

# Automatic Tissue Segmentation of breast biopsies imaged by QPI

Hassaan Majeed<sup>a</sup>, Tan Nguyen<sup>a</sup>, Mikhail Kandel<sup>a</sup>, Virgilia Marcias<sup>b</sup>, Minh Do<sup>c</sup>, Krishnarao Tangella<sup>d</sup>, Andre Balla<sup>b</sup>, and Gabriel Popescu<sup>\*a</sup>

<sup>a</sup>Quantitative Light Imaging Laboratory, Department of Electrical and Computer Engineering, Beckman Institute of Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; <sup>b</sup>University of Illinois at Chicago, Department of Pathology, Chicago, Illinois 60612, USA; <sup>c</sup>Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; <sup>d</sup>Christie Clinic and University of Illinois at Urbana-Champaign, Department of Pathology 1400 West Park Street, Urbana, Illinois 61801, USA.

## ABSTRACT

The current tissue evaluation method for breast cancer would greatly benefit from higher throughput and less inter-observer variation. Since quantitative phase imaging (QPI) measures physical parameters of tissue, it can be used to find *quantitative* markers, eliminating observer subjectivity. Furthermore, since the pixel values in QPI remain the same regardless of the instrument used, classifiers can be built to segment various tissue components without need for color calibration. In this work we use a texton-based approach to segment QPI images of breast tissue into various tissue components (epithelium, stroma or lumen). A tissue microarray comprising of 900 unstained cores from 400 different patients was imaged using Spatial Light Interference Microscopy. The training data were generated by manually segmenting the images for 36 cores and labelling each pixel (epithelium, stroma or lumen.). For each pixel in the data, a response vector was generated by the Leung-Malik (LM) filter bank and these responses were clustered using the k-means algorithm to find the centers (called textons). A random forest classifier was then trained to find the relationship between a pixel's label and the histogram of these textons in that pixel's neighborhood. The segmentation was carried out on the validation set by calculating the texton histogram in a pixel's neighborhood and generating a label based on the model learnt during training. Segmentation of the tissue into various components is an important step toward efficiently computing parameters that are markers of disease. Automated segmentation, followed by diagnosis, can improve the accuracy and speed of analysis leading to better health outcomes.

**Keywords:** Interferometry, histopathology, SLIM, image segmentation

## 1. INTRODUCTION

Breast cancer is a pressing global health problem with an average of 500,000 deaths among women worldwide in recent years. The World Health Organization (WHO) reports a rise in breast cancer incidence in recent years and recommends early detection as one of the most important steps towards addressing the problem [1,2]. Histopathological evaluation is the standard method for diagnosis and prognosis of breast cancer. After an abnormality is detected during a screening procedure such as mammography, a tissue biopsy is obtained from the patient and a pathologist observes the biopsy under a microscope after staining it with Hematoxylin and Eosin (H&E). Cancer is diagnosed based on the morphological abnormalities visible through the microscope due to the contrast generated by the H&E stain. Since this kind of evaluation is both qualitative and manual, it leads to inter-observer variation and low throughput [3,4]. Since early diagnosis and accurate prognosis are crucial in treating the patient effectively – there is a need to develop quantitative and automated evaluation methods.

gpopescu@illinois.edu; phone: 1 217 333-4840; www.light.ece.illinois.edu

Quantitative phase imaging (QPI) refers to a subset of optical microscopy techniques that use interferometric detection to measure the optical path length difference across the sample. The optical path-length difference is measured as an image by retrieving the phase  $\phi$  of the imaging field which is given by

$$\phi(x, y) = \frac{2\pi}{\lambda}(n_t(x, y) - n_m)t(x, y) \quad (1)$$

where  $n_t$  is the refractive index of tissue,  $n_m$  is the refractive index of the tissue mounting medium,  $t$  is the thickness of tissue and  $\lambda$  is the wavelength of light [5-8]. By measuring the phase map, which is related to refractive index and thickness of tissue, QPI provides quantitative information about the tissue biopsy. Furthermore, since the contrast in the image is generated by optical path length variation, QPI images are label-free. Since the pixel values in QPI images represent a physical property of tissue and will not vary due to inconsistency in staining, QPI images are suitable for the development of a quantitative, automated biopsy evaluation method that is calibration-free.

As shown in previous publications, QPI methods have been extensively applied for both quantitative and qualitative evaluation of tissue health. Wang et. al. used the mean and mode of the phase to separate benign and malignant areas in prostate tissue biopsies [9]. Sridharan et. al. showed that the a scattering signature, calculated using phase values, in the stroma adjoining a tumor in prostate tissue biopsies is a useful marker for predicting recurrence of cancer after prostatectomy [10,11]. Majeed et al. showed that QPI images of breast tissue biopsies are able to resolve the morphological markers required for diagnosing malignancy [12]. A number of publications have also explored refractive index signatures in epithelial cell nuclei, obtained using QPI, to address a number of questions related to cancer diagnosis and prognosis [13-16].

In this work, we show that image segmentation algorithms can be robustly applied to QPI images of breast tissue biopsies to separate different types of tissue. We employ a texton-based approach to segmentation outlined by Varma et. al. [17]. Segmentation is an important step in the development of a high-throughput quantitative diagnosis and prognosis method since different types of tissue need to be separated within images before parameters, that are indicative of patient health, can be computed over them.

## 2. EXPERIMENTAL PROCEDURES

### 2.1 Sample

For this study we used an unstained tissue microarray (TMA) comprising of biopsy cores from 400 different patients having different grades and stages. We obtained this TMA from our collaborator at the University of Illinois at Chicago (UIC) Dr. Andre Balla. Details about the sample preparation for this TMA have already been published elsewhere [12]. All the procedures related to human subjects were approved by the University of Illinois at Urbana-Champaign institute review board (Protocol # 13900).

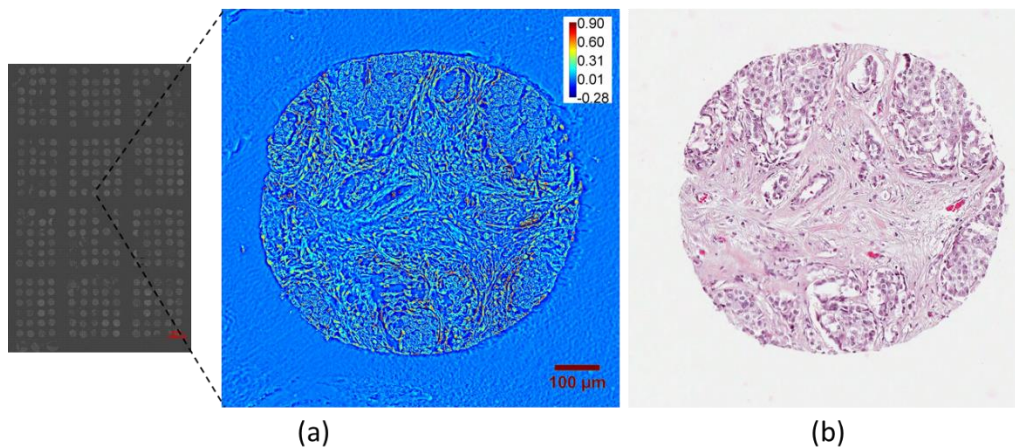


Figure 1 (a) Stitched phase image of an entire TMA slide (45 mm x 20 mm) and a single core. Color bar in radians.  
 (a) H&E stained tissue image of a corresponding core in a parallel section of tissue.

## 2.2 Optical Setup

The QPI modality chosen for this study is Spatial Light Interference Microscopy (SLIM). Details of the system have already been published in [18]. SLIM was chosen for this work because of its ability to provide diffraction limited resolution as well as sub-nanometer temporal and spatial path-length sensitivity [18]. Image segmentation on SLIM images of prostate tissue biopsies, using the methods discussed in this paper, have also been shown in a previous publication from our lab [19].

The imaging was carried out at 40x magnification using a Visual C++ based microscope slide-scanning platform developed in-house. As discussed in [12], with a scanning time per slide of roughly 2 hours at a sampling rate of 0.125  $\mu\text{m}/\text{pixel}$ , the speed of our slide scanner is comparable with that of commercial microscope scanners. The raw images obtained from the scanner were processed using a MATLAB based code and were stitched into a single slide image using a Visual C++ based stitching code [12]. Figure 1 illustrates the stitched phase image of an entire slide as well as the image of a single TMA core. Morphological comparison between the phase image for the core and the H&E stained tissue image for a core from a parallel tissue section is also shown.

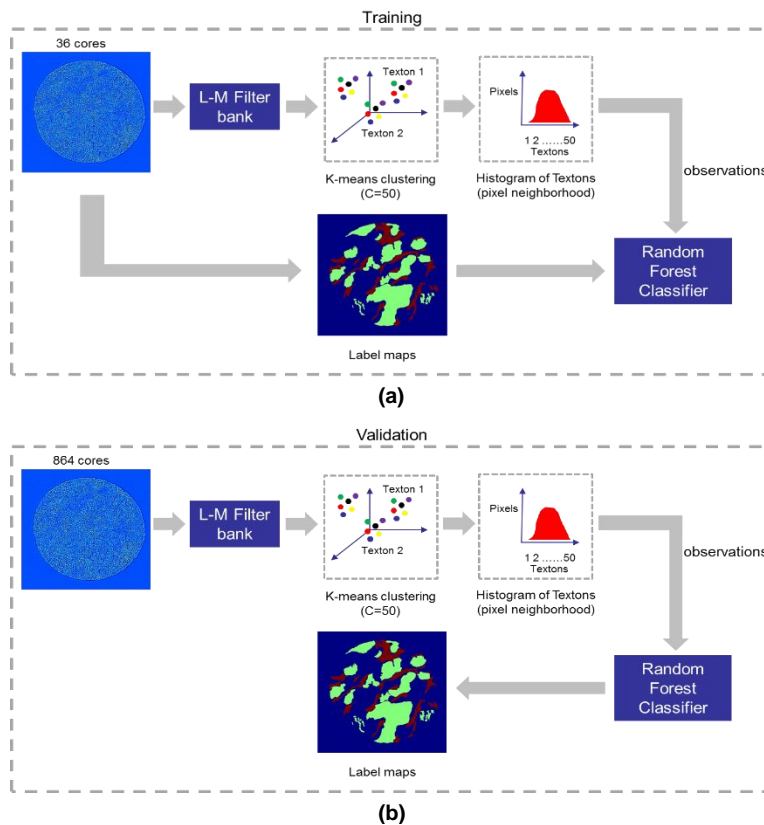


Figure 2. Flow chart representing the image segmentation method for (a) Training step (b) Validation step

## 2.3 Texton-based segmentation

Figure 2 summarizes the image segmentation algorithm. As stated earlier, we employed a texton-based approach to pixel classification originally described by Varma et. al [17]. We used the Random Forest Classifier in MATLAB to perform the classification of each pixel as either epithelial, stromal or luminal. The classifier was trained as follows. From the total dataset of around 900 cores, we randomly selected 36 cores for training. Using ImageJ, the pixels within the images of these 36 cores were labelled as either epithelial, stromal or luminal. These labels along with observations for each pixel in the training data were fed to the classifier which built a model relating the two. The observations themselves were generated as follows. A Leung-Malik (LM) filter bank was applied to each pixel in the phase image of each core

[17]. This filter bank computes the gradient along different orientations and at different spatial scales, for each pixel, and generates a response vector. The response vectors for all the pixels were next clustered using the K-means algorithm in MATLAB to find 50 cluster centers. The position vector corresponding to each center is called a ‘texton’ and characterizes the textural content of the images [17]. The histogram of these textons within the neighborhood of each pixel was next computed. This histogram is the observation for each pixel and, hence, the classifier performed regression to find the relationship between the pixel label and the pixel observation.

During validation, the test data was processed in the same way as the training data to generate the histogram or observation for each pixel. Based on the model learnt during training, the classifier then generated a label for each pixel to perform the segmentation.

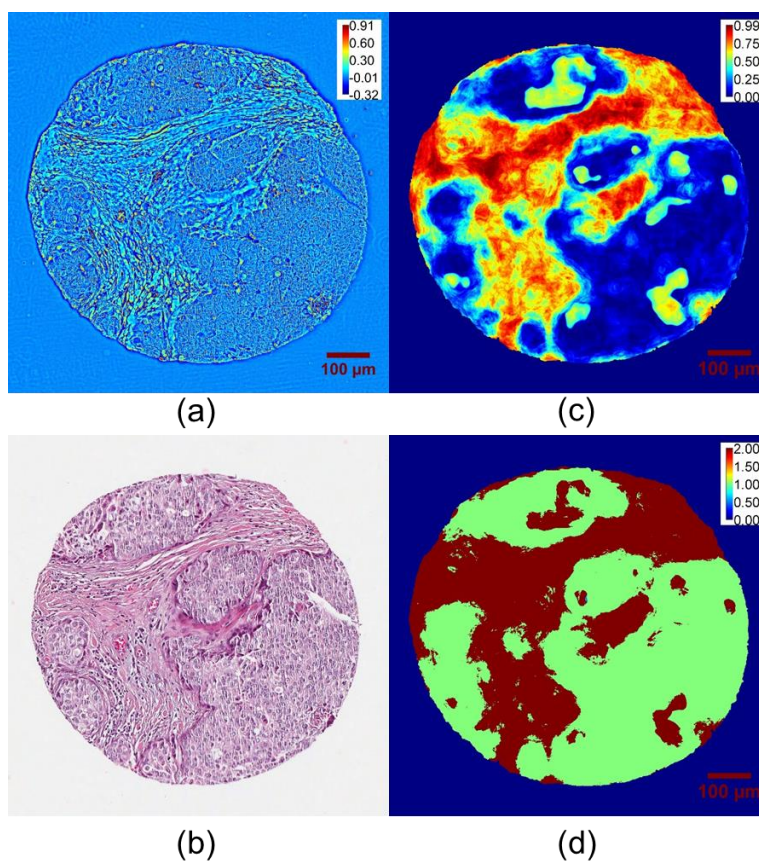


Figure 3. Segmentation results for one core. Comparison between (a) SLIM image (color bar in radians) (b) H&E image of a core from a parallel section of tissue (c) Probability map showing the probability of a pixel being stromal (d) Segmentation map with luminal pixels = 0, epithelial pixels = 1 and stromal pixels = 2.

### 3. RESULTS

Figure 3 illustrates the results of the segmentation algorithm for one of the non-training cores. Figure 3 (a) shows the phase image for the core while Figure 3 (c) shows the probability of a pixel within the core belonging to stromal tissue. In this probability map, a value of 1 represents certainty that the pixel is stromal, a value of 0 represents certainty that the pixel is epithelial and a value of -1 represents certainty that the pixel is luminal. As can be seen by comparing the two images, areas identified as stromal in the phase image show a higher probability of being labelled stromal in the probability map compared to areas identified as epithelial or luminal. Figure 3 (d) shows the final segmentation map generated by thresholding the probability map in Figure 3(c). The threshold is set at 0.5 and in the final label map stromal pixels have value 2, epithelial pixels have value 1 and luminal pixels have value 0.

To test the accuracy of our segmentation algorithm, we generated segmentation maps for the 36 training cores using the classifier and compared them pixel wise with the ground truth (manually labelled images). The classification accuracy for these training cores was 89.6%. This, however, does not represent the actual accuracy of the classifier which needs to be tested for overfitting by performing multi-fold validation. Multi-fold validation will require manual segmentation of a large number of phase images to serve as ground truth and, hence, has been left for future work.

#### 4. SUMMARY AND FUTURE OUTLOOK

We have discussed in this paper an approach for carrying out automatic tissue segmentation on phase images of breast tissue biopsies. This is an important first step in building an automated histopathology platform for cancer diagnosis and prognosis since different tissue types within a biopsy image need to be separated before quantitative parameters that are indicative of tissue health can be computed over them. Such a platform would go a long way towards addressing the current shortcomings in histopathology in terms of both throughput and observer bias.

While the segmentation results for non-training cores look qualitatively promising, we thus far have only been able to compute the (quantitative) segmentation accuracy for training cores. This is because the ground truth, for computing accuracy, at this point is only available for the training set. A rigorous examination of the segmentation accuracy for this classifier would require the ground truth for a larger number of cores. Therefore, our aim for the future is to perform three-fold validation in order to rigorously test the classifier. This will involve generating the label maps, through manual segmentation, of 150 cores. This dataset will then be partitioned into three equal parts of 50 cores each - one part will be used to train the classifier and the other two will be used for validation. This will be repeated twice more so that each of the three parts will have been used to train the data while the other two are used for validation. In this way the true accuracy of the classifier and its susceptibility to overfitting will be assessed.

#### REFERENCES

- [1] World Health Organization, "WHO position paper on mammography screening", (2014).
- [2] World Health Organization, "Latest world cancer statistics Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed", (2013).
- [3] Benard, A., Desmedt, C., Smolina, M. et al., "Infrared imaging in breast cancer: automated tissue component recognition and spectral characterization of breast cancer cells as well as the tumor microenvironment," *Analyst*, 139(5), 1044-1056 (2014).
- [4] Rakha, E. A., Reis-Filho, J. S., Baehner, F. et al., "Breast cancer prognostic classification in the molecular era: the role of histological grade," *Breast Cancer Research : BCR*, 12(4), 207-207 (2010).
- [5] G. Popescu, [Quantitative Phase Imaging of Cells and Tissues] McGraw Hill, (2011).
- [6] C. Mann, L. Yu, C.-M. Lo et al., "High-resolution quantitative phase-contrast microscopy by digital holography," *Optics Express*, 13(22), 8693-8698 (2005).
- [7] Kim, M., [Digital Holographic Microscopy: Principles, Techniques and Applications] Springer, New York (2011).
- [8] Wu, J., Yaqoob, Z., Heng, X. et al., "Harmonically matched grating-based full-field quantitative high-resolution phase microscope for observing dynamics of transparent biological samples," *Optics Express*, 15(26), 18141-18155 (2007).
- [9] Wang, Z., Tangella, K., Balla, A. et al., "Tissue refractive index as marker of disease," *Journal of Biomedical Optics*, 16(11), 116017 (2011).
- [10] Sridharan, S., Macias, V., Tangella, K. et al., "Prediction of Prostate Cancer Recurrence Using Quantitative Phase Imaging," *Scientific Reports*, 5, 9976 (2015).
- [11] Wang, Z., Ding, H. and Popescu, G., "Scattering-phase theorem," *Optics letters*, 36(7), 1215-1217 (2011).
- [12] Majeed, H., Kandel, M. E., Han, K. et al., "Breast cancer diagnosis using spatial light interference microscopy," *Journal of Biomedical Optics*, 20(11), 111210-111210 (2015).
- [13] Uttam, S., Pham, H. V., LaFace, J. et al., "Early Prediction of Cancer Progression by Depth-Resolved Nanoscale Mapping of Nuclear Architecture from Unstained Tissue Specimens," *Cancer Research*, 75(22), 4718-4727 (2015).
- [14] Wang, P., Bista, R., Bhargava, R. et al., "Spatial-domain low-coherence quantitative phase microscopy for cancer diagnosis," *Optics Letters*, 35(17), 2840-2842 (2010).

- [15] Wang, P., Bista, R. K., Khalbuss, W. E. et al., "Nanoscale nuclear architecture for cancer diagnosis beyond pathology via spatial-domain low-coherence quantitative phase microscopy," *Journal of Biomedical Optics*, 15(6), 066028-066028-8 (2010).
- [16] Bista, R. K., Wang, P., Bhargava, R. et al., "Nuclear nano-morphology markers of histologically normal cells detect the "field effect" of breast cancer," *Breast Cancer Research and Treatment*, 135(1), 115-124 (2012).
- [17] Varma, M. and Zisserman, A., "A Statistical Approach to Texture Classification from Single Images," *International Journal of Computer Vision*, 62(1-2), 61-81 (2005).
- [18] Wang, Z., Millet, L., Mir, M. et al., "Spatial light interference microscopy (SLIM)," *Optics Express*, 19(2), 1016-1026 (2011).
- [19] Nguyen, T. H., Sridharan, S., Macias, V. et al., "Prostate cancer diagnosis using quantitative phase imaging and machine learning algorithms." 9336, 933619-933619-10.