

Designing Transfer Functions for Exploring Hyperspectral Images

David Mayerich

Michael Walsh

Rohit Bhargava

Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign

Abstract—Spectroscopy is used in several fields to acquire chemical information about a substance for identification or analysis. Recent advances allow the acquisition of hyperspectral imagery for biomedical data. This has promising implications in the fields of biotechnology and medicine, where quantitative analysis of tissue can be performed by directly measuring spatially-resolved chemical information. However, this requires a significant amount of human intervention to identify spectral features for use in classification. We propose a method for designing transfer functions for hyperspectral images. This method allows researchers to interactively adjust parameters used to manipulate the input spectra in order to find metrics that can be used to classify features in the images.

Index Terms—biomedical imaging, medical imaging, spectroscopy, transfer functions

1 INTRODUCTION

Hyperspectral data is composed of a series of samples, generally taken across the electromagnetic spectrum. Different compounds exhibit a characteristic signature in the spectral vector, which can be used to determine the chemical properties of a sample. This technique is often used in the fields of biomedicine and forensics to determine the identity of chemical compounds.

Recent advances in detector technology allow hyperspectral signals to be resolved spatially, producing multidimensional images with an additional spectral component. These techniques have shown promise in biomedicine where sections of tissue can be classified by cell type based on the spatially resolved spectra from a mid-infrared image [2]. However, the interpretation of a spectral signature is a complex task that currently requires human intervention to identify spectral features for use in classification. In addition, a significant amount of preprocessing must be applied before identifying spectral components that correspond to structural and chemical information. This is because chemical and structural features of the tissue are not represented as independent spectral components. Therefore, determining chemical features often requires the removal of structural properties, such as cell density and scattering characteristics.

The goal of this work is to create a tool to aid spectroscopists in selecting features for tissue classification. Current automated methods, such as principle component analysis (PCA) and vertex component analysis (VCA) [5] operate on variance and are prone to capturing noisy features, such as spectra distorted through scattering. In addition, the components produced using PCA and VCA have broad support across the entire spectrum. Since IR imaging can be time-consuming, finding localized metrics is important for clinical applications, since they allow classification with fewer spectral samples. Accurate classification methods combined with fast and non-destructive chemical imaging can play an important role in clinical research, particularly for the diagnosis of cancer biopsies.

We present a method that allows interactive exploration of hyperspectral data sets by building transfer functions similar to those used for volume rendering. We demonstrate preliminary results using biological data acquired from mid-infrared spectroscopic imaging. We first define a transfer function showing the distribution of spectra in the data set. We then allow the user to specify color values in this domain.

- David Mayerich is with the Beckman Institute, University of Illinois at Urbana-Champaign, mayerich@illinois.edu.
- Michael Walsh is with the Beckman Institute, University of Illinois at Urbana-Champaign and Carle Foundation Hospital.
- Rohit Bhargava is with the Department of Bioengineering and the Beckman Institute, University of Illinois at Urbana-Champaign.

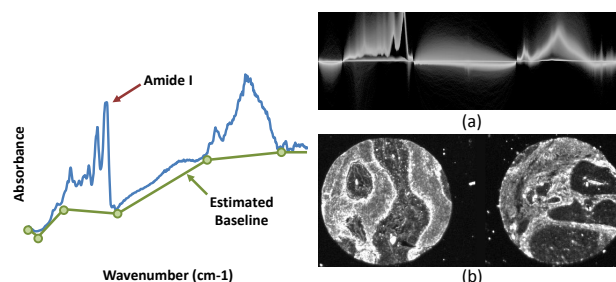


Fig. 1. Transformations applied to hyperspectral data in order to extract chemical information. (left) Baseline estimation (green) removes artifacts due to light scattering. Normalization to known peaks, such as Amide-I (red) compensates for tissue density and thickness. (a) Distribution of spectra that have been baseline corrected and normalized to Amide-I and (b) the 2D image of the Amide-I peak before normalization (after baseline correction). A threshold value is specified at the normalization wavelength in order to prevent the introduction of artifacts from division by small numbers.

This technique is similar to those published previously using spatial characteristics [4]. However, we also allow the user to dynamically adjust preprocessing parameters, thereby changing the distribution of features in the spectral domain.

1.1 Mid-Infrared Spectroscopy

Mid-infrared (IR) spectroscopic imaging is performed by measuring the amount of light absorbed by a tissue sample at wavelengths in the mid-IR region, generally between $2\mu\text{m}$ and $12\mu\text{m}$. The characteristic signature of a chemical compound is caused by the absorption of photons by molecular bonds found in organic molecules. The level of absorption at any point in the spectrum can be correlated with the existence of certain atomic bonds. These correlations can be used to identify the compound being analyzed. Other features, such as the scattering properties of the tissue sample and the cell density at a given spatial location, also affect the spectrum and may provide additional features useful for visualization.

1.2 Image Analysis

The absorbance spectrum measured at a particular spatial location in a tissue sample is affected by three major factors: (a) scattering of light through the tissue sample, (b) the tissue thickness and density, and (c) the chemical composition of the material. Scattering effects are caused by the tissue sample acting as an imperfect lens, distorting the light as it passes through the specimen. Since scattering is based on the wavelength of transmitted light, this distortion differs along the spectrum. In general, this results in a shift in the baseline and peak

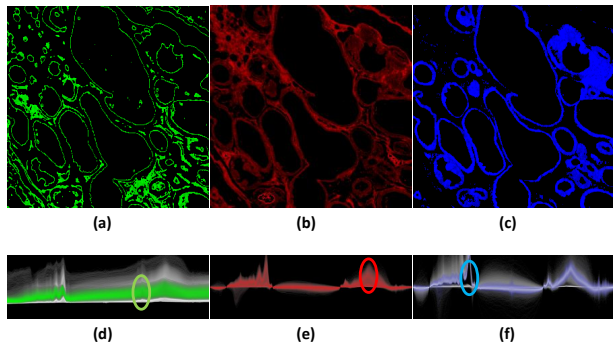


Fig. 2. Regions highlighted in a log-scale joint histogram of the spectral domain with various preprocessing steps applied. (a, d) Scattering effects allow edges to be highlighted by selecting spectra in the raw data in a region known to contain no chemical information. (b, e) After a baseline is applied, cell density is highlighted by setting a gradient along the O-H bonding region. (c, f) Normalization of the spectra to the Amide I peak allows selection of more complex chemical features, such as DNA-rich epithelial cells.

positions. While mathematical models have been proposed to compensate for scattering effects [1], they rely on an understanding of the expected output spectrum. For unknown tissue samples, a piecewise linear baseline correction is generally applied (Figure 1).

Tissue density and thickness cause scaling of the absorbance according to Beer’s Law:

$$A_{\lambda} = b \sum_i \epsilon_{i,\lambda} c_i \quad (1)$$

where A is the absorbance measured at a spatial location, b is the path length (tissue thickness), c_i is the concentration of compound i and $\epsilon_{i,\lambda}$ is the absorbance of compound i at wavenumber λ per unit of concentration [3]. Since Amide I (found at $\approx 1650\text{cm}^{-1}$) is a protein present in most tissue samples, spectra are often normalized to the value at this wavenumber for initial analysis. However, peak height ratios are often useful metrics, therefore normalization to other peaks for exploration is important.

2 TRANSFER FUNCTION DESIGN

We display the projection of the data set into the transfer function domain using a two-dimensional joint histogram of spectral distribution as a function of spectral component λ and signal amplitude. The histogram for each slice is computed on the graphics processor using CUDA. Each value from the raw data set is transformed based on user-specified parameters for baseline correction and normalization. The projection of the data set into the transfer function domain is computed interactively, allowing the user to dynamically change preprocessing parameters and view the corresponding changes in the distribution of spectra. This allows the user to segregate chemical and structural features of the tissue into localized regions in the spectral domain. The user can then select these features, which are rendered in a two-dimensional orthographic projection of the data set in the spatial domain (Figure 2 a-c). The distribution of selected spectra are also shown as an overlay in the spectral-domain image (Figure 2 d-f).

A typical search for classification metrics involves labeling known compounds in the spatial domain. This can be done using histology slides of neighboring sections as a reference. The user then tunes the baseline and normalization criteria to find parameters that localize the selected spectra at some point in the spectral domain. The user tests this region as a metric by highlighting it using a series of widgets [4] and validating the results in the 2D spatial-domain projection.

3 RESULTS

We demonstrate preliminary results of this technique by using dynamic hyperspectral transfer functions to segment physical and chem-

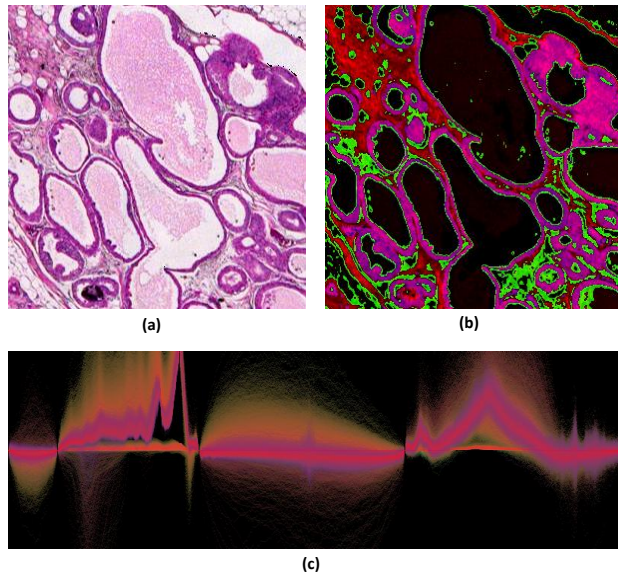


Fig. 3. (a) A tissue sample imaged using IR and stained with hematoxylin and eosin (H&E) for histology. The H&E section is compared to the same region visualized with a transfer function specified using the proposed method. Tissue features such as cell density (red), edge effects (green) and epithelium (blue) are selected independently and combined into a 2D visualization (b) and in the spectral domain (c). While the selection of cell types, such as epithelium, require baseline correction and normalization, these preprocessing steps eliminate other features from the spectra as shown by the overlapping regions in the transfer function (c).

ical characteristics of data from a breast biopsy. We show edge pixels prone to scattering artifacts by selecting them from the unprocessed spectra. By applying baseline correction, we determine tissue density based on IR absorption at the Amide I band. By normalizing to Amide I, we then locate a spectral region unique to epithelial cells that line lumen in the tissue sample. The segmented cells and corresponding spectral regions are shown in Figure 2 and the final classified image and spectral overlays are shown in Figure 3. As seen in the full spectral overlay (Figure 2b), independent selection of these components would be impossible without manipulating the preprocessing parameters.

Because of the size and complexity of these data sets, the ability to interactively search multiple parameter spaces provides a much more efficient method of metric selection. For future work, we will test metrics selected using this technique against those computed using other methods, such as PCA, VCA, and purely manual selection. In addition to providing metrics localized to small regions of the spectrum, we expect that the number of required metrics can be reduced, improving classification by limiting over-fitting.

REFERENCES

- [1] P. Bassan, A. Kohler, H. Martens, J. Lee, E. Jackson, N. Lockyer, P. Dumas, M. Brown, N. Clarke, and P. Gardner. RMieS-EMSC correction for infrared spectra of biological cells: Extension using full mie theory and GPU computing. *Journal of Biophotonics*, 3(8-9):609620, 2010.
- [2] D. C. Fernandez, R. Bhargava, S. M. Hewitt, and I. W. Levin. Infrared spectroscopic imaging for histopathologic recognition. *Nature Biotechnology*, 23(4):469–474, Apr. 2005.
- [3] P. R. Griffiths and J. A. De Haseth. *Fourier transform infrared spectrometry*. Wiley-Interscience, 2007.
- [4] J. Kniss, G. Kindlmann, and C. Hansen. Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets. *Proceedings of Visualization’01*, pages 255–562, 2001.
- [5] J. Nascimento and J. Dias. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):898–910, Apr. 2005.