Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Chemical Profiling of the Plant Cell Wall through Raman Microspectroscopy

Permalink

https://escholarship.org/uc/item/2tv6r1k5

Author

Han, Ju

Publication Date

2010-09-07

CHEMICAL PROFILING OF THE PLANT CELL WALL THROUGH RAMAN MICROSPECTROSCOPY

Ju Han[1,2], Seema Singh[3], Lan Sun[3], Blake Simmons[3], Manfred Auer[2] and Bahram Parvin[2]

[1]Energy Biosciences Institute, University of California, Berkeley, CA 94720, U.S.A.
[2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.
[3]The Joint BioEnergy Institute, Emeryville, CA 94608, U.S.A.

ABSTRACT

This paper presents a computational framework for chemical profiling of the plant cell wall through the Raman spectroscopy. The system enables query of known spectral signatures and clustering of spectral data based on intrinsic properties. As a result, presence and relative concentration of specific chemical bonds can be quantified. The primary contribution of this paper is in representation of raman profile in terms of fluorescence background and multiscale peak detection at each grid point (voxel). Such a representation allows efficient spatial segmentation based on the coupling between high-level salient properties and low-level symbolic representation at each voxel. The high-level salient properties refer to preferred peaks and their attributes for the entire image. The low-level symbolic representations are based on fluorescence background, spectral peak locations, and their attributes. We present results on a corn stover tissue section that is imaged through Raman microscopy, and the results are consistent with the literature. In addition, automatic clustering indicates several distinct layers of the cell walls with different spectral signatures.

Index Terms— Raman spectroscopy, multispectral analysis, spatial clustering

1. INTRODUCTION

Raman spectroscopy is a powerful method for quantifying chemical and morphological information in a biological specimen. It is based on an inelastic scattering process where incident photons, on tissue, transfer energy to or from molecular vibration modes. The Raman spectrum is chemical specific as a result of a unique energy level for each molecule, nondestructive, and label free [1]. It has been used for a wide array of applications, such as (i) quantifying the conversion efficiency from biomass to ethanol, and shown to require little sample preparation, which makes it amenable for high throughput screening of large reaction matrices [2]; (ii) identifying interspecies differences of pollen carotenoid content [3]; (iii) characterizing heterotrophic biofilm in combination with confocal microscopy [4]; and (iv) visualizing changes in the molecular composition of the secondary plant cell wall tissue in poplar [5]. Our objective is to map chemical composition of the plant cell wall; thus, identifying various sugar polymers (cellulose and hemicellulose) and potential barriers to their accessibility (lignin). Such a mapping enables a better understanding of the desirable properties of energy crops as a source of renewable energy with significant economic and environmental benefits [6]. Majority of current biofuels are derived from variant

of food crops, but future energy demands will require that biofuels be derived from a diversity of plant species [7]. A critical step in the biofuel conversion is the biomass (e.g., cellulose, hemicellulose, lignin) deconstruction and to gain fundamental understanding into the pretreatment process for process optimization. Our long term objective is to construct a database of chemical profiles of multiple tissues, for each energy crop, at different stages of development, and under different deconstruction processes. Such a time-varying chemical profiling can also be linked to phenotypic endpoints (e.g., stem size) or genomic manipulation (e.g., selective mutation). Ideally, we are interested in quantifying precise concentration of cellulose, hemicellulose, lignin at the highest possible resolution, given a general model of spectra based on pure compounds. However, there are several technological barriers, i.e., limited diffraction pattern of the optical system, potential interactions of chemical components within each voxel, and potential shift in the Raman spectra as a result of these interactions. Therefore, the current focus is on capturing the intrinsic properties of a tissue specimen based on spatial organization of chemical spectra. In this context, a number of issues needs to be addressed in the context of a database for chemical profiling since each pixel, at each voxel, is represented by a vector of real number corresponding to the chemical composition at that point. These issues are (i) how to capture the fluorescence background, (ii) how to detect peaks, and (iv) how to use intrinsic properties for spatial clustering.

2. TECHNICAL APPROACH

Steps for chemical profiling of the plant cell walls through intrinsic properties is shown in Fig. 1. Low-level spectral representation of each voxel includes the estimated fluorescence background, detected peaks, and their attributes. High-level salient spectral representation is derived from global peak distribution of the entire image. Subsequently, clustering provides automatic spatial segmentation.

2.1. Data acquisition

The Raman image was collected from sclerenchyma cells of a $100\mu m$-thick stem cross section of corn stover. The image stack contains 81×81 voxels with spatial resolution of 100nm, at total 1024 discrete wavenumbers.

2.2. Background correction

Plant species are highly autofluorescent. This autofluorescence signal is not limited to a single spectrum and acts as a non-specific and
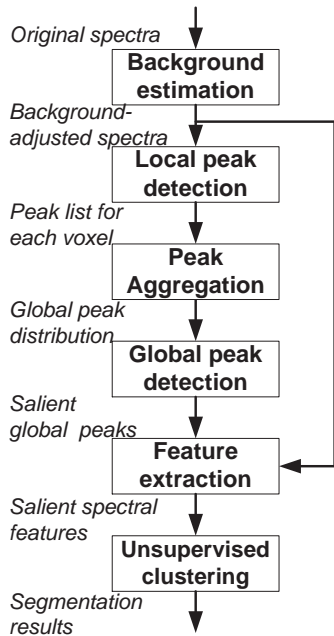
Fig. 1. Steps in chemical profiling of the plant cell wall involves capturing a rich high-level representation of each voxel.

spatially dependent additive signal [8]. Therefore, it needs to be extracted prior to peak detection. But it also needs to be characterized because it provides an important discriminative feature spatially.

There is a rich literature on background correction from multispectral data. These methods are based either on global or local fitting and representation of the spectral signals. An example of background estimation, based on global spectral properties, is through high order polynomial fitting of the original spectral signal. This method is sensitive to the fitting range and the order of polynomial. Lieber and Mahadevan-Jansen proposed a modified multi-polynomial fitting method to reduce the variation of using different parameters [9]. This method is iterative, where at each step the polynomial fitting is performed on the lesser of the original signal and previously fitted signal for each wavenumber. Zhao, *et al* [10] further improved this method by removing peaks and considering noise effects. Although this method improved polynomial fitting, it also has the potential to generate artificial peaks under specific conditions, as shown in Fig. 2(a). On the other hand, local fitting based on the rolling circle [11], eliminates such an artifact, as shown in 2(b). In this method, a circle, with fixed radius, is rolling under the spectral signal. At each wavenumber, the highest intersecting point of the circle with the spectral signal is assigned as the background signal. Although the results of rolling circle fitting also rely on the choice of the radius parameter, this method has less artifact than global fitting methods. This disadvantage has less impact on successive peak analysis in comparison with artificial peaks generated by polynomial fitting, and has been opted in our analysis.

2.3. Peak detection

In the analysis of Raman spectra, local peaks are of great importance since it reflects the abundance of specific chemical bonds. The spec-
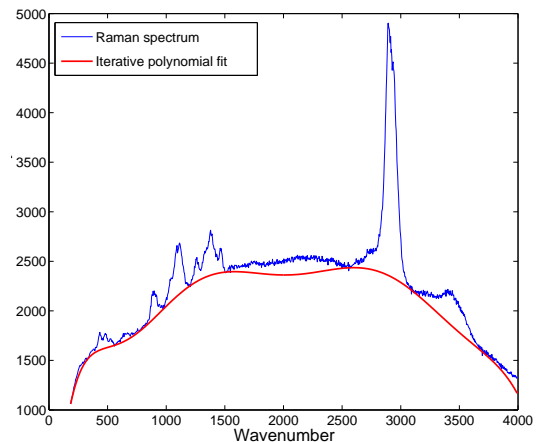
(a)
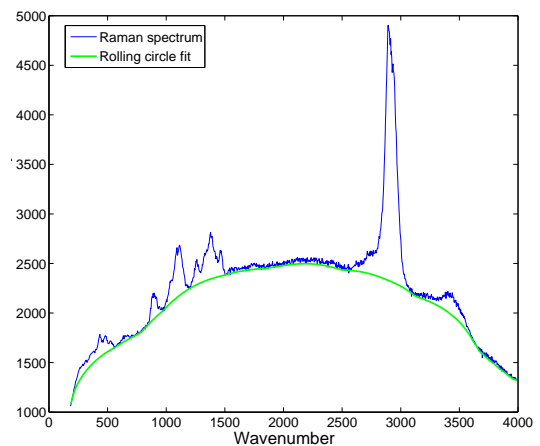
(b)

Fig. 2. Local background estimation performs better than iterative global fitting: (a) Global iterative polynomial fit creates a small artificial peak around 2200 wavenumber; and (b) local rolling circle fit does not create such an artifact.

tral peaks in background adjusted signals are detected through denoising based on undecimated discrete wavelet transform (UDWT) [12]. The UDWT is similar to the DWT except the signal is never down-sampled, that is, the number of the wavelet coefficients does not decrease with the increase of the transform levels. As a result, UDWT is redundant and shift-invariant compared with DWT. This redundancy improves the discrimination between the signal and noise, and, therefore, leads to effective denoising. The computational steps of peak detection are summarized as follows.

1. Compute the UDWT coefficients of the background adjusted spectral signal;

2. Estimate the noise as the median absolute deviation (MAD) of all coefficients divided by 0.67;

3. Compute the threshold as $t = \eta \frac{MAD}{0.67}$, where η is a thresholding parameter;

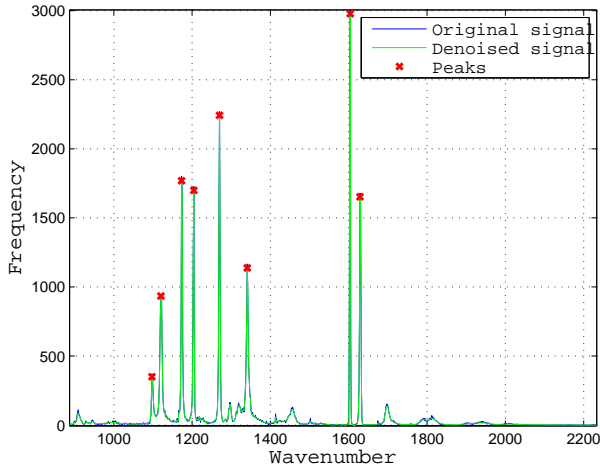4. Set any coefficient less than t to 0;

Fig. 3. Distribution of peaks of 81×81 voxels in a stem cross section of corn stover. Red 'x's indicate the significant peaks detected from this distribution.

5. Reconstruct the denoised spectral signal from the thresholded coefficients;

6. Find local maxima of denoised spectral signal and corresponding endpoints;

7. Compute the signal-to-noise ratio for each local maximum;

8. Assign local maxima with signal-to-noise ratio greater than ϕ are considered as peaks, where ϕ is a user-selected thresholding parameter.

2.4. Feature extraction and similarity analysis

The current focus of our initial research is on capturing the intrinsic properties of a tissue specimen based on spatial organization of chemical spectra. The major obstacle for intrinsic similarity analysis is the high dimension of the original spectral signals, e.g., $1,024$ wavenumbers per pixel. Principal component analysis (PCA) is one of the most popular methods for dimensional reduction. However, the mapped features, in the reduced dimensional space, lose their original meaning (e.g., indication of wavenumbers in this case). On the other hand, representation based on an ordered set of peak locations and attributes provide a compact representation for efficient comparative analysis.

Here, we represent each voxel by salient spectral features according to the global peak distribution of the entire image, where each feature represents the strength of Raman signals around a significant spectral peak. The computational steps are summarized as follows:

1. Estimate the spectral background for each voxel;

2. Detect peaks in the background removed spectral signal for each voxel;

3. Compute the global peak distribution for all voxels in the image;

4. Detect significant peaks and their ranges in the global peak distribution using the same methods as used in Section 2.3;
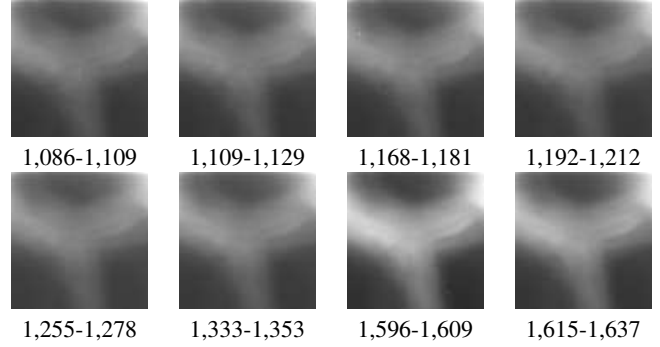

Fig. 4. Raman images of salient peaks from a stem cross section of corn stover. All units are in wavenumber.

5. Compute a feature-based representation of the spectral signal based on position of global peaks. One representation is the total signal as measured by the area under each peak.

Fig. 3 shows the global peak distribution of 81×81 voxels in the image. A K-means clustering algorithm is then applied for capturing spatial similarities and labeling. In our implementation, clustering is performed with multiple random set of seeds for enhanced stability. The results with the least amount of error is then used for final labeling.

3. EXPERIMENTAL RESULTS

3.1. Peak detection and representation

The experiment is performed on the data as described in Section 2.1. The peak list, computed from each voxel, is aggregated to construct a global preferred distribution of the peaks in the image, as shown in in Fig. 3. There are eight significant peaks and the corresponding spatial profiles, for each peak, are shown in Fig. 4. Six of these peaks can be manually curated, based on the known literature [5], for further analysis and association. The final representation is essentially an eight dimensional vector, at each voxel position, which offers a significant dimensionality reduction from the original data.

3.2. Intrinsic similarity analysis

The plant cell wall consists of a complex network of cellulose, lignin, hemicellulose, and pectin. It is of order of *100 nm* in width, and as a result, it is a major challenge for detailed optical microscopy. However, we can delineate different components based on their self similarities. Fig. 5 shows the segmentation results based on condensed spectral features (e.g., driven from peak locations) with two to six clusters. The Raman signals exhibits distinct layers of the cell walls for further analysis. Notice that the clustering is purely based on the spectral features, and no spatial information (neighborhood information among voxels) is involved. Each spectral signal is represented with an eight dimensional feature vector, and Fig. 6 shows the average feature profiles for each segmented region. We suggest that background corrected Raman signature can enhance final clustering and segmentation, and reduces potential artifacts introduced by the autofluorescent signal.
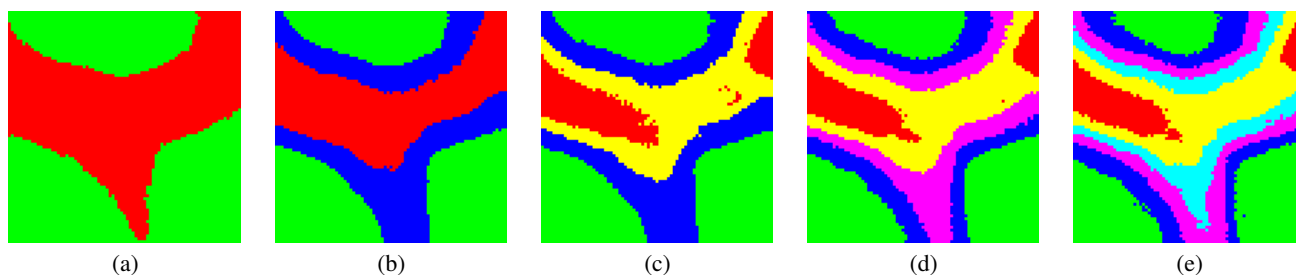
Fig. 5. Clustering results based on intrinsic properties of Raman profile for 2 to 6 clusters indicate potential layers of chemical composition for a stem cross section corn stover.
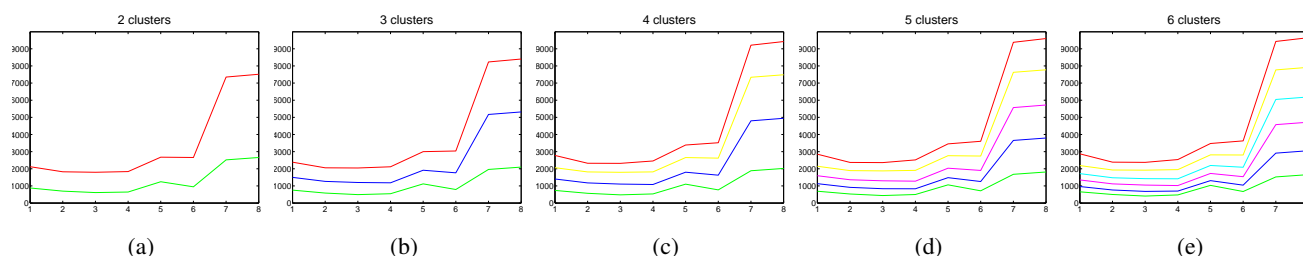

Fig. 6. Average feature profile for each segmented region (a)-(e) of Fig. 5. Horizontal axis indicates the feature index, and the vertical axis indicates the feature value.

4. CONCLUSIONS

In this paper, we presented a computational protocol for chemical profiling of plant tissue sections that are imaged through the Raman spectroscopy. First, the fluorescence background is estimated and the spectral peaks are detected from the background removed Raman Signal. Second, the Raman signal for each voxel is represented by features extracted from detected peaks over the entire image. The net result is a significant dimensionality reduction by focusing on information-preserving regions. Third, the image content is quantized through feature-based representation and k-means clustering. Proposed feature-based representation offers a significant computational efficiency based on meaningful characterization. Our current effort focuses on quantifying composition of the tissue section based on extrinsic properties of the chemical building blocks.

5. REFERENCES

[1] C.V. Raman and K.S. Krishnan, "A new type of secondary radiation," *Nature*, vol. 121, no. 4, pp. 501–502, 1928.

[2] C.J. Shih and E.A. Smith, "Determination of glucose and ethanol after enzymatic hydrolysis and fermentation of biomass using raman microscopy," *Anal Chim Acta*, vol. 653, no. 2, pp. 200–206, 2009.

[3] F. Schulte, J. Mader, L.W. Kroh, U. Panne, and J. Kneipp, "Characterization of pollen carotenoids with in situ and high performance thin layer chromotography supported raman microscopy," *Anal Chem*, vol. 81, no. 20, pp. 8426–8433, 2009.

[4] M. Wagner, N.P. Lvleva, C. Haisch, R. Niessner, and H. Horn, "Combined use of confocal laser scanning microscopy and raman microscopy: investigation on eps-matrix," *Water Research*, vol. 43, no. 1, pp. 63–76, 2009.

[5] N. Gierlinger and M. Schwanninger, "Chemical imaging of poplar wood cell walls by confocal raman microscopy," *Plant Physiology*, vol. 140, pp. 1246–1254, 2006.

[6] S.D. Mansfield, "Solutions for dissolution - engineering cell walls for deconstruction," *Current Opinion in Biotechnology*, vol. 20, no. 3, pp. 286–294, 2009.

[7] C.R. Somerville, "Biofuels," *Current Biology*, vol. 17, no. 4, pp. R115–119, 2007.

[8] A. Mahadevan-Jansen and R. Richards-Kortum, "Raman spectroscopy for cancer detection: A review," in *Proc. 19th International Conference IEEE/EMBS*, 1997.

[9] C.A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological raman spectra," *Appl. Spectrosc.*, vol. 57, pp. 1363–1367, 2003.

[10] J. Zhao, H. Lui, D.I. McLean, and H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical raman spectroscopy," *Appl. Spectrosc.*, vol. 61, pp. 1225–1232, 2007.

[11] I.K. Mikhailyuk and A.P. Razzhivin, "Background subtraction in experimental data arrays illustrated by the example of raman spectra and fluorescent gel electrophoresis patterns," *Instruments and Experimental Techniques*, vol. 46, no. 6, pp. 42–47, 2003.

[12] J.S. Morris, K.R. Coombes, J. Koomen, K.A. Baggerly, and R. Kobayash, "Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum," *Bioinfomatics*, vol. 21, no. 9, pp. 1764–1775, 2005.