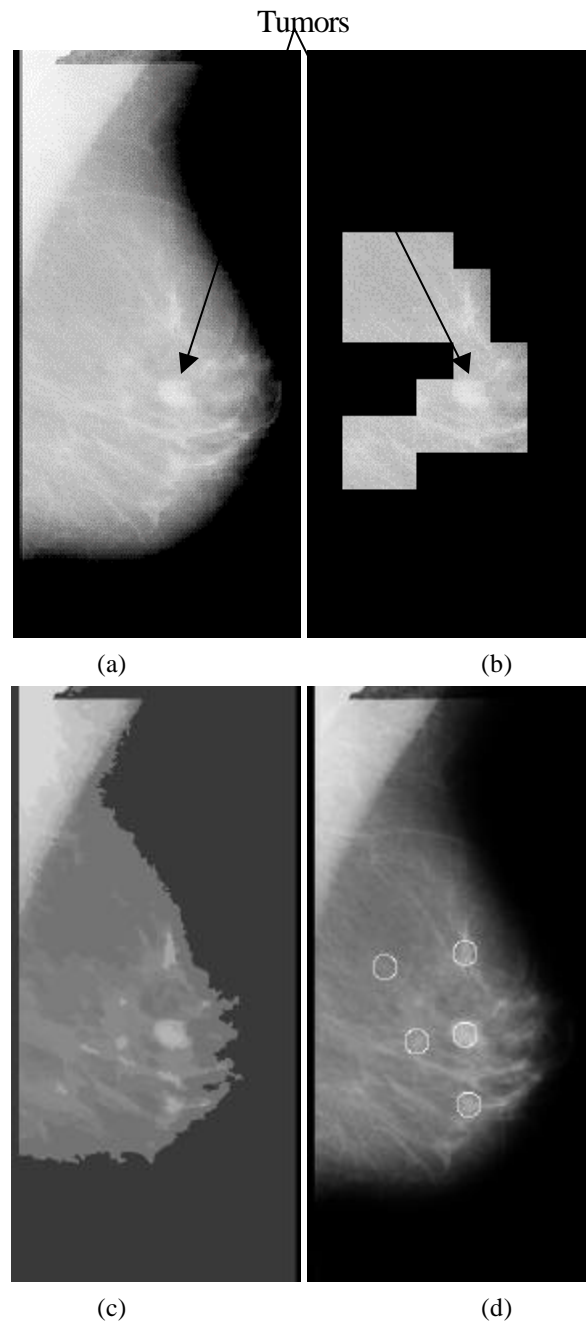


Figure 2. (a) The original image. (b) The output of fractal preprocessing of mammogram with  $D$  from 2.6~2.7. (c) The output of DWT based MMRF. (d) The final output.