Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Detecting Microcalcifications in Digital Mammograms using Wavelet Domain Hidden Markov Tree Model

Emma Regentova, Lei Zhang, Jun Zheng, and Gopalkrishna Veni

Abstract— **In this paper we investigate the performance of statistical modeling of digital mammograms by means of wavelet domain Hidden Markov Tree Model (WHMT) for its inclusion to a computer-aided diagnostic prompting system for detecting microcalcification (MC) clusters. The system incorporates: (1) gross-segmentation of mammograms for obtaining the breast region; (2) eliminating the pepper-type noise, (3) block-wise wavelet transform of the breast signal and likelihood calculation; (4) image segmentation; (5) postprocessing for retaining MC clusters. FROC curves are obtained for all MC clusters containing mammograms of mini-MIAS database. 100% of true positive cases are detected by the system at 2.9 false positives per case.**

I. Introduction

Breast cancer is known as a most common form of cancer registered among women in the U.S. It is a second major cause of mortality after the lung cancer [1]. Early diagnosis and treatment play a vital role for survival.

Huge amounts of mammography imagery obtained on the daily basis as a result of the recommended screening program increase the demand in radiologists whose work load under another pressure, i.e., limits of human resources, is highly intensive. Under these conditions, the automated systems are expected to compensate mistakes caused by the radiologists' fatigue. However, the complexity of the breast parenchyma, and a low contrast of mammograms due to the concerns of sustaining a low radiation dose complicate the task of computer vision.

Among other signs of breast cancer, radiologists look for microcalcifications (MCs), - microscopic grains of calcium produced by the cells as the result of some benign or malignant process. Most MCs result from some benign process. The calcifications produced by cancer cells are generally smaller than the benign calcifications and appear in clusters. Automatic detection of MC clusters (MCC) draws the considerable attention of researchers in the fields of image processing and pattern recognition. Both features of MCs and classification methods are studied thoroughly. The results obtained from different databases with a variety of methods indicate that efficiency in detecting clusters varies from 88% to 95% [2]. Because of the critical nature of the application, higher accuracies of detecting all true cases are demanded.

Emma Regentova, Lei Zhang, and Gopalkrishna Veni are with the Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154, USA. phone: 702-895-3187; fax: 702- 895-4075; e-mail: regent@ee.unlv.edu.

Jun Zheng is with the Computer Science Department, Queens College, City University of New York, Flushing, NY 11367, USA. e-mail: junzheng@ieee.org

In this paper, we investigate the efficiency of WHMT modeling [3] for automated detection of MCCs for its inclusion into a computer aided diagnostic prompting system. The method we designed is based on differentiation between edges created by the breast background texture and those generated by MCs. The likelihood values provide a first-hand estimate. Further processing steps allow for detecting all true clusters in digital mammograms of the mini-MIAS database [4] at the rate of 2.9 of false positives per true cluster. In the next section, we introduce the WHMT model, and then in Section III describe the system. Section IV gives the experimental results, and we conclude the paper in Section V.

II. WHMT Model

For the wavelet domain statistical modeling, there are two properties of wavelet decomposition which suggest a reasonable, tractable, and computationally effective statistical model. First, wavelet coefficients are not completely decorrelated that leads to a residual structure seen in the detail subbands. Coefficients also cluster together, that exhibit the same properties in an adjacent neighborhood. Secondly, the magnitudes of the coefficients across the scales display persistency. These properties are summarized in [3] in a model that combines mixture densities and probabilistic graphs. i.e., Hidden Markov Model (HMM). More specifically, the marginal probability of a wavelet coefficient $f(W_i)$ is modeled as a mixture density with a hidden state variable. The Markovian dependencies are introduced between the state variables. The model is graphically represented by a *forest* of trees wherein each tree has a root at the coefficient of the coarsest level; coefficients of the finest scale are the leaves, and the scaling coefficient stands above the root.

Each coefficient is modeled as being in one of two hidden states, i.e., S (*Small*) or L (*Large*), meaning the magnitudes, or energy. Then, two-state mixture model is simply being either in a state of high-variance or low-variance zero-mean Gaussian density. J-level wavelet transform of image I of size $l \times l$ yields a set of coefficients in three subbands $\{W_{LH}, W_{HL}, W_{HH}\}$ for each subtree as

$$W_{LH} = \{\omega_i^j | j = 1, 2, ..., J; i = 4^{J-j}\};$$
$$W_{HL} = \{\omega_i^j | j = 1, 2, ..., J; i = 4^{J-j}\};$$
$$W_{HH} = \{\omega_i^j | j = 1, 2, ..., J; i = 4^{J-j}\};$$

where $J = \log_2 l$, j denotes the scale, and i denotes the ordinal number.

Let ω_i^j belong to either of two hidden states: *Small* or *Large* meaning the magnitudes or energies of coefficients,

TABLE I

DISTRIBUTIONS OF TWO HIDDEN STATES

Small	$\omega_i^j \sim g(\mu_S, \sigma_S^2)$	$f(\omega_i^j \mid state = S) = \frac{e^{\left[\frac{(\omega_i^j - \mu_S)^2}{2\sigma_S^2}\right]}}{\sqrt{2\pi\sigma_S^2}}$
Large	$\omega_i^j \sim g(\mu_L, \sigma_L^2)$	$f(\omega_i^j \mid state = L) = \frac{e^{[\frac{(\omega_i^j - \mu_L)^2}{2\sigma_L^2}]}}{\sqrt{2\pi\sigma_L^2}}$

and each state has a Gaussian probability distribution as shown in Table I.

Given coefficient ω_i^j in scale j, and its parent coefficients is $\omega_{\rho(i)}^{j+1}$ which is in the scale $j+1$, the state transition probabilities between ω_i^j and $\omega_{\rho(i)}^{j+1}$ are $\varepsilon_{i,\rho(i)}^{j,SS}$, $\varepsilon_{i,\rho(i)}^{j,LS}$, $\varepsilon_{i,\rho(i)}^{j,SL}$ and $\varepsilon_{i,\rho(i)}^{j,LL}$, where $\varepsilon_{i,\rho(i)}^{j,SL}$ indicates the conditional probability of the state of ω_i^j is *Small* under the condition that the state of $\omega_{\rho(i)}^{j+1}$ is *Large* and $\varepsilon_{i,\rho(i)}^{j,SS} = p(S_i^j = S | S_{\rho(i)}^{j+1} = L)$. The goal is to derive the parameters Θ^{LH}, Θ^{HL}, Θ^{HH} for the three HMT models from wavelet coefficients in three subbands.

$$\Theta^{HMT} = \left\{ \begin{array}{c} p(S_i^j = S), p(S_i^j = L), \\ \varepsilon_{i,\rho(i)}^{j,SS}, \varepsilon_{i,\rho(i)}^{j,LS}, \varepsilon_{i,\rho(i)}^{j,SL}, \varepsilon_{i,\rho(i)}^{j,LL}, \mu_i^j(S), \sigma_i^j(S), \\ \mu_i^j(L), \sigma_i^j(L) | j = 1, 2, ..., J; i = 1, .., 4^{J-j} \end{array} \right\}$$

where HMT equals to LH, HL, or HH.

III. MC DETECTION ALGORITHM

A. Preprocessing

Before applying the model and processing the digital mammograms for likelihood calculation, we pre-process images for obtaining the breast part of it. Another step is involved for eliminating the dark spots of pepper noise type.

First, contrast limited adaptive histogram equalization (CLAHE) technique is applied to the whole image [5]. Then morphological operations are performed to remove the artifacts such as patient and imaging information labels. Specifically, morphological opening operation is done which is defined as Erosion followed by Dilation using the "disc" shaped structuring element of size 28.

Next, we process the enhanced image to eliminate the pepper type noise, i.e. dark spots of the same size as to MCs, i.e. 2-3 pixels size, that can be accidentally misclassified as being MCs in the course of evaluation of the edge statistics. We do neighborhood comparison and voting (NCV). In NCV, a window of 3×3 pixels slides throughout the image. If the intensity difference of the central pixel and its neighbor is greater than a fixed threshold (for example, 10), the decision about this neighbor is positive. The identification of whether the central pixel is dark is performed by voting among its 8 neighbors. If the number of positive votes is greater than a fixed threshold (for example, 5), the central pixel is considered a dark spot.

B. Maximum Likelihood (ML) Estimation

Modeling and processing are accomplished in the domain of 2-level wavelet transform using db3 mother wavelet,

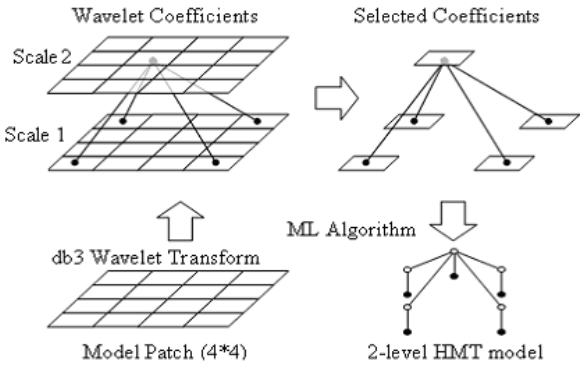

Fig. 1. HMT model setting

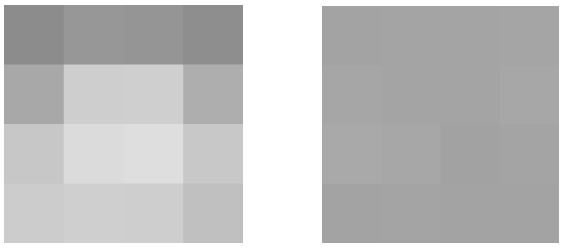

Fig. 2. (a) 4×4 region for MC model; (b) 4×4 region of background model

which is the best fit for capturing the edges created by MCs. HMT model is constructed on the tree structure shown in Fig. 1. Model parameters are derived by the known Baum-Welsch algorithm for two types of objects, i.e. MC containing region of 4×4 pixels and a background region of the same size. They are shown in Fig. 2. The regions are taken from a real mammogram and pertain to the respected objects. It has to be noted that modeling is accomplished with a single training object.

Given wavelet coefficients $\{W^{LH}, W^{HL}, W^{HH}\}$ of an image under observation and a J-scale HMT model $\Theta = \{\Theta^{LH}, \Theta^{HL}, \Theta^{HH}\}$, the likelihood value between the image and the model Θ is $f(W|\Theta) = f(W^{LH}|\Theta^{LH} f(W^{HL}|\Theta^{HL}) f(W^{HH}|\Theta^{HH})$. The steps to calculate the likelihood value for each wavelet coefficient tree are shown in below where the W^{LH} tree is used as an example.

Step 1: Set j = 1, and for each coefficient ω_i^1 compute $f(\omega_i^1|\Theta, S_i^1 = S)$ and $f(\omega_i^1|\Theta, S_i^1 = L)$:

$$\left\{ \begin{array}{l} f(\omega_i^1|\Theta, S_i^1 = S) = p(S_i^1 = S) \frac{e^{\left[-\frac{(\omega_i^1 - \mu_i^1(S))^2}{2(\sigma_i^1(S))^2}\right]}}{\sqrt{2\pi(\sigma_i^1(S))^2}} \\ f(\omega_i^1|\Theta, S_i^1 = L) = p(S_i^1 = L) \frac{e^{\left[-\frac{(\omega_i^1 - \mu_i^1(L))^2}{2(\sigma_i^1(L))^2}\right]}}{\sqrt{2\pi(\sigma_i^1(L))^2}} \end{array} \right.$$

Step 2: Set $j = j+1$, for each coefficient ω_i^j in this scale, compute the conditional likelihoods $f(\omega_i^j|\Theta, S_i^j = S)$ and

$f(\omega_i^j|\Theta, S_i^j = L)$:

$$f(\omega_i^j|\Theta, S_i^j = S) = p(S_i^j = S)\frac{e^{\left[-\frac{(\omega_i^j-\mu_i^j(S))^2}{2(\sigma_i^j(S))^2}\right]}}{\sqrt{2\pi(\sigma_i^j(S))^2}}\times \Pi_{k\in c(\rho(k)=j)}[\varepsilon_{k,\rho(k)}^{j-1,SS} f(\omega_k^{j-1}|\Theta, S_k^{j-1} = S) + \varepsilon_{k,\rho(k)}^{j-1,LS} f(\omega_k^{j-1}|\Theta, S_k^{j-1} = L)]$$

$$f(\omega_i^j|\Theta, S_i^j = L) = p(S_i^j = L)\frac{e^{\left[-\frac{(\omega_i^j-\mu_i^j(L))^2}{2(\sigma_i^j(L))^2}\right]}}{\sqrt{2\pi(\sigma_i^j(L))^2}}\times \Pi_{k\in c(\rho(k)=j)}[\varepsilon_{k,\rho(k)}^{j-1,SL} f(\omega_k^{j-1}|\Theta, S_k^{j-1} = S) + \varepsilon_{k,\rho(k)}^{j-1,LL} f(\omega_k^{j-1}|\Theta, S_k^{j-1} = L)]$$

Step 3: After obtaining the conditional likelihood values $f(\omega_i^j|\Theta, S_i^j = S)$ and $f(\omega_i^j|\Theta, S_i^j = L)$ of every node ω_i^j in the W^{LH} tree, the global likelihood value of the tree is obtained by summing up the logarithm of likelihood values of all nodes in the tree:

$$f(W^{LH}|\Theta^{LH}) = \sum_{j=1}^{J}\sum_{i=1}^{4^{J-j}} \log\left[f(\omega_i^j|\Theta, S_i^j = S) + f(\omega_i^j|\Theta, S_i^j = L)\right]$$

The likelihood values of other two subbands W_{HL} and W_{HH} are obtained in the same way and the global likelihood is given as

$$f(W|\Theta) = f(W^{LH}|\Theta^{LH}) + f(W^{HL}|\Theta^{HL}) + f(W^{HH}|\Theta^{HH})$$

C. Segmentation

The image under observation is processed in a sliding window fashion. That is, a 4×4 block is transformed and two likelihood values per block are calculated. Ideally, ML estimation is expected to provide the estimate of whether the image block exhibits the statistical properties of the model of MC or those of the background model. This way, each block can be labeled as one or another. However, the complexity of the background introduces false estimates. Therefore the weighting of likelihood values is introduced.

MCs are defined are bright small spots on a darker background. However, there is no general intensity threshold of either MC or background. So it is impossible to detect MCs by intensity difference merely. However, there is a specific relation between the background and MCs, i.e. if the background is dim, the intensity of MC must be higher than background. In the contrary, if the background is very bright, intensity of MCs are a bit dimmer. Hence the intensity of the background can be used for weighting of likelihood values.

Each image block under classification is labeled as 1 (MC) or 0 (background) according to the following logic

$$Label = Logic[(L_1 \times \alpha \times \beta) \geq L_2],$$

where L_1 and L_2 are likelihood values of an MC and the background, α is a weighting factor found experimentally, $\beta = f(I_B)$ is a function of I_B - the intensity of the background taken as a first of four median values of the quantized intensities of 4×4 block under processing. In our studies, β is defined as

$$\beta = 275/(275 - I_B) \quad (1)$$

Normalization parameter in Eq. (1) is taken as 275, otherwise in some cases when the intensity of the background is close to the maximum intensity, i.e., 255 then β tends to infinity. That might produce false MCs. The procedure results in a segmented mammogram image.

TABLE II
PARAMETER SETTING FOR PROCESSING STEPS

Area of a single MC = 64	
Eccentricity of a single MC = 6	
Cluster filter	CF2: At least 2 MC in 10×10 window
	CF3: At least 3 MC in 10×10 window
Eccentricity of MC cluster = 6	

D. Postprocessing

The post-processing step includes area, eccentricity, cluster size and the shape analyses. In our experiments this is performed with settings listed in Table II. The settings are derived from a criterion recommended by Kallergi et al. [5] for identifying MC clusters. A group of objects classified as MCs is considered to be a true positive (TP) cluster only if: 1) the objects are connected with the nearest-neighbor distances less than 0.2 cm; and 2) at least three true MCs are detected within an area of 1 cm^2. A group of objects classified as MCs is labeled as an FP cluster provided that the objects satisfy the cluster requirement but contain no true MCs.

One result of mammogram processing by the system is displayed in Fig. 3, where the top-left image is an original mammogram; the right-top image is an enlarged region of MCC; the bottom left is a gross-segmented breast image with the circled location of an MCC; and the right-bottom image is the enlarged region of the detected MCC.

To appreciate the difference between segmentation based on just a maximum likelihood evaluation and the weighted likelihood based labeling followed by postprocessing, we bring in Fig. 4.

IV. EXPERIMENTAL RESULTS

Experiments have been carried out on all images containing MC clusters in MiniMammographic Database of the Mammographic Image Analysis Society (MIAS) and compared to the provided ground-truth data. The best results for all 40 MCC are obtained with CF2 clustering, α = 0.3, i.e. TP = 100% with 2.94 false positive clusters per case. The FROC curves for CF2 and CF3 are shown in Fig. 5. Two mammograms in the database are "hard cases". In the mammogram 'mdb240.pgm', there is an MCC of two single MCs. However, the distance between them is greater than the threshold, i.e., 1 cm2. The difficulty of processing 'mdb226.pgm' is in that the intensity of MCs is almost same to the intensity of the background. Any adjustment for likelihood weighting for the sake of this case would lead to higher FP rates.

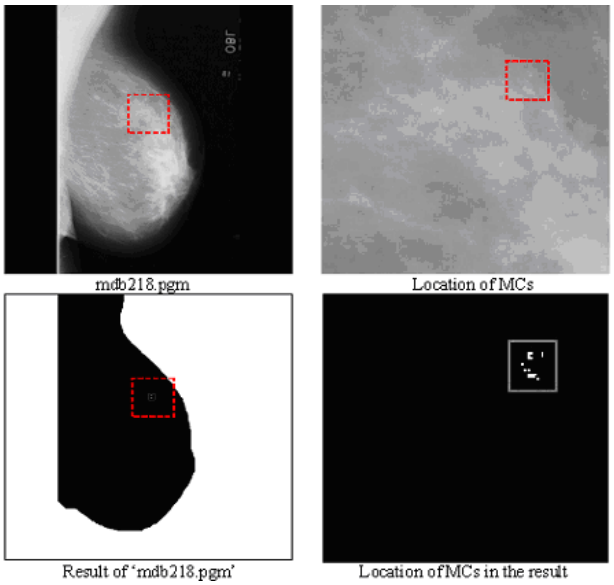

Fig. 3. Processing example

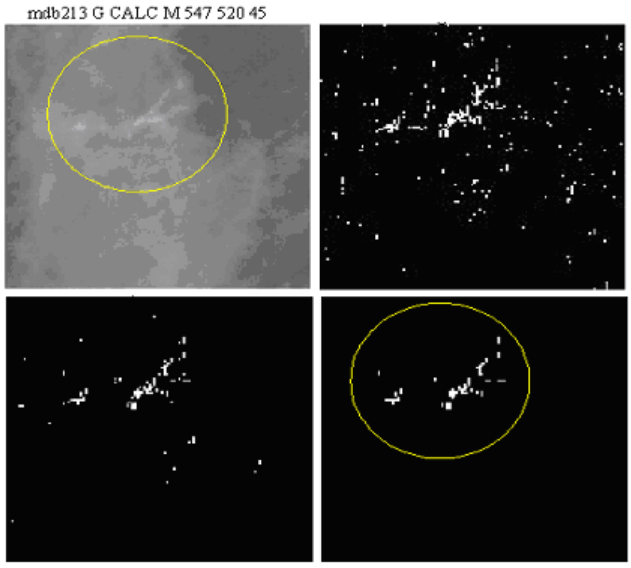

Fig. 4. Top-left: A part of the original image with a circled MCC; Top-right: MLE based segmentation; Bottom-left: Weighted MLE; Bottom-right: final result after postprocessing.

V. Conclusion

We designed a method for MC detection based on a segmentation algorithm that utilizes WHMT modeling. By using db3 filter, a specific tree structure and weighting of likelihood values in MLE method, the capability of the HMT model is greatly promoted for the aim. The method performs well in the framework of the system we have developed. A fair evaluation of the whole system has been performed with mini-MIAS database. The targeted accuracy of true positive cases is 100%, because the system is intended for diagnostic prompting. False positive cases which are kept at as low rates as possible can be further discriminated by radiologists. Thus, the goal of providing radiologists with all true MC clusters is attained.

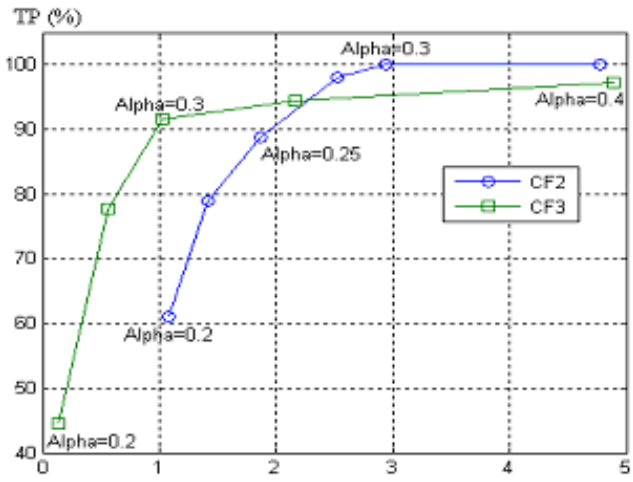

Fig. 5. FROC curves for different cluster filters

VI. Acknowledgements

This paper was supported by American Cancer Society Institutional Research Grant #IRG-103719.

References

[1] S. H. Landis, T. Murray, S. Bolden, and P. A. Wingo, "Cancer Statistics," Cancer J Clin., vol. 48, pp. 6-29, 1998.

[2] H. Schiabel, F. L. S. Nunes, M. C. Escarpinati, and R. H. Benatti, "Performance of a processing scheme for clustered microcalcifications detection with different images database", in Proc. of Annual Conference of IEEE Engineering in Medicine and Biology Society, 2000, pp. 1199-1202.

[3] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet Based Statistical Signal Processing Using Hidden Markov Models," IEEE Trans. Signal Proc., vol. 46, pp. 886-902, 1998.

[4] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis. I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic images analysis society digital mammogram database," Exerpta Medica. International Congress Series, vol. 1069, pp. 375-378, 199), available at http://www.wiau.man.ac.uk/services/MIAS/MIASmi

[5] M. Kallergi, G. Carney, and J. Gaviria, "Evaluating the performance of detection algorithms in digital mammography," Medical Physics, vol. 26, no. 2, pp. 267-275, 1995.