Texture analysis of optical coherence tomography images: feasibility for tissue classification

Kirk W. Gossage
Tomasz S. Tkaczyk
University of Arizona
Biomedical Engineering Program
Tucson, Arizona 85721-0104
E-mail: kgossage@email.arizona.edu

Jeffrey J. Rodriguez
Jennifer K. Barton
University of Arizona
Biomedical Engineering Program
and
Department of Electrical and Computer Engineering
Tucson, Arizona 85721

Abstract. Optical coherence tomography (OCT) acquires cross-sectional images of tissue by measuring back-reflected light. Images from in vivo OCT systems typically have a resolution of 10 to 15 μm, and are thus best suited for visualizing structures in the range of tens to hundreds of microns, such as tissue layers or glands. Many normal and abnormal tissues lack visible structures in this size range, so it may appear that OCT is unsuitable for identification of these tissues. However, examination of structure-poor OCT images reveals that they frequently display a characteristic texture that is due to speckle. We evaluated the application of statistical and spectral texture analysis techniques for differentiating tissue types based on the structural and speckle content in OCT images. Excellent correct classification rates were obtained when images had slight visual differences (mouse skin and fat, correct classification rates of 98.5 and 97.3%, respectively), and reasonable rates were obtained with nearly identical-appearing images (normal versus abnormal mouse lung, correct classification rates of 64.0 and 88.6%, respectively). This study shows that texture analysis of OCT images may be capable of differentiating tissue types without reliance on visible structures. © 2003 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1577575]

Keywords: image processing; speckle; brodatz; lung.

Paper 02034 received May 29, 2002; revised manuscript received Sep. 16, 2002, and Nov. 26, 2002; accepted for publication Jan. 6, 2003.

1 Introduction

1.1 Optical Coherence Tomography

In optical coherence tomography (OCT), cross-sectional images are created by measuring near-infrared light back-reflected from tissue.1,2 Typical in vivo OCT systems have a resolution on the order of 10 to 15 μm, and are therefore unable to image subcellular structures. Even with ultrahigh (better than 2 μm) resolution systems, subcellular level imaging of deep structures may be limited by phase distortions in turbid tissues. Therefore, most identifications of tissue type and pathology have relied on the presence or absence of structures and layers. For instance, in the gastrointestinal tract,3–6 bladder,7 and prostate,8 several groups have reported that OCT images of normal tissue show a well-organized composition, with layers such as mucosa, lamina propria, muscularis mucosae, and submucosa, and features such as colonic crypts and gastric pits. However, in pathologies, including Barrett-related and prostate adenocarcinoma, and bladder transitional cell carcinoma, common findings include disorganization and homogenization of the affected areas.

1.2 Speckle in OCT Images

OCT images of many normal tissues, for instance sclera or aorta, have few structures or none, and may appear to simply show an apparent exponential decrease in intensity with depth, assuming the signal is in the single-scattering regime. Therefore, it may appear that OCT would not be useful in classifying uniform-appearing tissues or pathologies in these tissues. However, examination of apparently homogeneous OCT images reveals that they display a characteristic texture that is due to speckle.

Speckles appear everywhere in space whenever an optically rough surface is illuminated with coherent light. We have observed that the random field distribution of speckle is stationary in time (for stationary objects) and is a function of spatial coordinates. The greater the size range of the surface elements, the greater will be the angular range of scattering produced. The structural size of individual speckles, however, is determined by the size of the illuminated area of the object in free-space propagation, or the aperture angle in an optical imaging system. The speckle pattern may be regarded as being produced by the coherent superposition of the interference fringes of the waves falling onto a plane of observation. This 2-D speckle theory for surface roughness can be extended to a 3-D model for OCT imaging.

The origin of speckle in OCT images was previously outlined by Schmitt.9 In his paper, it is proposed that two types of speckles appear in OCT images. The first is due to interference from multiply scattered photons. This type of speckle is generally small, typically a single pixel wide. The second type of speckle results from interference of the wavefronts from multiple scatterers within the OCT focal volume and is typically much larger.

The occurrence of speckle has been noted in similar fields, such as optical metrology with the use of coherent illumination. Numerous applications of speckle can be suggested, in-
cluding measurement roughness and shape, determination of strain and displacement, and monitoring movement. \textsuperscript{10–13} We hypothesize that variations in the quantity and size distributions of scatters could create uniquely different speckle patterns. Therefore it may be possible to use texture analysis to differentiate between uniform-appearing tissue types, based on features of their speckle.

1.3 Texture Analysis

Texture is a result of local variations in brightness within one small region of an image. If the intensity values of an image are thought of as elevations, then texture is a measure of surface roughness.\textsuperscript{14} Texture analysis techniques can be classified into three groups:\textsuperscript{15}

- Statistical technologies: based on region histograms and their moments, they measure features such as coarseness and contrast.
- Spectral technologies: based on the autocorrelation function or power spectrum of a region, they detect texture periodicity, orientation, etc.
- Structural technologies: based on pattern primitives, they use placement rules to describe the texture.

To the authors’ knowledge, texture analysis has not previously been performed on OCT images. However, a large body of literature exists for texture analysis of ultrasound, magnetic resonance imaging (MRI), computed tomography (CT), fluorescence microscopy, light microscopy, and other digital images. For example, in the eye, texture analysis of ultrasound images has been used to differentiate histological types of intraocular melanoma.\textsuperscript{16} In another study, posterior capsule opacification was assessed from digital camera images using the texture analysis techniques of variance measurement and co-occurrence.\textsuperscript{17} Texture analysis has also been used to evaluate ultrasound images of the prostate.\textsuperscript{18} Other optical imaging modalities have utilized texture analysis, such as fluorescence microscopy images of colonic tissue sections\textsuperscript{19} and light microscopy images of the chromatin structure in advanced prostate cancer.\textsuperscript{20}

Details of texture analysis procedures can be found in many signal-processing references. For example, see Refs. 14, 21, and 22. In this paper we describe a preliminary study to determine if two specific types of statistical texture analysis, spatial gray-level dependence matrices (SGLDM) and Fourier spatial frequency domain techniques, can be used to correctly classify OCT images of various tissues.

2 Methods

2.1 OCT System

The OCT system used in this study is similar to one described previously.\textsuperscript{23} The source was a superluminescent diode with a center wavelength of 1300 nm and a bandwidth of 49 nm. The light was carried to the sample arm by an optical fiber (SMF-28, Corning) with a 9-\textmu m core. The light was collimated into a 8.1-mm diameter beam and focused onto the sample by a 50-mm focal length lens with a 19-mm clear aperture. This gives the system an NA of 0.081. The axial and lateral resolutions of the system were 16 and 14 \textmu m in air, respectively.

The detector signal was coherently demodulated and the magnitude sampled by a data acquisition board. Images consisting of 1024 a-scans (columns) with 256 pixels/a-scan were obtained form an area 1 mm in lateral dimension and 0.25 mm deep. The images were taken from just under the air/tissue interface, so that they contained only the bulk tissue. Ten images of each tissue type were taken from various locations on the tissue specimens. Care was taken to ensure that all instrument settings, except detector gain, remained the same from image to image.

The original images had a pixel size of approximately 1 \times 1 \mu m. Modified images were created by averaging 4 \times 4 blocks of pixels. Thus the new image type consisted of 64 rows \times 256 columns. The pixel size (~4 \times 4 \mu m) in the modified images was still far less than the resolution of the system. Studies using unaveraged image data yielded less reproducible results than averaged images, perhaps because of the random nature of the noise and speckle from multiply-scattered photons.

2.2 Tissue Imaging

The specimens imaged in this study were freshly excised tissues from a single 15-week-old p53 double knockout mouse. The specimens included two portions of normal lung tissue, two portions of abnormal lung tissue (hyperplasia with extensive inflammation), a 1 \times 1 cm section of dorsal skin, and two portions of testicular fat. Under a protocol approved by the University of Arizona’s Institutional Animal Care and Use Committee, a technician surgically removed the tissue specimens. After removal, the specimens were covered with gauze, moistened with saline solution, placed in sealed containers, and refrigerated until imaged.

2.3 Image Enhancement and Grouping

The raw OCT image files contained floating-point numbers that corresponded to data acquisition board voltage levels. Three operations were performed on each OCT image file to improve contrast and achieve a uniform distribution of intensities over a restricted gray-scale range. First, the logarithm of each number was taken. Next, the brightness value of each pixel was scaled to fill integer gray levels between 0 and 255. To do this, the minimum and maximum values were recorded during the image data retrieval. The scaling was performed by

$$g = \left[ \frac{g_0 - g_{\min}}{g_{\max} - g_{\min}} \times 255 \right],$$

where $g$ is the integer-modified gray level of a given pixel, $g_0$ is the original float value of the given pixel, $g_{\min}$ and $g_{\max}$ are the minimum and maximum floating-point values present in the raw image data, and $\lfloor \rfloor$ represents the floor operator (the greatest integer not exceeding the operand). Third, histogram equalization was performed. The cumulative distribution function (CDF), $F(g)$, was computed as follows:

$$F(g) = \sum_{i=0}^{g} h(g),$$

where $h(g)$ is the gray-level histogram of the image. Then a new gray level $(g')$ was computed by
Two sets of texture features were extracted from the OCT tissue images. The first set of features was the SGLDMs, or co-occurrence matrices.\textsuperscript{24–26} The spatial gray-level dependence method is based on the estimation of the second-order joint conditional probability density functions, \(f(i,j|d, \theta)\), for \(\theta=0, 45, 90, \text{ and } 135\) deg.\textsuperscript{24} Each \(f(i,j|d, \theta)\) is the probability of a pixel with a gray-level value \(i\) being \(d\) pixels away from a pixel of gray-level value \(j\) in the \(\theta\) direction. If the image contains \(L\) gray levels, then an \(L\times L\) matrix, \(s_{g(i,j|d)}\), \(0 \leq i < L, 0 \leq j < L\), can be created from \(f(i,j|d, \theta)\), for each direction \(\theta\) and for a given distance \(d\). To help classify tissues using these SGLDMs, certain textural information can be calculated, including energy, entropy, correlation, local homogeneity, and inertia. These texture features give us twenty different parameters to discriminate tissue types since we calculate these five features for four different directions with \(d=1\). The SGLDM features are calculated as follows:\textsuperscript{5}:

\[
\text{energy} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [s_{g(i,j|d)}]^2, \tag{4}
\]

\[
\text{entropy} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} S_{g(i,j|d)} \log[s_{g(i,j|d)}], \tag{5}
\]

\[
\text{correlation} = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_x)(j - \mu_y)s_{g(i,j|d)}}{\sigma_x \sigma_y}, \tag{6}
\]

\[
\text{local homogeneity} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1+(i-j)^2}s_{g(i,j|d)}, \tag{7}
\]

\[
\text{inertia} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2s_{g(i,j|d)}, \tag{8}
\]

where \(s_{g(i,j|d)}\) is the \((i,j)\)th element of the SGLDM for distance \(d\), \(L\) is the number of gray levels in the image, and where

\[
\mu_x = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} s_{g(i,j|d)}, \tag{9}
\]

\[
\mu_y = \sum_{i=0}^{L-1} j \sum_{j=0}^{L-1} s_{g(i,j|d)}, \tag{10}
\]

\[
\sigma_x = \sum_{i=0}^{L-1} (i - \mu_x)^2 \sum_{j=0}^{L-1} s_{g(i,j|d)}, \tag{11}
\]

\[
\sigma_y = \sum_{i=0}^{L-1} (j - \mu_y)^2 \sum_{j=0}^{L-1} s_{g(i,j|d)}. \tag{12}
\]

The second set of features was derived from the two-dimensional discrete Fourier transform (DFT) of the images. As seen in Fig. 1, the 2-D DFT can be divided into regions based on frequency content. In this case, four regions were chosen, giving four texture parameters, where the innermost region represents the lowest spatial frequency range and the outermost region represents the highest spatial frequency range. Square regions were used instead of circular ones for computational efficiency. The magnitudes of all the spatial frequencies in each region were summed up and divided by the total frequency magnitude content in the complex 2-D FFT. This gives the percentage contribution of each region to the frequency content of the 2-D FFT.

2.4 DISTANCE MEASURES

The scales of the twenty-four different features varied greatly, so they were normalized. A feature vector was calculated for each of the selected regions within the images. From this set of data, a single combined mean feature vector (\(\bar{\mu}\)) and a single combined standard deviation vector (\(\bar{\sigma}\)) were calculated using the combined data from all of the classes (tissue types) in the training set. The data were averaged over all of the regions for each feature, leaving a single 24-feature vector (\(\bar{\mu}\)). Each feature vector of a given region in a known class, \(c\), was then normalized as follows:

\[
x = \frac{\bar{x} - \bar{\mu}}{\bar{\sigma}}. \tag{13}
\]

where \(\bar{x}\) is the raw feature vector for a given region. For each class, \(c\), the mean, \(\mu_c\), was computed by averaging the normalized feature vectors in that class. The covariance matrix used in the Mahalanobis distance was calculated for each class using the following equation:

\[
\sum_c = \frac{1}{n_c} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T
\]
where \(i\) and \(j\) both range from 1 to the number of features, in this case twenty-four (twenty SGLDM features plus four FFT features). The Mahalanobis distance, \(d_c\), between an unknown region’s normalized feature vector, \(\mathbf{x}\), to a specific class, \(c\), was calculated using the following equation:

\[
d_c = (\mathbf{x} - \mu_c)^T \Sigma_c^{-1} (\mathbf{x} - \mu_c),
\]

and

\[
\Sigma_c = \sum_i \sum_j (i \cdot (j - \mu_c)) (i \cdot (j - \mu_c))^T.
\]

2.5 Image Classification Program

An image classification program was developed to differentiate tissue types based on texture analysis. This program applies a minimum-error-rate Bayesian classification model. Given several sets of training images, the program extracts the best combination of three features, out of the twenty-four available, to distinguish the selected tissue types. Then, given an unknown region, the program measures the Mahalanobis distance between the 3-feature vector of that region and all the known classes in the training set. The program determines tissue type based on the shortest Mahalanobis distance to the classes in the training set.

To validate our texture analysis code, nine different Brodatz textures were used and are shown in Fig. 2. The Brodatz textures are a set of images commonly used as a standard in texture analysis algorithms. The 256×256 pixel gray-scale images used in this study were D4, D16, D19, D24, D29, D53, D68, D77, and D79. A 64×256 region of each Brodatz image was digitally processed in the same manner as the OCT images, with the exception of the log operation because they were already gray scaled (0 to 255 range) and did not have the large dynamic range of raw OCT images. This region appears as the horizontal stripe in the images shown in Fig. 2. The remaining parts of the images were not preprocessed.

Ten images of each tissue class were obtained from various locations in the tissue specimens, except abnormal lung with nine images. Six of the images from each class were randomly selected and used for training, and the remainder for classification. In each image, the program extracted ten 64×64 pixel regions randomly. Since the images were 64×256 pixels, there was significant overlap among the sixty training regions and among the forty classification regions. However, there was no overlap between the training and classification regions.

3 Results

Two types of speckle were noticed in OCT images, which is consistent with the observations of Schmitt.9 The appearance of the large (30 to 100 \(\mu\)m) speckles remained constant from image to image taken at the same tissue location. This is expected since our object is stationary in time. The small (pixel-sized) speckles did not remain constant. Since only the top 250 \(\mu\)m of tissue was imaged, and this region was fairly uniformly reflective, the average signal level was considerably higher than the system noise floor. Therefore, this noise was ignored in our analysis.

Representative images of mouse skin, testicular fat, normal lung, and abnormal lung are given in Fig. 3. The smaller speckle, on the order of a single pixel, was reduced by averaging, but the larger speckle was unaffected. A few structures were noticed (for instance, the horizontal linear structures in the fat image), but in general the images were dominated by speckle. To the unaided eye, very subtle differences in the shape of the speckle were seen between tissue types.
Table 1 shows the classification performance of the system when four separate experiments were conducted. In each experiment, twenty different runs of the program were performed to determine statistical repeatability. The differences between runs were due to the random selection of regions within the images.

The results of experiment 1 show that the nine Brodatz textures could be correctly classified 97 to 100% of the time with a standard deviation of 0.0 to 6.7%. The results from experiment 2 show the program could correctly differentiate skin and testicular fat 98.5 to 97.3% of the time, with standard deviations of 1.3 and 3.4%, respectively. Experiment 3 results show that the program was able to correctly differentiate regions of normal and abnormal lung tissue into their true class 89 and 64% of the time, with a standard deviation of 11.2 and 17.7%, respectively. Finally, the results of experiment 4 show that when all of the normal tissue types are analyzed together, the program can achieve correct classification rates for skin, testicular fat, and normal lung tissue of 37.6, 94.8, and 65.3%, with standard deviations of 17.7, 5.1, and 15.2%, respectively.

4 Discussion

Brodatz textures are a standard test case for texture analysis code. The Brodatz classification problem was much more difficult than the binary comparisons because nine different classes were being compared simultaneously, but the Brodatz textures represent an ideal case in texture analysis: an extremely regular, nearly noise-free image. The Brodatz textures also appear more differentiated from each other than the tissue images are when viewed with the unaided eye. The excellent results (97 to 100% correct classification rate) gave us confidence that our program was performing correctly.

The program then performed a binary comparison of mouse skin and testicular fat images. There are some structural differences that can be seen between the two classes with the unaided eye—predominantly horizontal striping in the fat images. The excellent classification results may be due to the presence of these features.

The program also performed a binary comparison between normal and abnormal mouse lung tissue. There are no structural differences that can be seen between the two classes with the unaided eye, but the normal lung tissue appears to have a slightly larger speckle size than the abnormal lung tissue. The program was able to correctly classify the abnormal lung tissue 64% of the time and normal lung tissue 88.6% of the time. The normal and abnormal lung tissue images were very similar, so it was expected that the program would yield poorer results than the Brodatz texture or the skin and fat results. This experiment suggests that an identifiable structure, such as in testicular fat, helps the program differentiate the tissue types, but that speckle alone can still yield reasonable classification rates. Basset et al. used SGLDM texture analysis on ultrasound images to differentiate normal versus cancerous prostate cancer and obtained sensitivities ranging from 41 to 83% and specificities ranging from 55 to 71%. Our correct classification results are comparable to those achieved in this study. Optical coherence tomography has a resolution approximately 10 times higher than ultrasound, making OCT better suited for visualizing the texture and structure of tissue volumes smaller than a millimeter.

The program was finally tested with all three of the normal tissue classes collected from the mouse, including skin, testicular fat, and normal lung. The purpose of this experiment was to see how the program’s performance decreased with more tissue types. To the unaided eye, the skin and normal lung images look very similar, lacking any apparent structures. Not surprisingly, the texture analysis code was able to differentiate the testicular fat (94.8%) much better than the other two classes. The program still did quite well with normal lung (65.2%), but skin results (37.6%) were just slightly better than chance (33.3%).

The relatively shallow depth of the images was chosen to minimize light attenuation and multiple scattering effects, and to maximize image contrast. The reduced scattering coefficient of human skin was found to be approximately 10 cm⁻¹ at 1300 nm. The optical scattering properties of mouse tissues at 1300 nm were not available, but assuming they are similar, and based on the work presented by G. Yao and L. V. Wang, our OCT signal should be dominated by single-scattered photons. Future studies are needed to determine the accuracy and repeatability of this method with deeper images and with images from varied depths.

There are other texture analysis approaches that were not investigated in this study. These include gray-level run-length matrices (GLRLM), gray-level difference matrices (GLDM), power spectral methods (PSM), filtering methods, and the
wavelet transform. The relative merit of three statistical approaches (GLRLM, SGLDM, and GLDM) and PSM was examined in Ref. 24. That study found that for all possible finite groupings of the distance between sampling pixels, SGLDM gives better results than either GLRLM or GLDM. Therefore, in this feasibility study we chose to focus on the SGLDM approach. Other groups have applied various feature extraction techniques, such as karyometry, to histology samples of various types of abnormal tissue, but these methods lack the ability to perform noninvasive discrimination techniques.29

In future work, our Bayesian classifier will be extended to handle real-world applications, such as cancer detection. We will attempt to take into account a priori probabilities of the possible classes and the magnitude of risk involved for various misclassifications. For example, a Bayesian classifier might be more likely to classify normal tissue as cancerous instead of classifying cancerous tissue as normal because the consequence of missing a cancerous lesion is severe. In the current study, all of the classes had an equal probability of occurrence, and there were no additional penalties for various misclassifications, so the Bayesian classifier was reduced to a minimum-error classifier.

To the authors’ knowledge, this is the first attempt to perform automated classification techniques on structureless OCT images. This study has shown that texture analysis, combined with OCT imaging, has the potential to provide an automated means of diagnostic differentiation of tissue.

Acknowledgments

This work was partially supported by grants from the Whitaker Foundation and the National Institutes of Health (CA83148). The authors appreciate helpful conversations with Drs. Steven Jacques and Zhongping Chen. We would also like to thank Mr. Xianfeng Li for his texture analysis contributions to our lab.

References