

An Artificial Intelligent Algorithm for Tumor Detection in Screening Mammogram

Lei Zheng and Andrew K. Chan*, *Senior Member, IEEE*

Abstract—Cancerous tumor mass is one of the major types of breast cancer. When cancerous masses are embedded in and camouflaged by varying densities of parenchymal tissue structures, they are very difficult to be visually detected on mammograms. This paper presents an algorithm that combines several artificial intelligent techniques with the discrete wavelet transform (DWT) for detection of masses in mammograms. The AI techniques include fractal dimension analysis, multiresolution markov random field, dogs-and-rabbits algorithm, and others. The fractal dimension analysis serves as a preprocessor to determine the approximate locations of the regions suspicious for cancer in the mammogram. The dogs-and-rabbits clustering algorithm is used to initiate the segmentation at the LL subband of a three-level DWT decomposition of the mammogram. A tree-type classification strategy is applied at the end to determine whether a given region is suspicious for cancer. We have verified the algorithm with 322 mammograms in the Mammographic Image Analysis Society Database. The verification results show that the proposed algorithm has a sensitivity of 97.3% and the number of false positive per image is 3.92.

Index Terms—Decision tree, discrete wavelet transform, fractal dimension analysis, Markov random field, tumor detection.

I. INTRODUCTION

BREAST cancer is one of the leading causes of cancer related mortality among American women. Although breast cancer can be fatal, people have the highest chance of survival if physicians can detect the cancer at its early stages. Thus, early diagnosis and treatment play the critical roles in increasing the chance of survival. Screening mammography has been recommended as the most effective method for early detection of breast cancer. Women over 40 years of age are recommended to obtain mammograms regularly for screening. This screening program will generate 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. In order to improve the diagnostic accuracy and efficiency of screening mammography, computer-aided diagnosis (CAD) has been introduced into the screening process.

The objective of this research is not to construct a high-tech radiologist. The algorithm is intended to serve as the first cut in the process and as a second opinion for the radiologists. We

focus our effort on the detection of cancerous masses from mammograms, which is one of the major types of breast cancer. Our goal is to locate suspicious regions in the mammogram for more detailed examination by the attending physicians. Currently, there are several image processing methods proposed for the detection of tumors in mammograms. Various technologies such as fractal analysis [9], multiresolution based image processing [3], [5] and Markov random field (MRF) [11] have been used. In [5], the wavelet analysis is used for image segmentation and microcalcification detection. In [3], Brzakovic and Neskovic described an algorithm for tumor detection from mammogram that is based on fuzzy pyramid linking and multiresolution segmentation. In [16], Petrick *et al.* reported a two-stage adaptive density-weighted contrast enhancement (DWCE) algorithm for tumor detection from mammograms. Although there are various tumor detection algorithms in the literature, the detection rate is still not high. For example, [11] reports a detection rate of 90% and [16] also reports a detection rate around 90% with over ten FPs per image. This is due to the high variance in size and shape of the tumors, and also due to the disturbance (noise) from the fatty tissues, veins and glands.

In order to achieve a high detection rate, we introduce our segmentation algorithm together with other technologies for detecting tumor in mammograms. The algorithm includes three major steps: preprocessing, segmentation, and classification. In the preprocessing step, all suspicious regions are singled out for further analysis. We have observed from many mammograms that the image roughness of the regions containing masses are usually different from that of normal tissue. We use a fractal technique to detect these differences. In the second step, a new segmentation algorithm, the discrete wavelet transform (DWT)-based multiresolution MRF (MMRF) segmentation algorithm is introduced. We have shown that the proposed wavelet-based MMRF segmentation is very efficient in removing image noise introduced by veins and fibers, to generate clean images for subsequent analysis. In the classification step, features are generated from the segmentation step and a binary decision algorithm is used to pinpoint suspicious areas.

In this research, we use mammograms from the Mammographic Image Analysis Society (MIAS) MiniMammographic Database [14] to evaluate the effectiveness of the proposed algorithm. The database contains 322 mammograms including normal, mass, and microcalcification cases. It indicates different classes of abnormalities such as calcification, well-defined circumscribed masses, speculated masses, ill-defined masses, architectural distortion, asymmetry and normal. Each of these abnormalities has been diagnosed and confirmed by a biopsy to indicate its severity: benign or malignant.

Manuscript received August 31, 1999; revised April 24, 2001. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Giger. Asterisk indicates corresponding author.

L. Zheng is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77840 USA.

*A. K. Chan is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77840 USA (e-mail: akchan@ee.tamu.edu).

Publisher Item Identifier S 0278-0062(01)05524-0.

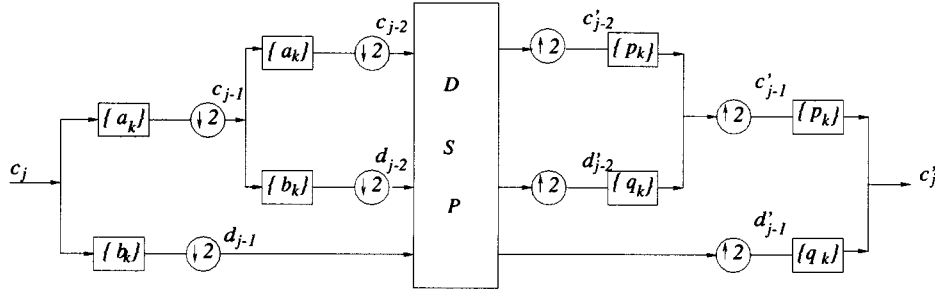


Fig. 1. One-dimensional signal wavelet decomposition and reconstruction.

II. THEORETICAL BACKGROUND

A. Discrete Wavelet Transform

In contrast to the continuous wavelet transform (CWT) which generates a continuous two-dimensional (2-D) function over the time-scale plane via the projection operator

$$Wf(t, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t') \psi\left(\frac{t' - t}{s}\right) dt' \quad (1)$$

the DWT uses the scaling function $\phi(t)$ and wavelet function $\psi(t)$ to separate a signal into components in octave frequency bands. Each of these two functions satisfies a two-scale relation

$$\phi(t) = \sum_k p_k \phi(2t - k) \quad (2)$$

$$\psi(t) = \sum_k q_k \phi(2t - k) \quad (3)$$

where two sequences $\{p_k\}$ and $\{q_k\}$ determine their time-domain representations.

On the other hand, the scaling function can be constructed by a linear combination of these two functions $\{\psi, \phi\}$ at a lower resolution as give by

$$\begin{aligned} \phi(2^j t - l) &= \sum_k a_{l-2k} \phi(2^{j-1} t - k) \\ &+ \sum_k b_{l-2k} \psi(2^{j-1} t - k), \quad l \in Z. \end{aligned} \quad (4)$$

The two pairs of sequences $(\{p_k\}, \{q_k\})$ and $(\{a_k\}, \{b_k\})$ are considered as four digital filters that decompose and reconstruct a signal. In these processes, decimators and interpolators are needed to keep the signal at critical sampling.

A signal $f_j(t)$ at the resolution “ j ” can be represented by a sum of two lower resolution signal $f_{j-1}(t)$ and $g_{j-1}(t)$ as

$$f_j(t) = f_{j-1}(t) + g_{j-1}(t). \quad (5)$$

Each of these functions can be written in terms of ϕ and ψ as

$$\begin{aligned} f_j(t) &= \sum_k c_k^j \phi(2^j t - k) \\ f_{j-1} + g_{j-1} &= \sum_k c_k^{j-1} \phi(2^{j-1} t - k) + \sum_k d_k^{j-1} \psi(2^{j-1} t - k). \end{aligned} \quad (6)$$

Using (4) and the decomposition relation (6) with the orthogonal properties between basis in different resolutions, we have

$$\begin{aligned} c_k^{j-1} &= \sum_l a_{l-2k} c_l^j \\ d_k^{j-1} &= \sum_l b_{l-2k} c_l^j. \end{aligned} \quad (7)$$

This is the DWT decomposition for the signal $f_j(t)$ and the c_k^{j-1} and d_k^{j-1} are coefficients of its lowpass and highpass components respectively. The signal $f_j(t)$ can be recovered from $f_{j-1}(t)$ and $g_{j-1}(t)$ by

$$c_k^j = \sum_l (p_{k-2l} c_l^{j-1} + q_{k-2l} d_l^{j-1}). \quad (8)$$

The process of multilevel wavelet decomposition and reconstruction of one-dimensional (1-D) signal from coefficients c_j is shown in Fig. 1 where a_k and b_k are the coefficients of a series representation as in equation (7). Without the digital signal processing block in the figure, the coefficients at both ends of the system should be the same, i.e., $c'_j = c_j$. This is the perfect reconstruction (PR) condition. To implement the DWT, the computation of c_{j-1} and d_{j-1} in Fig. 1 can be carried out using

$$\begin{aligned} c_k^{j-1} &= \sum_n [c_n^j] \cdot a[2k - n], \\ d_k^{j-1} &= \sum_n [c_n^j] \cdot b[2k - n] \end{aligned} \quad (9)$$

and the reconstruction of c_j can be expressed as

$$c_k^j = \sum_n ([d_n^{j-1}] \cdot q[2k - n] + [c_n^{j-1}] \cdot p[2k - n]). \quad (10)$$

For processing of images, the 1-D DWT is extended to a 2-D DWT. Using tensor product wavelets, one performs the 1-D decomposition process along both x and y axes. A total of four subband images, HH, HL, LH and LL, are generated. Different subband image contains different information and they may be processed individually using different algorithms. The decomposition procedure may be repeated on LL subband image to generate the next lower resolution subband images. Similar to the 1-D case, the reconstruction of 2-D DWT is simply the reverse operation of decomposition. For more information on 2-D DWT, the reader is referred to [4] and [6].

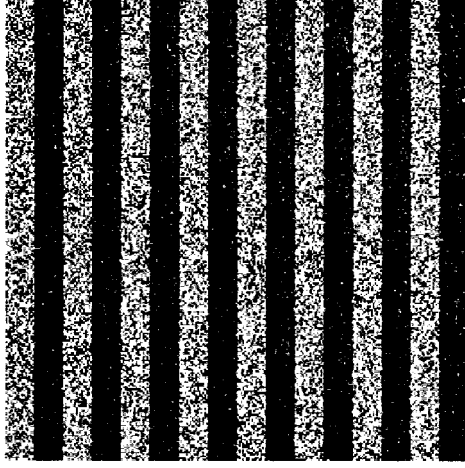
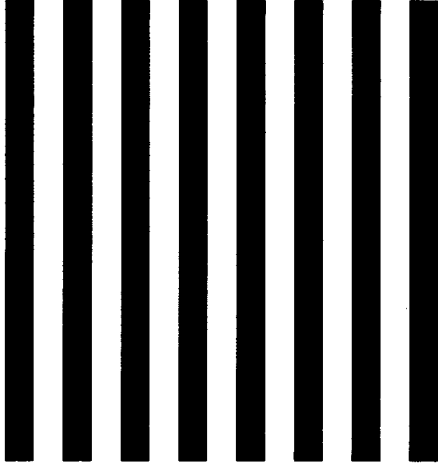

 (a) The observed image Y .

 (b) The original image X .

Fig. 2. The MRF relationship between the observed noisy image and the original unpolluted image.

B. Introduction to Markov Random Field

MRF plays a vital role in image segmentation and noise removal [7], [15]. Given a set of observed image data (original image plus noise), the goal of the MRF is to retrieve the original image from the observed data. To achieve this, the MRF uses the *Maximum a posterior* (MAP) estimation technique. Using the image in Fig. 2 as an example, Fig. 2(a) is the noisy image observed and Fig. 2(b) is its original image. The MRF segments the observed image Fig. 2(a) such that the resultant image has the maximum probability to be the original image Fig. 2(b).

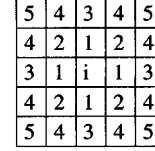
Mathematically, we denote the observed image as Y . The values of image pixel are represented by a 1-D vector $Y = \{y_i\}$, where $i = 1, 2, \dots, W \times H$. W and H are the width and height of image Y . After segmentation, the result is labeled as $X = \{x_i\}$, $i = \{1, 2, \dots, W \times H\}$. In addition, $x_i = k$ denotes the pixel at position i in the original image Y belongs to label k in the segmented image X . The index $k \in 1, 2, \dots, K$ is the index for the labels and K is the total number of distinct labels. If $X_{\setminus x_i}$ denotes all labels of image X except pixel x_i and



(a)



(b)



(c)

Fig. 3. The neighborhood systems defined in MRF. (a) The first-order neighborhood system. (b) The second-order neighborhood system. (c) The fifth-order neighborhood system.

X_{N_i} denotes the neighborhood pixels around pixel x_i , then the MRF theory requires X to have the following two properties:

$$P(X) > 0 \quad (\text{positivity}) \quad (11)$$

$$P(x_i | X_{\setminus x_i}) = P(x_i | X_{N_i}) \quad (\text{Markovianity}). \quad (12)$$

The positivity ensures that all probability is greater than zero and the Markovianity depicts that the label at a given pixel only depends on the neighboring pixels. In other words, the probability of any one specific pixel value is determined only by its neighboring pixels.

1) *Neighborhood Systems and Cliques in Markov Random Field*: One important concept in MRF is the *neighborhood system* and *cliques*. In MRF, a pixel x_i in a given image is related to other pixels via a neighborhood system N_i . Fig. 3 describes the first order, the second-order and the fifth-order neighborhood system for a pixel x_i . The first-order neighborhood system also called the four-neighborhood system where every pixel has four neighbors [Fig. 3(a)]. In the second-order neighborhood system (the eight-neighborhood system) has eight neighbors for each pixels [Fig. 3(b)]. Fig. 3(c) is a fifth-order neighborhood system. In this figure, the largest integer for n indicates the outermost neighboring pixels in the n th order of the neighborhood system.

A clique of a given neighborhood system is a subset of pixels within that neighborhood system. The single-pixel clique consists of one pixel x_i at the center of the neighborhood system in Fig. 3. A two-pixel clique has the center pixel x_i with one other pixel in the neighborhood system. A three-pixel clique consists of the center pixel x_i and two other pixels in the neighborhood system. These cliques denoted by C_1, C_2, C_3

$$C_1 = \{x_i | x_i \in X\} \quad (13)$$

$$C_2 = \{\{x_i, x_{i'}\} | x_{i'} \in X_{N_i}, x_i \in X\} \quad (14)$$

$$C_3 = \{\{x_i, x_{i'}, x_{i''}\} | x_{i'}, x_{i''} \in X_{N_i}, x_i \in X\} \quad (15)$$

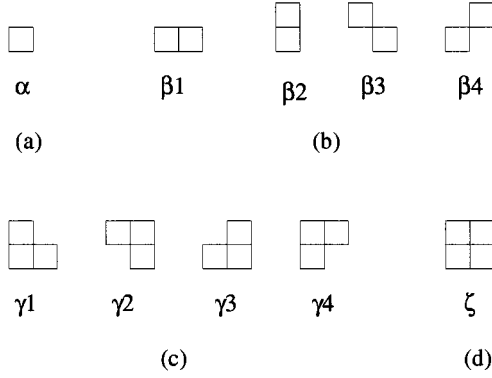


Fig. 4. The cliques for second-order neighborhood systems. (a) Single-pixel clique. (b) Double-pixel cliques. (c) Triple-pixel cliques. (d) Quad-pixel clique.

where

- X set of all pixels in the image;
- x_i center pixel of a given neighborhood system N_i ;
- $x_{i'}, x_{i''}$ pixels within that system;
- X_{N_i} set of all pixels in the neighborhood system.

Fig. 4 describes several clique systems.

Since a higher order neighborhood system introduces much higher computational load, we only use a second-order neighborhood system in this paper as described in Fig. 3(b). Our experimental results show that this neighborhood system is adequate for expressing the relationships between the pixels describing the suspicious masses. The sizes of cliques is up to two-pixel that include single-pixel clique and double-pixel cliques already shown in Fig. 4.

2) *Markov Random Field Theory*: According to the Hammersley–Clifford theorem [1], [8], the probability of X is given by a Gibbs distribution expressible by the following equations:

$$\begin{aligned} P(X) &= Z^{-1} \times e^{-(1/T)U(X)} \\ Z &= \sum_{X \in \mathbf{R}_X} e^{-(1/T)U(X)} \\ U(X) &= \sum_{c \in C} V_c(X) \end{aligned} \quad (16)$$

where Z is a normalization constant called *partition function* that ensures the $P(X)$ is always less than or equal to unity. T is a constant set to unity by default, and $U(X)$ is called the *energy function* that equals the summation of all clique potentials $V_c(X)$ where c is an integer over all possible cliques C .

The energy function of MRF is represented by a summation of all cliques potential functions grouped by their sizes

$$\begin{aligned} U(X) &= \sum_{\{i\} \in X} V_1(x_i) + \sum_{\{i, i'\} \in C_2} V_2(x_i, x_{i'}) \\ &+ \sum_{\{i, i', i''\} \in C_3} V_3(x_i, x_{i'}, x_{i''}) + \dots \end{aligned} \quad (17)$$

The potential function $V_c(X)$, $c \in Z$ depends on the local configuration of the clique. If the second-order neighborhood system is selected and the clique size is limited up to two, only

the first two sums in the above equation need to be considered. For the two-pixel clique used in this paper, it is defined as

$$V_2(x_i, x_{i'}) = \begin{cases} -\beta_{i'}, & \text{if } x_i = x_{i'} \text{ and } i, i' \in C_2 \\ \beta_{i'}, & \text{if } x_i \neq x_{i'} \text{ and } i, i' \in C_2 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where $\beta_{i'} > 0$ is a constant associated with pixel $x_{i'}$.

The potential function of single-pixel clique is defined as

$$V_1(x_i) = \alpha_k \quad \text{if } x_i = k \quad \text{and} \quad \text{pixel } i \in X \quad (19)$$

where α_k is a constant parameter associated with label k . The α_k s are different for a heterogeneous MRF system.

According to Bayes' theorem

$$P(X|Y) \propto P(Y|X)P(X) \quad (20)$$

where $P(X)$ is the *a priori* probability of the image X , and $P(Y|X)$ is the conditional probability of the observed image Y given the true (noise free) image X . According to equation (16), the *a posteriori* probability mass function for X can be expressed as

$$P(X|Y) = Z^{-1} e^{-U(X|Y)}. \quad (21)$$

Following the expression of (20), the corresponding energy function in equation (21) is

$$U(X|Y) = U(X) + U(Y|X) \quad (22)$$

where $U(X)$ is given by equation (17). Assuming that the observed image Y is the sum of the original image X and an independent Gaussian noise, we have $Y = X + \mathbf{e}$ where $\mathbf{e} = \{e_i, 0 < i < W \times H\}$ and $e_i \sim N(x_i, \delta^2)$ represent the additive Gaussian noise with zero mean and variance δ^2 .

In some MRF applications, the value of $\beta_{i'}$ s in Fig. 4 may change with the clique formats. The energy function will have a much larger value when two pixels in that clique have different labels and vice versa. In other words, to reduce the energy, the system inclines to set the same label to pixels along a given direction. Because the image pixel values of cancerous masses in a mammogram have no directional feature, the $\beta_{i'}$ s in Fig. 4 have the same value in our second-order neighborhood system.

III. ALGORITHM DEVELOPMENT FOR TUMOR DETECTION IN MAMMOGRAMS

Fig. 5 explains the algorithm for the detection of regions in a mammogram suspected for tumor. The algorithm consists of three stages: fractal analysis, image segmentation and classification of suspicious regions. The fractal analysis is a preprocessing procedure. In this stage, regions that do not contain tumors in the coarsest (lowest resolution in the decomposition process) LL DWT subband image will be removed. In the segmentation stage, a mammogram is segmented using the DWT-based MMRF segmentation. This procedure is applied only to the preprocessed mammograms. The dogs and rabbits clustering algorithm [13] is applied to this preprocessed subband image to initialize the segmentation. Initialization for segmentation at the next higher resolution comes from the segmented image at

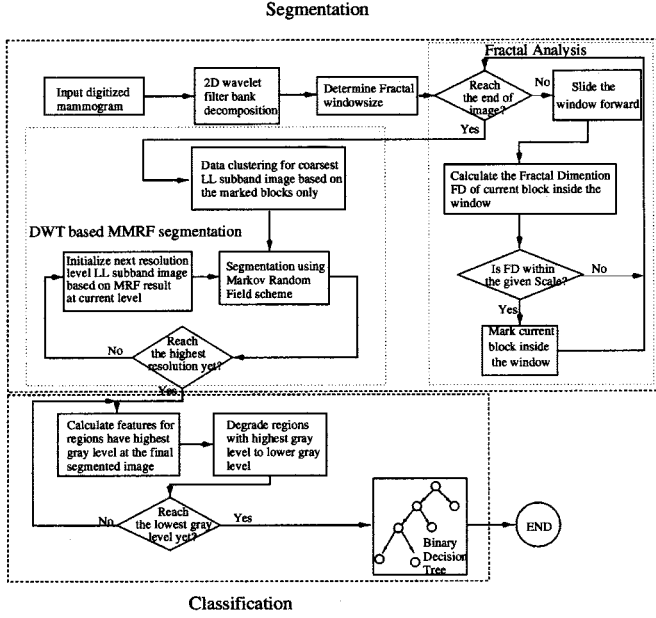


Fig. 5. The flow chart of Tumor Detection algorithm.

the current resolution. The classification stage uses six criteria to identify regions that are suspicious for tumor. The six criteria include: 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.

Since those regions in a mammogram corresponding to tumor tissues have different texture patterns and gray levels than the normal ones, it is possible to use conventional image segmentation methods to select these regions. However, image segmentation on a mammogram that contains cancerous tumors is reported to be very difficult because of large variance of size and texture and noise disturbances. We apply the DWT-based MMRF segmentation algorithm to solve the over-segmentation problem. In our approach, we first discard all those areas that certainly do not contain tumors using fractal analysis at the pre-processing stage.

A. Fractal Analysis Stage

For simplicity and efficiency, we subdivide the mammograms into blocks of 16×16 and calculate the fractal dimension of each block. In fractal analysis, the fractal dimension D measures the roughness of a $n \times n$ block. The fractal dimension of an image block containing tumor is computed and discovered to be within a certain range. Hence, the roughness of the pixel value may be used to locate the tumorous region in a given mammogram. The area of fractal surface can be expressed as

$$A_r = K \times r^{2-D} \quad (23)$$

where

- A_r surface area;
- r ruled area;
- K scaling constant;
- D fractal dimension indicating the roughness of a given region.

To compute the parameter D , the Blanket Method [12] is used. Let us assume the volume of an image region is u_r , and the gray value at position (i, j) is $p(i, j)$, then

$$u_0(i, j) = p(i, j) \quad (24)$$

$$u_r = \max \left\{ u_{r-1}(i, j) + 1, \max_{|(k, l) - (i, j)| \leq 1} u_{r-1}(k, l) \right\} \quad (25)$$

$$r \geq 1.$$

The blanket volume is obtained by

$$V_r = \sum_{(i, j)} [u_r(i, j) - p(i, j)]. \quad (26)$$

The blanket surface area at r is

$$A_r = V_r - V_{r-1}. \quad (27)$$

Finally, D can be computed from

$$\log(A_r) = (2 - D) \log(r) + K'. \quad (28)$$

For a curve, the value scale for D is between $1 \sim 2$. For a surface, D is between $2 \sim 3$. The larger the D is, the rougher the surface. For all the subdivided blocks of a mammogram, the blocks that have a very smooth surface or a very rough surface are discarded. Only the data within remaining blocks are used in the initialization procedure for the clustering.

B. DWT-Based MMRF Segmentation

The MMRF segmentation technique has been used for image segmentation by several investigators, and proven to be fast and robust [2]. By taking a different approach than the conventional algorithms that depend on decimation (downsampling) to acquire images at different resolutions, we use the DWT in our MMRF algorithm. By taking only the LL subband images for MRF segmentation, we show that our algorithm can effectively merge those unnecessary small regions and produce only the sizable regions for subsequent processing. Because only the LL subband is used in our algorithm, this algorithm can be extended to other lowpass filters (i.e., not necessary wavelet lowpass filters) and down-sampling algorithm to generate the segmented image at low resolution.

1) *Model of the Algorithm:* In MMRF, the input image is decomposed into subimages of different resolution levels that are expressed as $\{Y_0, Y_1, \dots, Y_i, \dots, Y_k\}$, where Y_0 is the original image and Y_k is a subimage at lower resolution level and its size is $1/2^k \times 1/2^k$ of the original image Y_0 . The MRF is performed only on the LL subimage at each resolution to achieve segmentation. The energy function $U_j(X)$ for resolution “ j ” in the MMRF can be expressed as

$$U_j(X) = \sum_{\{i\} \in C_{(j, 1)}} V_{(j, 1)}(x_i) + \sum_{\{i, i'\} \in C_{(j, 2)}} V_{(j, 2)}(x_i, x_{i'}) + \sum_{\{i, i', i''\} \in C_{(j, 3)}} V_{(j, 3)}(x_i, x_{i'}, x_{i''}) + \dots \quad (29)$$

where $C_{(j, \cdot)}$ is cliques defined at resolution j , $V_{(j, \cdot)}$ is the corresponding potential functions for a clique at resolution j .

Since the DWT is a linear transform, Gaussian noise remains as Gaussian. Hence, equation (20) can still be used to calculate $P(X|Y)$.

As mentioned in previous sections, an image may be decomposed to produce the LL subband image containing only low frequency information of the image at that particular resolution. In other words, small regions in the original image that contribute to the high-frequency information in the original image, have been smoothed out. These small regions are more likely to be removed during the MRF segmentation procedure when it is performed on all LL subband images.

2) *Initialization of the Coarsest Level:* A robust dogs and rabbits clustering is used to replace the k -means initialization algorithm for clustering [13]. Since the classification procedure uses parameters extracted from the segmented image, the clustering results will ultimately affect the MRF classification and the tumor detection procedure.

The dogs and rabbits is a stable clustering algorithm. It operates iteratively to place the cluster centers in a data set. 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 from each dog to the rabbit and find the shortest distance (closest dog);
4. Move individual dog toward the rabbit according to the distances between the dog and the rabbit;
5. If the dog is closest to the rabbit, increase the fatigueness of the dog;
6. Repeat step 2 \sim 5 until a convergence criterion on the dogs has been reached.

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

$$\vec{d}' = \vec{d} + \frac{2D}{(1+D)^f} (\vec{x} - \vec{d}) \quad (30)$$

and for the other dogs, they are determined by

$$\vec{d}' = \vec{d} + \frac{D}{A+D} \frac{2D}{(1+D)^f} (\vec{x} - \vec{d}) \quad (31)$$

\vec{d} and \vec{d}' are the previous and current positions of the dog respectively. $D = \|\vec{x} - \vec{d}\|$ is the Euclidean distance between the current dog and rabbit, and $f \geq 1$ is the fatigueness of the dog which determines how much the current dog should move to the appeared data (rabbit). $A > 0$ in (31) determines the inhibition of the movement for the other dogs except the one closest to the rabbit. After the dogs and rabbits clustering 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 stage.

3) *Initialization of the Finer Level:* The initialization between adjacent resolution in MMRF is based on the self-similarity map. That is, assuming that $X_{k-1}(i, j)$ is the segmentation label value of pixel (i, j) in the $(k-1)$ th resolution, 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 $\lfloor m/2 \rfloor = i$, and $\lfloor n/2 \rfloor = j$.

4) *The DWT-Based MMRF Segmentation for Mammogram:* 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 LL subband image size to be at least 128×128 . As introduced in the previous section, the MRF segmentation results are propagated according to the self-similarity mapping relations between different levels until to the finest level. 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. After the segmentation procedure, the analyzed mammogram will be segmented into different regions according to their gray-levels and texture. The features of each region will be generated in the classification step for tumor detection.

C. Tumor Classification

1) *Features for Tumor Detection:* To locate the regions that are suspicious of tumors, we must select certain features that can be used as criteria. The features selected in our approach are given as follows.

- *Area-A:* This parameter is the total number of pixels within a certain extracted region;
- *Compactness-cmp:* This quantity reflects the shape of the given region and equals to

$$cmp = \frac{\text{Area of given region}}{\text{Area of the smallest rectangle circumscribe the given region}} \quad (32)$$

- *Mean gradient within current region-mwg:* This parameter measures the average gradient of each pixel in the given region

$$mwg = \frac{1}{N} \sum_{k=1}^N g_k \quad (33)$$

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

- *Mean gradient of region boundary-mg:* This parameter indicates the sharpness of the region boundary

$$mg = \frac{1}{N'} \sum_{k=1}^N g'_k \quad (34)$$

where N' equals the total number of pixels on the boundary of the given region and g'_k is the gradient along the boundary of the given region;

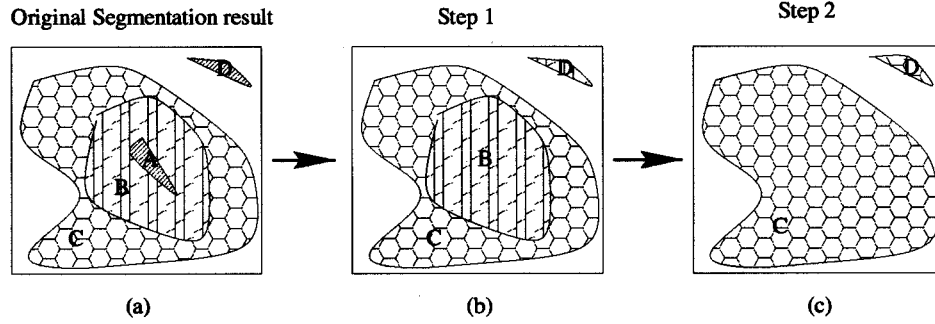


Fig. 6. The region merging procedure for feature generation. (a) The original segmentation result after DWT-based MMRF segmentation. (b) Region A and D are degraded. A merges with B but D is still stand along. (c) Region B and D are degraded again. B merges into C .

- *Gray value variance-var*: var measures the smoothness of the given region

$$\text{var} = \left(\frac{1}{N} \sum_{(i,j) \in A} (X(i,j) - \bar{X})^2 \right)^{1/2} \quad (35)$$

$$\bar{X} = \frac{1}{N} \sum_{(i,j) \in A} X(i,j) \quad (36)$$

where $X(i,j)$ is the gray level of each pixel within region A , and N is the total number of pixels in the given region.

- *Edge Distance Variance-edv*: edv measures the shape of the region and its rotational symmetry

$$\text{edv} = \frac{\left(\frac{1}{N} \sum_{k=1}^N (d_k - \bar{d})^2 \right)}{\bar{d}} \quad (37)$$

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 parameter measures the gray value difference between the values inside the region and those outside the region but inside the smallest rectangle cover the region

$$\text{diff} = \frac{1}{N_a} \sum_{(i,j) \in A} X(i,j) - \frac{1}{N_c} \sum_{(i,j) \in C} X(i,j) \quad (38)$$

where N_a is the total number of pixels in region A , N_c is the total number of pixels in a rectangle region C , which represents the region pixels covered by the rectangle but not inside the region A .

2) *Feature Generation and Region Merging for Segmentation*: The features generation procedure can be illustrated by Fig. 6. Fig. 6(a) is the original segmentation result. There are four different regions in this figure. Region A and D have the same gray level value that is the highest. B has a higher gray level value than C . The feature generation procedure starts from the regions with the highest gray level value. In Fig. 6, features from region A and D are generated first by using an iterative algorithm. In the next step, those regions with the highest gray level value will be degraded to the second highest gray level

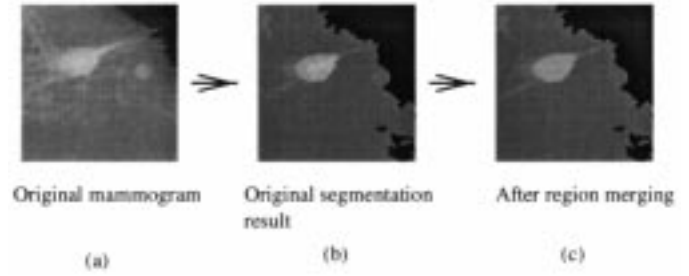


Fig. 7. The region merging result for mammogram segmentation result.

value. That is, gray level value of regions A and D is set to be the same as region B . The result is that these regions will either merged with those region surrounding them [which is like region A in Fig. 6(b)] or will still be standing alone [which is like region D in Fig. 6(b)]. Hence, region B and D will have the highest gray level value in the image after this step. The iterative algorithm is repeated to calculate features for regions with the highest gray level. This generates features for region B in Fig. 6(b). This procedure continues until all regions have merged with the background. In each iteration, the regions with highest gray level will be analyzed by the binary decision tree introduced in next section. If a given region is found to have met the set criteria, it will be singled out as a suspicious region. In the example shown in Fig. 6, this will be the region B .

The above procedure is necessary to generate features suitable for mammogram detection, as illustrated by Fig. 7. The Fig. 7(a) is part of a mammogram with a tumor in it. After segmentation, in Fig. 7(b), the region with highest gray value in the tumor area is very irregular due to background noise. But after region degrading and merging, in Fig. 7(c), the segmented region on the tumor area is round and mimic the shape of a tumor. The features generated from this region are helpful for the detection of the tumor in the original mammogram.

3) *Suspicious Region Classification*: All feature parameters of a given extracted region are input into the binary decision tree described in Fig. 8. Each node in this tree contains a threshold for the feature it represents and the input data will flow along the direction of the arrows based on the result at each node. When the signal reaches the end, a classification result has been made. Comparing with neural networks employed in many pattern recognition cases, the decision tree is much simpler with low computational overhead.

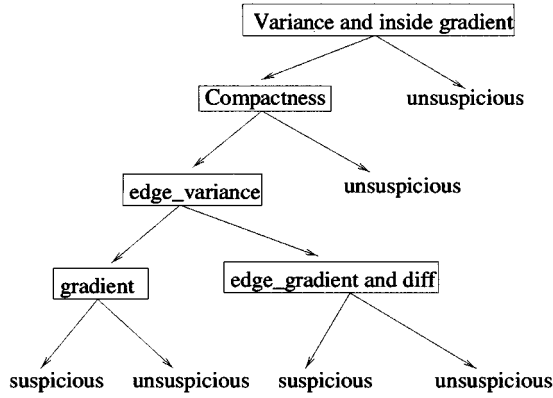


Fig. 8. The structure of binary decision tree and the threshold in each node.

IV. ALGORITHM PERFORMANCE AND CONCLUSION

Our algorithm is verified using a database from the MIAS. This data base contains a total of 322 mammograms. The total number of well-defined masses and other ill-defined masses in the database is 37. During the experiment, the fractal dimension used to preprocess the mammograms is between 2.6 and 2.7 and the block size during this step is 16×16 . Fig. 9(a) gives the example of one of the mammograms with a suspicious tumor in it. Fig. 9(b) is the output of the fractal model preprocessing procedure. Fig. 9(c) shows the segmentation result of DWT-based MMRF scheme. Here, we choose a three-level DWT decomposition. Fig. 9(d) shows the final output of the binary decision tree where the suspicious regions are marked. The parameter selection for each step is very important throughout the experiment. For example, if the range of the fractal dimension is too wide causing a large prelocated suspicious area, extraction of the tumorous region will be difficult. If the features in binary tree have too large a range, the false positive rate will be very high. The range we used in our algorithm are based on statistics of the MIAS database.

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

- Average number suspicious regions per image

$$\frac{\text{Total number of suspicious regions}}{\text{Total number of images}} = \frac{1298}{322} = 4.03. \quad (39)$$

- Number of false negatives per image

$$\frac{\text{Total number of undetected positive regions}}{\text{Total number of images}} = \frac{1}{322} = 0.0315. \quad (40)$$

- Number of false positives per image

$$\frac{\text{Total number of false alarms}}{\text{Total number of images}} = \frac{1298 - 36}{322} = 3.92. \quad (41)$$

- Sensitivity

$$\frac{\text{Detected True Positives}}{\text{Real number of Positives}} = \frac{36}{37} = 97.30\%. \quad (42)$$

The results of this algorithm show that both the well-defined and ill-defined masses are detectable. There is no special

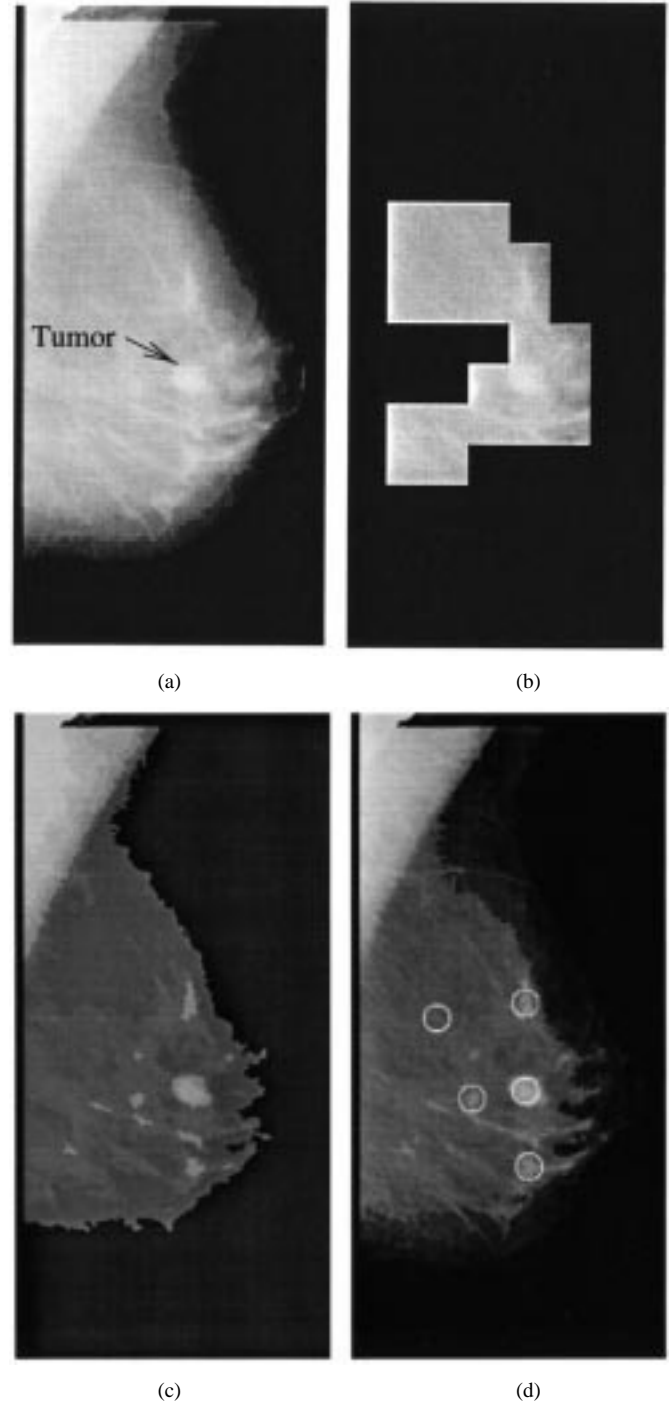


Fig. 9. (a) The original image. (b) The output of fractal preprocessing of mammogram with D from 2.6 ~ 2.7. (c) The output of DWT-based MMRF. (d) The final output.

process needed to detect the ill-defined mass because the DWT-based MMRF segmentation algorithm can effectively segment out tumor region. The one missed by the algorithm is very small with low gray level value that is almost mixed with the background. For future research, more work need to be done to lower the false positive rate. This can be done by putting more features in the binary decision tree in the classification step and/or adjusting the parameter ranges.

In this paper, we have described an algorithm that acts as a preprocessor for marking out suspicious tumor regions in a

mammogram. To increase the segmentation accuracy, fractal analysis, and the dogs and rabbits clustering 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 show that the proposed algorithm is efficient and very successful. The high detection accuracy with low false alarm rate make it a capable radiologist's assistant.

REFERENCES

- [1] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc.*, ser. B 36, pp. 192–236, 1974.
- [2] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 13, pp. 99–113, Feb. 1991.
- [3] D. Brzakovic and M. Neskovic, "Mammogram screening using multiresolution-based image segmentation," *Int. J. Pattern Recog. Artif. Intelligence*, vol. 7, no. 6, pp. 1437–1460, 1993.
- [4] C. K. Chui, *Wavelets: A Tutorial in Theory and Applications*. New York: Academic, 1992.
- [5] C. H. Chen and G. G. Lee, "On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis," *Graphical Models Image Processing*, vol. 59, no. 5, pp. 349–364, 1997.
- [6] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Soc. Ind. Appl. Math., 1992.
- [7] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. PAMI-9, pp. 39–55, Jan. 1987.
- [8] R. Kindermann and J. L. Snell, *Markov random fields and their applications*. Providence, RI: Amer. Math. Soc., 1980.
- [9] F. Lefebvre and H. Benali *et al.*, "A fractal approach to the segmentation of microcalcifications in digital mammograms," *Med. Phys.*, vol. 22, no. 4, pp. 381–390, Apr. 1995.
- [10] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. New York: Springer-Verlag, 1995.
- [11] 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 Trans. Med. Imag.*, vol. 14, pp. 565–576, Sept. 1995.
- [12] B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1977, p. 468.
- [13] P. McKenzie and M. Alder, "Initializing the EM algorithm for use in Gaussian mixture modeling," in *Proc. Pattern Recognition in Practice IV*, Vlieland, The Netherlands, 1994, pp. 91–105.
- [14] Mammographic Image Anal. Soc., Manchester, U.K.. [Online]. Available: <http://s20c.smb.man.ac.uk/services/MISA/MISAc.com.html>.
- [15] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field *a priori* model," *IEEE Trans. Image Processing*, vol. 6, pp. 549–565, Apr. 1997.
- [16] N. Petrick, H. P. Chan, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Trans. Med. Imag.*, vol. 15, pp. 59–67, Feb. 1996.