Spatial Image Segmentation for Breast Cancer Detection in Terahertz Imaging

Tanny Chavez, Nagma Vohra, Jingxian Wu, Magda El-Shenawee
Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701, USA
tachavez@email.uark.edu, nvohra@email.uark.edu, wuj@uark.edu, magda@uark.edu

Keith Bailey
Veterinary Diagnostic Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61802, USA
kbailey1@illinois.edu

2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting | 978-1-7281-6670-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/IEEECONF35879.2020.9330445

***Abstract*—This paper proposes a new spatial image segmentation algorithm for breast cancer detection in terahertz (THz) images of freshly excised human tumors. Region classifications of fresh tissue with 3 or more regions, such as cancer, fat, and collagen, remain a challenge for cancer detection. We propose to tackle this problem by exploiting the spatial correlation among neighboring pixels in THz images, that is, pixels that are close to each other are more likely to belong to the same region. The spatial correlation among pixels is modeled by using Markov random fields (MRF). A Gaussian mixture model (GMM) with expectation maximization (EM) is then used to represent the statistical distributions of the THz images in both the frequency and spatial domain. Experiment results demonstrated that the proposed spatial image segmentation algorithm outperforms existing algorithms that do not consider spatial information.**

I. Introduction

Terahertz (THz) imaging has been demonstrated as a promising technique in a plethora of biomedical applications, such as brain injury evaluation [1], glioma [2], prostate cancer [3], and breast cancer detection [4]–[6]. Various signal processing and machine learning techniques have been used in THz imaging. For instance, [2] applies fuzzy c-means clustering for the detection of *ex vivo* rat glioma with good accuracy. Other approaches explore machine learning techniques for THz image segmentation such as support vector machine (SVM) [1], [3], K-nearest neighbors [1], [6], random forest [1], and artificial neural networks [6].

Unlike the animal trials performed in [1], [2], this paper focuses on the detection of breast cancer within human tumors, which are far more heterogeneous and complex compared to their mice counterparts. Specifically, this paper presents results on freshly excised samples. In our previous work [5], a low-dimension ordered orthogonal projection (LOOP) method with Gaussian mixture model was proposed to perform image segmentation of THz imaging of freshly excised human samples. However, in [5], all pixels are assumed to be statistically independent of each other, which is in general not the case in practice. Since the pixels are collected by scanning the tumor with steps of $200\mu m$, it is natural to consider that neighboring pixels have a higher probability of belonging to the same region. Motivated by this fact, we propose a new spatial image segmentation algorithm that exploits the spatial correlation among pixels. The spatial correlation is modeled by applying Markov random field (MRF) on Gaussian mixture model (GMM), and the expectation maximization (EM) algorithm is then applied to the statistical models to classify the different regions in the THz image.

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

This work focuses on human sample # ND15588, which was obtained from National Disease Research Interchange (NDRI). The experimental setup for the collection of the reflected THz data per pixel can be found in [5]. Once the THz image is obtained, the low-dimension ordered orthogonal projection (LOOP) method [5] is first applied to the raw data to reduce the dimension of THz waveform for each pixel. In the LOOP algorithm, the high-dimensional signals in the frequency domain are projected onto a lower dimensional space with minimum loss of information. To achieve this, the LOOP algorithm utilizes a modified Gram-Schmidt process to create an orthonomal basis, which span a subspace containing the most important features in the THz dataset.

Unlike our previous work that assumed statistical independence among pixels in the spatial domain [5], the newly proposed spatial image segmentation approach considers the spatial correlation among pixels. Specifically, it is assumed that pixels within a certain neighborhood, that is, a cluster of pixels that are close to each other, are correlated with each other by following certain prior distributions.

Let $\boldsymbol{Y} = [y_1, \ldots, y_N]$ denote the classification labels for the N pixels in the THz image, where $y_i \in \{1, 2, \ldots, K\}$, and K represent the number of regions (e.g. cancer, fat, collagen, etc.). The spatial correlation among the pixels can be represented by a Gibbs prior to the labels as [7]

$$P(\boldsymbol{Y}) = \frac{1}{Z}\exp\left(-\sum_{c\in\mathcal{C}} V_c(\boldsymbol{Y})\right), \tag{1}$$

where Z is a normalization constant, $\mathcal{C}$ corresponds to the clique within the defined neighborhood, $V_c(y_i, y_j) = \beta(1 - I_{y_i,y_j})$, and $I_{y_i,y_j} = 1$ if $y_i = y_j$ or 0 otherwise. In this paper, we consider neighborhoods of sizes 4, 8, and 24 directly

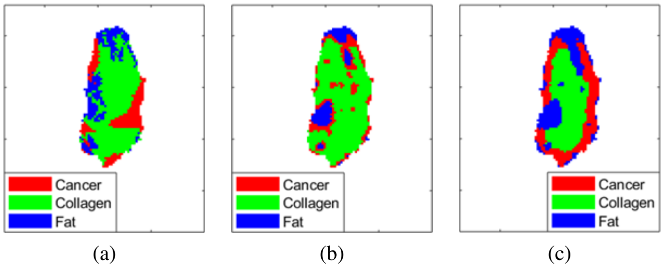

Fig. 1: Fresh sample ND15588. (a) Morphed pathology [5]. (b) Segmentation results from 1D MCMC [5]. (c) Segmentation results from 2D EM with 8-nearest neighbors.

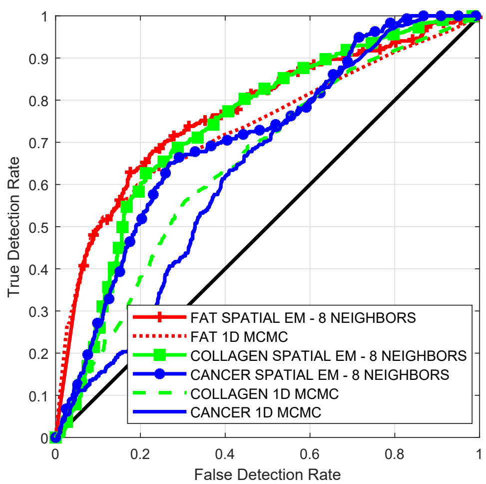

Fig. 2: ROC curves for fresh sample ND15588.

surrounding each pixel of interest. Finally, this new objective function is solved by using expectation maximization (EM) and Gaussian mixture model (GMM), as reported in [5].

III. Results

The results in this paper are obtained by applying the newly proposed spatial image segmentation algorithm to the same sample used in [5]. Fig. 1a shows the morphed pathology obtained through mesh morphing [4], which represents our ground truth. Fig. 1b presents the classification results obtained through a one dimension (1D) GMM with Markov chain Monte Carlo (MCMC) as described in [8]. Fig. 1c shows the classification results using the 2-dimensional (2D) spatial EM approach proposed in this paper. In these figures, we can observe that the spatial model presents better region correlation with the morphed pathology than the 1D MCMC model.

Fig. 2 shows the receiver operating characteristic (ROC) curves of the classification models, which provide the quantitative evaluation of these models in the form of true vs. false detection rate for each region within the tissue. As shown in Fig. 2, the spatial model performs better than the 1D MCMC approach for all regions. This can be further confirmed in Table I, which presents the areas under the ROC curves. As shown in Table I, we can observe that the area under these curves slightly increases as the number of neighbors increases.

TABLE I: Areas under the ROC curves for sample ND15588 fresh.

Method	Cancer	Collagen	Fat
1D MCMC	0.6338	0.6521	0.7372
Spatial EM - 4 neighbors	0.7092	0.7400	0.7721
Spatial EM - 8 neighbors	0.7099	0.7401	0.7726
Spatial EM - 24 neighbors	0.7126	0.7408	0.7750

IV. Conclusion

This paper presents a spatial image segmentation approach based on MRF and GMM for breast cancer detection in THz imaging. The proposed approach moderately improves the performance of breast cancer detection by exploiting the spatial correlation among neighboring pixels in THz images. Experimental results on freshly excised human tissue have demonstrated that the spatial GMM model achieves better detection rates when compared to 1D MCMC. Finally, it was proved that increasing the neighborhood size represents minimal improvement on the areas under the ROC curves for all the regions within the sample.

Acknowledgement

The authors would like to thank the Oklahoma Animal Disease Diagnostic Laboratory and Oklahoma State University for their collaboration in handling sample # ND15588. In addition, we acknowledge the use of tissues procured by the NDRI with support from the NIH grant U42OD11158.

References

[1] J. Shi, Y. Wang, T. Chen, D. Xu, H. Zhao, L. Chen, C. Yan, L. Tang, Y. He, H. Feng, and J. Yao, "Automatic evaluation of traumatic brain injury based on terahertz imaging with machine learning," *Opt. Express*, vol. 26, no. 5, pp. 6371–6381, Mar 2018. doi: 10.1364/OE.26.006371. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-26-5-6371

[2] Y. Wang, Z. Sun, D. Xu, L. Wu, J. Chang, L. Tang, Z. Jiang, B. Jiang, G. Wang, T. Chen, H. Feng, and J. Yao, "A hybrid method based region of interest segmentation for continuous wave terahertz imaging," *Journal of Physics D: Applied Physics*, vol. 53, no. 9, p. 095403, dec 2019. doi: 10.1088/1361-6463/ab58b6. [Online]. Available: https://doi.org/10.1088%2F1361-6463%2Fab58b6

[3] Y. V. Kistenev, A. V. Borisov, A. Knyazkova, O. Zakharova, E. Sandykova, L. Spirina, A. Gorbunov *et al.*, "Paraffin embedded cancer tissue 2D terahertz imaging and machine learning analysis," 2018.

[4] T. Chavez, T. Bowman, J. Wu, K. Bailey, and M. El-Shenawee, "Assessment of terahertz imaging for excised breast cancer tumors with image morphing," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 39, no. 12, pp. 1283–1302, Dec 2018. doi: 10.1007/s10762-018-0529-8. [Online]. Available: https://doi.org/10.1007/s10762-018-0529-8

[5] T. Chavez, N. Vohra, J. Wu, K. Bailey, and M. El-Shenawee, "Breast cancer detection with low-dimensional ordered orthogonal projection in terahertz imaging," *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 2, pp. 176–189, March 2020. doi: 10.1109/TTHZ.2019.2962116.

[6] H. J. Motlak and S. I. Hakeem, "Detection and classification of breast cancer based-on terahertz imaging technique using artificial neural network & k-nearest neighbor algorithm," *International Journal of Applied Engineering Research*, vol. 12, no. 21, pp. 10 661–10 668, 2017.

[7] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 809–823, March 2012. doi: 10.1109/TGRS.2011.2162649.

[8] T. Bowman, T. Chavez, K. Khan, J. Wu, A. Chakraborty, N. Rajaram, K. Bailey, and M. El-Shenawee, "Pulsed terahertz imaging of breast cancer in freshly excised murine tumors," *Journal of Biomedical Optics*, vol. 23, no. 2, p. 026004, 2018. doi: 10.1117/1.JBO.23.2.026004.