35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Automation of ROI Extraction in Hyperspectral Breast Images

B. Kim, N. Kehtarnavaz, *Fellow, IEEE,* P. LeBoulluec, H. Liu, Y. Peng and D. Euhus

Abstract— **The extraction of regions-of-interest (ROIs) in hyperspectral images of breast cancer specimens is currently carried out manually or by visual inspection. In order to address the labor-intensive and time-consuming process of the manual extraction of ROIs in hyperspectral images, an algorithm is developed in this paper to automate the extraction process. This is achieved by using a contrast module and a homogeneity module to duplicate the same manual or visual steps that an expert goes through in order to extract ROIs. The success of the automated process is determined by comparing the classification rates of the automated approach with the manual approach in terms of the ability to separate cancer cases from normal cases.**

I. Introduction

The most recent estimate of invasive breast cancer stands at about 226,870 cases with the death rate of 39,510 [1]. In fact, breast cancer is the second most occurring cancer in the United States [2]. Partial mastectomy (lumpectomy) is a surgical procedure that is used to remove cancerous cells and the surrounding tissue. In this procedure, it is desired to perform the removal such that there is a clear margin around cancerous cells. Positive margin refers to the situation where cancerous cells still reside on the edge of the removed tissue. Positive margins lead to additional surgeries in 20-50% of the initial surgeries [3].

In order to decrease the number of additional surgeries due to positive margins, a hyperspectral imaging approach has been developed [3-6]. As part of this effort, this paper presents an automatic extraction of regions-of-interest (ROIs) in hyperspectral images, while the current method for extraction of ROIs is done by visual inspection. The ultimate goal is to have a real-time imaging examination system in the surgery room to determine whether the removed tissue in partial mastectomy has positive margin so that if necessary additional tissue is removed.

In Section II, an overview of the hyperspectral imaging system utilized and the hypersepctral images collected is described. Section III covers the current manual or visual inspection approach to extract ROIs. The automated approach to extract ROIs is then covered in section IV. Section V includes the results of the automated ROI extraction. Finally, the conclusion is stated in section VI.

B. Kim and N. Kehtarnavaz are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (phone: 972-883-6838; fax: 972-883-2710; e-mail: bkim@utdallas.edu, kehtar@utdallas.edu).

P. LeBoulluec, and H. Liu are with the Department of Biomedical Engineering, University of Texas at Arlington, Arlington, TX 76019 USA (e-mail: peter.leboulluec@mavs.uta.edu, hanli@uta.edu)

Y. Peng, and D. Euhus are medical doctors with the University of Texas Southwestern Medical Center, TX 75390 USA.

II. Hyperspectral Imaging System

This section provides an overview of the hyperspectral imaging system used, see Fig. 1. A spectral light engine (OL 490) is used to diffuse light onto a DLP processor consisting of an array of micromirrors. Reflected light from a tissue sample is detected via a CCD array camera (CoolSNAP HQ2). A different amount of light gets reflected at different wavelengths. A computer program is then used to read and convert different reflections at different positions into a 2D image per wavelength.

A. Image acquisition

Breast cancer tissue specimens were collected from 19 patients who underwent mastectomy at the University of Texas Southwestern Medical Center. Three types of breast tissues (cancer, fibrous and adipose) were collected from each patient specimen. As a result, a total of 55 hyperspectral image sets were generated due to some cases having only one or two tissue types. A pathologist report was acquired providing tagged small specimen pieces corresponding to respective tissue types and imaged by the hyperspectral imaging system. Per pixel, a 101-wavelength spectral curve corresponding to the wavelength range of 380nm-780nm was acquired in 4nm increments. As a result, each hyperspectral image sample set

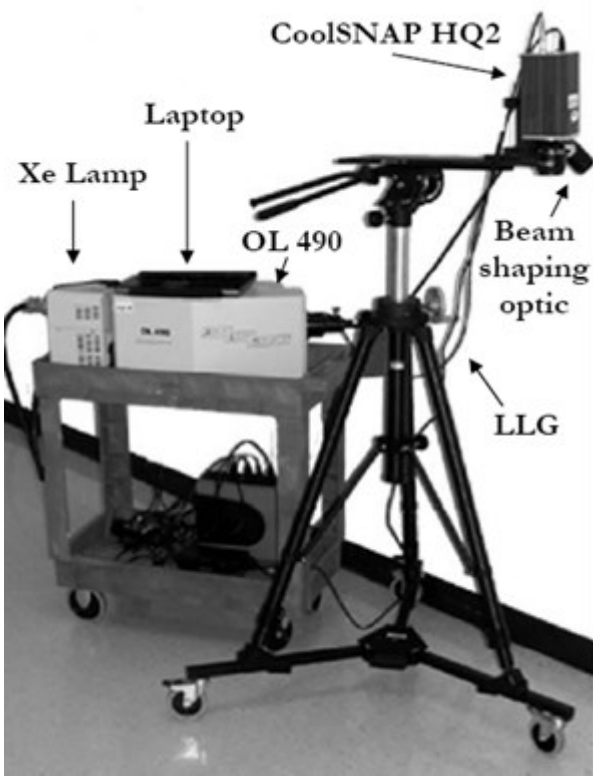


Figure 1. Hyperspectral imaging system [11]

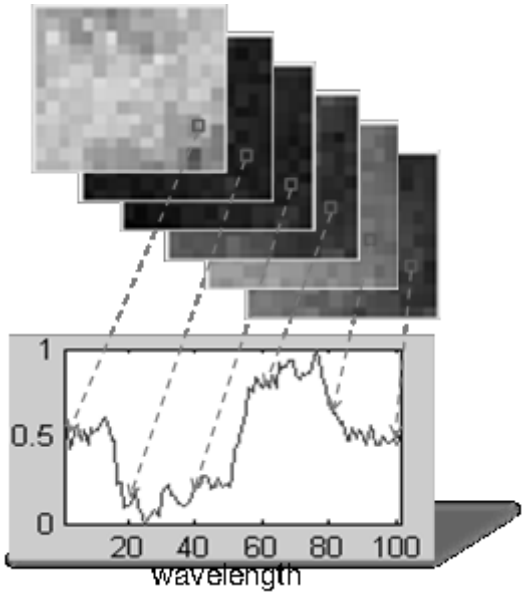


Figure 2. A sample set of hyperspectral images and a spectral curve associated with a pixel

consisted of 101 hyperspectral images.

The hyperspectral imaging system utilized allows generating an image per wavelength, thus the signal per pixel consists of a spectral curve. Figure 2 shows an example of a sample set and the corresponding spectral curve associated with a pixel.

III. Manual ROI Extraction

Detection or identification of ROIs from the hyperspectral images of breast tissues is currently done visually or manually, based on the pathology results performed on previously measured cancerous or non-cancerous (fibrous, adipose) tissues. The extracted ROIs are then used for further data analysis. The manual or visual approach needs to exclude several types of areas, including those contaminated with dye,

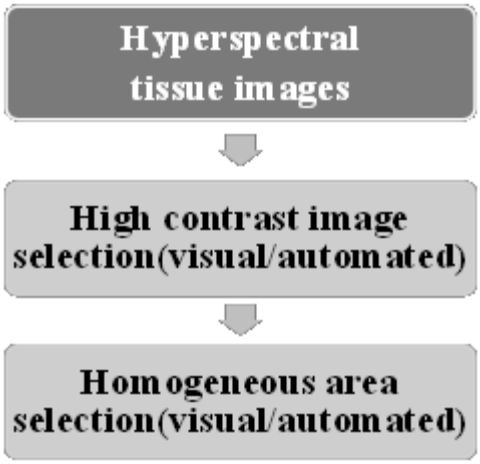
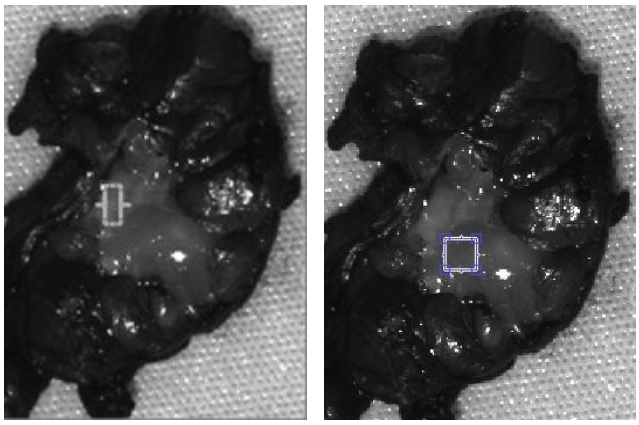
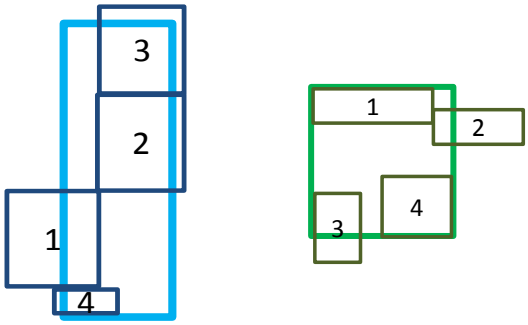


Figure 3. Manual or visual inspection steps duplicated in the automated approach



(a) Pathology-tagged area of fibrous tissue

(b) Pathology-tagged area of adipose tissue



(c) 4 ROIs extracted manually from (a)

(d) 4 ROIs extracted manually from (b)

Figure 4. Manually selected ROIs from pathology-tagged areas

contained blood or interfered by specular reflection.

The manual steps taken to extract ROIs from hyperspectral images that contain pathology-tagged cancerous or non-cancerous area are illustrated in Figure 3. The manual process starts by selecting one image at a single wavelength out of a large number of spectral images (in our case, a total of 101 images). This is done by visually inspecting that the selected image has minimal contamination from dye and specular reflection within the tissue under study. Figure 4 shows the visually extracted ROIs that correspond to two pathology-tagged areas. The extraction of ROIs is achieved by visually selecting those sub-areas from the tagged regions via a rectangular box; these sub-areas possess relatively uniform signal intensities and thus are free of contaminated tissue sub-areas which are used for further data processing. This is a labor-intensive and time-consuming process.

IV. Automated ROI Extraction

This section covers the steps for automation of the above manual ROI extraction process involving a large number of hyperspectral images. The developed automation process duplicates manual steps as aforementioned in Figure 3. Specifically, first, a contrast analysis is performed in order to select a hyperspectral image with a good discriminatory power among the large number of hyperspectral images. Then, within the selected image, local texture features are examined

per pixel to further select those ROIs or sub-areas which possess homogenous characteristics.

A. Contrast Automation

Selecting an image with the highest contrast among all the hyperspectral images allows the processing time to be significantly lowered. This step is achieved by examining between and within class variations in the image histogram. More specifically, each hyperspectral image is divided into two classes via the Otsu's optimal thresholding applied to its histogram [8]. Then, the separation of means between the two classes is computed. This procedure is repeated for all the images and the image producing the highest separation of classes or means is selected for further analysis. Basically, this selected image corresponds to the image with the highest contrast among all the images. In the manual process, the selection of image with the highest contrast in an image set is done visually and subjectively. Figure 5 shows the distribution of the selected wavelength using the contrast automation. The x-axis corresponds to 101 different wavelengths from 380nm to 780nm and the y-axis to the frequency or count of occurrence.

B. Homogeneity Automation

As the second step of the automation process, the image with the highest contrast is used to segment homogeneous ROIs or sub-areas. Six widely used texture features based on the image histogram were examined: mean, standard deviation, smoothness, 3rd moment, homogeneity and entropy [9] using five different window sizes of 3x3, 5x5, 7x7, 9x9 and 11x11. These particular texture features were chosen due to their computational efficiency. They are named T1 through T6 here,

$$\mathrm{T1} = \sum_{i=0}^{L-1} z_i p(z_i) \quad (1)$$

$$\mathrm{T2} = \sqrt{\mu_2(z)} = \sqrt{\sigma^2} \quad (2)$$

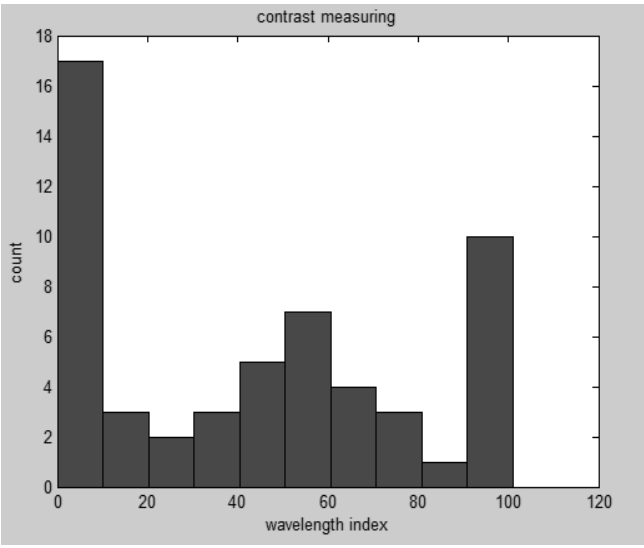


Figure 5. Distribution of all 55 image sets using the automated contrast approach

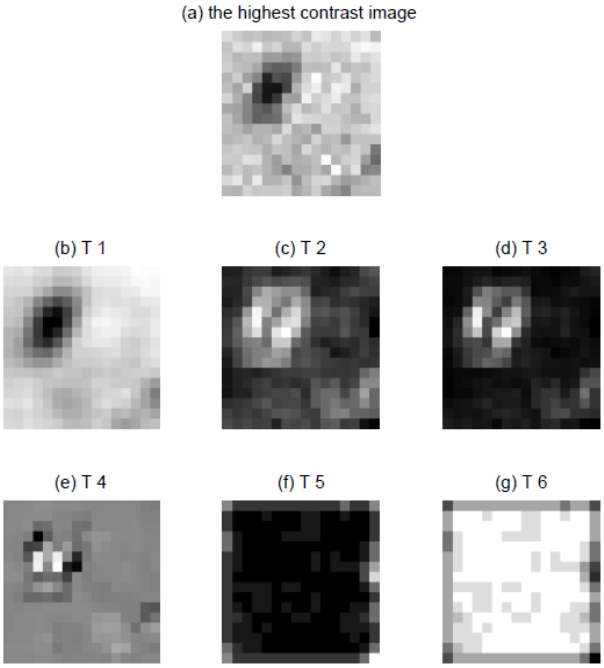


Figure 6. Six texture feature images using 3x3 windows for the hyperspectral image in (a)

$$\mathrm{T3} = 1 - \frac{1}{1+\sigma^2} \quad (3)$$

$$\mathrm{T4} = \mu_3 \quad (4)$$

$$\mathrm{T5} = \sum_{i=0}^{L-1} p^2(z_i) \quad (5)$$

$$\mathrm{T6} = -\sum_{i=0}^{L-1} p(z_i) log_2^{p(z_i)} \quad (6)$$

where, $\mu_n = \sum_{i=0}^{L-1}(z_i - m)^n p(z_i)$, L denotes the number of intensity levels, z intensity, $p(z_i)$ the histogram and m the average intensity computed as follows:

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \quad (7)$$

Figure 6(a) shows a sample image with the highest contrast obtained from the contrast automation step, and Figure 6(b)-(g) shows the corresponding set of six texture images using a 3x3 window.

In addition, the following two contrast and homogeneity texture features based on the co-occurrence matrix were examined with a distance of 1 along 0, 45, 90 and 135 degrees:

$$\mathrm{gContrast} = \sum_{i,j} |i-j|^2 p(i,j) \quad (8)$$

$$\mathrm{gHomogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (9)$$

where, $p(i,j)$ denotes the normalized co-occurrence matrix.

Table 1. Average classification rates of automated extraction of ROIs using texture features and manual extraction of ROIs

	T1	T2	T3	T4	T5	T6	Manual	
TPR	96.8	97.2	**97.3**	91.2	92.5	93.8	**98.7**	
TNR	94.7	95.7	**95.9**	89.5	88.2	90.3	**96.4**	
	gContrast0	gContrast45	gContrast90	gContrast135	gHomo0	gHomo45	gHomo90	gHomo135
TPR	97.2	97.2	97.1	97.3	96.6	94.6	96.3	96.6
TNR	95.5	95.6	95.6	95.7	96.7	94.6	94.5	94.7

V. Automated Extraction Results

Fifty-five hyperspectral image sets of breast tissues were examined. These image sets were provided by the University of Texas Southwestern Medical Center. The size of the images varied from 9x14 to 32x32 pixels. In this work, the most common type of breast cancer cases, or IDC (invasive ductal carcinoma), were separated from normal cases.

A support vector machine (SVM) classifier with a 3rd degree polynomial kernel was used to verify the effectiveness of the automated versus manual ROI extraction pipeline [10]. Two classes consisting of 17 cancerous tissues and 38 non-cancerous tissues (19 fibrous, 19 adipose) were examined. Table 1 presents the classification outcome based on the automated extraction of ROIs as well as based on the manual extraction of ROIs. The use of T3 feature provided the best outcome. Different combinations of the six texture features were used but none produced a better outcome than T3. TPR (true positive rate) and TNR (true negative rate) in this table exhibit the sensitivity and specificity, respectively. As indicated in the table, the window size of 3x3 produced the best outcome.

It is worth mentioning that the classification was repeated 500 times using a 10 fold cross validation method and the table includes the average classification rates. The P-value of the statistical significance test between T3 and the manual approach was found to be very close to 1. In other words, from a statistical standpoint, there was no difference between the manually extracted ROIs and the automatically extracted ROIs.

VI. Conclusion

Currently, the extraction of ROI in hyperspectral images of breast cancer is done manually or by visual inspection, which is cumbersome and time-consuming. This paper has presented an automation method to replace the manual extraction process by using the contrast and texture information in hyperspectral images. By examining hyperspectral images taken from ex-vivo breast cancer specimens, it was shown that the developed automated extraction of ROI closely matched the manual or visual extraction of ROI.

Acknowledgment

This work was funded in part by a grant from Texas Medical Research Collaborative.

References

[1] American Cancer Society, "Cancer Facts and Figures 2011-2012", 2012.

[2] Type of breast cancer, *Breastcancer.org*, [online] 2013, http://www.breastcancer.org/symptoms/types (Accessed: 5 February 2013).

[3] V. Sharma, S. Shivalingaiah, Y. Peng, D. Euhus, Z. Gryczynski, and H. Liu, "Auto-fluorescence lifetime and light reflectance spectroscopy for breast cancer diagnosis: potential tools for intraoperative margin detection," *Biomedical Optics Express*, vol. 3, no.8, pp. 1825-1840, 2012.

[4] N. Ramanujam, J. Brown, T. Bydlon, S. Kennedy, L. Richards, M. Junker, J. Gallagher, W. Barry, L. Wilke and J. Geradts, "Quantitative Spectral Reflectance Imaging Device for Intraoperative Breast Tumor Margin Assessment," *Proceedings of IEEE International Conference on Engineering in Medicine and Biology Society,* Minnesota, 2009.

[5] M. Keller, S. Majumder, M. Kelley, I. Meszoely, F. Boulos, G. Olivares, and A. Jansen, "Autofluorescence and Diffuse Reflectance Spectroscopy and Spectral Imaging for Breast Surgical Margin Analysis," *Lasers in Surgery and Medicine*, vol. 42, pp.15-23, 2010.

[6] M. Keller, S. Majumder, M. Kelley, I. Meszoely, F. Boulos, G. Olivares, and A. Jansen, "Autofluorescence and Diffuse Reflectance Spectroscopy and Spectral Imaging for Breast Surgical Margin Analysis," *Lasers in Surgery and Medicine*, vol. 42, pp.15-23, 2010.

[7] M. Martin, M. Wabuyele, K. Chen, P. Kasili, M. Panjehpour, M. Phan, B. Overholt, G. Cunningham, D. Wilson, R.C. Denovo and T. Vo-Dinh, "Development of an Advanced Hyperspectral Imaging (HSI) System with Applications for Cancer Detection," *Annals of Biomedical Engineering*, vol. 34, no. 6, pp.1061-1068, 2006.

[8] N. Otsu, "A Threshold Selection Method From Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp.62–66, 1979.

[9] M. Tuceryanand and A. Jain, "Texture Analysis," *Handbook of Pattern Recognition and Computer Vision*, 1993.

[10] B. Schölkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio and V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, 1997.

[11] R. Francis, "DLP Hyperspectral Imaging for Surgical and Clinical Utility," University of Texas at Arlington Thesis, 2009.