Cancer Classification of Freshly Excised Murine Tumors with Ordered Orthogonal Projection

Tanny Chavez, Tyler Bowman, Jingxian Wu, Magda El-Shenawee
Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701, USA
tachavez@uark.edu, tcbowman@uark.edu, wuj@uark.edu, magda@uark.edu

Keith Bailey
Oklahoma Animal Disease Diagnostic Laboratory
Oklahoma State University
Stillwater, OK 74076, USA
keith.bailey@okstate.edu

***Abstract*—This paper proposes a low dimension feature extraction method for cancer classification with terahertz (THz) images of freshly excised murine tumors. In THz imaging, the properties of various tissue regions, such as cancer, fat, and muscle, are embedded in the shape of reflected THz pulses. We propose to extract such information by projecting the frequency domain THz waveforms onto a series of ordered orthogonal basis. Experiment results demonstrated that the proposed algorithm can achieve reasonably accurate detection results for fresh tissues with three regions.**

***Index Terms*—Gram-Schmidt, basis expansion, dimension reduction, THz imaging, breast cancer**

I. Introduction

Terahertz imaging has proven to be a promising technique for breast cancer classification of excised tissues [1]–[4]. In THz imaging, each pixel is represented by an ultra-narrow pulse with bandwidth up to a few THz (0.1-4 THz). Different tissues, such as cancer, fat, and muscle, have different impacts on the reflected THz pulses at different frequency ranges. Consequently, rich information is embedded in the THz pulse regarding the physical properties of various tissues.

Various feature extraction methods have been employed in existing works. In [2], each THz pulse is summarized into a scalar by using the peak of the time domain pulse for block tissue, or the integration of the frequency domain waveform for fresh tissue. The one-feature technique shows promising results for tumor with two regions (cancer and fat), yet it cannot produce satisfactory results when three or more regions are present in the tumor [2]. Other heuristic features, such as the ratio of the minimum to the maximum point of the THz waveform [5], the time width of the pulse, and the area beneath the amplitude of the pulse [6] have presented fairly accurate classification results. The choice of these parameters are rather arbitrary and their effectiveness relies on the combination of two or more heuristic features. Principal component analysis (PCA) provides a systematic way of low dimension feature extraction that requires the normalization and centralization of all the waveforms [6], [7].

In this paper, we propose a new feature extraction algorithm, where the low dimension features of samples with three or more regions are extracted by consecutively projecting the frequency domain waveform onto a set of ordered orthonormal basis. The ordered orthogonal projection (OOP) is implemented by using a modified Gram-Schmidt (GS) algorithm. Unlike conventional GS algorithms that use arbitrary orders of basis [8], the proposed algorithm orders the basis based on its residual energy. As a result, the first few basis vectors captures the majority of the information, and it is sufficient to use a low dimension vector (e.g. 5-dimension) to represent an original waveform of thousands dimensions.

This work was funded by the National Institutes of Health under Award No. R15CA208798. It was also funded in part by the National Science Foundation under Award Nos. 1408007 and 1711087.

II. Method

Xenograft murine tumors were obtained using C57BL/6 black laboratory mice. These mice were kept on a high fat diet to generate fat deposits that would later be injected with E0771 murine-derived breast adenocarcinoma cells. Once the xenografts tumors reached 1 cm in diameter, they were excised and imaged at the pulsed THz imaging system in the University of Arkansas. After imaging, the samples were shipped to the Oklahoma Animal Disease Diagnostic Laboratory (OADDL) to obtain the histopathology report of the sample [2]. For this paper, the authors evaluated the proposed algorithm using sample 9 section B, which has three regions: cancer, fat, and muscle.

The proposed OOP algorithm utilizes the frequency domain signals collected through the pulsed THz system. In orthogonal projection, the selection of the basis is critical to the overall performance. The proposed algorithm minimizes the dimension of the feature vector by ordering the basis based on its residual energy. The first basis vector is chosen by normalizing the average of the waveforms from all pixels. Subsequent basis vectors are chosen as the ones that can minimize the residual energy after projecting the waveforms onto the space spanned by all previous and the candidate basis vectors. Such an approach can ensure the majority of the energy is stored in the first few basis vectors.

After the dimension reduction, the algorithm classifies each pixel using an unsupervised Bayesian learning method with Gaussian mixture model and Markov Chain Monte Carlo (MCMC) [9]. The classification results are evaluated by using the method presented in [1].

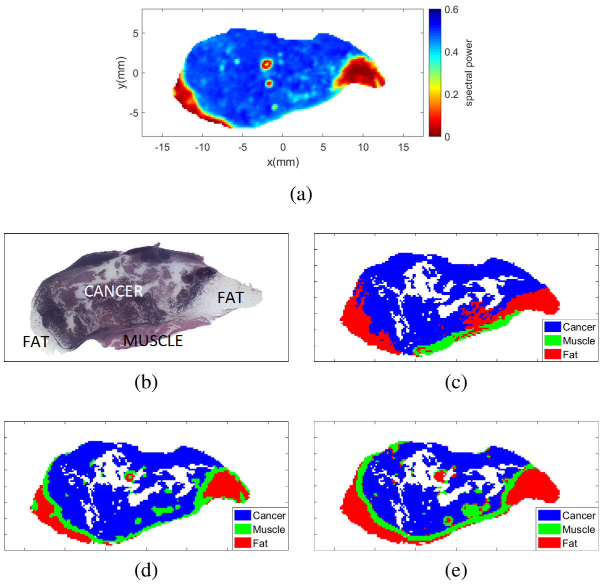

Fig. 1: Sample 9B fresh. (a) THz image. (b) Pathology image. (c) Morphed Pathology. (d) 1-D Gaussian mixture model. (e) 5-D Gaussian mixture model.

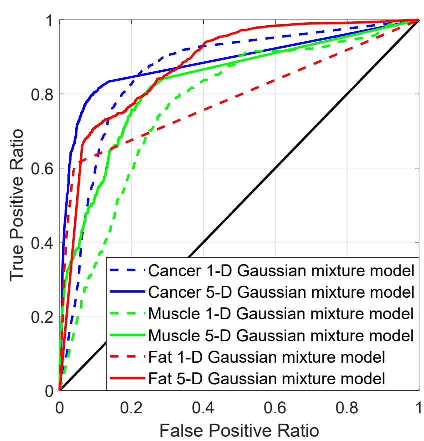

Fig. 2: ROC curves comparison: 1-D vs. 5-D Gaussian mixture model for sample 9B fresh.

III. Results

Fig. 1a represents the THz image of fresh sample 9B. Fig. 1b represents the pathology results for the sample once it was fixed in paraffin. Fig. 1c summarizes the pathology results into a mask with the help of image morphing [1]. Finally, Figs. 1d and 1e show the classification results using the 1-D and 5-D Gaussian mixture model, respectively. The 1-D result is obtained by using the same approach as in [1] and [2], and the 5-D results are obtained by using the proposed OOP algorithm. From Figs. 1c, 1d, and 1e, we can observe that there is a good correlation between the classification results and the morphed pathology.

To quantitatively evaluate the classification results, Fig. 2 shows the receiver operation characteristic (ROC) curves of sample 9B fresh. Table I presents the area under the ROC curves of sample 9B fresh. This area represents a measure of the capability of the classification method to distinguish an specific region from the rest of the tumor. Both Fig. 2 and Table I show that the 5-D Gaussian mixture model considerably outperforms their 1-D counterparts for all regions.

TABLE I: Area under the ROC curve for sample 9B fresh.

Region	1-D	5-D
Cancer	0.8630	0.8858
Muscle	0.7729	0.8292
Fat	0.7885	0.8810

IV. Conclusion

An OOP method has been presented in this paper for low dimension feature extraction in THz imaging for cancer classification. The algorithm provides a systematic way to extract salient information embedded in high-dimensional THz waveforms. Unlike other orthogonal projection algorithms, the proposed algorithm can minimize the dimension of the feature vector by ordering the orthogonal basis with respect to it's residual energy. Experiment results demonstrated that the proposed algorithm achieve reasonable performance gains over 1-D algorithm for samples with three regions.

Acknowledgment

The authors would like to thank Dr. Rajaram and his students for handling the mice and proving the tumor samples used in this paper.

References

[1] Tanny Chavez, Tyler Bowman, Jingxian Wu, Keith Bailey, and Magda El-Shenawee. Assessment of terahertz imaging for excised breast cancer tumors with image morphing. *Journal of Infrared, Millimeter, and Terahertz Waves*, 39(12):1283–1302, Dec 2018.

[2] Tyler Bowman, Tanny Chavez, Kamrul Khan, Jingxian Wu, Avishek Chakraborty, Narasimhan Rajaram, Keith Bailey, and Magda El-Shenawee. Pulsed terahertz imaging of breast cancer in freshly excised murine tumors. *Journal of Biomedical Optics*, 23(2):026004, 2018.

[3] T. Bowman, T. Chavez, K. Khan, A. Chakraborty, J. Wu, K. Bailey, and M. El-Shenawee. Statistical signal processing for quantitative assessment of pulsed terahertz imaging of human breast tumors. In *2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2, Aug. 2017.

[4] T. Bowman, K. Alhallak, T. Chavez, K. Khan, D. Lee, N. Rajaram, J. Wu, A. Chakraborty, K. Bailey, and M. El-Shenawee. Terahertz imaging of freshly excised breast cancer using mouse model. In *2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2, Aug. 2017.

[5] Anthony J. Fitzgerald, Vincent P. Wallace, Mercedes Jimenez-Linan, Lynda Bobrow, Richard J. Pye, Anand D. Purushotham, and Donald D. Arnone. Terahertz pulsed imaging of human breast tumors. *Radiology*, 239(2):533–540, 2006. PMID: 16543586.

[6] Anthony J. Fitzgerald, Vincent P. Wallace, Sarah E. Pinder, Anand D. Purushotham, Padraig O'Kelly, and Philip C. Ashworth. Classification of terahertz-pulsed imaging data from excised breast tissue. *Journal of Biomedical Optics*, 17(1), 2012.

[7] Zhen-Wei Zhang, Haishun Liu, and Cunlin Zhang. Terahertz pulse data dimensional reduction and classification for hepatic tissue samples. pages 1–3, 09 2018.

[8] Sumanta Guha and Ananta Raj Lamichhane. Document classification after dimension reduction through a modified gram-schmidt process. In K. R. Venugopal and L. M. Patnaik, editors, *Wireless Networks and Computational Intelligence*, pages 236–243, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[9] Jessica Franzén. Bayesian inference for a mixture model using the gibbs sampler. *MResearch Report*, 1, 2006.