Optical processing of light-induced autofluorescence for characterization of tissue pathology

Jianan Qu and Hanpeng Chang

Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

Shengming Xiong

Institute of Optics and Electronics, Chinese Academy of Sciences, Sichuan 610209, China

Received March 19, 2001

We describe an optical processing method for characterizing tissue pathology that is based on principal-component analysis of light-induced autofluorescence. A set of optical spectral filters, which are related to the principal-component loading vectors, is designed to process the autofluorescence signal optically and to generate principal-component scores from the autofluorescence spectra. The scores are then correlated with the tissue pathology. An optical processing system is designed that uses the *in vivo* fluorescence spectra recorded from nasopharyngeal tissues. We demonstrate that the system can differentiate nasopharyngeal carcinoma from normal tissue with a high degree of sensitivity and specificity and that the optical filters used in the system can be manufactured.

OCIS codes: 170.0170, 110.7050, 170.3880, 170.4580, 170.6510.

Light-induced fluorescence (LIF) spectroscopy has the capability to sense the biochemical and morphological differences between diseased and normal tissue noninvasively. The altered biochemical and morphological state of the diseased tissue is reflected in the measured autofluorescence spectral characteristics of the tissue. In the application of LIF spectroscopy for the classification of tissue it is of the utmost importance to develop a spectral analyzer that has an accurate algorithm with which to correlate the measured spectra with the histopathology.

Principal-component analysis (PCA) is an effective method for analyzing the statistical characteristics of spectral data. The spectra processed by the PCA technique are dimensionally reduced into an informative set of principal components that comprise a small number of variables that fully describe variations in spectral data within the limitations of noise. The principal-component (PC) scores of the autofluorescence spectra are then correlated with the tissue pathology to create a clinically useful algorithm. The PCA method has been used for differentiation of diseased tissue from normal tissue.[1–4]

In general, there are two steps involved in LIF spectroscopy of tissue by the PCA technique. The first step is to disperse the fluorescence signal and to record the spectrum. Next, the spectrum is processed with an algorithm based on the diagnostically relevant PCs to generate the diagnostic results. An ordinary spectrograph can record the spectrum of the signal collected from a single pixel on the tissue surface in one measurement. A sophisticated imaging spectrograph with a two-dimensional sensor can measure the spectra from one line at a time over the tissue surface.[5] However, a pixel-by-pixel or line-by-line examination may not be practical in the surveillance of large areas of tissue during clinical practice; an imaging approach is more desirable. There is, therefore, a need for new techniques that can process the spectral signal of each pixel rapidly over an entire imaged tissue surface. Here we present a method with which to process the autofluorescence optically and extract the clinically useful information based on the PCA.

A simple color system such as the human visual system consists of a set of spectral filters for color recognition. Any incidence with spectrum $S(\lambda)$ is projected onto this set of spectral filters as

$$C_i = \int F_i(\lambda) S(\lambda) \mathrm{d}\lambda , \tag{1}$$

where $F_i(\lambda)$ is the transmission of the ith spectral filter. This projection yields a set of color responses, C_i. Similarly, if a spectral analysis system is equipped with a filter set, $F_i(\lambda)$, that is related to the loading vectors of the PCs, the scores of the incident spectrum onto the PCs may be extracted from the system responses. A PC filter can be designed as $F_i(\lambda) = \alpha_i P_i(\lambda) + \beta_i$, subject to the condition that $0 \le F_i(\lambda) \le 1$. Here $P_i(\lambda)$ is the spectrum of the ith PC loading vector in wavelength space. α_i and β_i are two parameters that ensure the value of $F_i(\lambda)$ in the range from 0 to 100% because the value of $P_i(\lambda)$ at different wavelengths could be positive or negative. The ith PC score then can be calculated as

$$I_i = \frac{1}{\alpha_i} \int F_i(\lambda) S(\lambda) \mathrm{d}\lambda - \frac{\beta_i}{\alpha_i} \int S(\lambda) \mathrm{d}\lambda . \tag{2}$$

Here, $\int F_i(\lambda) S(\lambda) \mathrm{d}\lambda$ is the response of spectrum $S(\lambda)$ to filter $F_i(\lambda)$ and $\int S(\lambda) \mathrm{d}\lambda$ is the system respone of $S(\lambda)$ without any filter. Both terms are measurable. This means that a spectral analyzer with i PC filters needs in total $i + 1$ measurements to generate all the PC scores. In an imaging system with $i + 1$ channels

0146-9592/01/161268-03$15.00/0

of i PC filters, the spectral signals of all the pixels can be processed in parallel optically and the score images of all the PCs can be produced simultaneously. The image with diagnostic information can then be created by use of an algorithm built on the PC scores. A state-of-the art image-processing board such as the Genesis-Plus (Matrox Electronic Systems, Canada) can be used to process the multichannel images and generate the multiple PC score images in real time.

To evaluate the capability of the optical processing method discussed above to characterize tissue pathology, we designed a spectral analyzer to include a set of PC filters based on autofluorescence spectra collected from nasopharyngeal tissue *in vivo*. The details of the instrument for recording the *in vivo* spectral data have been described in Ref. 6. Briefly, the excitation source was a mercury-arc lamp filtered by a bandpass filter in the wavelength range 390–450 nm. The fluorescence and reflection over the tissue surface under examination were imaged by an endoscope. Seven optical fibers were evenly distributed in the image plane of the endoscope to collect fluorescence signals for spectral analysis. The fluorescence measurements were performed at the sites where the biopsy specimens were taken for histological analysis. The spectral range of the recorded fluorescence was 470–680 nm. The spectral response of the system was calibrated. In this study, 216 biopsy specimens were collected from 59 subjects. Of these specimens, 131 were found to be normal and 85 exhibited carcinoma. The typical raw LIF spectra for normal tissue and carcinoma are shown in Fig. 1.

First we calculated the PCs from the 216 spectra to identify those of diagnostic relevance. To prepare the data for PCA, we normalized each raw spectrum to its area to eliminate individual-to-individual variation.[4] We found that the first four PCs account for more than 98% of the total variance of all the LIF spectra of normal and carcinoma tissue. The higher-order PCs are probably the main contributors to the noise. We performed unpaired Student's t-tests of the scores of spectra of normal and carcinoma tissue on the first four PCs. The results showed a statistically significant difference in the first two PC scores between normal and carcinoma tissue (p value $\ll 0.001$); the statistical difference between normal and carcinoma tissue in the scores of the third and fourth PCs were not significant (p value > 0.28). We chose the first two PCs, PC1 and PC2, with which to develop the diagnostic algorithm. The spectra of PC1 and PC2 are shown in Fig. 2.

We used commercial software for optical thin films (Film Wizard from Scientific Computing International) to design the optical PC filters. We selected the values of the parameters α_i and β_i to keep the transmissions of the PC1 and PC2 filters in the range 10–90%. SiO_2 glass was chosen as the substrate material. The coating materials used were commonly SiO_2 for the low-index layer and TiO_2 for the high-index layer. The tolerance of thickness error for each layer of the coatings was set to ± 1 nm. By using the optimization methods (Simplex and Global Modified Levenberg–Marquardt) provided by the software, we determined the number of coating layers and the thickness of each layer that would produce the optimal match of the transmissions designed and calculated PC filters. The optimal numbers of coating layers for PC1 and PC2 filters were 17 and 36, respectively. The optimal thickness for the coating layers of PC1 and PC2 filters ranges from 7.3 to 291.3 nm and from 9.3 to 337.2 nm, respectively. As is shown in Fig. 3, we found that the transmissions of the optimal thin-film filters agreed well with those of the calculated PC1 and PC2 filters.

We used a cross-validation method to evaluate the performance of the optical processing method for differentiation of the carcinoma from the normal tissue because we had access to a limited sample size. In the validation procedure we held back one spectrum and designed the PC optical filters by using filters PC1 and PC2 calculated from the remaining 215 spectra. Based on Eq. (2), the ith PC scores of the normalized spectra can be calculated from measurements with and without a filter:

$$I_i = \frac{1}{\alpha_i}\left[\frac{\int F_i(\lambda)S(\lambda)\mathrm{d}\lambda}{\int S(\lambda)\mathrm{d}\lambda} - \beta_i\right]. \tag{3}$$

We used a line that separates the scores of the carcinoma spectra from those of the normal tissue in the plot of PC1 relative to PC2 to create the diagnostic

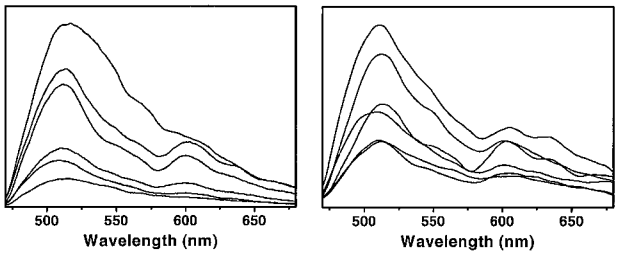

Fig. 1. Typical raw fluorescence spectra of normal tissue (left) and carcinoma (right).

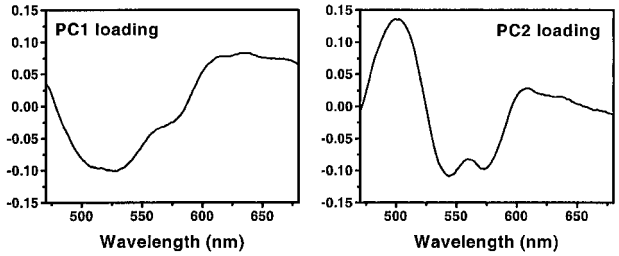

Fig. 2. Spectra of PC1 and PC2.

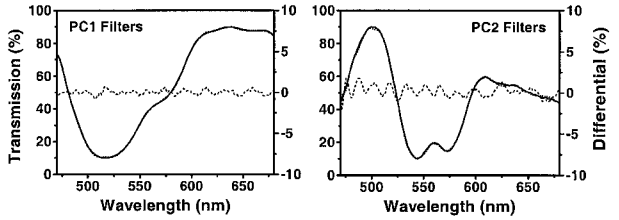

Fig. 3. Transmissions of PC1 and PC2 filters. Solid curves, designed optical filters; dotted curves, calculated filters; dashed curves (barely perceptible), differential transmissions between optical and calculated filters.

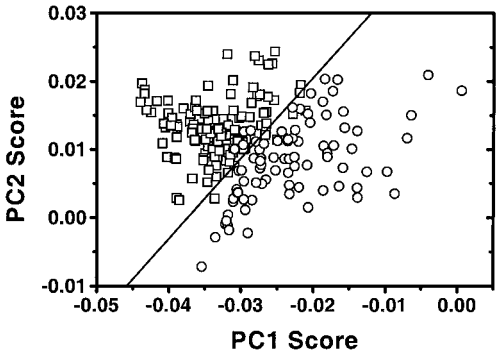

Fig. 4. Plots of the scores of PC1 and PC2 for normal tissue (squares) and carcinoma (circles). Solid line, algorithm classifying the spectra of carcinoma with a sensitivity of 95% and a specificity of 93%.

Table 1. Correlation of Sensitivity SE and Specificity SP for Three Diagnostic Lines

Diagnostic Line Number	Training Set Algorithm		Cross-Validation Algorithm	
	SE (%)	SP (%)	SE (%)	SP (%)
1	98	83	96	82
2	95	93	94	90
3	93	97	92	96

algorithm. Specifically, the diagnostic line was defined as $Y = aX + b$, where X and Y are the scores calculated from Eq. (3). The accuracy of the algorithm was controlled by choice of the parameters of a and b. Taking the remaining 215 spectra as the training set in each round of validation, we created three diagnostic lines to differentiate the spectra of carcinoma from those of normal tissue. The sensitivities of the lines were set to 98%, 95%, and 93%. We then performed an exhaustive search to determine the values of a and b that maximize the specificity that corresponds to each selected sensitivity. As an example, a diagnostic line that classifies the carcinoma spectra at a sensitivity of 95% and a specificity of 93% is depicted in Fig. 4. After the diagnostic lines were created, we used them to classify the spectrum that we had held back. This procedure was repeated until all the spectra were classified. In every round of cross validation, a new set of PC filters was created by use of the remaining 215 spectra. The sensitivity and specificity of the cross-validation procedure were calculated based on the classification results of 216 rounds of validation. The final results are presented in Table 1. The results for the training set are the selected sensitivities and the mean specificity over 216 rounds of validation. The sensitivities were fixed in all rounds of validation. The standard deviation of the specificity over the 216 rounds of validation was found to be less than 0.1% because the specificity changed almost imperceptably with any one of the 216 samples that had been held back. As a comparison, we developed PCA algorithms based on calculated PCs and found that the algorithms exhibited the same accuracy as those created by use of optical PC filters because of precise fitting of the transmissions of optical filters to those of the calculated filters shown in Fig. 3.

In conclusion, we have demonstrated that optical processing of autofluorescence to characterize tissue based on principal-component analysis is feasible. The parallel-processing method eliminates the need for a dispersion instrument or computer processing to generate the PC scores. This simplification could significantly increase the system's sensitivity because here the fluorescence signal is processed and recorded directly instead of being dispersed by a spectrograph to many pixels of the array sensor. With a multiple-channel imaging system such as one reported in Ref. 7, this optical processing technique may permit multivariate fluorescence imaging of tissue in real time. For a practical system, the system response should be incorporated into the design of the optical filters. It should be emphasized that the lesions studied here were visible or flat nasopharyngeal carcinomas. In the future we shall focus our research on the detection of early-stage cancers.

The authors acknowledge support from Hong Kong Research Grants Council grant HKUST6052/00M. Address correspondence to J. Qu (e-mail: eequ@ust.hk).

References

1. K. M. O'Brien, A. F. Gmitro, G. R. Gindi, M. L. Stetz, F. W. Gutruzzola, L. I. Laifer, and L. I. Deckelbaum, IEEE Trans. Biomed. Eng. **36,** 424 (1989).
2. N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, Lasers Surg. Med. **19,** 46 (1996).
3. Z. Ge, C. W. Brown, and H. J. Kisner, Appl. Spectrosc. **49,** 432 (1995).
4. G. Deinum, D. Rodriguez, T. J. Romer, M. Fitzmaurice, J. R. Kramer, and M. S. Feld, Appl. Spectrosc. **53,** 938 (1999).
5. C. A. Drumm and M. D. Morris, Appl. Spectrosc. **49,** 1331 (1995).
6. J. Y. Qu, W. P. Yuen, Z. J. Huang, D. R. Kwong, J. Sham, S. L. Lee, W. K. Ho, and W. I. Wei, Lasers Surg. Med. **26,** 432 (2000).
7. S. Andersson-Engels, J. Johansson, and S. Svanberg, Appl. Opt. **33,** 8022 (1994).