TR # 1311

Algorithms from Signal and Data Processing Applied to Hyperspectral Analysis: Discriminating Normal and Malignant Microarray Colon Tissue Sections Using a Novel Digital Mirror Device System

Mauro Maggioni*§, Gustave L Davis†§, Frederick J Warner*, Frank B Geshwind‡, Andreas C Coppi‡, Richard A DeVerse‡, Ronald R Coifman*§ Department of Mathematics
Yale University, New Haven, CT, 06510
Email: mauro.maggioni@yale.edu *Department of Mathematics
Yale University, New Haven, CT, 06520 USA †Yale Pathology
Yale University, New Haven, CT, 06520, USA ‡Plain Sight Systems
Hamden, CT, 06514, USA §Program in Applied Mathematics
Yale University, New Haven, CT, 06510, USA

Abstract— **Hyperspectral imaging is an important tool in various fields, notably geosensing and astronomy, and with the development of new devices, it is now also being applied also in medicine. Concepts and tools from signal processing and data analysis need to be employed to analyze these large and complex data sets.**

In this paper, we present several techniques which generally apply to hyperspectral data, and we use them to analyze a particular data set.

With light sources of increasingly broader ranges, spectral analysis of tissue sections has evolved from 2 wavelength image subtraction techniques to Raman near infra-red microspectroscopic analysis permitting discrimination of cell types and tissue patterns. We have developed and used a unique tuned light source based on micro-optoelectromechanical systems (MOEMS) and applied algorithms for spectral microscopic analysis of normal and malignant colon tissue. We compare the results to our previous studies which used a tunable liquid filter light source.

I. Introduction

Hyperspectral imaging is an important tool in various fields, notably geosensing and astronomy, and with the development of new devices, it is now being applied also in medicine. Tools from signal processing and data analysis need to be employed to analyze these large and complex data sets.

A first important challenge is to compress and reduce the dimensionality of the data available, without discarding relevant information. Depending on the data and any assumption/model for it, one can employ different signal processing techniques in order to efficiently compress the data.

The second challenge is usually a classification or a regression task. It is important to look for features which enhance discrimination among different classes or which serve as good inputs for some regression algorithm.

In this paper we present various techniques from signal processing, wavelet analysis, spectroscopy and data analysis in general that we consider to be useful in the study of hyperspectral data. We then combine these a selection of these tools to obtain an automated classification algorithm for a dataset of hyperspectral images of stained normal and malignant colon tissues.

The application of hyperspectral imaging to medicine, and pathology in particular, while not new, is becoming more widespread and powerful. With light sources of increasingly broader ranges, spectral analysis of tissue sections has evolved from 2 wavelength image subtraction techniques to hyperspectral analysis. A variety of proprietary spectral splitting devices, including prisms and mirror [1], interferometers [2], [3], variable interference filter-based monochromometers [4] and tuned liquid crystals [5], mounted on microscopes in combination with CCD cameras and computers have been used to discriminate among cell types, tissue patterns and endogenous and exogenous pigments [6]. The increasing power of these methods holds promise for developing automatic diagnostics, though the increased volume of the data collected requires more efficient algorithms to analyze the data in a short time. Moreover, in addition to amount of data collected, the dimensionality of such data has increased dramatically, which makes the extraction of statistically useful and reliable information much harder.

We use a unique prototype tuned light source (from Plain Sight Systems), based on a digital mirror device (DMD), to collect hyperspectral images of normal and neoplastic samples from tissue microarrays. Analysis of the data is done with a combination of algorithms from the fields of spectroscopy, chemometry and signal processing. The goal is to evaluate the diagnostic efficiency of hyperspectral microscopic analysis of normal and neoplastic colon biopsies prepared as microarray tissue sections [7]–[10].

II. Signal Processing and Data Analysis

We introduce a number of techinques and tools of general applicability in dealing with hyperspectral data. The specifics of the algorithm applied in the present situation will be detailed in a later section.

A. Compression and De-nosing

In many important applications, such as astronomy or medical imaging, there is a great amount of a priori knowledge on the type of hyperspectral data at hand. For example there are good models for each layer $I_i(x, y)$, and for each spectrum $s(x, y) = I.(x, y)$. Each I_i is an image to which standard image compression and denoising techniques can be applied. Among these, wavelet and wavelet packet [11] techniques are now classical and have been proven to be very effective. Novel multiscale techniques, especially curvelets and ridgelets [12]–[14] are also very promising and have been applied to astronomical imaging [15]. All of these techniques can be applied to each layer I_i for compression or denoising of the layer.

On the other hand, each spectrum $s(x, y)$ is naturally expected to be a smooth function, hence amenable to one-dimensional compression and denoising techniques. These could be Fourier-transform based or, better, wavelet based.

Another possibility is to do a full 3-D analysis, by using for example 3-D wavelet packets, or more recent tools such as 3-D curvelets. On our dataset, for example, 3-D wavelet packet based denoising performed very well.

B. Dimensionality reduction

The problem or reducing the dimensionality of the data, while preserving relevant structures of the data, is extremely important and has received a great amount of attentionin the last few years from many researchers across different disciplines.

To be concrete, let us consider the example of the hyperspectral images at hand. Each spectrum is a point in 128-dimensional space. If we look at the ensemble of all spectra in a data cube, or even in all data cubes, we do not expect them to be scattered in 128 dimensions. In fact, we could easily argue that they should be highly concentrated around a submanifold of low intrinsic dimension. For example, one parameter may be the total energy of the spectrum, and two others could the absorbance of hematoxylin and eosin stains, respectively, at particular wavelengths. These three parameters already would contain most of the information about each spectrum, and in fact this is more or less the only information that a pathologist has, looking at one of the samples in the microscope. It just happens that we are measuring 128 numbers, but far fewer parameters would sufficient to identify a spectrum. The natural questions are then: how do we discover the parameters and how many do we really need?

We will describe a few of the many approaches suggested to solve this problem, broadly subdividing them into linear and nonlinear techniques. Linear techniques in general project the data on some low-dimensional subspace, so that important features of the data are preserved, where the measure of importance has to be defined, and is often application-specific. Linear techniques include random projections, principal component analysis, partial least squares, several variations of these, and many others. In the first two techniques, the important features of the data that one seeks to preserve are essentially the pairwise distances, in the third a function (e.g. labels) on the the data is given and is taken into account in the computation of the subspace.

1) Local Discriminant Bases: Local Discriminant Bases (LDB) of Coifman and Saito [16], [17] apply naturally to a family of labeled vectors that represent smoothly varying functions, for example spectra and sounds. The labels of these vectors may correspond to more or less well-defined clusters in the data, though determining those labels via clustering or other unsupervised non-linear separation methods can be very expensive, if not unfeasible, due to the high dimensionality of the vectors. The goal of LDB is to find directions in these high dimensional spaces such that the data projected onto these directions are still well-discriminated. Then discriminating the low dimensional projections of the data should be almost as good as discriminating in high dimension with all the advantages and tools available in lower dimensional spaces. At the same time, while discriminating features are preserved, confounding features are removed, thus denoising the data with respect to the discrimination task at hand.

The search for features in high dimensional spaces is notoriously difficult. One way LDB alleviate some aspects of the "curse of dimensionality" is by searching sub-optimal projections among hierarchically well-organized dictionaries of wavelet or Fourier packets. There are fast algorithms with which perform such a search and to compute the projections onto ensembles of these patterns. We use a version of LDB that uses arbitrary Haar packet decompositions of the phase-space, but other, even less flexible, wavelet dictionaries would work as well.

Nonlinear techniques include local linear embeddinge (LLE) [18], Laplacian Eigenmaps [19], Hessian Eigenmaps [20] and Diffusion maps [21]–[27], which together have received a lot of attention in the last few years. Many of these techniques are based on the idea that the data lies on some manifold in a high dimensional space, but with the intrinsic dimensionality of the manifold actually being quite low due to constraints in the data allowing for a description by few parameters.

Here we would like to illustrate the use of Principal Component Analysis and Diffusion Maps applied to this particular dataset. Similar results would be expected in the analysis of other types of hyperspectral data, for example astronomical hyperspectral data.

We consider a data cube with the spectra centered around their mean, and we compute the principal components of the centered spectra contained in the cube. This is computationally quite expensive, so in practice we select a random subset of spectra and we compute the principal components for that subset. We immediately discover that the top 20 principal components capture almost 95% of the energy of the data. In particular, inner products and pair-wise Euclidean distances could be computed on the projection onto the top few principal

components with very good precision (and less sensitivity to noise!). The projection onto the principal components, when viewed layer by layer, is represented in Figure 1. Despite eliciting some structure within the data, the projections are not particularly informative.

If the low-dimensional set on which the spectra actually lie is quite nonlinear, in general there will not be a linear subspace onto which the data can be meaningfully projected. So while the principal components analysis does show that the intrinsic dimensionality of the spectral data is rather small, it does not help in extracting good parameters and understanding clusters in the spectral space.

We adopt a nonlinear technique based on diffusion [26], [27] in order to better understand the structure of the data. Instead of looking at the directions of maximum variability, as principal component analysis does, this technique looks at each spectrum, and at the connections between each spectrum with its very closest neighbors. It then looks at how these connections allow a random walker to explore the data. The idea is that the connections inside each cluster will be numerous and strong, and connections across clusters will be fewer and weaker. It is then possible to construct a map from spectral space to Euclidean space such that the Euclidean distance between two points measured in the range is equal to the "diffusion distance" between those two points on the original data set. Moreover, this map has the form

$$s \mapsto (\Phi_1(x), \Phi_2(x), \ldots, \Phi_k(x)) \ ,$$

where the functions Φ are defined on the set of spectra, and are eigenfunctions of a Laplacian defined on the set of spectra itself, interpreted as a graph. Details and discussion of these ideas can be found in references [19], [22], [23].

When we apply this technique to the spectra in a data cube, we get a much more meaningful descriptions of the data, and in fact various eigenfunctions Φ_i separate very well between different tissue types. This is a consequence of the staining, which we had reasonably expected as being one of the most important parameters (see Figure 2 and 3).

The parameters discovered with this algorithm allow one to "virtually stain" the biopsy, and could be mapped from biopsy to biopsy in order to resolve normalization issues that greatly affect global distances between points.

This technique can be used effectively for segmentation of the data cube. Spectral features, together with spatial features (for example filter responses to various texture or edge filters) can be clustered using the eigenvectors of the diffusion process on these features, and effectively find clusters corresponding to segmentations of the data cube. See for example [28], [29].

C. Classification and Regression Techniques

The goal of the analysis of a set of hyperspectral images may be classification or regression. Depending on the application and/or goal, one may want to classify single spectra, or groups of spectra around particular locations. For example, in astronomy one may want to classify galaxy types based on their spatial configuration and spectral characteristics. In our example, we want to discriminate between normal nuclei and abnormal (malignant) nuclei in various regions of the tissue.

In general, seeking features in the full 3-dimensional data cube can lead to the best results, since the various spatial and spectral correlations help in denoising the data, and can be used to define features of local aggregates which can be much more meaningful than features of a single spectrum.

Without giving a full discussion of all details, we present here the techniques that will be the building blocks of the final algorithm.

1) Nearest Neighbor Classifier: Given a set of points $\{x_i\}_i$ with corresponding labels $\{l_i\}_i$, and given a test point y, the k-nearest neighbor classifier assigns the label l_{i^*} to y as follows. The k closest points $\{x_{i_1}, \ldots, x_{i_k}\}$ to y are found. Then the most frequent label l_{i*} among $\{l_{i_1}, \ldots, l_{i_k}\}$ is assigned to y. Ties are broken randomly. The 1-nearest neighbor classifier has many good theoretical properties, and performs extremely well when the number of training points is large. When the number of points is small for the dimension in which the points are given, then k-nearest neighbor classifiers with $k > 1$ may be preferable since the choice $k > 1$ corresponds to regularizing the data in a particular way.

2) Partial Least Squares: Given a set of points $\{x_i\}_i$ and a function f defined on these points, and an integer $k > 0$, PLS [30]–[34] computes a set of orthonormal vectors $\{v_1, \ldots, v_k\}$, and a k-dimensional vector w, and then extrapolates f at y by first computing $P(y)$, the projection of y onto the subspace spanned by $\{v_1, \ldots, v_k\}$ and then letting $f(y) =< y, w >$. The computation of the vectors $v_1, \ldots, v_k$ is done in the following way. Once the first i vectors $v_1, \ldots, v_i$ have been constructed, v_{i+1} is the vector that solves the problem

$$\max_{\substack{||v||=1 \\ v \perp \{v_1, \ldots, v_i\}}} \mathrm{Corr}^2 \left(f, \sum_l (v)_l x_l \right) \mathrm{Var} \left(\sum_l (v)_l x_l \right)$$

where $(v)_l$ denotes the l-th coordinate of v, and where for $i = 0$ we consider the maximization over all v of norm 1. Once the k vectors $\{v_i\}_i$ have been found, the data is projected onto the subspace spanned by these vectors and linear regression is used in that subspace. The projection has the goal of denoising the data but in such a way as to preserve directions that have strong linear correlation with the function to be predicted.

III. Experimental Design

A. Platform

The prototype tuned light source [35] transilluminates hematoxylin and eosin (H & E) stained micro-array tissue sections with arbitrary combinations of light frequencies, ranging from about 440 nm to about 700 nm, through a Nikon Biophot microscope. To collect the initial data we used the flexibility of the light source to implement a randomized version of the standard Hadamard multiplexing for spectra measurement, in order to reduce noise and biases in the signal-to-noise ratios of the collected data. Hyperspectral tissue images are collected with a CCD camera (Sensovation) and are analyzed mathematically with a PC, using algorithms written in Matlab.

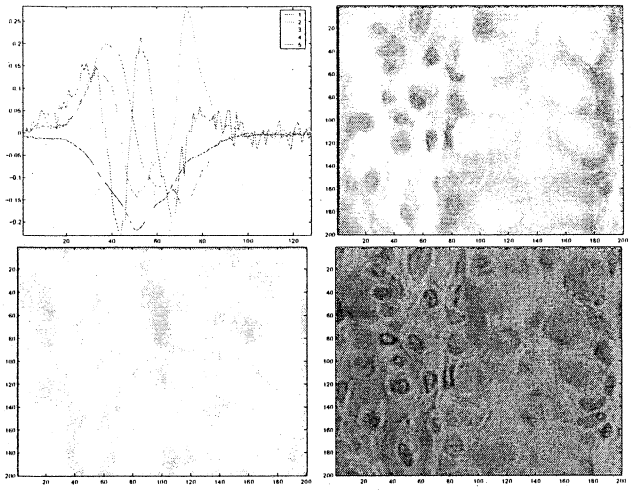

Fig. 1. Top left: The first few principal components of the spectra of a sample, top right and bottom are the coefficients of the projection of each spectrum onto the first, second and third principal components.

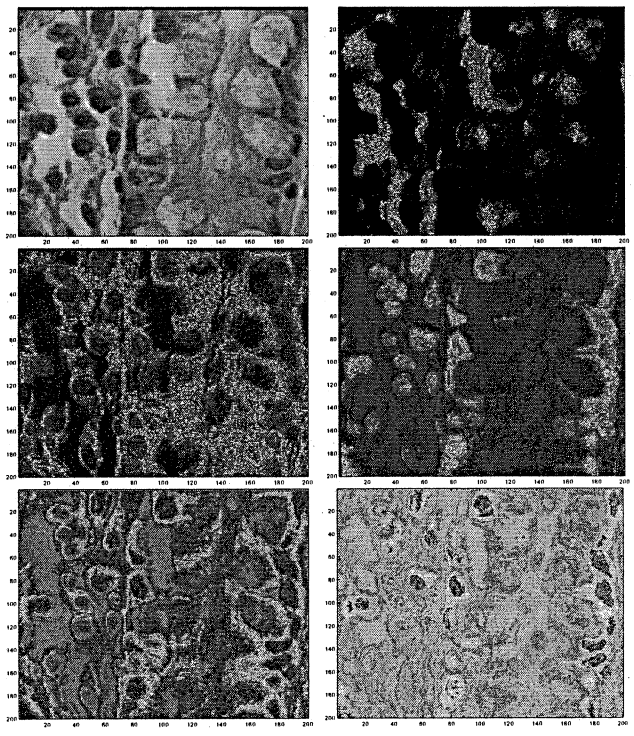

Fig. 2. Moving left to right, top to bottom, the eigenfunctions φ_k of the diffusion on the data set of spectra, for $k = 1, 2, 3, 4, 5, 6$ respectively, are mapped to the colors.

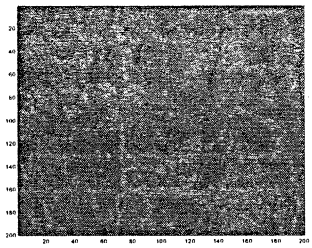

Fig. 3. The vector $(\varphi_1(x), \varphi_2(x), \varphi_3(x))$ is normalized and mapped onto RGB colors.

B. Image source

137 (66 normal, 71 malignant) hyperspectral gray scale images at 400x magnification are derived from, respectively, 58 and 62 different tissue microarray biopsies.

C. Data Acquisition

Each measurement yields a data cube C, which is a set $\{I_i\}_{i=1\ldots128}$ of images, each of which 495 by 656 pixels. The intensity of the pixel $I_i(x, y)$ ideally represents the transmitted light at location (x, y) when the i-th light pattern is shone through the sample. The measurement of the hyperspectral image is subject to noise, which is roughly independent of the intensity of light shown through the sample. In order to maximize the signal-to-noise ratio of the measurement of each I_i, given a fixed integration time, one needs to maximize the amount of light shone through the sample. The flexibility of the instrument allows for shining arbitrary patterns ψ_i of light, in the form

$$\psi_i(\nu) = \sum_{j=1}^{N} \epsilon_{ij} \delta_j(\nu) \tag{1}$$

where $\epsilon_i \in \{0, 1\}$ and δ_i represents approximately a δ-function at frequency of index i (and $N = 128$ in our experiment, but the instrument would allow up to $N = 1024$).

Hence we can think of $I_i(x, y)$ as the value of the inner product

$$< f(x, y, \nu), \psi_i(\nu) >_\nu ,$$

where $f(x, y, \nu)$ is the transmittance of the sample at location (x, y) and frequency ν.

A *raster scan* consists in shining the sequence $\{\psi_i\} = \{\delta_i\}$. In this case the energy of light shone through for each I_i will be of the order of

$$\frac{E_0}{N},$$

E_0 being the intensity of the light source. Hence, reasonable signal-to-noise ratios could be obtained only by integrating for a long time.

Multiplexing allows a much faster scan, for a given signal-to-noise ratio, and consists in shining a sequence of *Hadamard patterns* $\{\psi_i^H\}_{i=1,\ldots,N}$. These patterns have the property that for each i there are $\frac{N}{2}$ non-zero ϵ_i's in (1) (so that the energy of the light shining through the sample is about $\frac{E_0}{2}$ for the measurement of each I_i), and also these patterns are quite independent. These patterns have a multiscale structure, in the sense that the index set $\{1, \ldots, N\}$ can be split into subsets

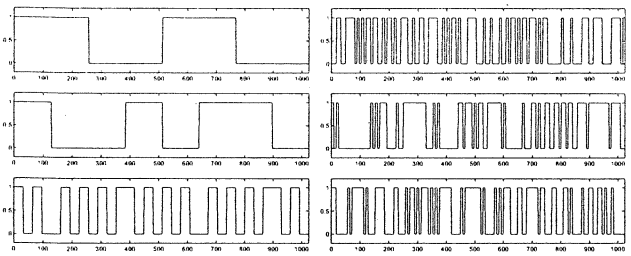

Fig. 4. Examples of Hadamard patterns (left), and randomized Hadamard patterns (right).

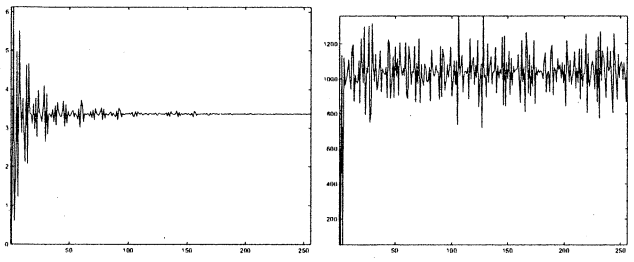

Fig. 5. Example of expansion of a smooth spectrum on standard Hadamard patterns (left) and on random Hadamard patterns (right).

$\{J_1, \ldots, J_{\log_2 N}\}$ such that the patterns in each subsets are constant on dyadic intervals at a certain scale. However, it turns out that in this way the signal-to-noise ratio is not uniformly distributed. This is a consequence of the smoothness of the spectra to be measured and of the structure of the system of Hadamard functions, which implies a priori a decay of $| < f, \psi_j^H > |$ as a function of j. To spread the signal-to-noise ratio uniformly among the coefficients, we consider randomized Hadamard functions, which we obtain by building a random bijection $m : \{1, \ldots, N\} \to \{1, \ldots, N\}$ and considering $\psi_i^{RH}(\nu) = \psi^H(m(\nu))$. We compute this random bijection once and use the induced shuffling in all of our measurements. Of course, the change of variable m simply induces an orthogonal transformation between $\{\psi_i^H\}$ and $\{\psi_i^{RH}\}$. In Figure 6 we show a collected spectrum and the corresponding physical spectrum. It is clear how the size of all the collected coefficients $\{< f, \psi_i^{RH} >\}$ is roughly constant in i. The indices 2 and 4 are special since the corresponding mirror patterns are set entirely to 0 in order to measure background.

IV. Algorithm

The algorithm aims at discriminating between normal and abnormal data cubes. In fact, it would be more useful to be able to classify normal and abnormal (malignant) regions in each sample. This would be particularly important in order to be able to spot abnormal (malignant) regions, which are small and/or only partially present in the sample in question.

The way a trained pathologist would work in analyzing these samples is mainly through pattern recognition. He would look for characteristic structures of large ensembles of cells, such as the structure of glands, their shape, size, and to smaller details such as the shape, size, density and granularity of the

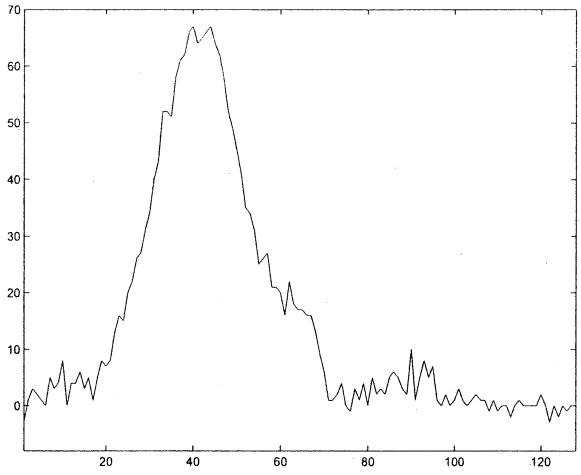

Fig. 6. A typical physical spectrum

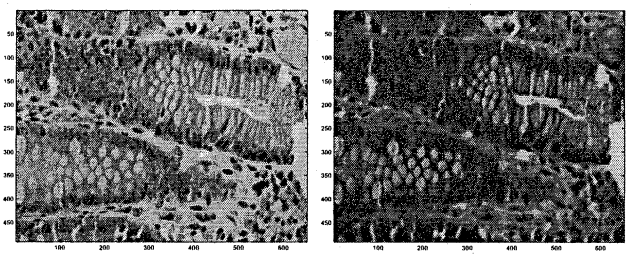

Fig. 7. Two different spectral slices of a normal sample.

nuclei. This kind of pattern analysis is mainly based on rather large scale features, and it could yield inaccurate results on smaller regions.

Our algorithm will generate a classifier for square regions, or patches, with edges of a certain length l which are "admissible", in the sense that contain a certain density of nuclei, as specified below. Each data cube or sample will contain several such "admissible" patches, each of them roughly centered around a nucleus, and of size about the size of the nucleus. Each patch can be viewed as a cloud of l^2 spectral vectors in R^{128}. We will then classify a whole slide by voting among the classifications of the patches in that slide.

We naturally divide the algorithm into the following building blocks:

(1) Nuclei identification

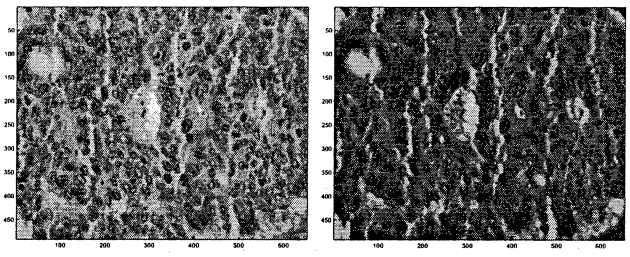

Fig. 8. Two different spectral slices of a very malignant sample.

(2) Collection of admissible patches.
(3) Construction of the classifier based on the mean of the nuclei spectra in each patch.

Step 1. Nuclei identification The first task is to extract the nuclei spectra from a data cube. This is essentially a tissue classification task, and we expect this to be easy since the H&E stain used for the preparation of the slides differentiates between nuclei and the other tissue components. We seek spectral signatures that allow one to discriminate the spectra of nuclei from all the other spectra. In order to do this, we selected about $3,000$ spectra from two distinct datacubes, about one third of which belonged to each of three different classes: {nuclei, cytoplasm, lamina propria/other}. Let these samples be denoted by

$$\{\{v_{i,l}\}_i\}_{l\in\{\text{nuclei,cytoplasm,lamina propria}\}} \subset \mathrm{R}^{128}.$$

We normalize this set of spectra so that each spectrum has $\mathcal{L}^2$-norm, or energy, equal to 1:

$$\tilde{v}_{i,l} = \frac{v_{i,l}}{||v_{i,l}||_2}.$$

We then used LDB on $\tilde{v}_{i,l}$ to find features that best discriminate among the different classes. We found that 4 spectral signatures (see Figure 9) are enough to discriminate among the various tissue types, in particular they are enough to discriminate well the nuclei spectra from all the others. We project our (normalized) training set onto these 4 features, see Figure 10, and to classify a spectrum from any data cube we normalize it and project onto these 4 features. On this projection, we use a 15-nearest-neighbor classifier to identify to which of the three classes the spectrum belongs. Notice that the dimensionality reduction has a de-noising effect on the spectra, thus regularizing the distance computations used by the nearest-neighbor algorithm used to classify in the appropriate low-dimensional subspace. Let us denote by C_{tissue} the classifier that computes this projection and classifies into tissue types as just described. The performance of the classifier C_{tissue} is quite good, uniformly over all datacubes. Mistakes are isolated and can be easily removed by voting among the spatial (x, y) neighbors. From now on we declare that a spectrum is a nucleus spectrum if it is classified as a nucleus spectrum by C_{tissue}.

It is important to remark at this point that the instrument is able to directly measure the projection of the spectrum onto the LDB light patterns by shining exactly these 4 patterns of light through the sample. The results of these measurements can be provided immediately to the nearest neighbor classifier. This saves the millions of CPU operations necessary to project the data onto these features. The flexibility of the device essentially allows one to move these computations from the computer to the instrument itself, essentially performing an ad hoc experiment that measures exactly the quantities of interest.

Step 2. Collection of admissible patches

Now that the nuclei spectra are identified, we define the patches we want to classify as follows. A patch is a subset of a datacube of the form $Q^l_{x_0,y_0} \times S$ where $Q^l_{x_0,y_0}$ is a square of side l pixels long, centered at (x_0, y_0) and S denotes the complete spectral range. A patch is *admissible* if it contains at least $\frac{8}{10}l^2$ nuclei pixels. From now on we will consider each patch simply as a collection of the nuclei spectra it contains, hence as a cloud in $\mathrm{R}^{|S|}$ (with $|S| = 128$ in our specific case).

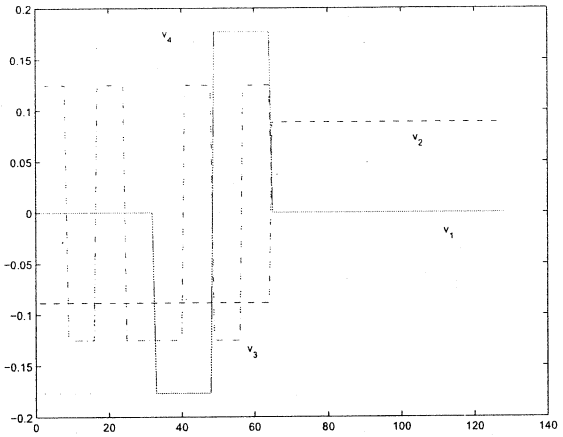

Fig. 9. Top 4 discriminant vectors for tissue classification found by LDB

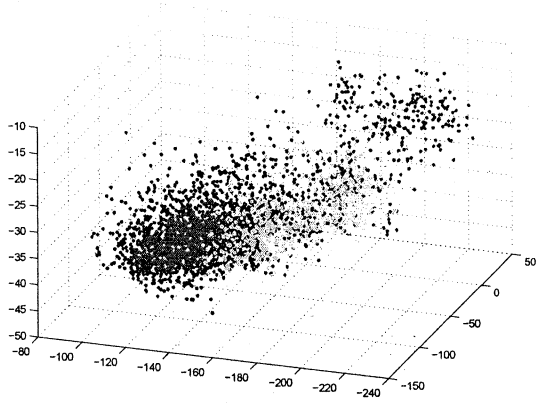

Fig. 10. Projection of the training set for tissue type classification onto its top 3 LDB vectors. Red, green and blue denote respectively nuclei, cytoplasm, lamina propria.

We considered different sets of patches, corresponding to $l = 32, 64$, and 128, and the results improved with the patch size. However, since they are already very good for $l = 32$ (this size corresponds roughly to the size of a single nucleus), we present here the results corresponding to $l = 32$.

The set of patches we consider consists of 2440 patches of size $l = 128$, collected randomly, 30 per slide. We denote by $\{N_{i,k}\}_{k\in K_i}$ the set of nuclei spectra in the i-th patch P_i.

Step 3: Construction of the classifier based on the mean of the nuclei spectra in each patch

For each admissible patch P_i collected, we compute the mean of the nuclei spectra $\{N_{i,k}\}_k$ and we normalize it to unit energy. The label (normal or abnormal) attached to the patch is transferred to the corresponding mean nucleus spectrum. We used PLS, keeping $k = 15$ top vectors, and we ran 50 rounds of 10-fold cross-validation to make sure we are not overfitting. The confusion matrices of the classifiers thus obtained are

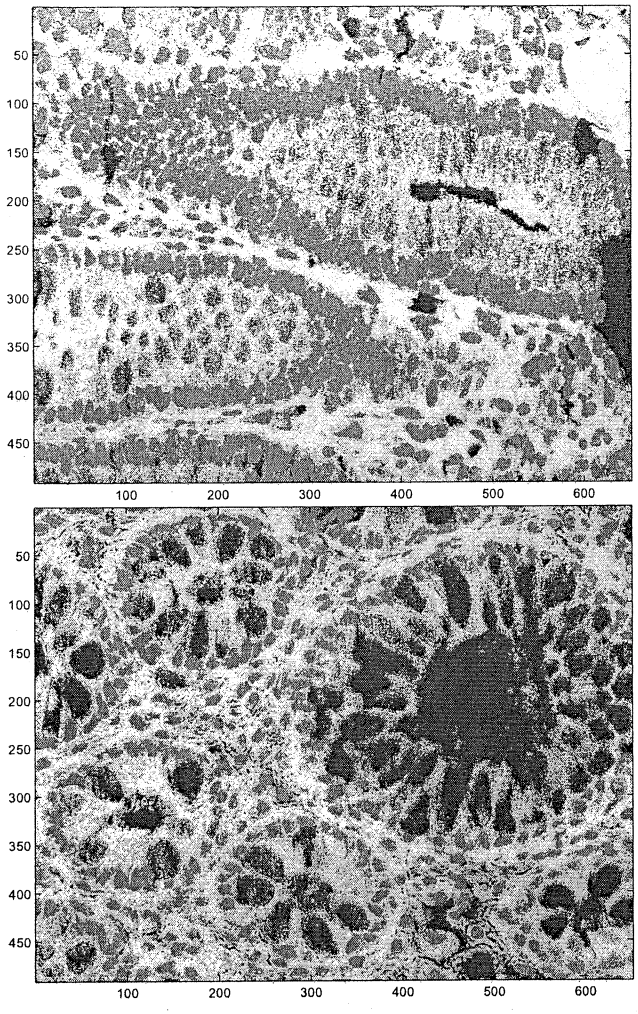

Fig. 11. Two examples of tissue classification.

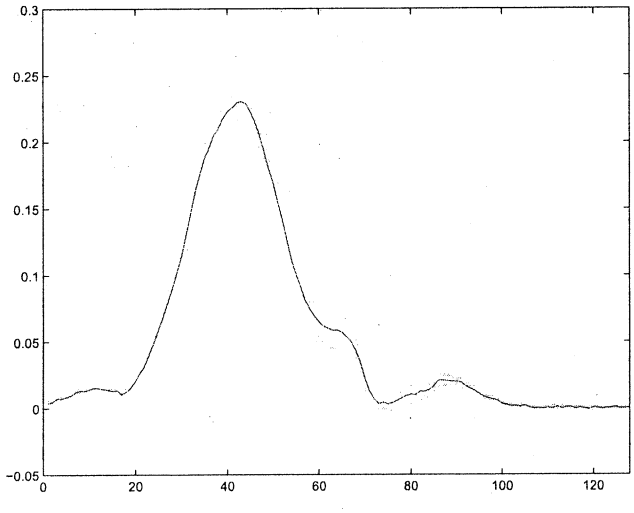

Fig. 12. Local average of nuclei spectra, with 1-standard deviation bars.

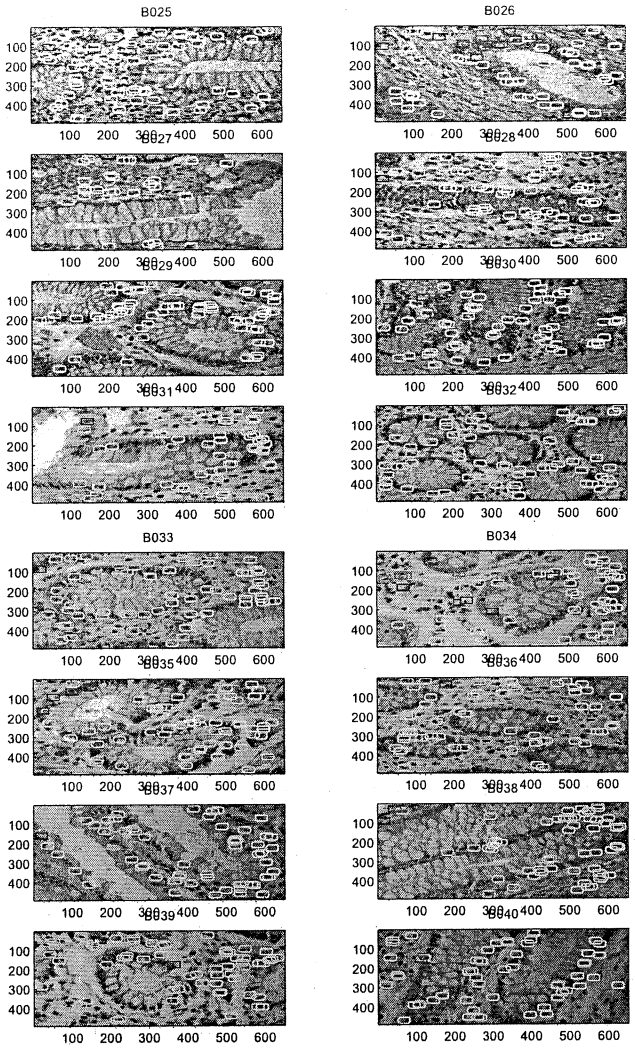

Fig. 13. Classification of nuclei patches of normal samples: green, blue and red mean classified respectively as normal, not-classified, abnormal (malignant).

summarized in table I for classifiers of patches of size 32.

We tried various normalizations of the spectra (energy, baseline) and various subband selections and wavelet packet compression techniques for dimensionality reduction before applying PLS, and to check the stability of the algorithm, but all these attempts gave very similar results. This is due to the large number of samples available for training, even under cross-validation.

V. Classification of Biopsies

To classify biopsies, we collect several admissible random patches from the biopsy, and classify each of them. A biopsy is considered normal if the majority of patches are classified as normal, and the biopsy is deemed malignant if a minimum number (fixed and validated under cross-validation) of patches is deemed abnormal. Of course, more conservative choices could be made, depending on the weight that is chosen for biopsies classified as false positives or false negative. Since

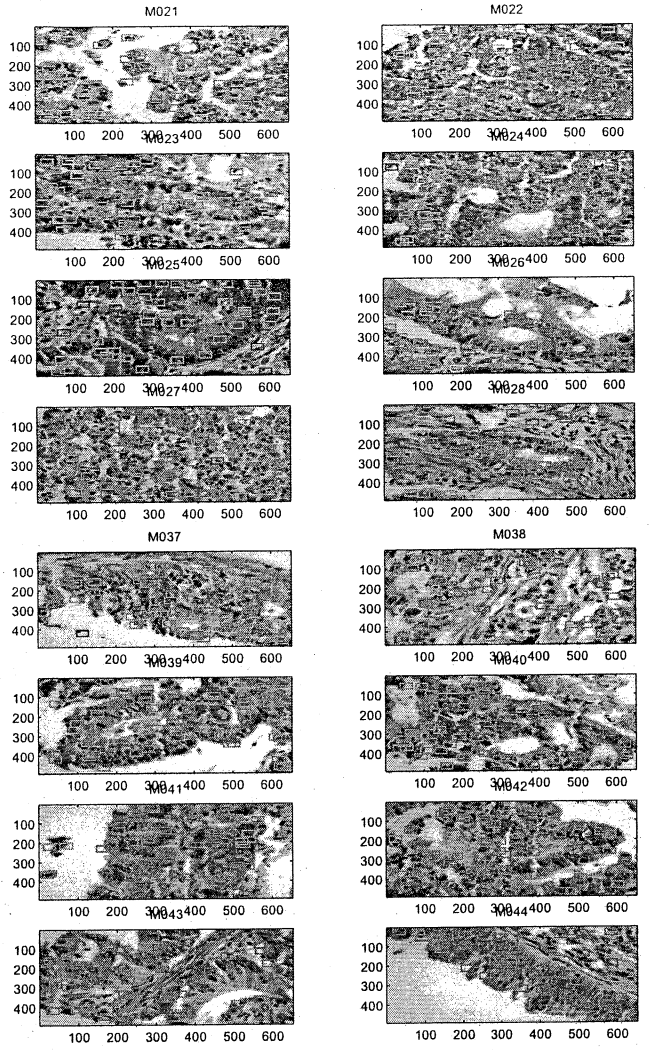

Fig. 14. Classification of nuclei patches of abnormal (malignant) samples: green, blue and red mean classified respectively as normal, not-classified, abnormal (malignant).

TABLE I

CONFUSION MATRIX OF PREDICTIONS OF CARCINOMA VS. NORMAL NUCLEI PATCHES OF SIZE 32 BY 32, AVERAGE 10-FOLD CROSS-VALIDATED ERROR.

	TP	TN
PP	94.0%	7.3%
PN	6.0%	92.7%

Sensitivity	93.5%
Specificity	93.5%
False Negative	6.6%
False Positive	6.6%
Pred Val Pos	94.1%
Pred Val Neg	92.7%
Diag Eff	93.4%

TABLE II

CONFUSION MATRIX, CROSS-VALIDATED, OF PREDICTIONS OF CARCINOMA VS. NORMAL ON 136 BIOPSIES, BY MAJORITY VOTE OF THE CLASSIFIER ON 32 BY 32 NUCLEI PATCHES.

	TP	TN
PP	100%	0%
PN	0%	100%

Sensitivity	100.0%
Specificity	100.0%
False Negative	0.0%
False Positive	0.0%
Pred Val Pos	100.0%
Pred Val Neg	100.0%
Diag Eff	100.0%

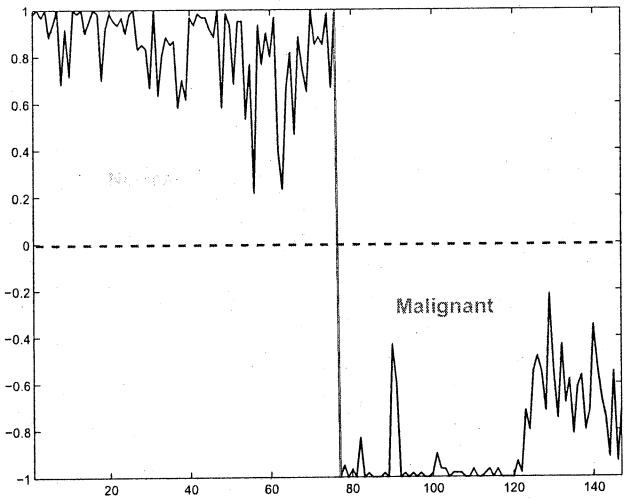

Fig. 15. Classification result for the various biopsies, represented as mean classification of the nuclei patches. The separation between normals (left, around $+1$) and malignant (right, around -1) is sharp.

the classification of nuclei patches is quite accurate, one could, for example, conservatively call a slide malignant if a minimum number m (e.g. 10) of nuclei patches are classified as malignant. In our dataset this did not make any difference in the classification of the slides, for most reasonable values of m (e.g. $m \geq 10$). In any case, no matter which criterion of voting among the classified patches we choose, we have no errors in classifying each data cube (biopsy).

VI. CONCLUSIONS

Hyperspectral imaging presents many challenges from the point of view of data and signal analysis due to the large amount and high-dimensionality of the data. In this paper we present some tools from statistical data analysis, multiscale signal processing and nonlinear dimensionality reduction, that are effective in analyzing hyperspectral data. We demonstrate a successful (100% diagnostic efficiency) application of a combination of these techniques to the problem of automatically discriminating between normal and abnormal (malignant) hyperspectral images of colon biopsies. Further research will involve abnormal but non-malignant samples, and discriminating between these and the malignant abnormal samples. Discriminating between these and the malignant abnormal samples can permit automatic diagnosis.

VII. Acknowledgements

We thank Boaz Nadler for his interesting comments on the application of the PLS method.

Partially funded by National Science Foundation: NSF DMS-01399; Defense Advanced Research Projects Agency (DARPA).

Results presented in part at the United States and Canadian Academy of Pathology, Annual Meeting, March 2004, Vancouver, BC, and the Fourth Inter-Institute Workshop on Optical Diagnostic imaging from Bench to Bedside at the National Institutes of Health, 20-22 Sep. 2004, Bethsea MD.

References

[1] S. Vari, G. Muller, J. Lerner, and R. Naber, *Telepathology and imaging spectroscopy as a new modality in histopathology*, P. K. et al ICS Press, Ed., 1999.

[2] Y. Garini, N. Katzir, D. Cabib, R. Buckwald, D. Soenksen, and Z. Malik, *Spectral bio-imaging*, New York, 1996.

[3] C. Rothman, I. Bar-Am, and Z. Malik, "Spectral imaging for quantitative histology and cytogenetics," *Histol Histopathol*, vol. 13, pp. 921–6, 1998.

[4] A. Papadakis, E. Stathopoulos, G. Delides, K. Berberides, G. Nikiforides, and C. Balas, "A novel spectral microscope system: Application in quantitative pathology," *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 207–17, 2003.

[5] R. Levenson and D. Farkas, "Digital spectral imaging for histopathology and cytopathology," *Proc SPIE*, vol. 2983, pp. 1123–35, 1997.

[6] ——, "Digital spectral imaging for histopathology and cytopathology," *Proc. SPIE*, vol. 2983, pp. 1123–35, 1997.

[7] G. Davis, M. Maggioni, R. Coifman, D. Rimm, and R. Levenson, "Spectral/spatial analysis of colon carcinoma," *Mod Pathol*, vol. 16, p. 320:3321A, 2003.

[8] C. Angeletti, N. Harvey, V. Khomitch, R. Levenson, and D. Rimm, "Detection of malignant cells in cytology specimen using genie hybrid genetic algorithm," *Mod Pathol*, vol. 17, p. Suppl 1:350A, 2004.

[9] C. Angeletti, R. Jaganth, R. Levenson, and D. Rimm, "Spectral analysis: A novel method for classification of urine cytology," *Mod Pathol*, vol. 16, p. 57A, 2003.

[10] T. Barry, A. Gown, H. Yaziji, and R. Levenson, "Use of spectral imaging analysis for evaluation of multi-color immuno- histochemistry," *Mod Pathol*, vol. 17, p. Suppl 1:350A, 2004.

[11] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Info. Theory*, 1992.

[12] E. Candès and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation of objects with edges," Stanford University, Tech. Rep., 1999, preprint.

[13] ——, "Continuous curvelet transform i: Resolution of the wavefront set," Stanford University, Tech. Rep., 2004, preprint.

[14] ——, "Continuous curvelet transform ii: Discretization and frames," Stanford University, Tech. Rep., 2004, preprint.

[15] J. Starck, F. Murtagh, E. Candes, and D. Donoho, "Astronomical image representation by the curvelet transform," *Astronomy and Astrophysics*, vol. 398, pp. 785–800, 2003.

[16] R. Coifman and N. Saito, "Local discriminant bases and their applications," *J Mathematical Imaging Vision*, vol. 5, pp. 337–58, 1995.

[17] R. Coifman, F. Geshwind, N. Saito, and F. Warner, "Discriminant feature extraction using empirical probability density estimation and a local basis library," *Pattern Recog*, vol. 35, pp. 2841–52, 2002.

[18] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[19] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 6, no. 15, pp. 1373–1396, June 2003.

[20] D. L. Donoho and C. Grimes, "Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sciences*, pp. 5591–5596, March 2003, also tech. report, Statistics Dept., Stanford University.

[21] R. R. Coifman and S. Lafon, "Geometric harmonics," Yale University, Dept. Comp. Sci., Tech. Rep., 2003.

[22] R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comp. Harm. Anal.*, 2004.

[23] ——, "Geometric harmonics," *Appl. Comp. Harm. Anal.*, 2004 2004, submitted.

[24] R. Coifman and M. Maggioni, "Diffusion wavelets," *Tech. Rep. YALE/DCS/TR-1303, Yale Univ., Appl. Comp. Harm. Anal., submitted*, Sep. 2004.

[25] J. Bremer, R. Coifman, M. Maggioni, and A. Szlam, "Diffusion wavelet packets," *Tech. Rep. YALE/DCS/TR-1304, Yale Univ., Appl. Comp. Harm. Anal., submitted*, Sep. 2004.

[26] R. R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data. part i: Diffusion maps," *Proc. of Nat. Acad. Sci.*, no. submitted, October 2004.

[27] ——, "Geometric diffusions as a tool for harmonic analysis and structure definition of data. part ii: Multiscale methods," *Proc. of Nat. Acad. Sci.*, no. submitted, October 2004.

[28] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Tran PAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[29] E. Sharon, A. Brandt, and R. Basri, "Fast multiscale image segmentation," pp. 70–77. [Online]. Available: citeseer.ist.psu.edu/sharon99fast.html

[30] H. Martens and T. Naes, *Multivariate Calibration*. Wiley, 1988.

[31] T. Naes, T. Isaksson, T. Fearn, and T. Davies, *User-friendly Guide to Multivariate Calibration and Classification*. NIR Publications, 2002.

[32] S. Wold, M. Sjostrom, and L. Eriksson, "Pls-regression: a basic tool of chemometrics," *Chem. Int. Lab. Sys.*, vol. 58, pp. 109–130, 2001.

[33] D. Haaland and E. Thomas, "Partial least-squares methods for spectral analysis. 1.relation to other quantitative calibration methods and the extraction of qualitative information," *Anal. Chem.*, vol. 60, pp. 211–228, 1988.

[34] A. Hoskuldsson, "Pls regression methods," *J. Chem.*, vol. 2, pp. 211–228, 1988.

[35] R. DeVerse, R. Coifman, A. Coppi, W. Fateley, F. Geshwind, R. Hammaker, S. Valenti, F. Warner, and G. Davis, "Application of spatial light modulators for new modalities in spectrometry and imaging," *Proc SPIE*, vol. 4959, pp. 12–22, 2003.