Distinguishing autofluorescence of normal, benign, and cancerous breast tissues through wavelet domain correlation studies

Anita H. Gharekhan
Siddharth Arora
Ashok N. Oza
Mundan B. Sureshkumar
Asima Pradhan
Prasanta K. Panigrahi
Distinguishing autofluorescence of normal, benign, and cancerous breast tissues through wavelet domain correlation studies

Anita H. Gherekhan, Siddharth Arora, Ashok N. Oza, Mundan B. Sureshkumar, Asima Pradhan, and Prasanta K. Panigrahi

Abstract. Using the multiresolution ability of wavelets and effectiveness of singular value decomposition (SVD) to identify statistically robust parameters, we find a number of local and global features, capturing spectral correlations in the co- and cross-polarized channels, at different scales (of human breast tissues). The copolarized component, being sensitive to intrinsic fluorescence, shows different behavior for normal, benign, and cancerous tissues, in the emission domain of known fluorophores, whereas the perpendicular component, being more prone to the diffusive effect of scattering, points out differences in the Kernel–Smother density estimate employed to the principal components, between malignant, normal, and benign tissues. The eigenvectors, corresponding to the dominant eigenvalues of the correlation matrix in SVD, also exhibit significant differences between the three tissue types, which clearly reflects the differences in the spectral correlation behavior. Interestingly, the most significant distinguishing feature manifests in the perpendicular component, corresponding to porphyrin emission range in the cancerous tissue. The fact that perpendicular component is strongly influenced by depolarization, and porphyrin emissions in cancerous tissue has been found to be strongly depolarized, may be the possible cause of the above observation.

Keywords: fluorescence spectroscopy; wavelet transform; singular value decomposition.

Paper 10591P1RR received Nov. 2, 2010; revised manuscript received Jun. 7, 2011; accepted for publication Jun. 10, 2011; published online Aug. 1, 2011.

1 Introduction

Of late, fluorescence spectroscopy is emerging as a preferable method for early detection of diseases, because of its sensitivity, high speed, and safety. A number of fluorophores have been identified in human tissues, which differ in their emission properties in diseased and non-diseased conditions.1–6 Fluorescence characteristics of cancer tissues and their differences with the normal counterparts have been under intense investigation, since the last few decades. Significant biochemical and morphological changes in the tumor tissues affect light propagation properties, which can introduce detectable differences in the channels which are parallel and perpendicular to the incident plane polarized light. Since tissue is a turbid and complex medium, the intrinsic spectra is substantially modified by the medium before its detection,2–4,7,8 both in normal and dysplastic tissues. All these provide avenues for studying the nature and condition of the tissue through the characteristic fluorescence emissions, as well as through the statistical features of the emitted spectra, arising due to randomization effect of this turbid medium.

Here, we carry out a systematic investigation of the wavelet domain correlation characteristics of the autofluorescence of normal, benign, and cancerous human breast tissues in the 500 to 700 nm range. The tissues are excited by a 488 nm wavelength plane polarized light from an Ar-ion laser and the components of fluorescence light which are parallel (co-) and perpendicular (cross-) to the incident polarized light were measured in the 500 to 700 nm wavelength, with the dominant fluorophores in the above range corresponding to flavins and their derivatives and porphyrin. We make use of the multiresolution ability of wavelets and dimensional reduction nature of singular value decomposition (SVD) to identify statistically robust parameters, capturing correlated spectral variations in the co- and cross-polarized channels, at different scales. This corresponds to both fluorescence emissions of known fluorophores, as well as spectral variations. The localization property of wavelets reveal the possible fluorophores responsible for the observed spectral activity, whereas the eigenvectors, corresponding to the dominant eigenvalues of the correlation matrix in SVD, exhibit clear differences between the tissue types. The copolarized component, being more sensitive to intrinsic fluorescence, shows different behavior for normal, benign, and cancerous tissues in the emission domain of known fluorophores, while the perpendicular component, being more prone to the diffusive effect due to scattering, points out differences in the standard deviation of...
percentage fluctuations, which distinguishes between malignant, normal, and benign tissues.

Interestingly, the significant distinguishing feature among tissue types manifests in the perpendicular component, corresponding to porphyrin emission range in the cancerous tissue. The fact that cross polarization component is strongly influenced by depolarization and porphyrin emissions in cancerous tissue has been found to be strongly depolarized9–11 may possibly explain the above observation.

2 Materials and Method
The study involved 22 patients with benign growths, and 23 patients with histopathologically confirmed malignant growths. Pathologically characterized fresh breast tissue samples with their normal counterparts were obtained within 2 h of surgery and were kept in the refrigerator until used. The age of patients spanned over a broad range, from 16 to 85 years, coming from varied economic backgrounds. The collected tissue samples were kept in moist saline and frozen (4°C) immediately after biopsy. After the biopsy, a part of the tissue sample was sent for histopathology and the other part was used for fluorescence measurements. The experiment was performed within a few hours of the surgery, after thawing the sample, without any chemical treatment. During experiments, the tissue was at room temperature and placed on a quartz plate of size 3 cm×1 cm×2 mm. Details about the measurement procedure are presented in Refs. 4, 12, and 13.

The tissue samples were excited by 488 nm wavelength plane polarized light from an Ar-ion laser (Spectra Physics 165, 5W) and the parallel (co-) and perpendicularly (cross-) polarized fluorescence light were measured from 500 to 700 nm. The polarized fluorescence spectra were collected in the right angle geometry, using triplemate monochromator (SPEX-1877E) and photomultiplier tube (RCA C-31034). Keeping the excitation polarizer vertical, fluorescence was recorded with the emission polarizer in both the vertical (||) and horizontal (⊥) positions to obtain the co- (parallel) and cross- (perpendicular) polarized states, respectively. Typical plots of the co- and crossed-components of fluorescence spectra of normal, benign, and cancerous tissues are shown in Fig. 1. Figure 1(d) shows the nonnormalized autofluorescence intensity of the cross component in cancer tissue is higher than in the healthy one as the depolarization in cancerous tissue is strong.

Polarization is defined as the ratio of the linearly polarized component’s intensity divided by the natural light component’s intensity. In an ideal system, polarization is measured only by the vertically polarized excitation with the horizontal and vertical emission components. These measurements are designated as $I_{VH}$ or $I_{∥}$ and $I_{HV}$ or $I_{⊥}$, respectively; the first subscript indicating the position of the excitation polarizer and the second subscript indicates the emission polarizer. Vertically oriented polarizers ($V$) are said to be at 0 deg (from normal) and horizontal polarizers ($H$) are said to be at 90 deg. All the spectra were taken in an L-format polarizer system. L-format utilizes two polarizers with the emission polarizer rotated between horizontal and vertical polarizations for measurements. The entrance and exit polarizers are fully automated and adjustable to within 1 deg rotations. Insertion and removal of polarizers from the optical path is controlled by the computer.

![Normalized fluorescence intensity plots](https://www.spiedigitallibrary.org/journals/Journal-of-Biomedical-Optics)
Table 1 T-test (paired samples test).

<table>
<thead>
<tr>
<th>Pair</th>
<th>Lower</th>
<th>Upper</th>
<th>t</th>
<th>df</th>
<th>sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal - Benign Co-</td>
<td>−576.258</td>
<td>−404.065</td>
<td>−11.162</td>
<td>3617</td>
<td>.000</td>
</tr>
<tr>
<td>Normal - Benign Cross-</td>
<td>−381.609</td>
<td>−259.934</td>
<td>−10.338</td>
<td>3617</td>
<td>.000</td>
</tr>
<tr>
<td>Normal - Cancer Co-</td>
<td>−1591.878</td>
<td>−1353.649</td>
<td>−24.240</td>
<td>4622</td>
<td>.000</td>
</tr>
<tr>
<td>Normal - Cancer Cross</td>
<td>−1206.119</td>
<td>−1010.756</td>
<td>−22.246</td>
<td>4622</td>
<td>.000</td>
</tr>
</tbody>
</table>

The value for the $G (I_{HV}/I_{HH})$ factor, the ratio of the sensitivity of the instrument to vertical and horizontally polarized light, was measured for the entire fluorescence wavelength region, 500 to 700 nm. The $G$ factor improves the S/N ratio for weak signals. The $G$ factor is defined as:

$$G = G(\lambda_{EM}) = \frac{I_{HV}}{I_{HH}}.$$  \hspace{1cm} (1)

In the experiments performed, measurements of $I_{VV}$, $I_{VH}$, $I_{HV}$, and $I_{HH}$ were all taken and finally used in the intrinsic fluorescence model, taking into account the $G$-factor.

We have employed paired t-test on benign and cancer tissues with their normal counterpart for co- and crossed-fluorescence intensity. The result at 95% confidence interval for the difference is presented in the Table 1. The result shows that the difference between normal and tumor intensities differ significantly from zero. We have also done analysis of variance to test the significance difference for the mean of both co- and crossed-component of autofluorescence intensity of tumor and their normal counterparts. The test shows significant results. Then the post hoc analysis is done by using Tukey HSD. The test gives homogeneous subgroups shown in Table 2.

Table 2 Homogeneous sub-groups.

<table>
<thead>
<tr>
<th>Normal Co-⊥</th>
<th>Tumor Co-⊥</th>
<th>Normal Cross-∥</th>
<th>Tumor Cross-∥</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>No. of samples</td>
<td>Group</td>
<td>No. of samples</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>4</td>
</tr>
</tbody>
</table>
3 Extracting Spectral Features Through Wavelets and SVD

The autofluorescence spectra are studied through wavelet transform, for pin-pointing local spectral features, as well as global statistical characteristics. For analysis of wavelet coefficients, we make use of the dimensional reduction ability of singular value decomposition, which also enables one to identify correlated domains in the tissue spectra. Wavelet transform is well suited for identifying multiscale properties of nonstationary processes, which enables one to explore aspects of data that other analysis techniques miss. These include finding trends, breakdown points, discontinuities at higher derivatives, and self-similarity in fluctuations, to mention a few. Differing from the Fourier analysis, wavelet transform is used for representing general functions in terms of simple, fixed blocks at different scales and positions. These blocks, forming a basis set, are a family of wavelet functions, generated from a prototype function, called “mother” wavelet, by translation and scaling operations. The mother wavelet needs to satisfy certain admissibility conditions, and is selected so that the translations and dilations make it possible to obtain a complete frequency domain representation of the function or data under study.

Any finite energy signal $f(t) \in L^2(R)$, in discrete wavelet transform, can be expanded as

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi(t) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi_{j,k}(t).$$ (2)

Fig. 2 Probability density of PC1 for (a) and (b) benign, (c) and (d) normal, and (e) and (f) cancer tissue types. The probability density is estimated on 100 equally spaced points that span the range of values observed in the principal component using a kernel smoothing technique, whereby a normal kernel type was chosen for estimation.
Fig. 3 Probability density of PC1 of level-1 low pass coefficients for (a) and (b) benign, (c) and (d) normal, and (e) and (f) cancer tissue types.

Fig. 4 Probability density for PC1 of (a) intensity and (b) low pass coefficients (level-1).
Fig. 5 Scatter plot of PC1 and PC2 for (a) and (b) benign, (c) and (d) normal, and (e) and (f) cancer tissue types. PC1 (eigenvector corresponding to highest eigenvalue) captures the highest proportion of variance in the time series, while PC2 captures the second highest proportion of variance.

Fig. 6 Probability density for PC2 of (a) intensity and (b) low pass coefficients (level-1).
Fig. 7 Correlation matrices of benign, normal and cancer tissue samples of (a), (d), and (g) original intensity, (b), (e), and (h) using PC1, and (c), (f), and (i) using PC2 of parallel component of polarized fluorescence data. x- and y-axis corresponds to covariance $C_{ij}$.

Here $c_j$'s are the low pass and $d_j$'s are the high pass coefficients. High frequency and low frequency components at multiple scales are known as high pass and low pass coefficients, respectively. High pass coefficients represent the variations or deviation from the trend, whereas low pass coefficients provide average behavior or trend of the data over corresponding window sizes. For our analysis, we have used the Haar wavelet, because of its symmetric nature, as also for physically transparent interpretation of the wavelet coefficients. The low-pass coefficients in this case correspond to averaging over appropriate window sizes, depending on scale and the high-pass coefficients corresponding to differences of averages of the data under analysis.

4 Singular Value Decomposition and Principal Component Analysis

In order to extract the robust diagnostic content in the wavelet domain decomposition of the autofluorescence spectra of the cancer, benign, and normal human breast tissues, and unravel their correlation properties, we take recourse to singular value decomposition. This procedure dimensionally reduces the spectral data into a smaller orthogonal set of linear combinations of the wavelet coefficients that account for most of the variance. In the process, the correlation behavior in the wavelet coefficients also gets highlighted through SVD. For analysis, the SVD starts with an $m \times n$ matrix of real-valued data $X$, whose decomposition is given by

$$X = USV^T,$$

where $U$ is an $m \times m$ matrix, $S$ is an $m \times n$ diagonal matrix, and $V^T$ is also an $n \times n$ matrix. The columns of $U$ are called the left singular vectors and rows of $V^T$ contain the element of the right singular vectors. The elements of $S$ are only nonzero on the diagonal and are called the singular values. For a square symmetric matrix $X$, singular value decomposition is equal to diagonalization or solution of the eigenvalue problem. There is a direct relation between principal component analysis (PCA)
Fig. 8 Correlation matrices of benign, normal, and cancer tissue samples of (a), (d), and (g) original intensity, (b), (e), and (h) using PC1, and (c), (f), (i) using PC2 of perpendicular component of polarized fluorescence data. x- and y-axis corresponds to covariance $C_{ij}$.

and SVD, in the case where principal components are calculated from the covariance matrix.

The covariance matrix $C$ to be investigated is defined as:

$$C = (A^T A)/N,$$

where $A^T = \delta I \delta(k)$ is an $m \times n$ rectangular matrix and $T$ denotes matrix transposition. $N$ is the normalization factor. We have concentrated on the dominant eigenvalues. Since the first and second principal components (PC1 and PC2) capture the highest proportion of variance present in the data, we only analyze these two highest principal components. The correlated structures in the correlation matrix manifest in the eigenvectors, whose entries are studied using probability density functions.

5 Kernel-Smoother Density

Kernel density estimation is a nonparametric way of estimating the probability density function of a random variable. The kernel is usually chosen to be symmetric, nonzero, and continuous.

If $x_1, x_2, \ldots, x_n \sim f$ is an independent and identically-distributed sample of a random variable, then the kernel density approximation of its probability density function is

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right),$$

where $K$ is some kernel and $h$ is a smoothing parameter called the bandwidth. Quite often $K$ is taken to be a standard Gaussian function. The variance is controlled indirectly through the parameter $h$:

$$K \left( \frac{x - x_i}{h} \right) = \frac{1}{\sqrt{2\pi}h} e^{-\left(x-x_i\right)^2/2h^2}.$$

Some of the other common choices for kernel include - uniform, triangular, and epanechnikov.
6 Results and Discussion

Here we present the results of a systematic analysis of the kernel-smoother (KS) density, employed to principal components. It is observed that the nature of the KS density of the eigenvector corresponding to the highest eigenvalue entries, i.e., PC1 (Fig. 2), show remarkable difference between benign, normal, and cancer tissues. A secondary peak is seen to emerge, along with the primary peak in all three type of tissues. With the smoothing of data, by considering the low-pass coefficients, through Haar wavelets, the secondary peak becomes more prominent (Fig. 3 and 4), in the case cancer crossed component as the intensity of the perpendicular component is not only affected more by scatterers, but is also quite sensitive to absorption, since the path traversed by the same in the tissue medium is more. It is found that the KS density shows significantly better discrimination. The scatter plot of PC1 versus PC2 (Fig. 5) reveals considerable flattening in benign and normal tissue samples as compared to the cancer ones. It is significant to observe that the perpendicular component shows much better differentiation between the tissue types.\textsuperscript{16}

Significantly, the probability density of second principal component, i.e., PC2 of low pass coefficients clearly separate out cancer tissue from benign and normal [Fig. 6(b)]. This motivates one to identify the correlation domains, corresponding to these two components. For this purpose, first we computed the correlation matrix C and obtained its eigenvectors and eigenvalues, then reconstructed the correlation matrix using eigenvectors corresponding to two highest eigenvalues independently. Note that PCA transformation is linear and orthogonal and the principal components corresponding to the larger eigenvalues capture a higher proportion of variance in the data. Since PC1 and PC2 capture the highest proportion of variance present in the data, we had analyzed and used only the two highest principal components in this work. To obtain PC1 and PC2, we had multiplied the two eigenvectors of the correlation matrix (corresponding to two largest eigenvalues) with the data matrix independently. We had then obtained the reconstructed data matrix, computed using any one principal component, by now multiplying the transpose of the eigenvector matrix (comprising only one highest eigenvectors) with the transformed data matrix. The reconstructed correlation matrix [Figs. 7(b), 7(c), 7(e), 7(f), 7(h), 7(i), and 8(b), 8(c), 8(e), 8(f), 8(h), 8(i)] was obtained using the reconstructed data matrix. Though we had focused on the two largest principal components in reconstructing the data, one could extend this analysis by incorporating more principal components. One way to formally select the total number of principal components is via scree plots.\textsuperscript{17}

Two distinct domains associated with the two principal components, i.e., PC1 and PC2, are observed capturing different spectral features (Figs 7 and 8). In parallel component, the large domain associated with PC1 [Figs. 7(b) and 7(e)] is at lower wavelength (510–650 nm) in the case of benign and normal tissues, whereas this domain shifts toward higher wavelength (550–700 nm) [Fig. 7(h)] in the case of cancerous tissues. As at lower wavelength, flavin and its derivatives are the active fluorophores, whereas porphyrin emits at higher wavelength; hence this shifting of domain may be attributed to more porphyrin emission in the case of cancer case, as the parallel component is more sensitive to the emission of fluorophores present in the tissue. In the visible wavelength regime, flavin adenine dinucleotide (FAD) and porphyrin are the major fluorophores which fluoresce, with peak intensities at 530 and 630 nm, respectively. These fluorophores are considered as contrast agents for cancer detection.\textsuperscript{1,18} Porphyrin is a weak fluorophore. Ferrochelatase is the enzyme required for conversion of protoporphyrin IX (PpIX) to heme. In tumors, the deficiency of ferrochelatase results in accumulation of PpIX relative to the normal ones.\textsuperscript{18} Such accumulation changes the relative concentration of these fluorophores, thus altering the fluorescence spectra significantly. The local environment surrounding the native fluorophores (flavins and porphyrins) at their binding site within normal and cancerous tissue give rise to the fluorescence depolarization.\textsuperscript{9}

Large eigenvalues carry the information about dominant fluorophores. In the case of eigenvector corresponding to the second highest eigenvalue, i.e., PC2 [Figs. 7(c), 7(f), and 7(i)], one sees large contributions at lower wavelengths in cancer tissue, whereas in normal and benign tissues contribution is more at higher wavelengths. For analysis of this we have made use of a kernel smoother. Use of the same allows us to isolate the distributions responsible for the features in the correlation matrix. It is worth emphasizing that the most significant distinguishing feature between the three tissue type manifests in the perpendicular component through its low pass coefficients of level-1 [Figs. 4(b) and 6(b)]. This is due to the fact that a perpendicular component is strongly influenced by depolarization, absorption, and porphyrin emission at 630 nm in cancerous tissues, and the presence of scatterers.\textsuperscript{9} We suspect that this strong overlap is reflective of this fact and the porphyrin emission at 630 nm.

7 Conclusion

In conclusion, the use of wavelet transform in conjunction with the singular value decomposition led to a transparent distinction between the tissue types. Wavelets allowed a scale dependent separation of average behavior, which is less prone to statistical and experimental uncertainties. Singular value decomposition then enabled one to achieve dimensional reduction and pinpoint the dominant distinguishing correlated features between the tissue types. One clearly observes two distinct domains in the correlation matrix of the SVD, highlighting two dominant spectral features. The corresponding eigenvectors capture this information, which gave significant differences between the cancer, normal, and benign tissues. Use of a kernel smoother efficiently extracted the distribution of the entries of the dominant eigenvectors, which contained information about the correlation aspects of the fluorescence data. Significantly, clear distinction between the tissue types emerged in the second principal component, which corresponds to the smaller correlated domain in the correlation matrix. It is also interesting to note that better tissue differentiation is achieved in the perpendicular component, which shows that the present method extracts the characteristic medium effect, since the perpendicular component is more sensitive to the same, due to its relatively larger travel path, whereas the perpendicular component shows the medium effects, since this component travels the longer route. Significantly, the strongest distinguishing feature, originating from the perpendicular component, corresponded to the porphyrin emission domain in cancerous tissues.
Acknowledgments

We acknowledge Ganesh Shankar Vidyarthi Memorial Medical College, Kanpur, India for providing us the tissue samples and Dr. Manish Thakar of M. G. Science Institute, Ahmedabad, India for helping in statistical analysis.

References


