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

Classification of the Stages of Hyperplasia in Breast Ducts by Analyzing Different Depths and Segmentation of Ultrasound Breast Scans into Ductal Areas

Ezgi Taslidere*, Fernand S. Cohen*, and Georgia Georgiou*
*Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, USA
*European Patent Office, The Hague, The Netherlands

Abstract— **In this paper, we study in depth the potential of detection of epithelium hyperplastic growth in the breast ducts leading to early breast cancer detection. Towards that end, we use a stochastic decomposition algorithm of the RF echo into its coherent and diffuse components that yields image parameters related to the structural parameters of the hyperplastic stages of the breast tissue. Previously, we proved that the two parameters, in particular the number of coherent scatterers and the Rayleigh scattering degree show very high ability to discriminate between various stages of hyperplasia even in cases of low resolution and low SNR values. In this paper, the discrimination power of the other parameters is studied further considering different depths using a point scatterer model simulator that mimics epithelium hyperplastic growth in the breast ducts. Significant improvement is obtained in the performance with the newly adopted method considering depth. Values of A_z up to 0.974 are obtained when discriminating between pairs of stages using the parameter residual error variance. In addition, this paper presents a fast nonparametric segmentation procedure to locate the ducts illustrated using phantom data. The performance of the segmentation procedure is obtained as Az>0.948 for various regions of breast scans.**

I. INTRODUCTION

Early breast cancer detection is crucial as the likelihood of survival exceeds 95%. Early changes causing cancer in the breast occur in the form of excessive cell growth (hyperplasia) 90% of which originates in the ductal epithelium [2]. Ultrasonic imaging can be used to classify various stages of hyperplastic evolution in the breast tissue in order to avoid painful procedures performed on patients. In our previous work, we obtained a quite good performance for detection of different stages of hyperplasia formation in the ducts [3]. The RF echo reflected from the breast ducts is simulated as there is no database of ultrasound images available where the cases of hyperplasia of different stages are present. The tissue related parameters are extracted using Continuous Wavelet Transform Decomposition (CWTD) algorithm [4]. The extracted parameters are used for the classification of various stages of hyperplastic evolution in the breast ducts. The work in this paper uses our previous work [3][5], and investigates parameters further considering simulations at different depths. A normalization procedure depending on depth is presented in order to increase the performance of the parameters. Significant improvement is expected for the performance of the parameters. In addition, a segmentation method is presented for detecting and segmenting duct regions in ultrasound breast scans. The segmentation scheme is tested on phantom data that was used by Donohue and was reported in [6]. This segmentation step would be a first step before classification of hyperplastic stages in breast ducts. Results based on simulations and phantom data are presented in the form of Receiver Operating Characteristics (ROC) curves, and the area under the ROC curve is used as the performance metric.

II. SIMULATION OF RF ECHO FROM BREAST DUCTS AT DIFFERENT DEPTHS

In our previous work, we performed simulation of RF Echo from breast ducts mimicking epithelium hyperplastic growth [3]. Now, we perform simulations considering different depths. We numbered the stages of hyperplasia formation inside the breast ducts starting from the normal duct [3]. Stage 1, 2, 3, 4, and 5 correspond to normal duct, hyperplasia, atypical hyperplasia, ductal carcinoma in situ, and infiltrating ductal carcinoma, respectively. In the simulation presenting the reflection of RF echo from breast ducts, the following parameters are considered: the scatterer densities; the scatterer spatial/temporal distributions; the frequency dependent attenuation $\alpha(f)$; the central frequency of the pulse; the SNR (ratio of the average coherent to diffuse energy); the sampling frequency; the propagation velocity of sound in tissue; and the resolution. The values of the parameters used in the current paper are presented in Table 1. 60 different realizations of the reflected RF echo are generated for each hyperplastic stage.

In the simulator, different depths are considered with the intention that the performance of the features will increase with the normalization according to depth. In medical ultrasound, a pulse is emitted into the body and is scattered and reflected by density and propagation velocity perturbations [7]. One fundamental problem in ultrasound imaging is that soft tissue attenuates ultrasound roughly linearly with frequency [8]. A number of mechanisms cause loss of energy in the transmitted ultrasound signal. When sound travels through a medium, its intensity diminishes with distance. The combined effect of scattering and absorption is called attenuation. The intensity decreases

exponentially according to the equation

$$I = I_0 e^{-\alpha(f)x}$$

where I represents the intensity, $\alpha(f)$ is the attenuation coefficient, x is the round-trip distance, and f is the frequency. Attenuation is $-10\log_{10}(I/I_0)$ dB.

TABLE 1. SIMULATOR SPECIFICATIONS

Number of diffuse scatterers/resolution cell outside the cells	Poisson (μ=32)
Number of diffuse scatterers/resolution cell inside the cells	Poisson (μ=48)
Diffuse scatterers spatial distribution	Uniform
Distribution of strength of diffuse scatterers inside the cells	Gamma (σ/μ=0.20)
Distribution of strength of diffuse scatterers inside the duct	Gamma (σ/μ=0.35)
Distribution of strength of diffuse scatterers outside the duct	Gamma (σ/μ=0.40)
Distribution of strength of coherent scatterers	Gamma (σ/μ=0.1)
Pulse center frequency	7.5 MHz
Sampling rate	20 MHz
Propagation velocity	1540 m/s
Resolution (mm)	0.25 – 0.80
SNR (dB)	-10 – 0

The human breast attenuates ultrasound by about 0.5dB/MHz/cm [9]. The attenuation coefficient $\alpha(f)$ is experimentally found to obey a frequency power law function [10]

$$\alpha(f) = \alpha_0 f^{y} \ , \ y \in [0,2]$$

where α_0 (attenuation constant) and y (frequency-power exponent) are empirical media-dependent parameters. For most solid and highly viscous materials, y is close to 2, while for most of human tissues, y ranges from 1 to 1.7. Several researchers tried to estimate the sound speed, attenuation constant and frequency-power exponent for different mediums [11][12][13]. The values of these parameters differ for different tissue types. In the simulations, the speed of sound is taken as 1540m/s, the attenuation coefficient is taken as 0.5dB/MHz/cm, and the frequency power exponent is taken as 1.

III. EXPLORING THE DECOMPOSITION METHOD FOR FEATURES RELATED TO DEPTH

The decomposition technique of RF echo was thoroughly described and tested on simulated RF data [4] and Breast RF data [5]. The method is named as Continuous Wavelet Transform Decomposition (CWTD) [4]. The features related to depth are considered in this paper. These parameters are:

Mean energy of the coherent scatterers, E: This parameter reflects differences in acoustic impedance between the large structures of breast (ducts and lobules i.e., parenchyma) and the surrounding tissue. The differences in the acoustic impedance are analogous to differences between the materials in the two sides of the parenchyma-surrounding tissue boundaries. This parameter turned out to be a good parameter to differentiate between fibrocystic and cancerous tissue for the subclass of tissue with a coherent component [5].

Residual error variance of the diffuse component, σ^2: The equivalent of the energy of the coherent scatterers for the diffuse component of the RF echo is the unconditional variance. The unconditional variance is related to the residual error variance (conditional variance of the diffuse component) through the coefficients of the autoregressive model (AR parameters), the values of the unconditional variance are reflected mainly on the residual error variance. This parameter was shown to possess excellent discrimination ability between benign, normal and malignant tissues [5].

We calculated the two features for each realization. The classification performance of each feature is evaluated separately. The parameter E is found to show no significant discriminative ability since only the duct region is considered for the window size used. σ^2 performed well for discrimination between only early and late stages of hyperplasia [3], however it did not perform well for discrimination between other pairs of stages. Hence, a normalization procedure is presented in order to increase the performance of the features. The results without applying the normalization procedure are given in Table 2.

TABLE 2. AREA UNDER ROC CURVE (A_z) USING FEATURE σ^2 FOR DISCRIMINATION BETWEEN TWO STAGES FOR RESOLUTION 0.25MM AND SNR 0DB

	A_z
Stage1 vs Stage2	0.729
Stage1 vs Stage3	0.837
Stage1 vs Stage4	0.878
Stage1 vs Stage5	0.903
Stage2 vs Stage3	0.557
Stage2 vs Stage4	0.651
Stage2 vs Stage5	0.665
Stage3 vs Stage4	0.613
Stage3 vs Stage5	0.623
Stage4 vs Stage5	0.529

A. Normalization based on depth

The normalization of the diffuse component based on depth is performed in order to improve on performance of the parameter residual error variance. The normalization is done on the diffuse component. The normalization scheme is similar to the one proposed in [6]. Normalization is performed on the diffuse component in each depth prior to finding the residual error variance σ^2.

$$\bar{d}(i,j) = \frac{d(i,j)}{\sqrt{\sum_i |d(i,j)|^2}} \quad \text{for} \quad i = 1,2,...,N \quad j = 1,2,...,M$$

where $d(i,j)$ is the diffuse component, N is the number of scan lines and M is the total number of segments within one scan line.

The area under ROC curve for differentiation of all stages is increased from 0.652 to 0.843 when compared to the one without normalization. A_z values are much better with the normalization scheme. We report on the ability of residual

error variance of the diffuse component to differentiate between the various stages of hyperplasia. This is given in Table 3 in terms the area under the ROC between pairs of stages. It can be concluded that the normalization scheme significantly increased the performance of the parameter residual error variance and it is a strong parameter to differentiate between early and late stages of hyperplasia. The normalization is also performed for different depths, such as 6mm, 10mm, 14mm, 20mm, and 30mm. The A_z value of the feature residual error variance stays in the range 0.8.

TABLE 3. AREA UNDER ROC CURVE (A_Z) USING FEATURE σ^2 FOR DISCRIMINATION BETWEEN TWO STAGES FOR RESOLUTION 0.25MM AND SNR 0DB AFTER NORMALIZATION SCHEME

	A_z
Stage1 vs Stage2	0.741
Stage1 vs Stage3	0.847
Stage1 vs Stage4	0.927
Stage1 vs Stage5	0.974
Stage2 vs Stage3	0.761
Stage2 vs Stage4	0.902
Stage2 vs Stage5	0.970
Stage3 vs Stage4	0.728
Stage3 vs Stage5	0.884
Stage4 vs Stage5	0.723

The normalization scheme is also applied to the coherent component in order to increase the performance of the parameter *E*, mean energy of the coherent scatterers. However, no significant increase in the performance is observed. Since the number of coherent scatterers is much less than the diffuse scatterers, the normalization scheme did not work very well on increasing the performance of the parameter *E*.

IV. SEGMENTATION OF BREAST SCANS INTO DUCTAL AREAS

We have also run preliminary experiments on phantom data to show how we can segment breast scans into ductal areas. We have used the phantom which is reported in [6]. In this experiment, we used a modified version of the hierarchical segmentation algorithm that we have developed and reported in [14] as the basis for detecting the ducts in the phantom scan. The data is divided into non-overlapping windows of size 50x5, i.e., 1.875mm in axial direction and 1.1mm in lateral direction. For each window, detection of coherent scatterers is performed by checking the hypothesis of Rayleigh scattering, which can be achieved by using the non-parametric KS test for color field [15]. If the hypothesis of Rayleigh scattering is accepted, then it is concluded that there is no coherent component in the RF echo, and only the diffuse component is present which means that the region is not a probable duct area. Otherwise, it is taken as a probable duct area.

RF echo *y(n)* is modeled by a 5th order AR as in [15], i.e.,

$$y(n) = \sum_{s=1}^{5} a_s y(n-s) + w(n)$$

The 5 AR parameters a_s, s=1…5 , and the residual error variance σ^2 are estimated using the fast Burg's algorithm described in [19]. In the Rayleigh scattering case, *y(n)* is Gaussian, which implies that the white innovation sequence *w(n)* is also Gaussian. The goodness of fit KS test is performed on the white normalized innovation sequence $(1/\sigma)\, w(n)$, where the innovation sequence is computed by

$$w(n) = y(n) - \sum_{s=1}^{5} a_s y(n-s)$$

The KS statistic denoted by *D* is then calculated by [15]:

$$D = \sup_{-\infty<w<\infty} |P_n(w) - G(w)|$$

where P_n is the empirical distribution of *w(n)* and G is the hypothesized Gaussian distribution. A test at the α significance level is obtained by choosing a number k_α such that

$$p(D_n > k_\alpha \mid H_0) = \alpha$$

and rejecting the hypothesis H_0 if a value of D_n greater than k_α is observed. The false alarm, α , is taken as 0.1.

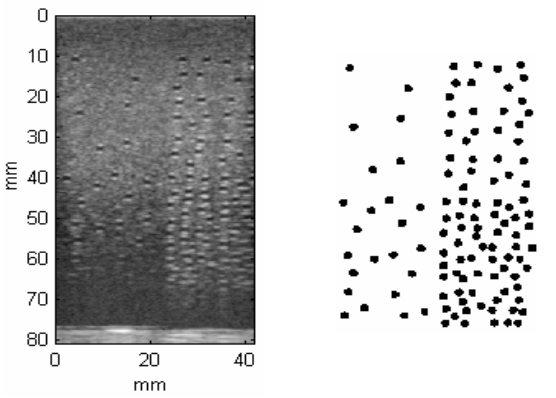
We report the results of segmentation for the four different quadrant images in Figure 1, which have ductal area densities of 0.187 (lower right quadrant), 0.1385 (upper right quadrant), 0.085 (lower left quadrant) and 0.0187 (upper left quadrant). In order to test the performance of our segmentation method, the density of the duct is used as the performance metric. The relative error is defined as

$$\delta d = \frac{|d_0 - d|}{d}\ ,\ 0 < d_0 < 2d$$

where d_0 is the estimated duct density and *d* is the true duct density. The performance of the system is defined as $p=1-\delta d$. This is given for the four sub images in Table 4 in percentage. The performance of the system is in the range 90%-98% for different densities. The result of the segmentation is shown in Figure 2 for a false alarm α=0.1. By and large, we can see that most of the duct areas were properly picked up. This could be followed by a finer segmentation by running the wavelet decomposition algorithm to fine the boundary of the duct. We also report on the A_z values for the four quadrants based on the KS distance used in the segmentation and duct detection algorithm. These values are tabulated in Table 4. The A_z values are in the range 0.948-0.964 which is quite satisfying. The segmentation scheme performed quite well for the phantom data.

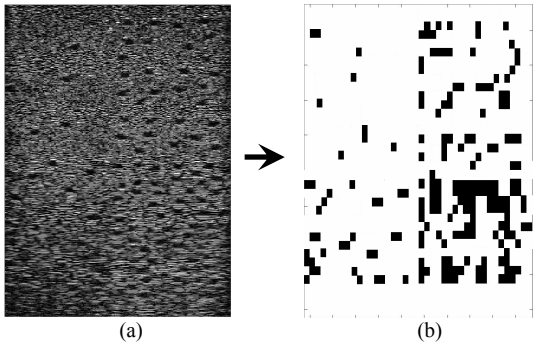
TABLE 4. PERFORMANCE OF THE SYSTEM AND AREA UNDER ROC CURVE (A_Z) FOR DIFFERENT DUCT DENSITIES

Duct Density	Performance (%)	A_z
0.2770	98.0%	0.964
0.1385	94.3%	0.948
0.0831	93.8%	0.951
0.0277	90.0%	0.953



(a) (b)

Fig. 1. (a) Phantom scan density image with the duct size 1.83mm in the transverse direction. (b) Distribution of the ducts in the transverse direction in the phantom.



(a) (b)

Fig. 2. (a) Data to be segmented, (b) Segmentation using a window size of 50x5, i.e., 1.875mm in axial direction and 1.1mm in lateral direction at a false alarm rate of 0.05.

V. CONCLUSION

The ultrasound RF Echo is simulated representing various stages of hyperplasia formation in breast ducts considering different depths. The features related to depth are considered. A normalization scheme according to depth is presented in order to increase the performance of the considered features. Compared to our previous work [3], a significant improvement is observed in the performance of the feature, residual error variance of the diffuse component σ^2. Our study indicates that the classification between various stages of hyperplasic evolution in the breast tissue can be done reliably using the residual error variance σ^2 using the presented normalization scheme. We report on the ability of the parameter to differentiate between the various hyperplasia stages, the A_z values varied from 0.723 to 0.970 for differentiation between different pairs of stages.

We also presented a segmentation scheme in order to extract the ductal areas in breast scans which will serve as an initial step to be applied on breast scans before classification of ducts. It is shown that our segmentation scheme is able to segment ductal regions. The A_z values are given for different duct densities as A_z >0.948. The reported work has the potential to assist and complement the radiologist in extracting tissue structure information for very early breast cancer detection from ultrasound B-scan images where the likelihood of survival exceeds 95%.

REFERENCES

[1] M. F. Insana, T. J. Hall, J. G. Wood, and Z. Y. Yan, "Renal Ultrasound Using Parametric Imaging Techniques to Detect Changes in Microstructure and Function," *Investigative Radiology*, vol. 28, pp. 720-725, 1993.

[2] D. B. Kopans, *Breast Imaging*: J.B. Lippincott Company, 1998.

[3] E. Taslidere, F. S. Cohen and G. Georgiou, "Detecting the stages of Hyperplasia formation in the breast ducts using Ultrasound B-Scan Images," in press**,** 2006 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI-06), Arlington, Virginia, April 6-9, 2006.

[4] G. Georgiou and F. S. Cohen, "Tissue characterization using the continuous wavelet transform Part I: Decomposition method," *Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 48, pp. 355-363, 2001.

[5] G. Georgiou, F. S. Cohen, C. W. Piccoli, F. Forsberg, and B. B. Goldberg, "Tissue characterization using the continuous wavelet transform part II: Application on breast RF data," *Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 48, pp. 364-373, 2001.

[6] L. Huang, K. D. Donohue, V. Genis, and F. Forsberg, "Duct detection and wall spacing estimation in breast tissue," *Ultrasonic Imaging*, vol. 22, pp. 137-152, 2000.

[7] J. A. Jensen, "A model for the propagation and scattering of ultrasound in tissue," *The Journal of the Acoustical Society of America*, vol. 89, pp. 182-190, 1990.

[8] R. S. Snyder and R. J. Conrad, "Ultrasound Image Quality," *Hewlett-Packard J.*, vol. 34, pp. 34-39, 1983.

[9] C. R. Hill, *Physical Principles of Medical Ultrasonics*: Ellis Horwood Limited, 1986.

[10] T. L. Szabo, "Time-Domain Wave-Equations for Lossy Media Obeying a Frequency Power-Law," *Journal of the Acoustical Society of America*, vol. 96, pp. 491-500, 1994.

[11] F. S. Foster, Strban M., and A. G., "The ultrasound macroscope: initial studies of breast tissue," *Ultrason Imaging*, vol. 6, pp. 243-261, 1984.

[12] F. T. D'Astous and F. S. Foster, "Frequency dependence of ultrasound attenuation and backscatter in breast tissue," *Ultrasound Med Biol*, vol. 12, pp. 795-808, 1986.

[13] C. M. Moran, N. L. Bush, and J. C. Bamber, "Ultrasonic Propagation Properties of Excised Human Skin," *Ultrasound in Medicine and Biology*, vol. 21, pp. 1177-1190, 1995.

[14] G. Georgiou and F. S. Cohen, "Unsupervised segmentation of RF echo into regions with different scattering characteristics (vol 45, pg 779, 1998)," *Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 45, pp. 1573-1575, 1998.

[15] G. Georgiou and F. S. Cohen, "Statistical characterization of diffuse scattering in ultrasound images," *Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 45, pp. 57-64, 1998.

[16] K. A. Wear, R. F. Wagner, and B. S. Garra, "A Comparison of Autoregressive Spectral Estimation Algorithms and Order Determination Methods in Ultrasonic Tissue Characterization," *Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 42, pp. 709-716, 1995.