Fuzzy Clustering of Raman Spectral Imaging Data with a Wavelet-Based Noise-Reduction Approach

YU-PING WANG,* YONG WANG, and PAULETTE SPENCER

School of Computing and Engineering (Y.-P.W.) and School of Dentistry (Y.W., P.S.), University of Missouri-Kansas City, Missouri 64110

Raman spectral imaging has been widely used for extracting chemical information from biological specimens. One of the challenges is to cluster the chemical groups from the vast amount of hyperdimensional spectral imaging data so that functionally similar groups can be identified. In this paper, we present an approach that combines a differential wavelet-based data smoothing with a fuzzy clustering algorithm for the classification of Raman spectral images. The preprocessing of the spectral data is facilitated by decomposing them in the differential wavelet domain, where the discrimination of true spectral features and noise can be easily performed using a multi-scale pointwise product (MPP) criterion. This approach is applied to the classification of spectral data collected from adhesive/dentin interface specimens where the spectral data exhibit different signal-to-noise ratios. The proposed wavelet approach has been compared to several conventional noise-removal algorithms.

Index Headings: **Raman imaging; Wavelets; Multi-spectral image; Raman spectroscopy; Fuzzy c-means clustering; Image classification.**

INTRODUCTION

Raman spectral imaging can provide qualitative and quantitative information about complex biological samples. One of the challenges of Raman imaging is to condense the vast amount of spectral information into compact, easily visible, and more meaningful data, which would facilitate the identification of chemical groups with similar expression patterns across one or more spectra. The solution to this problem involves high-dimensional data clustering. Conventional multivariate approaches such as principal component analysis (PCA), least squares curve fitting, discriminant analysis, and multivariate curve regression have been effective in processing the spectral data.[1]

In this paper, we applied a fuzzy-based clustering approach to classify the Raman spectral data collected from the components that make up the adhesive/dentin interface. In multi-spectral data classification, the high dimensional spectral representations of a pixel are usually used as features to classify the pixel into a certain chemical group. The conventional clustering approaches label a pixel as a particular class by using a binary membership value of either 0 or 1. The fuzzy-based approaches,[14] however, will assign the pixel a membership degree between 0 and 1 considering the fact that the chemical groups are not completely well separated. These realistic models are more appropriate for clustering the Raman spectral data.

Because of the intrinsic connection between the input features and a classifier, the redundancy in the spectral data, including the noise, should be reduced in order to improve the classification accuracy. For Raman imaging data, both interfering noise and fluorescence background could adversely affect the accuracy of the classification. A number of smoothing approaches have been used for preprocessing the Raman spectral data.[2] These methods include Savitzky–Golay (SG) smoothing[3] and derivative extraction, spline fit subtraction (SFS),[4] and low-pass Fourier transform (FFT) based filtering. These traditional smoothing approaches, however, suffer from a number of limitations. Both the SG- and SFS-based approaches use a polynomial and spline regression, which could attenuate or distort the true spectral features. Because most salient features occur at the peak of the spectral data, the FFT-based method cannot discriminate the feature from the noise by frequency analysis. These limitations have prompted our research efforts to apply a wavelet-based approach to the processing of Raman spectral imaging data.

Wavelet transform is a powerful mathematical transform that has already found wide-spread applications in image compression, vision analysis, and statistical data analysis.[5] Wavelet functions are distinguished from other transforms such as Fourier transform because they not only dissect signals into their component frequencies, but also vary the scale at which the component frequencies are analyzed. As a result, wavelets are exceptionally suited for applications such as data compression, noise reduction, and singularity detection. There are reports in the literature on the application of wavelets in processing spectroscopic data,[6] especially for Raman spectral data denoising.[7–10] In these studies, denoising was performed using conventional orthogonal or bi-orthogonal wavelet transforms followed by thresholding approaches. However, it has been shown that the denoising using a standard wavelet transform can exhibit Gibbs phenomena[11] and that a translation-invariant wavelet transform is more favorable. Furthermore, no objective evaluation of these denoising approaches is available in the literature. We believe that in order to evaluate noise-removal approaches one should test whether they can improve the subsequent data analysis such as the clustering.

In this study, we applied a special family of differential wavelets[12] that was previously designed by one of the authors. These wavelets provide a redundant representation of a signal,

Received 24 October 2005; accepted 23 April 2006.

* Author to whom correspondence should be sent. E-mail: wangyup@umkc.edu.

0003-7028/06/6007-0826$2.00/0

© 2006 Society for Applied Spectroscopy

which is advantageous in image denoising and enhancement.[13] The evaluation of a data denoising process is usually performed by measuring the increase in the signal-to-noise ratio (SNR). However, measuring the noise can be challenging, which makes the SNR difficult to calculate. A more objective and easy to use criterion is that the noise removal should improve the subsequent data clustering or classification.

THEORY

Translation-Invariant Differential Wavelet Transform. A family of wavelets designed by one of the authors[12] is translation invariant, which offers advantages when applied to signal feature extraction, enhancement, and denoising. This approach has been previously discussed[15] and used for chromosome image enhancement.[13] These wavelets are simply taken as the first- and second-order derivatives of spline functions:[12]

$$\psi^n(x) = \frac{d}{dx}\beta^{n+1}(x) \tag{1a}$$

or

$$\psi^n(x) = \frac{d^2}{dx^2}\beta^{n+1}(x) \tag{1b}$$

where $\beta^n(x)$ is the B-spline of order n.

If we define the smoothing and wavelet transforms of a signal f at the dyadic scales as $S_{2^j}f$ and $W_{2^j}f$, we can compute the smoothing operation and wavelet transforms using a fast iterative algorithm:

$$\begin{cases} S_{2^j}f = S_{2^{j-1}}f * h_{\uparrow 2^{j-1}} \\ W_{2^j}f = S_{2^{j-1}}f * g_{\uparrow 2^{j-1}}, \quad j = 1, 2, \ldots, J \end{cases} \tag{2}$$

where $\{h\}$ and $\{g\}$ are the low-pass and high-pass filters, respectively. $\uparrow 2^j$ is the up-sampling operation. Conversely, the signal can be recovered by

$$S_{2^{j-1}}f = S_{2^j}f \cdot \tilde{h}_{\uparrow 2^{j-1}} + W_{2^j}f \cdot \tilde{g}_{\uparrow 2^{j-1}} \tag{3}$$

where $\{\tilde{h}\}$ and $\{\tilde{g}\}$ are the reconstruction filters. This iterative algorithm is called a pyramid-like algorithm.[12]

Unlike the conventional pyramid algorithm for orthogonal or bi-orthogonal wavelet decompositions,[5] no down-samplings are performed in Eq. 2 because the continuous spline wavelet transforms are discretized along the dyadic scales. For the same reason, the transform is translation invariant.[12] This property is ideal for analyzing spectral patterns. When the spectral signal is decomposed in the differential wavelet domain, the true features or patterns still maintain strong correlations across different resolutions while noise components do not. The second advantage is that the wavelets are actually the derivative operations followed by spline smoothing, which is extremely suitable for identifying peaks in the Raman spectral data. The third advantage of the decomposition equation (Eq. 2) is that the filters $\{h\}$ and $\{g\}$ are binomials and difference operations. Therefore, only additions are needed when they are implemented. Filters of any order can be found in the previously published study.[12] In the following we list several filters of lower orders that are used in our experiments.

TABLE I. FIR filters for decomposition and reconstruction based on the 0th-order splines.

Taps	h	$\tilde{h}$	g	$\tilde{g}$
−1	1/2	1/2	−1	1/4
0	1/2	1/2	1	1/4

TABLE II. FIR filters for decomposition and reconstruction based on the 1st-order splines.

Taps	h	$\tilde{h}$	g	$\tilde{g}$
−2				1/16
−1	1/4	1/4	−1	5/16
0	1/2	1/2	1	−5/16
1	1/4	1/4		1/16

Haar-like Wavelets. The Haar-like wavelets are derived when the order of the spline is taken as 0. Table I lists the finite impulse responses (FIRs) of the decomposition and reconstruction filters. These filters (except for the normalization constant) are identical to the conventional Haar filters for orthogonal wavelet transform. The difference between them is that no down-sampling is performed in the decomposition formula (Eq. 2).

Linear and Cubic Differential Wavelets. The FIRs of linear and cubic differential spline wavelets, which have been used to evaluate the proposed algorithms, are listed in Tables II and III. These filters are derived from linear and cubic B-splines. Different orders of spline filters will have different denoising effects.

Denoising Based on Multiscale Point-wise Products. Wavelet-based denoising is usually performed by a hard- and soft-thresholding approach.[16] It has been shown that a translation-invariant transform would produce better results.[11] The specific differential wavelet transforms (Eq. 2) can facilitate the pattern correlation analysis between different scales. Because there is no down-sampling, the feature patterns such as peaks of the signal will have strong spatial alignment (correlation) in the multiscale domain, while noise will not. A quantitative way to measure the cross-scale correlation of the spectral features is by using the multiscale point-wise products (MPPs).[17] These are often used in computer vision and image analysis.[18,19] The MPP criterion is defined as

$$MPP_K(n) = \prod_{j=1}^{K} W_{2^j}f(n) \tag{4}$$

where $\{W_{2^j}f\}$ is the wavelet decomposition at scale j. This measurement can exploit the multiscale correlation due to the presence of desired peaks.

TABLE III. FIR filters for decomposition and reconstruction based on the cubic splines.

Taps	h	$\tilde{h}$	g	$\tilde{g}$
−4				1/256
−3				9/256
−2	1/16	1/16		37/256
−1	1/4	1/4	−1	93/256
0	3/8	3/8	1	−93/256
1	1/4	1/4		−37/256
2	1/16	1/16		−9/256
3				−1/256

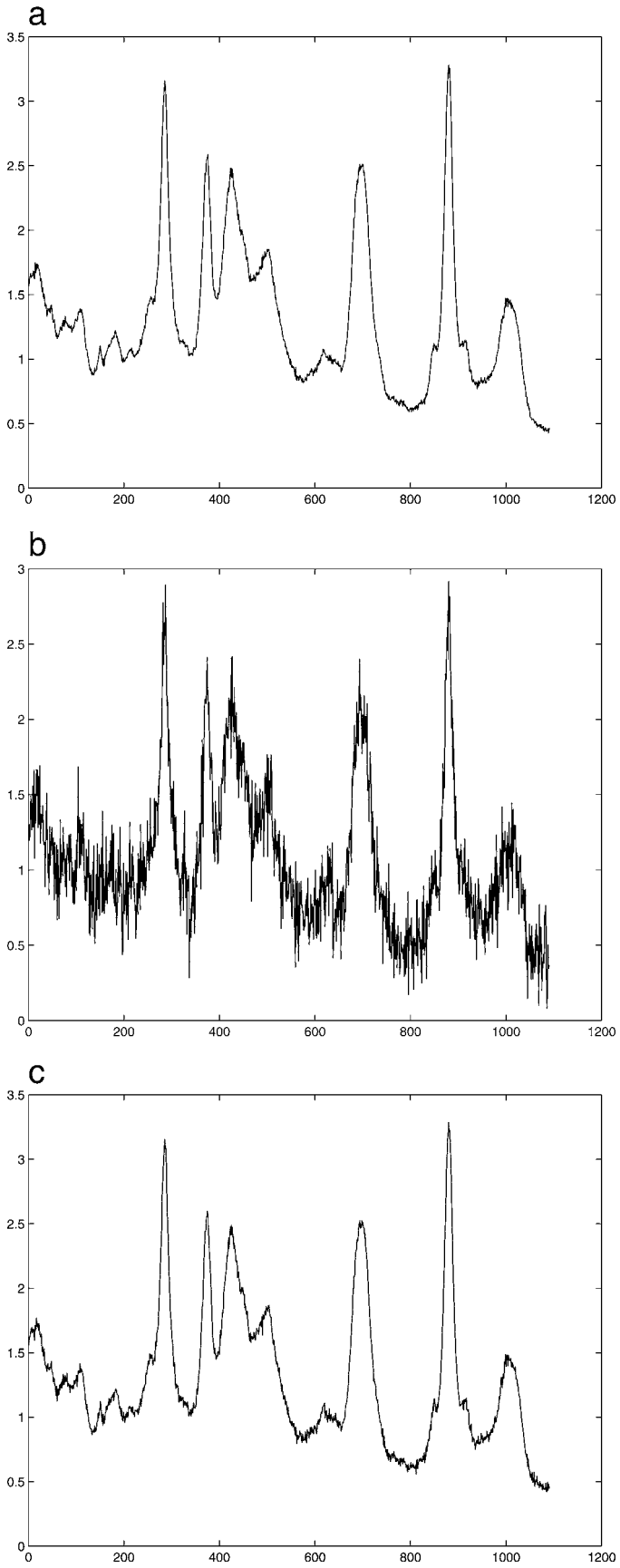

FIG. 1. (**a**) A spectral channel signal. (**b**) 20% noise is added to the signal. (**c**) The spectrum is denoised using MPP with thresholding.

In order to illustrate the correlation analysis using MPP, we use a spectral signal displayed in Fig. 1a. Noise is added to this signal (Fig. 1b). Figure 2 displays the multiscale wavelet decompositions of the spectral signal shown in Fig. 1b from scale 1 to 5. The last row is the smoothing component of the signal. From Fig. 2 it can be seen that the noise is mainly left in the high resolution while peak patterns spread over several scales. Figure 3 shows the values of the MPPs that are computed between two neighboring scales (1–2, 2–3, 3–4, and 4–5). It is evident that the MPPs have larger values in the neighborhood of peaks corresponding to true spectral features but take smaller values at other locations corresponding to noise. We have previously used MPPs to enhance the chromosomal banding patterns in a cell image.[13] The determination of the first-order probability distribution function (PDF) of the MPP was introduced in Ref. 17.

Let $W_{2^j}f$ and $W_{2^j}f$ be joint Gaussian with the zero-mean and the covariance matrix

$$C = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{5}$$

where ρ_{12} is the correlation coefficient. Then, the two-scale point-wise product $MMP_2(K)$ has the PDF:

$$PDF(x) = \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}}\exp(\rho_{12}\sigma_1\sigma_2 x)K_0(\sigma_1\sigma_2|x|) \tag{6}$$

where K_0 is the modified Bessel function of the second kind and zero order. The PDF of the MPP is generally non-Gaussian heavy tailed.[17] There is a sharp peak around the origin of $x=0$, which indicates that the majority of the MPP values are zeros. The larger values of MPP correspond to a small probability (outliers), implying the rare occurrence of the spectral peaks.

Based on the fact that the majority of MPP values correspond to the noise and only a few MPP values correspond to the occurrence of peak patterns, we can design a thresholding algorithm to identify the peaks. This can be formulated as a statistical inference problem. We denote the null hypothesis of no peak being present by H_0 and give an upper limit α to the probability of erroneously rejecting H_0 when in fact it is true. Then, the decision rule is given by the following: if $MPP(x) \geq MPP(\alpha)$, we can reject the null hypothesis with confidence $1-\alpha$; otherwise, we fail to reject it. In this decision rule, $MPP(\alpha)$ is the horizontal axis value for which the area under the right tail of the PDF is α. The threshold level T is taken as this value, i.e.,

$$\int_T^\infty PDF(x)\mathrm{d}x = \alpha \tag{7}$$

The threshold is actually determined by the confidence level $1-\alpha$. Using this threshold value, we can perform the following thresholding to magnify the wavelet coefficients for feature patterns while setting the coefficients for noise to be zero:

$$\tilde{W}_{2^j}f(n) = \begin{cases} W_{2^j}f(n) & \text{if} \quad MPP(n) > T \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

It is critical to choose a proper threshold level because the choice of a low threshold may render some spikes as peaks, while choosing a high threshold might skip some of the important peaks. In other words, higher threshold values can make the algorithm insensitive and would sometimes skip valid peaks, while too low a threshold would detect peak patterns

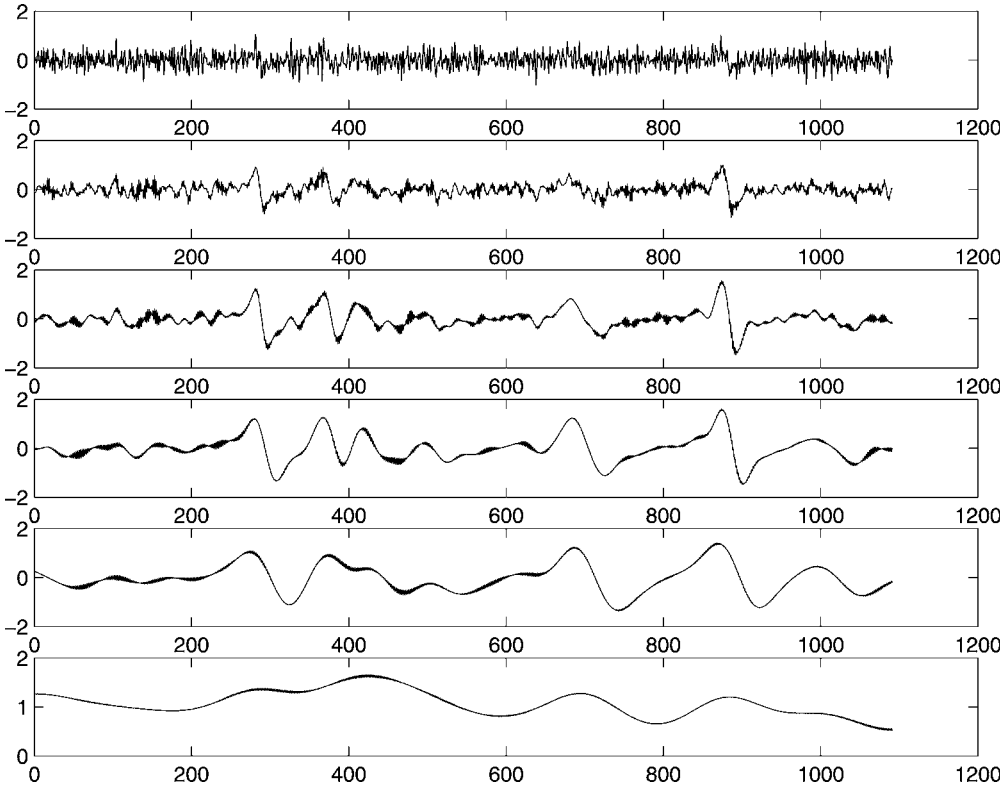

FIG. 2. The decomposition of the spectral signal in the wavelet domain. The figure in the last row shows the smoothing component of the signal.

with weak strength that might be noise. In practice, we normalize the MPP value to be within [0, 1]. The threshold is specified in terms of the percentage of the maximum MPP value. The wavelet thresholding procedure can be summarized in the following steps.

Wavelet-Based Noise-Removal Procedure.

(1) Decompose hyperspectral data into the differential wavelet domain according to Eq. 2.
(2) Perform the thresholding of wavelet coefficients in terms of Eq. 8.
(3) Reconstruct the spectral data using Eq. 3 to obtain a clean signal.

Fuzzy-Based Clustering of Spectral Imaging Data. The goal of clustering is to assign each pixel in the spectral sample to belong to different clusters based on the multi-spectral

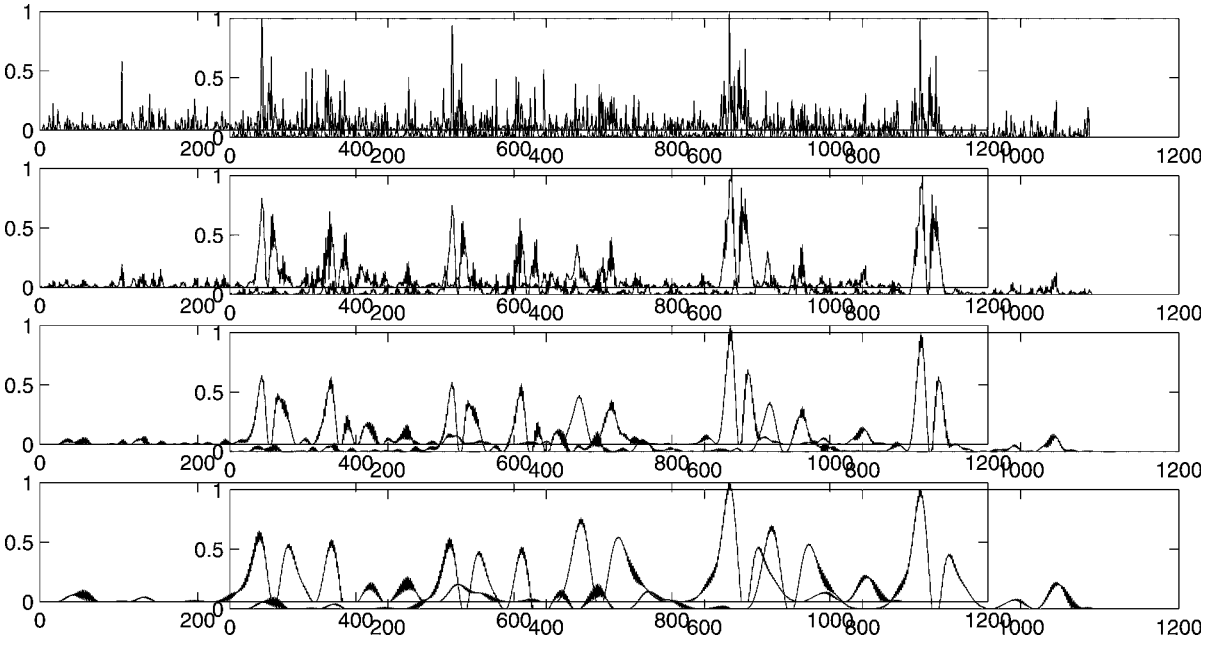

FIG. 3. The plot of multiscale point-wise products (MPPs) between the two neighboring scales of the spectral signal (shown at the top), which measure the cross-scale correlations of the signal.

representations collected from the biological specimen. In this section, we introduce a fuzzy-based unsupervised approach to this problem. This more sophisticated approach assumes that a pixel can belong to more than one class but varies in degree from 0 to 1. The fuzzy-based approach offers advantages over the conventional hard clustering technique, which has been widely used in many biomedical imaging problems.[14]

K-Means or Hard C-Means Clustering. Clustering is a technique to divide a data set into clusters or classes of similar attributes. In conventional cluster analysis, each class or sample is assigned to exactly one cluster. Thus, we obtain a crisp partitioning with sharp boundaries between the clusters. Fuzzy clustering relaxes such restriction; it allows a sample to belong to more than one class in terms of degree of similarity between 0 and 1.

The traditional k-means or hard c-means clustering is obtained by minimizing a dissimilarity (or distance) function given by

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{x_k \in G_i} d(x_k, c_i) \tag{9}$$

where c_i is the centroid of cluster i and $d(x_k, c_i)$ is the distance between the ith centroid c_i and the kth data point x_k. Typically, the Euclidian distance is used as the dissimilarity measure and the overall dissimilarity function is expressed as

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{x_k \in G_i} ||x_k - c_i||^2 \tag{10}$$

With the minimization of this objective function, partitioned groups can be defined by a $c \times N$ binary membership matrix U, where its element u_{ij} is 1 if the jth data point x_j belongs to the group G_i, and 0 otherwise.

$$\mu_{ij} = \begin{cases} 1, & \text{if } ||x_j - c_i||^2 \le ||x_j - c_k||^2, \quad \text{for each } k \ne i \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

The minimization according to this membership matrix leads to the k-means clustering algorithm or the hard c-means (HCM) clustering.

Fuzzy C-Means Clustering. The fuzzy c-means clustering (FCM) represents an improvement over the HCM since it employs a fuzzy partitioning so that a data point can belong to all groups with different membership grades between 0 and 1. The objective function used in the FCM to measure the overall data dissimilarity is given by

$$J(U, C_1, C_2, \ldots, C_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^m d_{ij}^2 \tag{12}$$

where u_{ij} is between 0 and 1, c_i is the centroid of cluster i, and d_{ij} is the Euclidian distance between the ith centroid c_i and the jth point. m is the fuzzifier that controls the degree of fuzziness; higher values make the boundary between the clusters softer, while lower values make it harder. Moreover, the membership values are required to satisfy the following constraint:

$$\sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij} = 1 \tag{13}$$

If additional terms are introduced to regularize the above minimization problem so that

$$J(U, C_1, C_2, \ldots, C_c) = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^m d_{ij}^2 + \sum_{i=1}^{c} w_i \sum_{j=1}^{N} (1 - m_{ij})^m \tag{14}$$

one can obtain the possibilistic c-means clustering (PCM).

The class of the chemical groups to which the pixel j belongs is determined by the maximum value of the membership function u_{ij}. The object functions in Eqs. 10, 12, and 14 are usually minimized by a two-step numerical optimization algorithm.[14] For the FCM, the iteration equations are

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m}, \qquad u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{d_{ij}}{d_{ik}}\right]^{2/(m-1)}} \tag{15}$$

For PCM, the equations are

$$c_i = \frac{\sum_{j \ne 1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m}, \qquad u_{ij} = \frac{1}{1 + \left[\frac{d_{ij}^2}{w_i}\right]^{1/(m-1)}} \tag{16}$$

The membership u_{ij} and the cluster centroids c_i are updated in an alternating approach according to the above equations until the change of the membership function values is less than a predefined threshold.

MATERIALS AND METHODS

This section evaluates the performance of the proposed algorithms when applied to real Raman spectral imaging data.

In order to validate the algorithm, we prepared adhesive/dentin interface specimens containing three classes of chemical components. This model has been studied previously.[20]

The experimental apparatus used to collect the Raman spectra was a Jasco NRS 2000 Raman spectrometer equipped with Olympus lenses and a liquid nitrogen cooled charge-coupled device (CCD) detector. Confocal Raman microscopy offers the unique opportunity to both image the interface and obtain chemical information. It allows visual identification of the position at which the Raman spectrum is obtained. The excitation source was an argon laser operating at 514.5 nm. The estimated power at the laser was 100 mW. After passing through the band pass filter and condensing optics, an approximately 3 mW power laser was incident upon the sample. Instrument fluctuation was evaluated by comparing spectra from standards such as silicon, and each spectrum was frequency calibrated and corrected for chromatic variations in the spectrometer system detection. The adhesive/dentin specimen was placed at the focus of a 100× objective with a numeric aperture of 0.95 and a working distance of 0.3 mm. The spectra were acquired at positions corresponding to 1 μm intervals across the adhesive/dentin interface using the computer-controlled x-y-z stage with a minimum step width of 50 nm. The focus of the laser beam in conjunction with a 25 mm confocal aperture provided a spatial resolution of 1 μm.

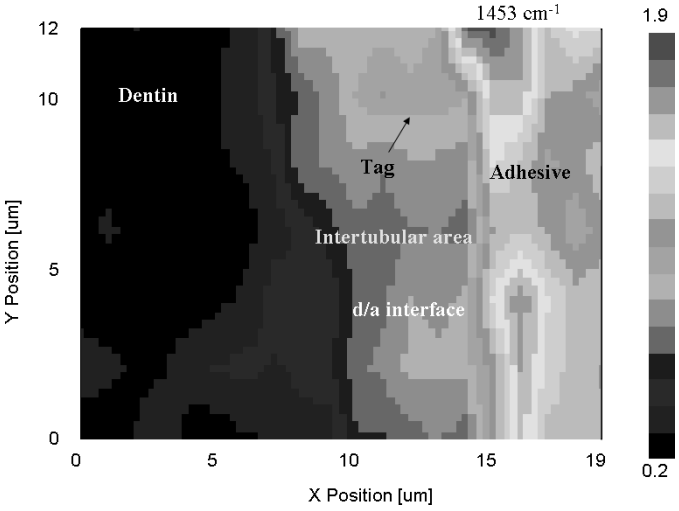

FIG. 4. The Raman spectra data at the 1453 cm^{-1} position is shown as a pseudo-color image (printed in gray scale). More specifically, every pixel in the image that corresponds to the peak of the spectral data is drawn in pseudo-color (printed in gray scale).

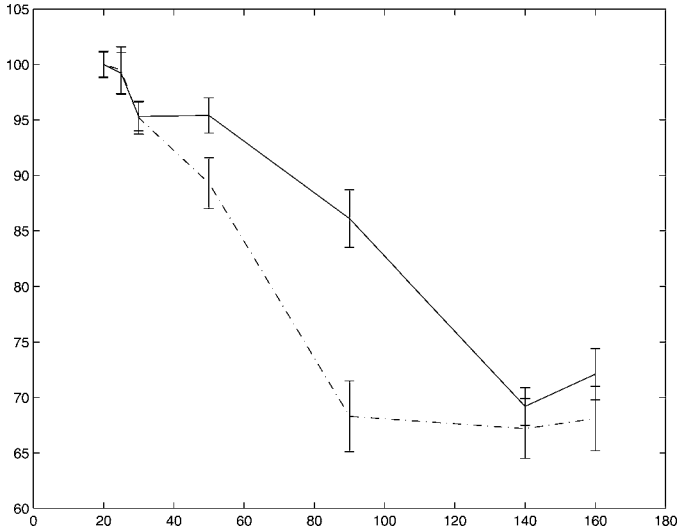

FIG. 5. A plot of the mean and standard deviation of classification accuracy as a function of noise levels after noise removal using Haar wavelet and Savitzky–Golay filter of order 3. The result of using wavelet noise removal is shown as a solid line.

Well-defined areas as small as 1 μm in diameter can be probed. Spectra were obtained at a resolution of 6 cm^{-1} over the spectral region of 875-1785 cm^{-1}, with an integration time of 60 s. Six specimens were used. A digital image of the adhesive/dentin interface with demarcations identifying the position of each spectrum was recorded simultaneously. Multiple sites across the interface of each specimen were examined spectroscopically. Overlap of the spectra from these sites confirmed the reproducibility of the technique.

As an example of illustration, the chemical image of the adhesive/dentin interface at 1453 cm^{-1} is shown in Fig. 4. The regions of the three chemical components of the specimen are shown in pseudo-color (printed in gray scale). Each pixel in the image represents the peak intensity value of the spectral data. The left portion of the figure is the dentin (C), the right part is the adhesive (A), and the middle part is the adhesive/dentin interface (B).

The Raman spectral data set can be represented by an $N \times P$ matrix $\{x_n^p\}$, where $n = 1, 2, \ldots, N$ is the index of pixels in the image, and $p = 1, 2, \ldots, P$ denotes the spectral channel or the wavenumber. In the present experiment the number of pixels in the image was taken to be $N = 100$. The number of spectral bands was $P = 1091$. Therefore, each pixel in the image was represented by a spectrum spanning over the 1091 wavelengths. These spectral data were used as features to classify each pixel into one of the three classes or regions, i.e., adhesive (class A), dentin/adhesive interface (class B), and dentin (class C).

RESULTS

In order to test the performance of the algorithms, we added different levels of noise to the spectral data that we collected. The noise is assumed to be a Gaussian white noise. The level of the noise is specified by a percentage of the maximum peak intensity, i.e., 10, 20, etc.

Evaluation of the Preprocessing Algorithm Using Different Classifiers. To evaluate whether the noise removal can improve the classification, we tested the classification of the Raman spectral data with and without wavelet-based smoothing. Table IV lists the comparison of classification using the HCM approach with and without wavelet preprocessing. Table V compares the classification results for the same data using FCM. These results indicate that significant improvement of the classification accuracy can be obtained after noise removal, independent of the clustering approaches. This further confirms that the preprocessing must be completed before the spectral data clustering or classification is applied.

The experiments also demonstrated that the fuzzy c-means (FCM) based classification, in general, can produce better classification over the hard c-means (HCM) in terms of average classification error and standard deviation, with or without spectral smoothing.

Comparison Between Different Noise-Removal Approaches. An objective way to compare the performance of different denoising algorithms is to evaluate whether the subsequent data clustering or classification can be improved after the data preprocessing. Because the FCM, on average, led to better accuracy, we used the FCM for comparison when using several different noise-removal approaches. Extensive tests have been conducted. As an example of illustration, Table VI lists the accuracy of the algorithm on one data set to which seven different noise levels were added. For the statistical analysis, we also performed ten runs at each of the noise levels

TABLE IV. A comparison of the pixel-wise classification accuracy (in percentage) using HCM with and without denoising.

Noise levels	20	25	30	50	90	140	160	Average	Std
Without denoising	66.67	70.00	63.33	61.67	58.33	31.66	31.66	54.76	15.00
With denoising	100.00	100.00	98.33	93.33	90.00	66.68	66.68	87.86	13.82

TABLE V. A comparison of the pixel-wise classification accuracy (in percentage) using FCM with and without denoising.

Noise levels	20	25	30	50	90	140	160	Average	Std
Without denoising	91.67	70.00	70.00	70.00	73.30	53.33	46.67	67.85	13.47
With denoising	100.00	100.00	96.67	96.67	85.00	70.00	70.00	88.33	12.50

TABLE VI. An example of comparison of classification accuracy (in percentage) between several different denoising approaches. The last rows are the average and standard deviation. The spline and Savitzky–Golay filter of order 3 were used for the test.

Noise levels	No denoising	Spline filter	Savitzky–Golay	Wavelet
20	91.67	91.67	100.00	100.00
25	70.00	91.67	100.00	100.00
30	70.00	91.67	95.00	96.67
50	70.00	70.00	91.67	93.33
90	73.30	70.00	70.00	85.00
140	53.33	73.30	70.00	70.00
160	46.67	68.33	70.00	70.00
Average	67.85	79.52	85.24	87.86
Std	14.55	11.46	14.54	13.22

to calculate the mean and standard deviation of the different methods. For example, Fig. 5 compares the mean and standard deviation of the classification accuracy as a function of noise levels between the Haar wavelet filter and the Savitzky–Golay filter of order 3.

From Table VI and Fig. 5, it can be seen that (1) any denoising approach can improve the classification, and (2) the wavelet, spline, and Savitzky–Golay based denoising methods give comparable results; however, the wavelet approach provides the highest average accuracy from a noise level between 50% to 120%.

DISCUSSION AND CONCLUSION

Spectroscopic imaging is an important modality for the study of the spatial relationships and distributions of functional or chemical groups. The challenge associated with this technique is to condense the vast amount of data into a compact, easily visible, and more meaningful format. The presence of noise in the spectral data can make the problem more difficult. In this paper, we have applied a differential wavelet-based approach for noise reduction in combination with fuzzy clustering. We have tested this algorithm on adhesive/dentin interface specimens with different noise levels and have shown that noise removal can improve the classification accuracy. When comparing wavelet filtering to spline and Savitzky–Golay filtering, comparable results were observed. However, the wavelet-based approach, on average, provides better accuracy. Furthermore, because of the localization property, wavelet filters produce visually more appealing results in the neighborhood of peak patterns, which usually carry the most important information. We have also shown that FCM outperforms HCM in terms of the average classification accuracy. Because FCM is an unsupervised approach, no training data are needed to find the cluster centers. This is especially useful when studying the composition of complex biological materials for which training data are often not available.

In this paper we have taken advantage of a special family of differential wavelets that are translation invariant.[12] These representations enable us to discriminate between the true feature patterns and the noise. In particular, we made use of the multiscale point-wise product to characterize the cross-scale correlations of the spectral signals. We are aware of other wavelet denoising approaches that have been proposed.[7–10] In these approaches, the thresholding was performed in the standard orthogonal wavelet transform domain. However, as discussed in Ref. 11, a translation-invariant wavelet will produce better results. Denoising using orthogonal wavelets usually exhibits visual artifacts that are caused by the Gibbs phenomena in the neighborhood of discontinuities. This is extremely undesirable for Raman imaging data because the discontinuities in the Raman spectra usually contain the most important information. The proposed translation-invariant wavelets can suppress the Gibbs phenomena, which makes the data smoothing more appealing. In addition, these wavelets offer computational advantages. As a powerful mathematical approach, the application of wavelets to spectral data has not been fully exploited.[7] In addition to the denoising, wavelets can also be applied to other problems such as compression of the hyperspectral data and baseline correction. We are currently exploring the application of wavelets to other spectroscopic imaging problems.

ACKNOWLEDGMENTS

This work was supported by the NIH DE 15735 and the University of Missouri Research Board. The graduate student, Aadil Shaikh, prepared some of the data. We thank the reviewers for detailed comments that helped improve the paper.

1. K. E. Shafer-Peltier, et al., J. Cell. Biochem. Suppl. **39,** 125 (2002).
2. Z. W. Yu, J. Liu, and I. Noda, Appl. Spectrosc. **57,** 1605 (2003).
3. A. Savitzky and M. J. E. Golay, Anal. Chem. **36,** 1627 (1964).
4. J. Steinier, Y. Termonia, and J. Deltour, Anal. Chem. **44,** 1906 (1972).
5. I. Daubechies, *Ten Lectures on Wavelets* (SIAM Philadelphia, 1992).
6. L. Shao, X. Lin, and X. Shao, Appl. Spectrosc. Rev. **37,** 429 (2002).
7. W. Cai, L. Wang, Z. Pan, J. Zuo, C. Xu, and X. Shao, J. Raman Spectrosc. **32,** 207 (2001).
8. F. Ehrentreich and L. Summchen, Anal. Chem. **73,** 4364 (2001).
9. P. M. Ramos and I. Ruisanchez, J. Raman Spectrosc. **36,** 848 (2005).
10. T. Cai, D. Zhang, and D. Ben-Amotz, Appl. Spectrosc. **55,** 1124 (2001).
11. R. R. Coifman and D. L. Donoho, in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. (Springer-Verlag, Berlin, 1995).
12. Y.-P. Wang, IEEE Trans. Image Proc. **8,** 1757 (1999).
13. Y.-P. Wang, Q. Wu, K. Castleman, and Z. Xiong, IEEE Trans. Medical Imaging **22,** 685 (2003).
14. J. C. Bezdek, M. A. Sutton, and T. C. Cahoon, *Handbook of Medical Imaging* (Academic Press, New York, 2000), Chap. 2.
15. S. Mallat and S. Zhong, IEEE Trans. Pattern Anal. Machine Intell. **14,** 710 (1992).
16. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, New York, 1998).
17. B. M. Sadler and A. Swami, IEEE Trans. Information Theory **45,** 1043 (1999).
18. A. Rosenfeld, Proc. IEEE **58,** 814 (1970).
19. Y. Xu, B. Weaver, D. M. Healy, and J. Lu, IEEE Trans. Image Proc. **3,** 747 (1994).
20. Y. Wang and P. Spencer, J. Biomed. Mater. Res. **59,** 46 (2002).