Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Stochastic Decomposition Method for Detection of Epithelium Dysplasia and Inflammation using White Light Spectroscopy Imaging

Ezgi Taslidere and Fernand S. Cohen
Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, USA

Abstract— **In this paper, we present a stochastic decomposition method (SDM) that allows the detection of dysplasia in epithelial tissue using white-light spectroscopy imaging. The main goal is to extract the data from the decomposition which will lead to the construction of a feature parameter space corresponding to changes in the tissue morphology related to formation of dysplasia and inflammation. These parameters include the number and mean energy of coherent scatterers; deviation from Rayleigh scattering; residual error variance of the diffuse component; and normalized correlation coefficient. The tests are performed on tissue-mimicking phantom data and tissue data collected from mouse colon in vitro. The obtained results demonstrate effectiveness of the method in differentiating between tissue structures with different cell morphologies. The results are shown by fusing all the estimated parameter set together and also using each parameter separately. Combination of all the features results in an Az value higher than 0.927 for the phantom data. For the tissue data, the best performances for differentiation between pairs of various levels of inflammation are 0.859, 0.983, and 0.999.**

I. Introduction

More than 90% of cancers arise in epithelial tissues [1]. Malignancies occurring in the epithelium lining the internal surfaces of organs typically are discovered late in the course of the disease, usually because they are "hidden" from the physician, and are not usually found on routine physical exam. Most of the time, malignant neoplastic changes follow pre-existing dysplastic changes. Dysplasia refers to a precancerous condition. Detection of dysplasia is a challenge since it does not form macroscopically observable structures and is invisible to the eye; hence there is a need of going beyond the visible to the expert. Dysplasia is recognized by alterations in the appearance of cells. As the dysplasia progresses the size of the nucleus increases. The scattered light provides an objective measure of epithelial nuclear enlargement and crowding which are the most significant characteristics of dysplasia and early cancer [1].

Our goal is to increase the likelihood of selecting a precancerous sample during biopsy by providing a tissue characterization technique that is sensitive to changes in the morphology of the tissue as it undergoes changes from normal tissue, to inflammation to ontogenesis. We previously presented a variation of our tissue characterization method using ultrasound as modality for breast cancer detection [7]. The model was proved to be quite effective for discrimination between benign, normal and malignant breast tissues (the Az values were in the range 0.862-0.999). Now, we adopted our method, namely Stochastic Decomposition Method (SDM), to be used on scattered light in order to track down the characteristics of dysplasia as it develops in epithelium and various levels of inflammation by seeking parameters that are capable of differentiating between the diseased and normal cells. One of the original contributions of this paper is the application of our model to optical data in order to detect dysplasia formation and various levels of inflammation. SDM separates the specular form the diffuse scatterers, and estimates the tissue related parameters for both scatterers individually. We investigate the parameters that are able to differentiate between normal and dysplastic cells and detect various levels of inflammation. We performed simulations of light scattered from the cells in order to mimic signatures of dysplasia using a stochastic simulator based on Mie scattering. We reported the details of the simulations and the performance results in [2]. In this paper, we collected phantom data representing the signatures of dysplasia formation, and also we collected real tissue data from mouse colon in vitro representing various levels of inflammation since we also aim to detect changes beyond nuclear crowding. We show the results on tissue-mimicking phantom data and tissue data collected from mouse colon in vitro. Some of the parameters prove to be extremely strong parameters to discriminate between normal and diseased tissue.

II. Methodology

A data collection technique is presented, which includes the collection of spectrum at equal intervals by scanning the sample. The use of single intensity for a specific wavelength for each data point forms an image of the sample. Our method enables greater sensitivity than the current the current state-of-art reflectance and fluorescence measurements by scanning the tissue through an XY stage in order to construct a 2-D "image" of the cellular structure through the use of collecting point measurements. We used every line of this image in order to extract information from the sample. Also, the same procedure is followed for various wavelengths. The wavelength giving the best performance is investigated. Due to our data collection technique, the modality can be named as *white-light spectroscopy imaging*. Our proposed system is fairly simpler than the diffused optical tomography (DOT) and optical coherence tomography (OCT) systems, since it is not our goal to

reconstruct the entire morphology of the sample. Instead, the goal is to extract the data which will lead to the construction of a feature parameter space corresponding to changes in the tissue morphology. A distinctive feature of our technique compared to DOT, OCT, and light scattering spectroscopy (LSS) is the formation of an image, rather than a spot, from which pertinent data is extracted that may lead to discovery of the hidden correlations that are obtained only by interrogating a wide sample area. In this fashion, the more data that is obtained can lead to a more confident calculation of the relevant features, such as nuclear size distribution, since our decomposition technique discovers the hidden correlations that are obtained only by interrogating a wide sample area. In contrast to point spectroscopy methods such as LSS, we rely on data vectors consisting of many reflectance intensity points in order to decide the most likely parameter of the scatters that have given rise to such a vector.

A. Spectroscopy Imaging of Tissue Mimicking Phantom

In order to mimic normal and dysplastic tissue we use polystyrene latex micro spheres (Polysciences, Inc) suspended in deionized water of sizes 3 μm, and 10 μm respectively. The micro spheres exist in a 2.5% aqueous suspension with a refractive index of 1.59-1.60 at 589nm. Micro spheres of 3 μm and of 10 μm at a magnification of 50x are shown in Figure 1.

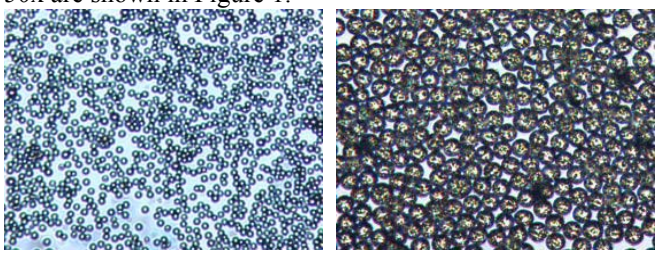


Fig. 1. 3 μm and 10 μm microspheres at 50x magnification.

The experimental system setup is shown in Figure 2. We use a bifurcated reflectance probe (Ocean Optics, Inc), consisting of 1 central fiber used for light delivery and 6 surrounding fibers used for light collection, each with a core of 200μm. The probe has a special 30° window at the end in order to reduce the specular reflectance from the sample surface. A tungsten halogen lamp (LS-1 Ocean Optics, Inc) is used as the light source and a high resolution spectrometer (HR4000, Ocean Optics, Inc) is for collection of spectrum. The spectrometer output contains 3648 points with a wavelength step size of 0.12nm. For data collection, the fiber bundle is placed in close proximity to the sample tissue. To create the image, the fiber is placed on a XYZ stage, which allows the fiber to be moved and data to be collected at different points on the sample. 1D-scans of the entire sample (4500μm by 4500μm) are taken. Each scan consists of 51 points. The step size is taken as 90μm. At each data point, the whole spectrum is recorded. After the spectrum is collected at all the data points, the intensity image is reconstructed by examining one single wavelength from which key parameters are extracted for each performance evaluation.

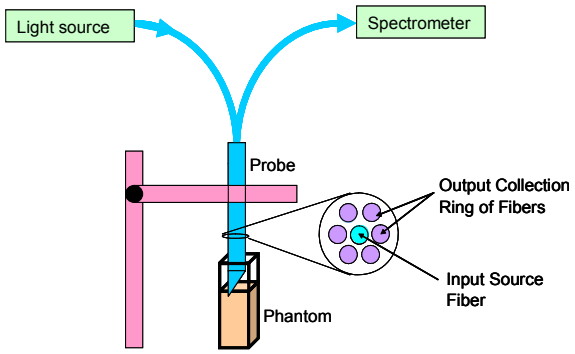
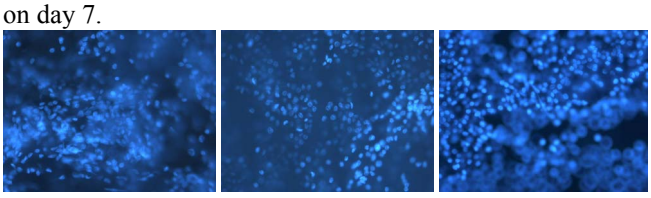


Fig. 2. The experimental setup used for phantom data collection.

B. Spectroscopy Imaging of Mouse Colon in Vitro

In the following, we present the collection of tissue data from mouse colon, in vitro. The study involved the use of mid colon obtained from three mice fed with 4% Dextrin Sulfate Solution (DSS). In our model, what we were measuring was uncontrolled acute inflammation. Colon was excised on days four (4) and seven (7) of DSS feeding. Clinically, after four days, the disease activity index was 1.3, which increased to 2.7 on day seven. In our experiments, the inflammation is certainly increased from day to day while on DSS feed. The excised colon was mounted in a Petri dish containing phosphate buffered saline. After the imaging scans were completed, the colon was stained with DAPI (4'-6-Diamidino-2-phenylindole), which is known to form fluorescent complexes with natural double-stranded DNA and for the visualization of the nucleus. The result of DAPI strain in the control animal is illustrated in Figure 3-a, the Day 4 animal is illustrated in Figure 3-b, and the Day 7 animal is illustrated in Figure 3-c. We denote these samples, mouse1, mouse2, and mouse3, respectively. DAPI staining at 40x magnifications in the control mouse Figure 3-a shows normal appearance of the tissue with no inflammation. Figure 3-b represents increased inflammation with the presence of increased neutrophils (polymorph nucleus) on day 4 and Figure 3-c represents further increase in inflammation with the possible presence of mucosal erosions on day 7.



(a) (b) (c)

Fig. 3. The resulting images of DAPI strain in the (a) Control Animal (mouse1), (b) DSS Fed Animal on Day 4 (mouse2), and (c) DSS Fed Animal on Day 7 (mouse3). (50x magnification)

1D-scans of the entire piece of the colon (10,000μm by 3,000μm) were taken in a zigzag fashion using a micrometer stage within 30 minutes after tissue mounting. . Each scan consists of 50 points. The step size is taken as 200μm. The data collection process followed a similar protocol as with the phantom data collection. A 3-D figure of the

experimental setup for in vitro data collection is shown in Figure 4.

III. Stochastic Decomposition Method (SDM) and the Extracted Features

The SDM first checks the existence of coherent scatterers by testing the hypothesis of Rayleigh scattering which can be achieved by using the non-parametric Kolmogorov-Smirnov (KS) test for color field. The rejection of the Rayleigh scattering hypothesis indicates the existence of a coherent component. If a coherent component exists, the signal is decomposed into its coherent *c(n)* and diffuse *d(n)* components using the wavelet decomposition described in [5]. This decomposition is consistent with the general decomposition of regular stochastic fields into predictable (the specular field) component and unpredictable (the diffused field) component, known in the literature of stochastic processes as the WOLD decomposition [4]. The SDM consists of taking the wavelet transform of the scattered signal and thresholding its energy. The decomposition uses the continuous wavelet transform and was thoroughly described and tested on simulated ultrasound Radio Frequency (RF) data [5][6]. After the decomposition, we fit the Auto-Regressive (AR) model to the diffuse part and extract the parameters of the diffuse part as well as the coherent part. If the test of presence of a coherent component fails, the signal is modeled as an AR process. The features extracted from the coherent component are: *number of coherent scatterers, N_c* and m*ean energy of the coherent scatterers, E.* And, the features extracted from the diffuse component are: *residual error variance of the diffuse component,σ^2,* and *Rayleigh scattering degree of the diffuse component, D.* The details of the feature extraction algorithm and the description of the features are explained in [6].

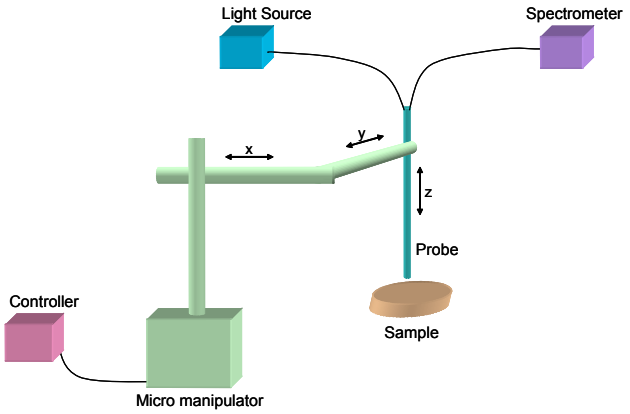


Fig. 4. 3-D figure of the experimental setup used for data collection.

IV. Results

A. Performance Evaluation on Phantom

Using the SDM, we calculated four features, *N, D, σ^2, E* for each 1D-scan at a given wavelength. The area under the ROC curve is used as the metric for performance. The classification performance of each feature is evaluated separately as well as by combining all the feature set. The combination of performance parameters is in accordance with the joint likelihood ratio statistics (Neyman Pearson statistics), which insures the highest detection rate for a fixed alarm rate over any other classifier including the NP based on the individual parameters in isolation. The results are shown in Table I and II for a diffuse scatterer only model versus a diffuse plus coherent scatterer model. The last row in both tables show the results of classification by fusing all the imaging parameter set together. Using a combination of features, the global detector that optimally combines the local detectors yields an outstanding performance with an overall A_z values varying from 0.927 to 0.994. It is shown that the performances of the individual parameters remain consistent with the simulation results [2] and the overall performance of the system is extremely high (Az>0.927).

Table I. Performance of the system in terms of A_z values for diffuse only model for Phantom Data

	500.04 nm	550.05 nm	571.67 nm	650.00 nm
D	0.694	0.551	0.580	0.519
σ^2	**0.981**	**0.979**	**0.982**	**0.924**
ρ_N	0.500	0.576	0.561	0.504
all	**0.990**	**0.982**	**0.989**	**0.927**

Table II. Performance of the system in terms of A_z values for coherent+diffuse model for Phantom Data

	500.04 nm	550.05 nm	571.67 nm	650.00 nm
N_c	0.551	0.587	0.517	0.575
E	0.986	0.967	0.979	0.865
D	0.733	0.530	0.718	0.616
σ^2	**0.984**	**0.979**	**0.982**	**0.924**
ρ_N	0.572	0.538	0.508	0.510
all	**0.994**	**0.988**	**0.991**	**0.931**

B. Performance Evaluation on Mouse Colon in Vitro

The classification is performance is shown in Table III. The best performance using one wavelength (625.24 nm) was at 0.859 combined A_z value (i.e., using all features) for differentiating between normal tissue and 4 day inflammation, whereas it was at 0.999 combined A_z value for differentiating between normal tissue versus day 7 inflammation at the same wavelength of 625.24 nm. We have also run the classification algorithm to see if we can differentiate between the 4-day inflammations versus the 7-day ones. The best performance using one wavelength was at 0.987 combined A_z value for differentiating between 4-day inflammation tissue versus 7-day inflammation tissue at the wavelength of 575.02 nm. The best performance of the algorithm was obtained at the wavelengths which are around the peak values of the spectrum. For the best two performing parameters *E* and σ^2, the wavelengths that they show the best detection performances show consistency with each other for all the classification pairs. Also, they show consistency with the corresponding wavelengths for the overall performance (performance generated by fusing all the features).

TABLE III. PERFORMANCE OF THE SYSTEM IN TERMS OF A_Z VALUES FOR TISSUE DATA (MOUSE COLON-IN VITRO) FOR CLASSIFICATION BETWEEN CASES (A) MOUSE1- MOUSE2, (B) MOUSE1-MOUSE3, (C) MOUSE2-MOUSE3

(a)

	500.33	525.11	550.02	560.24	575.02	600.10	625.24
N_c	0.568	0.549	0.559	0.523	0.590	0.542	0.559
E	**0.811**	**0.721**	**0.741**	**0.797**	**0.743**	**0.757**	**0.843**
D	0.515	0.552	0.505	0.561	0.563	0.575	0.565
σ^2	**0.809**	**0.752**	**0.703**	**0.750**	**0.718**	**0.736**	**0.821**
ρ_N	0.510	0.509	0.561	0.520	0.506	0.503	0.513
all	**0.812**	**0.786**	**0.770**	**0.799**	**0.798**	**0.776**	**0.859**

(b)

	500.33	525.11	550.02	560.24	575.02	600.10	625.24
N_c	0.541	0.501	0.511	0.479	0.549	0.562	0.508
E	**0.997**	**0.997**	**0.997**	**0.993**	**0.994**	**0.998**	**0.999**
D	0.551	0.619	0.581	0.695	0.538	0.651	0.584
σ^2	**0.997**	**0.996**	**0.998**	**0.994**	**0.995**	**0.999**	**0.993**
ρ_N	0.555	0.551	0.560	0.536	0.516	0.513	0.589
all	**0.998**	**0.998**	**0.998**	**0.995**	**0.996**	**0.999**	**0.999**

(c)

	500.33	525.11	550.02	560.24	575.02	600.10	625.24
N_c	0.541	0.543	0.596	0.525	0.596	0.556	0.571
E	**0.938**	**0.901**	**0.961**	**0.917**	**0.974**	**0.949**	**0.906**
D	0.556	0.607	0.578	0.681	0.509	0.690	0.622
σ^2	**0.816**	**0.741**	**0.896**	**0.772**	**0.887**	**0.755**	**0.784**
ρ_N	0.573	0.587	0.591	0.535	0.501	0.512	0.615
all	**0.940**	**0.935**	**0.968**	**0.932**	**0.987**	**0.961**	**0.913**

In Figure 5, we show the scattered data for the best two features for the 3 mice at 575.02 nm. The ROC construction and the A_z values are based on the scattered data points. We can easily see from the scattered data how well separated the data is between the normal and 7-day inflammation, and how it is less so between the normal and the 4-day inflammation, leading to high A_z value under the ROC curve for the former case and lower for the latter. The separation between the 4-day and the 7-day inflammations is also quite pronounced leading to high A_z value under the ROC curve. Also, from the figure we observe that the given features are directly proportional to the inflammation level of the animals. These results thus indicate the strong discrimination capability of the proposed technique.

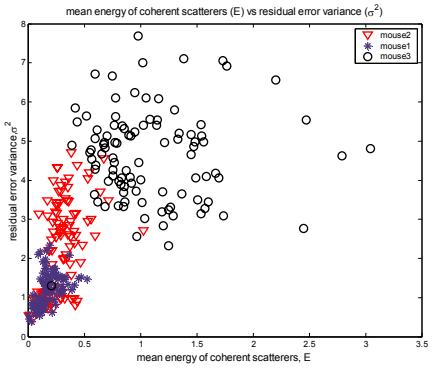


Fig. 5. Scatter plot of mean energy of coherent scatterers vs. residual error variance for the three animals studied.

V. CONCLUSION

A methodology for imaging the tissue is proposed and an SDM model is presented for modeling the obtained data. The methodology is named as white light spectroscopy imaging. The features are extracted from the obtained image in order to track down the characteristics of dysplasia as it develops in epithelium and detect various levels of inflammation. The performances of the extracted features are evaluated on tissue-mimicking phantom data as well as tissue data collected from mouse colon in vitro. Phantom data represents the signatures of dysplasia formation whereas real tissue data from mouse colon in vitro represents various levels of inflammation. The mean energy of coherent scatterers E, and the energy of the diffuse component σ^2 (the residual error variance) are found to be valuable parameters for detection of epithelium dysplasia formation and inflammation at its various stages. The results are shown by fusing all the imaging parameter set together as well as by a single parameter. The performances are reported as Az>0.927 for tissue-mimicking phantom data, and Az>0.859 for tissue data collected from mouse colon in vitro. Our system has a potential to be useful for clinical applications. Our results verify the ability of our system for detection of inflammation and signatures of dysplasia.

ACKNOWLEDGMENT

Animal model used in this paper is developed by Dr. Sreekant Murthy from Colllege of Medicine, Drexel University, Philadelphia, PA.

REFERENCES

[1] R. S. Cotran, S. L. Robbins, and V. Kumar, Robbins Pathological Basis of Disease. Philadelphia, PA: W. B. Saunders, 1994.

[2] F. S. Cohen, E. Taslidere, and D. S. Hari, "Tissue Characterization and Detection of Dysplasia using Scattered Light," in *Proc. 3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano,* Arlington, Virginia, April 6, 2006, pp. 590-593.

[3] Kullmann, F., Messmann, H., Alt, M., Gross, V., Bocker, T., Scholmerich, J. and Rüschoff, J., "Clinical and histopathological features of dextran sulfate sodium induced acute and chronic colitis associated with dysplasia in rats". Int. J. Colorectal Dis. 16(4):238-46, 2001.

[4] F. S. Cohen, G. Georgiou, and E. Halpern, "WOLD decomposition of the backscatter echo in ultrasound images of soft tissue organs." IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 44, no. 2, pp. 460-472. Mar. 1997.

[5] G. Georgiou and F. S. Cohen, "Tissue characterization using the continuous wavelet transform Part I: Decomposition method," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 48, no. 2. pp. 355-363, Mar. 2001.

[6] G. Georgiou. F. S. Cohen, C. W. Piccoli, F. Forsberg, and B. B. Goldberg, "Tissue characterization using the continuous wavelet transform, Part II: Application on breast RF data," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 48, no. 2, pp.364-373, Mar. 2001.