Breast Cancer Detection Using Haralick Features of Images Reconstructed from Ultra Wideband Microwave Scans

Blair D. Fleet[1,2(✉)], Jinyao Yan[1,2], David B. Knoester[1,3], Meng Yao[4], John R. Deller Jr.[2], and Erik D. Goodman[1]

[1] BEACON Center, Michigan State University, East Lansing, USA
{fleetbla,yanjinya,dk,goodman}@msu.edu
[2] Department of Electrical and Computer Engineering,
Michigan State University, East Lansing, USA
deller@msu.edu
[3] Department of Computer Science and Software Engineering,
Miami University, Oxford, USA
[4] East China Normal University, Shanghai, People's Republic of China
mengyao@msu.edu

Abstract. Microwave scanning of the breast would provide a technology for cancer detection and screening that is significantly safer than current methods involving radiation. This research focuses on finding the best way for accurate characterization of cancerous signals and normal signals using clinical data collected from a previously developed ultra wideband (UWB) antenna, BRATUMASS (Breast Tumor Microwave Sensor System). BRATUMASS which detects changes in dielectric constants within the breast. The signals collected from the microwave scanning procedure are reconstructed into a single, informative representation of the breast via diffraction tomography. This representation contains the information of the breast's conductivity and the change in dielectric constants. We illustrate the feasibility of using Haralick features to make distinctions among breasts with a malignant tumor present and breasts with no malignancy in data collected from Shanghai Sixth People's Hospital and Shanghai First People's Hospital.

Keywords: Microwave near-field imaging · Breast cancer detection · Haralick features

1 Introduction

Ultra high frequency (UHF) microwaves in the band $300\,\mathrm{MHz}$–$3\,\mathrm{GHz}$ are of increasing interest for their ability to penetrate through obstacles and perform

---

precise localization and tracking of objects in indoor environments [5]. Moreover, due to their low cost and minimal radiation, UHF band microwaves are being researched for medical imaging [11]. In breast cancer detection, UHF band antennas offer the ability to focus power from the antenna through the breast tissue to localize malignant tumors. Microwave antennas are able to detect very low power signals in the presence of noise and interference, which is important when the target is small and of low-contrast. Microwave breast imaging has the potential to obviate unnecessary biopsies, increase patient comfort, and increase the effectiveness with which breast cancer can be detected.

Detection and classification using microwave breasts imaging has concentrated on simulation studies [2,16], rather than the use of clinical data. Simulated classification studies have used various approaches including support vector machines [4,14], and neural networks [9]. Recently, the feasibility of lesion classification based on contrast-aided UWB breast imaging using simulations was demonstrated [3]. There has been a movement toward conducting more clinical experiments [8], but availability of clinical data is limited.

In this paper, a detection algorithm is developed to detect the differences between cancer subimages (Class 1), and normal subimages within a normal breast which has no evidence of maligancy present (Class 3) based on Haralick's textural features [6]. The uniqueness of this research is the UHF microwave clinical data being used for the detection analysis. The purpose of this detection algorithm is to find the optimal set of features that accurately distinguish between the two classes. Most classification techniques have been performed using mammograms or MRI data, due to the access to a variety of databases [7,10].

2   Background

2.1   Microwave Technology - BRATUMASS

The BRATUMASS developed by Yao [18] is monostatic, meaning one transmitter and one receiver are co-located; the device emits low power on the order of 6.0 mW, and transmits a chirp signal through an impedance matching medium that concentrates the signal for transmission into the breast. The measurements of interest for this system is dielectric constants. The distribution of water content throughout a cancerous breast will differ from that of a normal breast since in areas of cancer, the water content will be more highly concentrated [15]. This will lead to higher dielectric constants in that area. The dielectric constants of a malignant tumor area ($\epsilon \approx 50$) are significantly higher than that of a normal breast area ($\epsilon \approx 10$) at an intermediate frequency of 1.575 GHz [17]. Our research is novel in the fact that (1) the device transmits an intermediate frequency of only 1.575 GHz, (2) that signal and image processing techniques are being applied to clinical data and (3) BRATUMASS offers portability and the safety necessary to allow extensive, longitudinal studies of patients.

2.2 Clinical Data and Data Collection

Clinical data were collected at hospitals in Shanghai under a protocol approved by East China Normal University in accordance with Chinese regulations. In a procedure sanctioned by the Michigan State University IRB, the breast scans and diagnostic data are being used in the U.S. without patient-identifying information. The clinical dataset includes 11 diagnosed cancer patients, with quadrant specific ground truth from the clinician. Figure 1(right) illustrates the transceiver antenna used for data collection, which spans 50 mm. An example of the BRA-TUMASS positioned at the 6 o'clock around the breast boundary is illustrated in Fig. 1(left). As a patient lies on her back, a clinician uses the transceiver to collect data from 16 different positions. At each antenna position, a pulsed microwave signal is sent from the transmitter (A) in the direction of the metal coin slice. The receiver (B) collects information about the changes in dielectric properties from the reflections and scattering of the microwave pulses within the breast, and the clipboard (C) connects (A) and (B). The sent and received signals are passed through a frequency mixer. The output of the mixer is further processed to map the changes in frequency to time delay distributions. These 16 processed signals are used as the signal data to reconstruct a 2D image of a patient's breast.
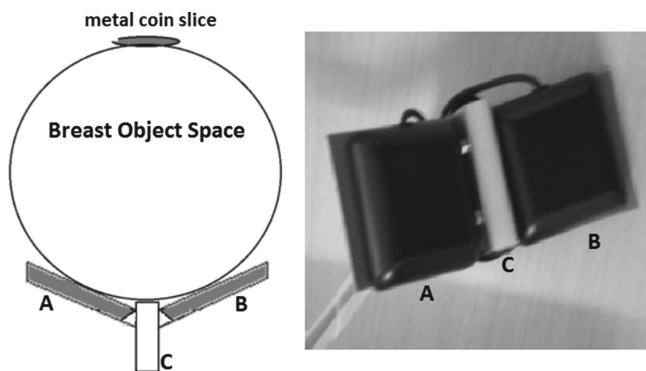

Fig. 1. Overview of BRATUMASS antenna setup, where the transmitter (A), and the receiver (B) are joined together by clipboard (C)

2.3 Haralick Features

Haralick features (HF) have been used to successfully capture textural patterns in images. HF are statistical computations that describe the overall texture of an image using measures such as entropy and sum of variance. Each feature uses information from the gray level co-occurence (GLCO) matrix, which is crucial in computing HF. In this study, the GLCO matrix characterizes the spatial dependence between two neighboring horizontal pixels. In addition to the traditional

set of HF, two more features were included, trace median and trace mean, due to the success of using those features for classification of malignant tumors found in mammograms [1]. This results in a total of 15 features.

3   Methods

3.1   Image Reconstruction and Feature Computation

The breast images are reconstructed using diffraction tomography [12,13]. Each data point is represented by a series of intersecting arcs at each antenna position. Each reconstructed image is mapped to a 160px by 160px space. Refer to Fig. 2 for the reconstruction of both breasts referring to patient ID 1, (PID 1). The box in the upper right area indicates the area where the cancer is present in PID 1. HF are calculated on non-overlapping subimages of the reconstructed image. Only those subimages that are not the background or center of the breast are used in this analysis. The background refers to the subimages that represent the air around the breast, and the center subimages are those near the center of the breast that represent the nipple area. Each image is divided into $5\text{px} \times 5\text{px}$ subimages, over which all 15 features are calculated. The subimage size was chosen after testing a variety of sizes. It was concluded that a tradeoff between subimage size and the amount of retained information is inevitable. If the subimage size is too small, the information in the GLCO will not be retained because the probability of detecting the desired pattern has been limited. If the subimage size is too big, the background effects near the breast boundary will dominate the information in those subimages, even if the tumor is present in that subimage.

3.2   Class Label Generation

The subimage can belong to one of the three different classes. A comparison between (Class 1) 'cancer,' (Class 2) 'normal' and (Class 3) 'normaln' subimages for the sum variance feature is depicted in Fig. 3. It shows the average and standard error of the sum variance (SV) feature for each possible class, across all 11 cancer patients. The sum variance feature is computed using

$$SV = \sum_{i=2}^{2N_g}(i + \sum_{i=2}^{2N_g} p_{x+y}(i) \log p_{x+y}(i))^2 p_{x+y}(i) \tag{1}$$

where N_g is the number of distinct gray levels, $p_{x+y}(i)$ is $\sum_{j=2}^{N_g} \sum_{k=2}^{N_g} p(j,k)$, and $p(j,k)$ corresponds to the probability distribution generated by i, the position entries in the GLCO matrix. In this experiment, the ground truth is quadrant specific, which means the diagnosed cancer location is localized to a quadrant. The key is to detect abnormalities between the normal breast and cancerous breast. Though there are different stages and types of cancer, we are currently
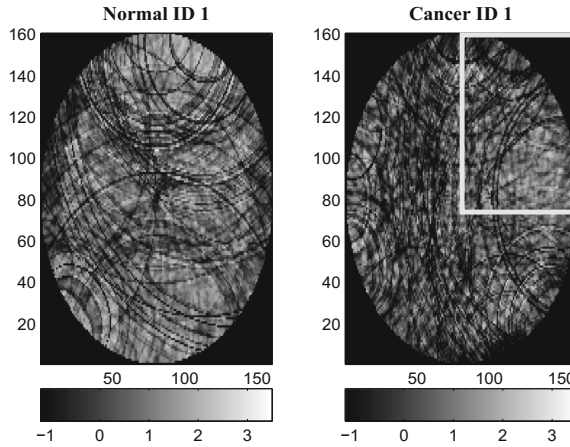
Fig. 2. Reconstructed breast images for PID 1, with normal breast pictured left and breast containing cancer pictured right with a rectangular box indicating the quadrant of the cancer location
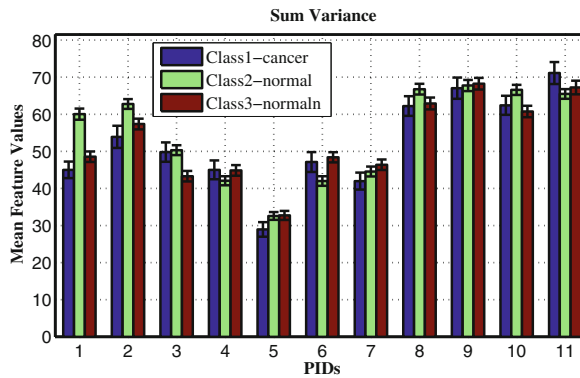
Fig. 3. Mean and standard error of sum variance feature

interested in depicting whether or not there are general differences between breasts with cancer and breasts without any cancer present. Though each breast in its entirety can be labeled as containing 'cancer' or 'normal' subimages in those defined quadrants are labeled 'cancer,' 'normal' or 'normaln' based on the additional quadrant information provided by the clinician. Ultrasound was used to cross check the clinician's diagnosis. Since the only accessible ground truth includes a general area of one quadrant location and tumor size, the difficulty with defining localized ground truth is clearly evident. Creating accurate ground truth is an active area of research.

3.3   Performance Measures

We determined which set of HF best discriminates between 'cancer', and 'normaln' subimages by generating all possible combinations, 2^{15}, of features and selecting that combination for which linear regression resulted in the highest classification performance. The performance is based on two criteria: (1) minimizing the error rate, and maximizing the posterior probability between classes and the (2) Matthew Correlation Coefficient (MCC) [14] score, which serves as a performance measure that normalizes class sizes, and incorporates true positives (TP), and false positives (FP) into the score. This score takes on values between -1 and $+1$, with $+1$ indicating perfect prediction. Cost were incorporated by weighting the false positive rate (FPR) by 0.9 and the false negative rate (FNR) by 0.1. As for the MCC score, for every one true cost incurred the final FP was increased by the five to represent a heavier weight. Finally, five-fold cross validation was performed on the two-class dataset, which includes instances of 200 'cancer' subimages, and 748 'normaln' subimages. The best set of features that yielded the highest classification performance by means of classifier performance and MCC score was tracked and recorded.

4   Results

Two examples are provided of the current images used in the HF analysis are shown in Fig. 2. Notice that the distribution of arcs is different between the two breasts within the same patient. In the cancerous breast, there is less uniformity and more scattering than in the normal breast. The lighter pixels indicate stronger changes in dielectric constant. Both the normal and cancer breast contain lighter pixels; however, in the cancerous breast, there is a concentration of lighter pixels surrounded by darker pixels, which is different from other quadrants within the breast. HF are strongly dependent on the images used, which means any slight change in the image reconstruction process can drastically effect the HF numerical measurements, which means extreme care has to be taken with the images selected for analysis. The best feature set can be represented using 0's for exclusion and 1's for inclusion of that feature in the set, which resulted in 010001111111111. This means that four of the features, 'energy,' 'correlation,' 'sum of variances' and 'inverse difference moment' were not used in the feature set. The features included were 'contrast,''sum average,''sum variance,' 'sum entropy,' 'entropy,' 'difference variance,' 'difference entropy,' 'information measures of correlation I,' 'information measures of correlation II,' 'trace mean' and 'trace median.' The inclusion of these features strongly influence the performance measures. The performance measures for the best linear combination of features can be found in Table 1. The accuracy is lower than desired because of the low resolution of ground truth labeling in the subimage domain and the imbalance in class sizes. The more specific and accurate the ground truth, the better the classification that would result. In order to address the issues with class imbalance, the MCC score was used which illustrated more favorable results because of the normalizing capability built inherently in the MCC Score computation.

Table 1. Performance measures of best feature set

Averaged test performance measures	
Classification Accuracy %	71.7
MCC Score	.889

Upon further investigation of features for all 11 patients, it can be seen that for certain patients there are significant differences between classes, as depicted in Fig. 3. For example, for PIDs 1, 2, 6, 8, and 10, the (Class 2) case is significantly different from (Class 1) and (Class 2). This suggests the importance of patient-specific techniques.

5   Conclusions

In this paper, a procedure for selecting discriminating features within clinical data using a HF detection algorithm was developed. The feasibility of using HF to make distinctions between 'cancer', and 'normaln' subimages within a patient was investigated. Due to the differences among patients, it is more beneficial to focus on patient-specific techniques versus across all patient techniques. That is, comparing the two breasts of a given patient to detect possibly cancer-indicating differences may show more promise than comparing each breast singularly with a broad reverence standard. Moving toward a more patient-specific approach is the next area of pursuit.

For future work, extensive studies will be conducted to finalize the most accurate and informative reconstructed images. Most clinicians would doubtless rather have images that clearly indicate the presence of cancer than having to analyze nonintuitive feature representation, so effective image reconstruction is crucial in making procedures straightforward for clinicians. Experimenting with other feature sets and feature extraction techniques, as well as increasing the number and variety of features is another future endeavor. Research is also being done in the signal domain, where classification and detection can be made prior to image reconstruction. Performing detection in the signal domain can inherently reduce the noise introduced in transition from the signal domain to the image domain. Image reconstruction is still needed to serve as a visual aid for clinicians who would rather see an image than a set of microwave signals.

References

1. Aroquiaraj, I.L., Thangavel, K.: Feature extraction analysis using mammogram images: a new approach. J. Comput. Sci. Appl. **3**(1), 33–44 (2011)
2. Bond, E., Li, X., Hagness, S., Van Veen, B.: Microwave imaging via space-time beamforming for early detection of breast cancer. IEEE Trans. Antennas Propag. **51**(8), 1690–1705 (2003)

3. Chen, Y., Craddock, I., Kosmas, P.: Feasibility study of lesion classification via contrast-agent-aided UWB breast imaging. IEEE Trans. Biomed. Eng. **57**(5), 1003–1007 (2010)
4. Conceicao, R., O'Halloran, M., Glavin, M., Jones, E.: Support vector machines for the classification of early-stage breast cancer based on radar target signatures. Prog. Electromagnet. Res. B **23**, 311–327 (2010)
5. Cruz, C., Costa, J., Fernandes, C.: Hybrid UHF/UWB antenna for passive indoor identification and localization systems. IEEE Trans. Antennas Propag. **61**(1), 354–361 (2013)
6. Haralick, R., Shanmugam, K.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973)
7. Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr., P., Moore, R., Chang, K., Munishkumaran, S.: Current status of the digital database for screening mammography. In: Karssemeijer, N., Thijssen, M., Hendriks, J., van Erning, L. (eds.) Digital Mammography. Computational Imaging and Vision, vol. 13, pp. 457–460. Springer, Heidelberg (1998)
8. Klemm, M., Craddock, I., Leendertz, J., Preece, A., Gibbins, D., Shere, M., Benjamin, R.: Clinical trials of a UWB imaging radar for breast cancer. In: 2010 Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP), pp. 1–4. IEEE (2010)
9. McGinley, B., O'Halloran, M., Conceicao, R., Morgan, F., Glavin, M., Jones, E.: Spiking neural networks for breast cancer classification using radar target signatures. Prog. Electromagnet. Res. C **17**, 79–94 (2010)
10. Nie, K., Chen, J., Yu, H., Chu, Y., Nalcioglu, O., Su, M.: Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. Acad. Radiol. **15**(12), 1513–1525 (2008)
11. Nikolova, N.: Microwave imaging for breast cancer. IEEE Microwave Mag. **12**(7), 78–94 (2011)
12. Slaney, M.: Imaging with diffraction tomography. Ph.D. thesis, Purdue University (1985)
13. Tao, Z., Pan, Q., Yao, M., Li, M.: Reconstructing microwave near-field image based on the discrepancy of radial distribution of dielectric constant. In: Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2009, Part I. LNCS, vol. 5592, pp. 717–728. Springer, Heidelberg (2009)
14. Viani, F., Meaney, P., Rocca, P., Azaro, R., Donelli, M., Oliveri, G., Massa, A.: Numerical validation and experimental results of a multi-resolution SVM-based classification procedure for breast imaging. In: 2009 IEEE International Symposium on Antennas and Propagation Society, APSURSI'09, pp. 1–4. IEEE (2009)
15. Wilson, C., Lammertsma, A., McKenzie, C., Sikora, K., Jones, T.: Measurements of blood flow and exchanging water space in breast tumors using positron emission tomography: a rapid and noninvasive dynamic method. Cancer Res. **52**(6), 1592–1597 (1992)
16. Winters, D., Shea, J., Kosmas, P., Van Veen, B., Hagness, S.: Three-dimensional microwave breast imaging: dispersive dielectric properties estimation using patient-specific basis functions. IEEE Trans. Med. Imaging **28**(7), 969–981 (2009)
17. Yao, M., Tao, Z., Han, Z.: The detection data of mammary carcinoma processing method based on the wavelet transformation. In: Wavelet Transforms and Their Recent Applications in Biology and Geoscience. InTech, pp. 77–92 (2012)
18. Yao, M., Tao, Z., Han, Z., Yao, Y., Fleet, B., Goodman, E.D., Wang, H., Deller, J.: Breast tumor microwave sounding, imaging and system actualizing. Adv. Inf. Sci. **1**(1), 1–21 (2013)