Fractal analysis for classification of breast carcinoma in optical coherence tomography

Amanda C. Sullivan
John P. Hunt
Amy L. Oldenburg
Fractal analysis for classification of breast carcinoma in optical coherence tomography

Amanda C. Sullivan,a John P. Hunt,b and Amy L. Oldenburgc
aUniversity of North Carolina at Chapel Hill, Department of Physics and Astronomy, Phillips Hall, Chapel Hill, North Carolina 27599-3255
bUniversity of North Carolina School of Medicine, Department of Pathology and Laboratory Medicine, Chapel Hill, North Carolina 27599-7525
cUniversity of North Carolina at Chapel Hill, Department of Physics and Astronomy, Biomedical Research Imaging Center, Phillips Hall, Chapel Hill, North Carolina 27599-3255

Abstract. The accurate and rapid assessment of tumor margins during breast cancer resection using optical coherence tomography (OCT) has the potential to reduce patient risk. However, it is difficult to subjectively distinguish cancer from normal fibroglandular stromal tissues in OCT images, and an objective measure is needed. In this initial study, we investigate the potential of a one-dimensional fractal box-counting method for cancer classification in OCT. We computed the fractal dimension, a measure of the self-similarity of an object, along the depth axis of 44 ultrahigh-resolution OCT images of human breast tissues obtained from 4 cancer patients. Correlative histology was employed to identify distinct regions of adipose, stroma, and cancer in the OCT images. We report that the fractal dimension of stroma is significantly higher than that of cancer ($P < 10^{-5}$, t-test). Furthermore, by adjusting the cutoff values of fractal dimension between cancer, stroma, and adipose tissues, sensitivities and specificities of either 82.4% and 88.9%, or 88.2% and 81.5%, are obtained, respectively, for cancer classification. The use of fractal analysis with OCT could potentially provide automated identification of tumor margins during breast-sparing surgery.

Keywords: breast cancer; optical coherence tomography; fractal dimension; box-counting method.

Paper 10620R received Nov. 19, 2010; revised manuscript received Mar. 20, 2011; accepted for publication Apr. 20, 2011; published online Jun. 2, 2011.

1 Introduction

According to the American Cancer Society, approximately 261,100 women in the United States will be diagnosed with breast cancer and 39,480 will die from it in 2010 alone, making breast cancer the second leading cause of cancer death in the United States.1 Reducing local recurrence after treatment correlates with a decrease in the overall breast cancer mortality rate.2 The local recurrence rate following breast-conserving therapy, which involves local resection (also referred to as “lumpectomy”) with or without adjuvant chemotherapy or radiation therapy, can be reduced considerably by obtaining histologically negative margins.2,3 The assessment of surgical margins at the time of the patient’s surgical procedure is therefore an important goal that can be addressed by optical techniques which have the potential to increase the speed and accuracy of cancer detection. Current research in this area includes the use of elastic scattering spectroscopy,4 Raman spectroscopy,5 fluorescence,6 and diffuse reflectance spectroscopy.7 The use of optical coherence tomography (OCT) is of interest for margin assessment because, unlike other optical techniques, it provides depth-resolved micrometer-scale imaging of tissue structure that can be registered with subsequent histology.8,9 The ability for OCT to spatially map tissue properties may be particularly advantageous in heterogeneous tissues, or for detecting small metastases.

OCT is a real-time, noninvasive imaging modality that employs interferometry to detect backscattered near-infrared light to render two-dimensional (2D) or three-dimensional (3D) images of tissue penetrating ~1 to 2 mm beneath the tissue surface.10,11 In addition, it can be combined with a catheter or endoscope to image interior tissues.12,13 The axial resolution of OCT is micrometer-scale, and is determined by the coherence length of the light source which is inversely proportional to the bandwidth.10 The high resolution and millimeter penetration depth of OCT provides images on a scale similar to histology, and as such, it has been dubbed “in vivo optical biopsy.”12 These features make OCT an ideal candidate for aiding in the detection and treatment of breast cancer.8,9

One of the most significant ways OCT can improve breast cancer treatment is through the intraoperative assessment of surgical margins.8,9,13 The current method used during a lumpectomy involves sending resected tissue to a pathologist for a frozen section analysis while the patient remains on the operating table. This is seldom performed because of technical and artifactual issues that limit morphologic assessment. Typically, histopathological margin assessment is performed postoperatively, and the finding of positive margins requires the patient to undergo an additional surgical procedure. With OCT, the imaging and analysis can be done in real-time, potentially completing the process more quickly and consequently reducing patient risk.8,9,13 Potentially, OCT can be used to rapidly scan the entire surgical margin directly in the patient, in contrast to...
the usual pathological evaluation with a limited sampling of the specimen margins. Therefore, OCT has the potential to provide increased sensitivity and to reduce recurrence.

Before this can become common practice, an accurate correlation between the OCT images and the histological standard needs to be developed. There are two main tissue types in a normal female human breast. The majority of the breast is composed of adipose tissue, which stores fat in cells called adipocytes. The rest of the breast contains systems of ducts and lobules, along with the fibrous connective tissue that supports them. This connective tissue is known as fibro glandular stroma (hereafter referred to as stroma). Several studies aimed at determining how normal and cancerous tissue types appear in OCT images have previously been performed, with limited results. In general, these studies have found that identifying adipose is fairly straightforward due to its unique pattern of scattering light. However, more difficulty is found in differentiating stroma from carcinoma, either by inspection or by computational methods. Since determining tumor margins requires knowing which areas contain cancer, it is necessary to be able to distinguish between cancer and stroma.

In this paper, we investigate the use of fractal dimension to characterize different tissue types. Fractal dimension, which is based largely on Mandelbrot’s work with the properties of fractals in nature, is a measure of an object’s complexity and self-similarity. The concept of fractal dimension has been used as a way of quantifying the texture, or physical appearance, of an image. Because cancer is characterized by disorder and irregularity in tissues, it is reasonable to expect that fractal analysis may have utility in cancer identification. Without a quantitative metric such as fractal dimension, texture analysis of breast tissue can be a subjective study, with different observers often producing different results. Fractal analysis has previously been employed to detect the presence of cancer in mammography and microscopy images. Unlike these previously employed imaging modalities, in OCT texture is dominated by coherent speckles, which may exhibit a different fractal property. Fractal analysis has recently been applied to OCT images of arterial tissue and has shown promising results in differentiating tissue layers. In this study, fractal analysis is applied to OCT images of histologically correlated breast tissue types to determine if their fractal dimensions provide an accurate way of distinguishing them. Unlike the previous study of arteries, we employ an ultrahigh resolution (~2 to 3 μm) OCT system to provide a finer scale for resolving the pattern of scatterers in normal and cancerous breast tissues.

2 Methods

2.1 Tissue Imaging and Pathology

Breast tissue specimens were obtained from mastectomy patients through the Biomedical Research Imaging Center according to protocols approved by the Institutional Review Board at the University of North Carolina at Chapel Hill, fixed with 10% formalin for preservation, and deidentified. (We note that a previous study has shown that formalin causes tissue shrinkage, resulting in increased scattering. While we do not expect it to affect the disorder in the texture of the scattering, potential differences between fresh and fixed tissue require future investigation.) Slices with lengths of ~15 mm and widths of 3 to 5 mm were cut from the samples and imaged individually. They were soaked in dulbecco’s phosphate-buffered saline (Mediatech, Inc.) to prevent dehydration during imaging. A total of nine 3D OCT image sets were acquired from tissue samples from four different patients.

The OCT system employed in this study is a spectral-domain system based upon the use of a line camera spectrometer. Briefly, the OCT system consisted of a diode laser-pumped Ti:Sapphire femtosecond laser (Griffin, KMLabs, Inc.) producing light with a center wavelength of ~810 nm and bandwidth of ~125 nm, corresponding to an axial resolution of ~3 μm. A single-mode fiber directed light into a free-space interferometer with reference and sample arms. The sample arm consisted of x- and y-lateral scanning mirrors and an imaging lens providing ~16 μm of lateral resolution. We note that this larger lateral resolution affords a Rayleigh range of 250 μm, which maintains a relatively constant coherence volume over twice this length, and may be important for avoiding depth-dependent variations in the speckle contrast. The output of the interferometer was directed using a single-mode fiber into a custom spectrometer comprised of a transmission grating, imaging lens, and high speed line CCD camera (Dalsa Piranha 2). The spectral resolution of the spectrometer was designed to provide 2.05-mm imaging depth in free space (corresponding to nominally 1.5-mm depth in tissues). The camera axial scan rate of 5 kHz and exposure time of 190 μs were chosen to maximize the signal while minimizing the amount of autocorrelation artifacts and saturation from specular reflections. Sequential B-mode (x-z) OCT scans were performed while incrementing in y to produce 3D images. Images were sampled over 3 to 5 mm × 3 to 5 mm × 1.5 mm (x × y × z) into 1000 to 2500 × 121 to 201 × 1024 pixels, respectively. With these sampling parameters, the pixel size was 1.46 μm in z corresponding to ~2× oversampling, which was constant throughout the subsequent box counting analysis.

After acquiring the OCT images, the tissue samples were inked with different colors of tissue dye from the Davidson Marking System (Bradley Products, Inc.) on four sides in order to maintain the orientation. The samples were processed by the Translational Pathology Laboratory (TPL) at the University of North Carolina at Chapel Hill, where they were paraffin embedded, sectioned in the same plane as the OCT B-mode images into 5 μm slices, and stained with hematoxylin and eosin (H&E) for analysis with a light microscope. The resulting microscopy slides were analyzed by a pathologist who classified the tissue and cancer types that were seen in each sample. Since the majority of the slides contained more than one tissue type, subregions within the slides were analyzed individually. The pathologist indicated which sections of each slide were adipose, stroma or in situ or invasive cancer, so they could be compared to their corresponding 3D OCT image sets. OCT images that had good agreement with histology were picked out first. Agreement was determined by comparing the positions and orientations of large structures, such as blood vessels and ducts, as well as regions of adipose, since they were easily identifiable in both the microscopy slides and the OCT B-mode images. From those that agreed with histology, OCT images containing large contiguous regions consisting of a single tissue type were chosen and classified as the predominant tissue type in them, based on the pathologist’s analysis. Attempts were made to choose areas of
tissue that were completely distinct from one another, so that no region contained multiple tissue types. However, this was difficult to achieve, particularly in regions of carcinoma, because there were few contiguous regions large enough for analysis. Therefore, the regions with only small portions of the nondominant tissue types were chosen. Regions of stroma were chosen from six different tissue samples taken from three of the patients, while regions of adipose and cancer were chosen from seven and eight tissue samples, respectively, taken from all four patients. In the end, 12 regions of adipose, 15 regions of stroma, and 17 regions of cancer, for a total of 44 distinct regions, were chosen for analysis.

2.2 Image Processing and Fractal Analysis

Next, the contrast in each OCT image was adjusted to achieve a constant signal-to-noise ratio as follows. A background level \( B \) was obtained from the mean value within an empty top portion of the image, and the maximum signal \( S \) was obtained from the mean value within the most highly scattering portion of the image. The image pixel values \( I \) were then rescaled to an 8-bit grayscale value \( I' \) according to \( I' = \frac{255(1-B)}{S-B} \).

The fractal dimensions within each region of interest were then calculated using a one-dimensional box-counting method.\(^{16,17}\) Each column within a given region was first divided into intervals of a specific length, \( l_i \), with the first interval starting at the top surface of the tissue and the last interval ending 768 pixels below the surface. The next step consisted of counting the number of intervals, \( N_i \), that contained any pixels above a certain threshold value. A threshold value of 70 was chosen by examining the distribution of pixel values in the images and finding a level above which pixels are predominantly signal distinguished from background noise. The threshold value was held constant for all of the regions analyzed. After counting, the interval length was decreased by a factor of 2 and the process repeated. The lengths of the intervals ranged from an initial size of 128 pixels down to the final size of one. The fractal dimension of the column was then calculated as the slope of the least squares regression line of \( \log(N_i) \) versus \( \log(1/l_i) \). A sample plot with the regression line is shown in Fig. 1. Since each column is a one-dimensional section of the image, all of the fractal dimension values should be between zero, the dimension of a point, and one, that of a line. We note that this is different from a previous OCT-based fractal analysis where the axial scan line contour was box-counted in two dimensions.\(^{24}\) The fractal dimension of each region of interest was assigned as the mean of the individual column fractal dimensions.

2.3 Statistical Analysis

In order to associate fractal dimension values with adipose, stroma, and cancer, the mean and standard deviation of the fractal dimensions were tabulated for the regions classified as each tissue type. A two-tailed \( t \)-test was used to determine if the mean fractal dimensions for all three tissue types were significantly different from one another. The \( t \)-values and associated \( P \)-values were calculated to test the null hypothesis between each pair of tissue type: adipose-cancer, adipose-stroma, and cancer-stroma. Since three \( t \)-tests were performed on the same data set, a Bonferroni correction was applied by increasing the \( P \)-values by a factor of 3.

The sensitivity and specificity were computed to determine how effective a test using the box-counting algorithm would be at classifying cancer. We note that, due to the selection of homogeneous tissue regions for analysis, these values may not reflect the true predictive utility of this technique in heterogeneous tissues, and further study is needed to establish this. Sensitivity was defined as the percentage of actual cancer-positive tissues that are correctly identified (true positive rate), while specificity was defined as the percentage of actual cancer-negative tissues (stroma and adipose) that were correctly identified (true negative rate). These were computed by choosing upper and lower cutoff values of fractal dimension for classifying cancer compared to noncancer (adipose or stroma). A receiver operating characteristic (ROC) curve was generated by varying the cutoff values over the entire range of fractal dimensions (0.516 to 0.900), and the cutoff values providing the highest sensitivity at each value of specificity were recorded.

3 Results and Discussion

The box-counting algorithm was used to calculate the fractal dimension of each column within a region of interest, providing a distribution of values from which the mean and standard deviation was computed. An example histogram of the fractal dimension distributions for each of the three tissue types (adipose, stroma, and cancer) is shown in Fig. 2, along with the corresponding OCT and microscopy images. As expected, all of the fractal dimensions were between zero and one. The width of each distribution, quantified by its standard deviation, provides an indication of the homogeneity of tissue in the particular region. Since it was not always possible to choose regions with only a single tissue type, some of the regions contained small areas of tissue with a different classification than the predominant tissue type. The regions with more of these small areas had wider distributions resulting in higher standard deviations. Conversely,
the more homogeneous regions had narrower distributions since the fractal dimension values were more uniform.

The image sections containing adipose had the lowest fractal dimensions, with a mean of 0.68 ± 0.06. Based on visual inspection of the OCT images, this result is expected because the interior of adipose (fatty) cells do not scatter light as strongly as either cancer or stroma, but the boundaries between cells and extracellular space exhibit stronger scattering due to the refractive index mismatch between fat and water. In comparison to stroma and cancer, the individual columns in a region of adipose have a much more ordered pattern, with larger contiguous regions corresponding to the interiors of the adipocytes. Stroma and cancer both have cells with internal structures that produce a more disordered (complex) scattering pattern. Because the OCT images of adipose tissues are less complex than either cancer or stroma, their fractal dimension values are lower.

The fractal dimensions for the regions of cancer and stroma were closer together, with a mean of 0.79 ± 0.04 for cancer regions and 0.85 ± 0.03 for stroma regions. The larger fractal dimension of stromal tissue is contrary to what is expected based upon the histological characteristic of tumors as having increased disorder.19 Further study is needed to understand why light microscopy, which contrasts tissue attenuation, exhibits a different fractal pattern for cancer tissue than OCT, which contrasts tissue backscattering.

Table 1 summarizes the results of the two-tailed $t$-tests. The low $P$-values indicate that the fractal dimensions for each tissue type are significantly different from one another. The difference between adipose and either of the other two tissue types was expected based on the visible distinctions in the OCT images and results from previous studies. However, the difference in the fractal dimensions of carcinoma and stroma is more significant than that found from other methods used previously.9,15 While these previous methods were often able to identify suspicious regions of tissue, they were not always successful at differentiating between cancer or stroma in these regions.9,15 The box-counting method appears to be able to separate the two tissue types with more certainty.

Since the mean cancer fractal dimension value is intermediate between adipose and stroma, we identified upper and lower cutoff values between which a sample was classified as cancer (positive), and outside of which the sample was classified as normal (negative). Using cutoff values of 0.75 and 0.83, the sensitivity of the box-counting method is found to be 82.4% and the specificity is 88.9%. Alternatively, a higher sensitivity of 88.2% can be obtained at the expense of a lower specificity of 81.5% when increasing the upper cutoff value to 0.85. A ROC curve (Fig. 3) illustrates this tradeoff between sensitivity and specificity. The area under the curve was calculated to be

<table>
<thead>
<tr>
<th>Tissue pair</th>
<th>T-value</th>
<th>Degree of freedom</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adipose–cancer</td>
<td>5.323</td>
<td>27</td>
<td>2.06E-4</td>
</tr>
<tr>
<td>Adipose–stroma</td>
<td>9.038</td>
<td>25</td>
<td>7.62E-7</td>
</tr>
<tr>
<td>Cancer–stroma</td>
<td>6.182</td>
<td>30</td>
<td>2.66E-6</td>
</tr>
</tbody>
</table>

Table 1 Results of a two-tailed $t$-test for each of the tissue type pairs.
Fig. 3 ROC curve for a sensitivity and specificity test for classifying cancer versus adipose or stroma based on fractal dimension in OCT images. The area under the curve is calculated to be 0.938.

0.938 using the trapezoidal rule. This value for the area is close to the ideal value of 1, indicating that the fractal dimension is a promising new metric for discriminating between cancer and normal tissue. We note that the specificity by this method is significantly improved from a previous method for automated cancer detection in OCT images (which was 56% to 58%), however the sensitivity in this study was lower. The fractal method is also capable of higher specificity, at loss of sensitivity, compared to a subjective analysis of OCT breast cancer images. The particular advantages of automated methods like this one includes the ability to adjust the balance between sensitivity and specificity by adjusting the parameter range for classifying cancer, and to assess large areas of tissue more rapidly than a human operator.

Given that the number of patients in this study was limited, we calculated the power of this study using the number of patients as the degrees of freedom, rather than the total number of samples that were used in the t-test above. This addresses the possibility that interpatient variability is a larger source of error than variability between samples from a single patient. Assuming that the actual distribution of fractal values are a normal distribution with mean and standard deviation given by the values measured in this study, and using a significance level $\alpha = 0.05$, we find the power is 80.7% for stroma and cancer, and 82.4% for adipose and cancer, which represents the probability of measuring a statistically significant difference between these groups. Importantly, the fractal dimension distributions measured in this study will motivate the choice of sample size in future study designs.

4 Conclusion

Due to variations in biological structure, the optical scattering properties of breast tissue differ among tissue types, which affects the texture in OCT images. Since these variations in OCT signal are often difficult to detect through visual inspection of the images alone, we demonstrate a quantitative method for distinguishing the different tissue types. We found that the fractal dimensions of OCT images of breast tissue, determined using a one-dimensional box-counting algorithm, are significantly different depending upon classification of the tissue region as adipose, cancer, or stroma using correlative histology. Therefore, fractal analysis could potentially be used to locate carcinoma in OCT images. The sensitivity and specificity of a test based on the fractal dimensions were 82.9% and 88.9%, respectively. Although higher values for both sensitivity and specificity are needed to provide assessment similar to histology, these first results are very promising. Due to the limited number of samples, we did not tabulate the various types of carcinoma present in the tissues, which may have caused increased variability in the results. Future studies involving a greater number of tissue samples could aid in improving upon these results by allowing us to control for different types of carcinoma. Of particular interest would be determining if differences between carcinoma and benign fibroglandular stroma would be present in both invasive and in situ carcinoma as well as assessing differences between ductal and lobular carcinomas. It may also be productive to incorporate other metrics into the algorithm used previously in OCT, such as the depth-dependent attenuation rate, Fourier spatial frequency analysis, and spatial gray-level dependence matrices. While the lower fractal dimension of adipose tissue was expected based on the ordered texture of this tissue type in OCT, it is not yet understood why stromal tissue exhibits a higher fractal dimension than cancer. A better understanding is needed of the morphology of cancerous and stromal tissues and how they relate to the observed differences in the complexity of the OCT images. It is also important to note that high resolution is important for extracting information from speckle in coherence imaging. Therefore, the results in our study may be partially enhanced by the use of an ultrahigh resolution OCT system. However, future study is needed to determine the resolution necessary for effective fractal analysis in breast cancer. Because OCT can provide real-time visualization over a broad tumor margin during surgical resection, the development and refinement of these margin assessment techniques can positively impact breast cancer treatment.

Acknowledgments

We would like to acknowledge Etta D. Pisano and Elodia Cole of the Medical University of South Carolina Department of Radiology, for providing breast specimens for this study. This project was supported by startup funds from the Department of Astronomy and Physics at the University of North Carolina at Chapel Hill.

References