Application of multivariate spectral analyses in micro-Raman imaging to unveil structural/chemical features of the adhesive/dentin interface

Ranganathan Parthasarathy
Ganesh Thiagarajan
University of Missouri, Kansas City
School of Computing and Engineering
Kansas City, Missouri 64108

Xiaomei Yao
University of Missouri, Kansas City
School of Dentistry
Kansas City, Missouri 64108

Yu-Ping Wang
University of Missouri, Kansas City
School of Computing and Engineering
Kansas City, Missouri 64108

Paulette Spencer
Yong Wang
University of Missouri, Kansas City
School of Dentistry
Kansas City, Missouri 64108
E-mail: wangyo@umke.edu

Abstract. This study presents the application of multivariate analyses to analyze micro-Raman spectral imaging data in reference to the adhesive/dentin interface as well as comparison with univariate analysis. The univariate statistical methods, such as mapping of specific functional group peak intensities, do not always detect functional group positions and quantities due to peak overlapping. A comprehensive chemical analysis of the adhesive/dentin interface, along with the multivariate statistical methods, principal component analysis, and fuzzy c-means clustering, is studied. Compared to univariate analysis, multivariate methods present the entire hyperspectral information from the specimen in a concise and uncorrelated way. Apart from the ease with which information can be extracted and presented, multivariate methods also highlight minute and often important variations in the spectra that are difficult to observe using univariate methods. The results show for the first time the clear chemical and structural classifications in the adhesive/dentin interface at successively greater resolutions. [DOI: 10.1117/1.2857402]

Keywords: dentin; adhesive; Raman spectroscopy; imaging; principal component analysis; fuzzy c-means clustering.

Paper 07244R received Jul. 6, 2007; revised manuscript received Sep. 10, 2007; accepted for publication Sep. 11, 2007; published online Feb. 27, 2008.

1 Introduction

Polymethacrylate-based dental composites have received widespread clinical acceptance as alternative restorative materials to dental amalgams. The popularity of the composite materials is attributed primarily to their esthetic appearance and the lack of mercury release. However, composites have a failure rate of about 2 to 3 times that of traditional amalgams.[1–3] The failures in composite restorations are commonly found at the adhesive/dentin interface.[4] Composite restoration procedures include enamel and dentin preparation using rotary instruments, acid etching, application of adhesives and methacrylate-based dental composites, and light curing processes of the adhesives and composites. Dentin contains 50% mineral, 30% collagen, and 20% water by volume.[5] Following acid etching, the mineral phase in dentin is removed from a zone that measures up to ~10 um of the bonding surface. Removal of the mineral phase literally suspends collagen fibers in water. On drying, the exposed collagen fibers will collapse, which reduces the space among fibers and inhibits adhesive resin penetration. Hence, a wet bonding technique has been used to achieve interfacial bonding between adhesive and dentin.[6] Characterization and imaging of the chemical structure of the several-micron-thick adhesive/dentin interface of the tooth has thus been an area of great interest.[7]

Researchers have successfully applied micro-Raman imaging spectroscopy toward the characterization of the chemical structure and composition of adhesive resins, collagen, and minerals, as well as to the adhesive-dentin interface.[8–11] Using micro-Raman spectroscopy, the degree of adhesive polymerization, the extent of dentin demineralization, and the amount of adhesive penetration at the adhesive/dentin interface could be measured. 2-D Raman mapping/imaging of the specimen enables us to obtain spectra at points on the specimen separated by 1 μm. Thus the data consists of three dimensions—the x and y coordinates at every pixel, and the intensity value at each Raman shift of the spectrum at that pixel. The data thus obtained are known as hyperspectral data, in view of their large size in each dimension. Analysis of this hyperspectral data is necessary to give a spatial distribution of the chemical makeup of the specimen. Therefore, analysts employ different techniques like univariate and multivariate analyses to filter, understand, and interpret the data. Previous micro-Raman studies on the adhesive/dentin interfaces have used univariate methods of analysis to arrive at the spatial relationships and distributions of the desired functional groups in the specimen.[11,12]

Univariate methods consider one wave number of the Raman shift at a time, thereby providing information about the

Address all correspondence to Yong Wang, Dept. or Oral Biology, Univ. of Missouri at Kansas City, 650 E 25th St, Kansas City, Missouri 64108 United States of America; Tel: 816–235–2043; Fax: 816–235–5524; E-mail: wangyo@umkc.edu

1083-3668/2008/13(1)/014020/9/$25.00 © 2008 SPIE

characteristic functional group corresponding to the wave number/peak concerned. Wave numbers corresponding to the adhesive monomer, collagen, and mineral have been used to characterize the distribution of these components in the interface.[11] However, univariate methods, while presenting the information at a particular functional group, do not represent chemical components in the system as a whole. Further, it becomes very difficult to keep track of minor changes in spectra across the specimen. These subtle spectral differences may be the key to the chemical differentiation of areas such as carious and caries-affected regions.

Multivariate analysis techniques serve to analyze the hyperspectral data by treating each spectrum (pixel) as a whole, rather than considering individual peaks in each spectrum. Various methods of multivariate analysis like principal component analysis (PCA),[13–15] and clustering—k-means and fuzzy c-means[16,17]—have been established by researchers for analyzing 2-D data. Techniques like PCA bring out the correlation between individual pixels, while clustering techniques serve to separate out the specimen pixels into regions based on likeness of their spectra. Also, PCA helps us reduce the size of our data from a broad range of wave numbers to a few scores, each of which describes a principal component spectrum. Often, as few as four to five principal components serve to bring out the major variations in the data, as opposed to a range of Raman shift values numbering in the hundreds. This dimensional reduction of the data not only helps us to immediately recognize the number of components but also remove noise, which would be difficult using univariate analysis alone, unless clear *a-priori* information about the nature of the specimen is present. Clustering techniques give us the option of being able to divide the specimen into chemically different regions in successively increasing detail. Fuzzy c-means clustering has been used for its ability to display overlap of different clusters representing various chemical constituents at a single region. In this study, the fuzzy c-means clustering algorithm uses the scores given by PCA as the basis data. The use of the PCA results as an input to the cluster analysis would provide chemical and structural classifications at successively greater cluster resolutions, as described later.

Multivariate algorithms have been used for various medical diagnostics.[17–20] Although molecular structural information is repressed, the capability to reveal structurally and chemically altered regions in a heterogeneous specimen is enhanced. Bonding adhesive to dentin relies on the extraction of mineral phase and infiltration of adhesive into the voids left by the mineral. Since acid etching of dentin will produce a very complex matrix with variable composition and the presence of water, which could be detrimental for adhesive penetration and polymerization, it is expected that the adhesive/dentin interface, is both chemically and structurally heterogeneous. The purpose of this study was to apply univariate and multivariate spectral analyses in micro-Raman imaging to reveal structural/chemical features of the adhesive/dentin interface, and to demonstrate the importance of multivariate imaging analysis in heterogeneous interface studies.

2 Materials and Methods

2.1 *Adhesive/Dentin Interface Specimen Preparation*

Specimen preparation was done at the University of Missouri, Kansas City (UMKC), School of Dentistry. Six extracted human molars were collected at the Oral Surgery Clinic at UMKC. Based on a protocol approved by the UMKC adult health sciences institutional review board, the teeth were collected after the patient's informed consent. The teeth were then stored in separate vials containing 0.9% normal saline and 0.002% sodium azide and stored at 4 °C. The specimen preparation has been described in greater detail in some earlier publications.[11,12,21]

First, dentin disks were formed by cutting off the roots at the cementum enamel junction by using a water-cooled low-speed diamond saw obtained from Buehler Limited (Lake Bluff, Illinois). The occlusal one-third of the crown was sectioned perpendicular to the long axis of the tooth. Using 600-grit silicon carbide under water, a uniform smear layer was created on these fractions. The dentin was then etched with 35% phosphoric acid for 15 s. Single bond (SB) adhesive was applied to the etched dentin according to the manufacturer's instructions and polymerized for 30 s using a visible light source (Spectrum Light, Dentsply, Milford, Delaware). The "wet bonding" technique was used throughout the bonding procedure.[6,22] The specimens were stored for a minimum of 24 h in water at 25 °C before further sectioning. The treated dentin surfaces were sectioned perpendicular and parallel to the surface using a water-cooled low-speed diamond saw. The final dimension of the slab was 10 mm long, 2 mm thick, and 1.5 mm wide.

2.2 *Data Acquisition*

The Jasco NRS 2000 micro-Raman spectrometer (Jasco Incorporated, Easton, (Maryland) used an argon ion laser beam (514.5 nm) focused through an 60× Olympus Plan Neofluor water-immersion objective [numerical aperture (NA) 1.2] to a ~1.5-μm beam diameter. Raman back-scattered light was collected through the objective and resolved with a monochromator. The spectra were recorded with a software-controlled charge-coupled device (CCD) array. The CCD array goes from left to right across a row of the specimen and the same way in the next row. Laser power was approximately 3 mW. An imaging system and high-resolution monitor were used to visually identify the points at which the Raman spectra are collected. The Raman shift was calibrated using the known peaks of silicon and neon.

The adhesive/dentin specimen was placed at the focus of a 60× water immersion objective and covered with distilled water in preparation for micro-Raman spectroscopic analysis. Spectra were acquired at positions corresponding to 1-μm intervals across the adhesive/dentin interface using the computer controlled *x-y-z* stage with minimum step width of 50 nm. Spectra were obtained at a spectral resolution of ~6 cm^{-1} over the spectral region of 875 to 1785 cm^{-1} and with an integration time of 60 s.

2.3 *Data Preprocessing*

The Raman spectra collected are often contaminated with various interfering signals. The sources of interference in-

clude fluorescence and stretching of spectral intensity amplitudes. The amplitude stretching could be due to specimen inclination, variation in specimen thickness, and absolute intensity of light. To eliminate interfering signals and bring the spectra to a common platform of comparison, certain preprocessing operations are performed. Data preprocessing was carried out by use of procedures written in the laboratory, in MatLab (The Mathworks, Natick, (Massachusetts). As in referenced literature,[23] a baseline correction is used to get rid of the fluorescence spectra. A third-degree polynomial baseline fit has been subtracted out of each spectrum for this purpose.

To correct fluctuations in the intensity amplitudes, signal normalization is performed using a standard normal variate (SNV) at every pixel in the specimen. This is a standard preprocessing technique used in earlier literature.[23] The SNV is given as

$$\mathrm{SNV}_j = \frac{(X_{ij} - \bar{X}_j)}{\left[\sum_{i=1}^{n} \frac{(X_{ij} - \bar{X}_j)^2}{n-1}\right]^{1/2}}, \tag{1}$$

where SNV_j is the standard normal variate of the spectrum at the j'th pixel. X_{ij} is the spectral intensity at the i'th Raman shift of the spectrum at the jth pixel. $\bar{X}_j$ is the average spectral intensity at the chosen pixel, and n is the number of wave numbers in the Raman shift range.

The disadvantage of the SNV is that if the intensity amplitude fluctuations are due to real concentration differences, these concentration differences will be masked, thus limiting quantitative analysis. This admission does not detract from the results of this study, since chemical and structural classification, not quantification, is attempted. After data normalization, we applied the PCA to extract significant features and then use fuzzy c-means clustering to identify chemically important constitutes. There are multi-variant statistical approaches, which are well suited for hyperspectral data analysis.

2.4 *Principal Component Analysis*

PCA is a widely used statistical technique to compress and denoise large datasets.[13] The principal components represent the best set of axes that can be used to view the data. The data are uncorrelated in the new set of axes as opposed to the old one. This shows an important characteristic of the principal components in that they form a basis for the data. Therefore, it is easy to make out patterns in the data. Further, the principal components of the data are arranged in decreasing order of their contribution to the variance. Hence, we can select those components that contribute the most to the variance. There are several algorithms to perform principal component analysis. The algorithm using the noniterative partial least squares (NIPALS) algorithm has been used in the current study, as it performs only the required number of principal components desired. The NIPALS algorithm gives the scores and principal components (PCs) out of the 2-D dataset X. The NIPALS algorithm is based on the fact that the dataset X can be expressed in terms of linearly independent components.

A random vector t from X is chosen as the initial assumption for the first (most important in terms of contribution to variance) principal component. The corresponding score is calculated based on the principle that it is the projection of t on X, or the vector of covariances of t with the other spectra using Eq. (2).

$$p'_{\mathrm{old}} = \frac{t'X}{t't}. \tag{2}$$

Then the score is normalized according to Eq. (3).

$$p' = \frac{p'_{\mathrm{old}}}{\|p'_{\mathrm{old}}\|}. \tag{3}$$

The principal component, which is the projection of the scores on X corresponding to this score, is then calculated as follows.

$$t'_{\mathrm{new}} = \frac{Xp}{p'p}. \tag{4}$$

If the difference between t_{new} and t is less than a tolerance value, the calculation is terminated or else t is set to be equal to t_{new}. When the direction of t is in the direction of maximum variance, this loop converges. In other words, p is now the first (most important in terms of contribution to variance) eigenvector.

After obtaining the first principal component t and the corresponding set of scores p, tp' is now the contribution of principal component 1 to the data and is subtracted from the data.

$$E = X - tp'. \tag{5}$$

The algorithm described is now applied on the remnant E of the dataset X. Proceeding in this way, the other principal components (PCs) t are extracted until they start appearing to be noise. The scores p represent the contribution of each principal component to each vector of X. Since the PCs form a basis for the data, the scores represent the data contained in each vector of X. Henceforth, the set of the principal components are represented as T and the set of score by P.

The 2-D matrix X considered in this study is the matrix consisting of the vector spectra. Therefore the PCs t represent "spectra." The principal components are not true spectra, since they also contain negative intensities. They are linear combinations of true spectra and as such are not expected to be similar to the pure component spectra.[13] Since the principal component spectra now form the basis spectra for the data, the scores of these principal components represent the data. Hence, these scores can be used in future calculations in place of the data. The PCs can be scrutinized based on the equivalence to the pure component spectra. If they are equivalent, then the correlated scores can be represented as 2-D images, which usually correlate to chemical images.

To determine whether the principal component obtained is noise or represents minute variation within the spectral data, a simple visual evaluation of the variation among the PCs[24] is performed. There should be a distinct difference between the PCs that associate with signal as opposed to those associating with noise. Usually, the number of principal components is much smaller than the number of spectra in the data. In this

work, the number of principal components to be used to describe the data was decided by looking at the plots of the first few components.

2.5 *Fuzzy C-Means Clustering*

Cluster analysis is an unsupervised statistical method of grouping the data, without *a priori* knowledge of its nature, purely based on its mathematical attributes. It is a particularly useful technique to characterize hyperspectral data. Moreover, it has been applied quite often in the analysis of medical data.[25] There are many types of clustering algorithms available in the literature like k-means, fuzzy c-means, hierarchical, etc. Fuzzy c-means clustering was chosen over k-means because of its capability to assign pixels to different clusters.

The hyperspectral data in our case are the set of scores p taken at each pixel. Initially, the membership function U is initialized as random numbers.

$$U = (u_{ij}). \tag{6}$$

Here, u_{ij} is a membership value of the i'th cluster at the j'th pixel. Using this initial assumption of the membership functions, the centroids of the clusters are calculated as

$$c_{iq} = \frac{\sum_{j=1}^{M} u_{ij}^{m} p_q}{\sum_{j=1}^{M} u_{ij}^{m}}, \tag{7}$$

where i represents the cluster number; q specifies the score number in the centroid; and M is the number of pixels. Also, $m \in (1, \infty)$ is a fuzziness index. The value of m has been taken as 2 in this work.

Using the centroids obtained, the new set of membership functions is calculated as

$$u_{ij} = \frac{1}{\sum_{k=1}^{NC} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}, \tag{8}$$

where d_{ij} is the distance of the pixel j from the cluster i, k is the cluster number, and NC is the number of clusters chosen.

Using this new set of distance values and the centroids calculated from the earlier set of membership functions, the dissimilarity function between the old set of membership functions and the new set of centroids is calculated as

$$J = \sum_{i=1}^{NC} \sum_{j=1}^{M} u_{ij}^{m} d_{ij}^{2}. \tag{9}$$

The new set of membership functions is now input in Eq. (7) to calculate the new set of centroids. This process is continued until the dissimilarity function J converges to within a set tolerance.

The membership functions u_{ij} here describe the contributions of each centroid score set c_{iq} to a particular pixel j. Each centroid score set can be converted into the corresponding centroid spectrum using Eq. (10).

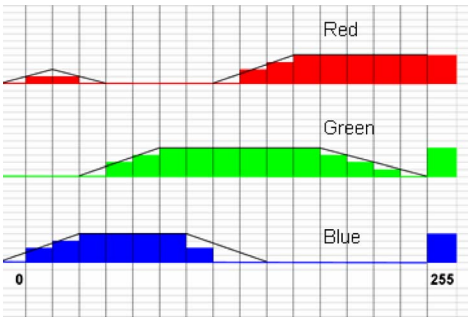

Fig. 1 The color graph ranging from 0 to 255. (Color online only.)

$$C_i = \sum_{q=1}^{S} t_q c_{qi}. \tag{10}$$

The specimen is divided into regions based on the centroid score set c. The centroid spectra C can be used to interpret the regions chemically. Using the membership functions, a pseudo color image is produced according to the following procedure. Every cluster i is represented by a particular color C_i that is picked from a spectrum ranging from blue to red. The graph used for picking the colors is shown in Fig. 1. However, it is noted that in the algorithm, the cluster numbers are assigned randomly; the assignment of colors to clusters is not always in the same manner each time a run is performed. The distribution of the membership function across the specimen is shown in the form of pseudo color images that are developed according the rule:

$$C_j = \sum_{i=1}^{NC} C_i u_{ij}, \tag{11}$$

where C_j is the color of the pixel j, and C_i is the color used to represent the cluster i.

3 Results and Discussion

3.1 *Univariate Analysis*

Univariate analysis results are presented first so that they can be compared and contrasted with the multivariate analysis. There are three major components in this specimen: adhesive, dentin, and the adhesive/dentin interface. The representative spectra of these three components are extracted from the specimen and shown in Fig. 2. The SB adhesive is a bonding agent containing both hydrophilic hydroxylethyl methacrylate (HEMA) and hydrophobic bisphenol A diglycidyl methacrylate (BisGMA) components.[11] The intense peaks related to the SB adhesive occur at 1720 cm^{-1} (carbonyl), 1609 cm^{-1} (phenyl C═C), 1453 cm^{-1} (CH_2 def), and 1113 cm^{-1} (C-O-C). These peaks are related to methacrylate monomers in the bonding agent. Particularly, the peaks at 1609 and 1113 cm^{-1} are related to the BisGMA monomer. In the dentin spectrum, the peaks related to collagen occur at 1667 cm^{-1} (amide 1), and 1245 cm^{-1} (amide 3); the peak related to mineral occurs at 961 cm^{-1} (P-O). The spectrum of the interface shows the contribution from the adhesive and dentin [Fig. 2(c)], indicating adhesive has penetrated into the partially demineralized dentin layer.

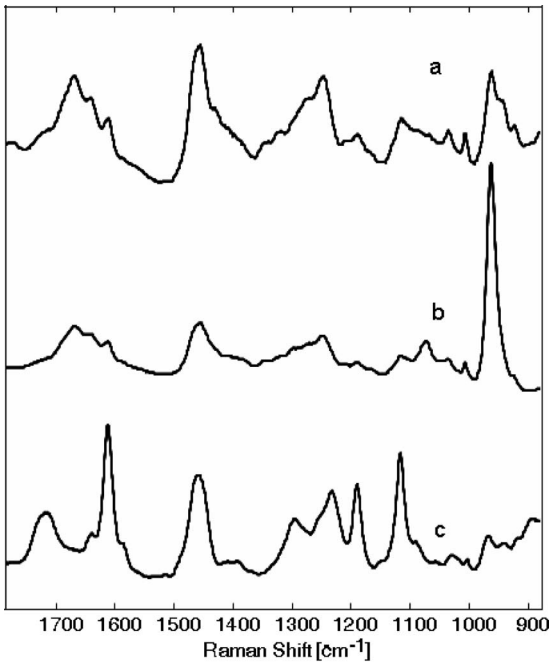

Fig. 2 Three fundamental spectra present in the specimen, that of the (a) interface, (b) dentin, and (c) adhesive, in that order.

The simplest approach of micro-Raman spectral data analysis is the display of specific functional group intensities related to the chemical components of the specimen as a function of position. The univariate intensity maps of the Raman shift values of 961, 1113, 1453, and 1667 cm^{-1}, respectively, are shown in Fig. 3 (the related light image of the adhesive/dentin interface can be seen in Ref. 11). The intensities at these positions are used to display the distributions of the phosphate mineral, BisGMA monomer, SB adhesive (both BisGMA and HEMA), and collagen as a function of position across the adhesive/dentin interface. Black represents the lowest intensity, while red represents the highest. By comparing the spectra shown in Fig. 2, it can be seen that the peaks at 961 and 1113 cm^{-1} do not overlap with other peaks, and that they can be considered to be "unique" peaks for chemical mapping of mineral phosphate and BisGMA. The univariate images based on these two parameters provide reliable information on the spatial distribution of the mineral and BisGMA monomer. However, it is noted that the CH_2 peak (1453 cm^{-1}) of the adhesive overlaps with the CH_2 peak of the collagen. Hence, the univariate image does not give an accurate picture of the description of either adhesive or collagen without spectral subtraction. By the same reasoning, it is observed that Fig. 3(b) shows high intensity of CH_2 in the interface region due to the summing up of the peaks of adhesive as well as collagen. Also, the amide 1 region more or less overlaps the carbonyl and C═C region of the adhesive. Therefore, these images too cannot be used directly for composition interpretation. In summary, the univariate imaging approach does not always precisely identify functional group distribution due to interference from neighboring peaks.

3.2 *Principal Component Analysis*

To evaluate the number of components that can be detected in the mapped spectra data, PCA was applied. The first six principal components from the PCA on the Raman data *X* are

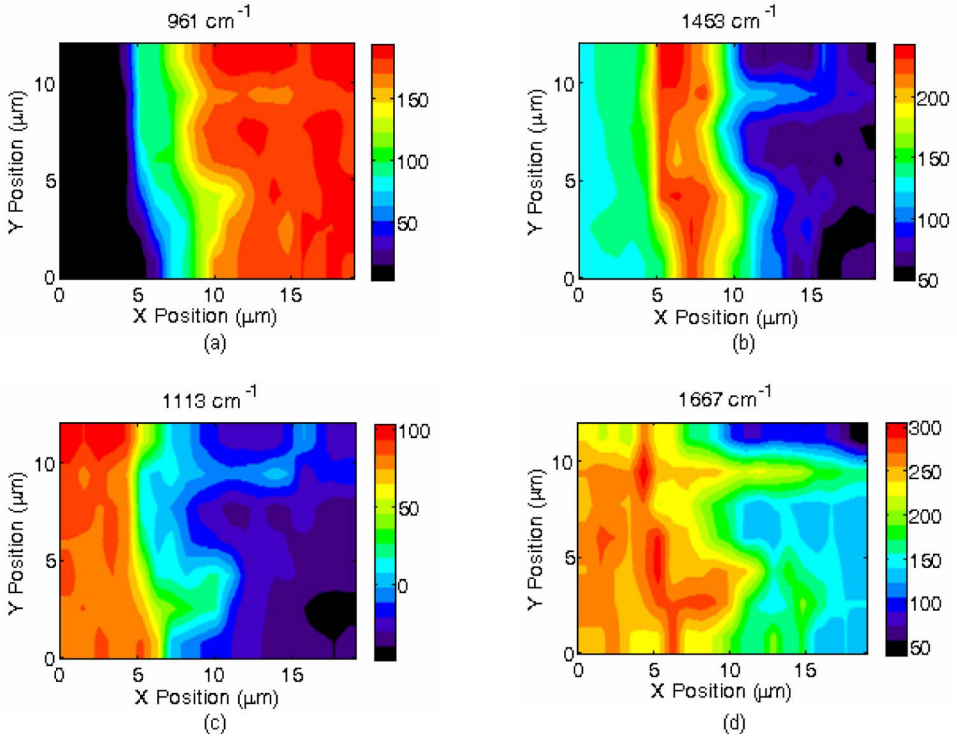

Fig. 3 The pseudo color representations of the distributions of (a) phosphate mineral, (b) adhesive, (c) BisGMA monomer, and (d) amide 1 in the specimen. (Color online only.)

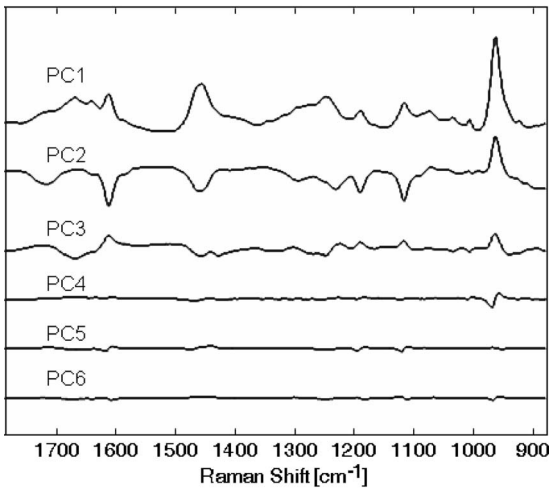

Fig. 4 Principal components of Raman spectral data in decreasing order of their contribution to the variance.

shown in Fig. 4. It is shown that the first three or four components are the ones that contribute to the part of the spectrum comprising the dentin and/or adhesive. The other principal components could either represent minor changes in the chemical structure or could indicate cosmic or other noise. Since it is difficult to ascertain whether a component represents noise or chemical structure change, different numbers of components are tested for results. An important point to be noted is that principal components can sometimes have negative values for their intensities. It is observed that they sometimes are combinations of positive and negative spectra. Hence, the components have to be considered with both the positive and negative signs. As described previously, the scores of these principal components can be used as the basis for the Raman data rather than the intensities at different Raman shift values.

The score of the first principal component shown in Fig. 5(a) represents the partially demineralized dentin in the specimen. The partially demineralized dentin is visible in the region from about 10 to 15 μm in the x direction. The regions from about 5 to 10 μm in the x direction represent the demineralization of dentin in the specimen resulting from acid etching. The region beyond 15 μm represents dentin that is still in the mineralized form. Earlier work[7] used the ratio of 961/1454 cm^{-1} to arrive at the distribution of demineralization. However, the subtraction process necessitates the subtraction of the CH_2 from the adhesive before taking the ratio. Also, principal component analysis shows us that there are two components, namely, partially demineralized dentin and mineralized dentin. This separation of components is difficult to obtain by univariate methods.

Figure 5(b) represents the score of the second principal component, which shows the complementary relationship between the adhesive and the mineral distribution. It represents adhesive in a negative sense and mineral in a positive one. The mineral distribution is seen to correspond closely with that obtained from the univariate imaging shown in Fig. 3(a). We note that the adhesive infiltrates those regions of the interface that are void of mineral. Looking at score 2 in a negative sense helps us to note the decrease in concentration of the adhesive as it penetrates into the demineralized dentin. This result is consistent with that obtained using univariate peak ratio methods.[11,21] It is noted that the adhesive distribution as determined by the earlier methods required the subtraction of the CH_2 peak of the collagen from the CH_2 peak of the adhesive.[11,21]

As shown in Fig. 5(c), score 3 describes the distribution of the collagen in the interface region. This distribution is difficult to obtain by univariate methods, because the amide 1 and 3 peaks of collagen overlap the vibrational peaks of the adhesive.[26] The thickness of the demineralized layer as well the extent of penetration of the adhesive is critical to the effectiveness of the bond between the dentin and adhesive. These characteristics, i.e., depth of dentin demineralization and adhesive infiltration, are directly obtained, noise free from the principal component analysis results, as they are indepen-

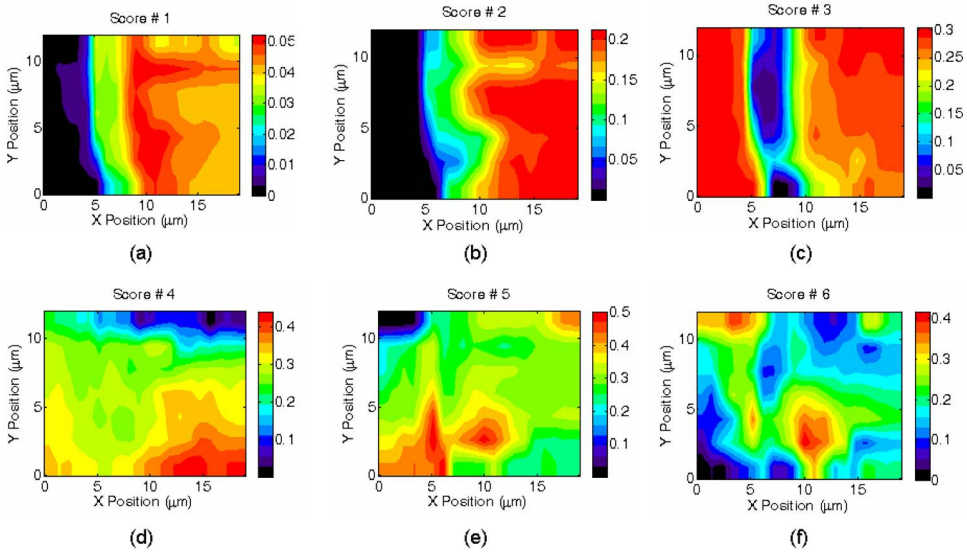

Fig. 5 The scores of the six principal components used in describing the specimen.

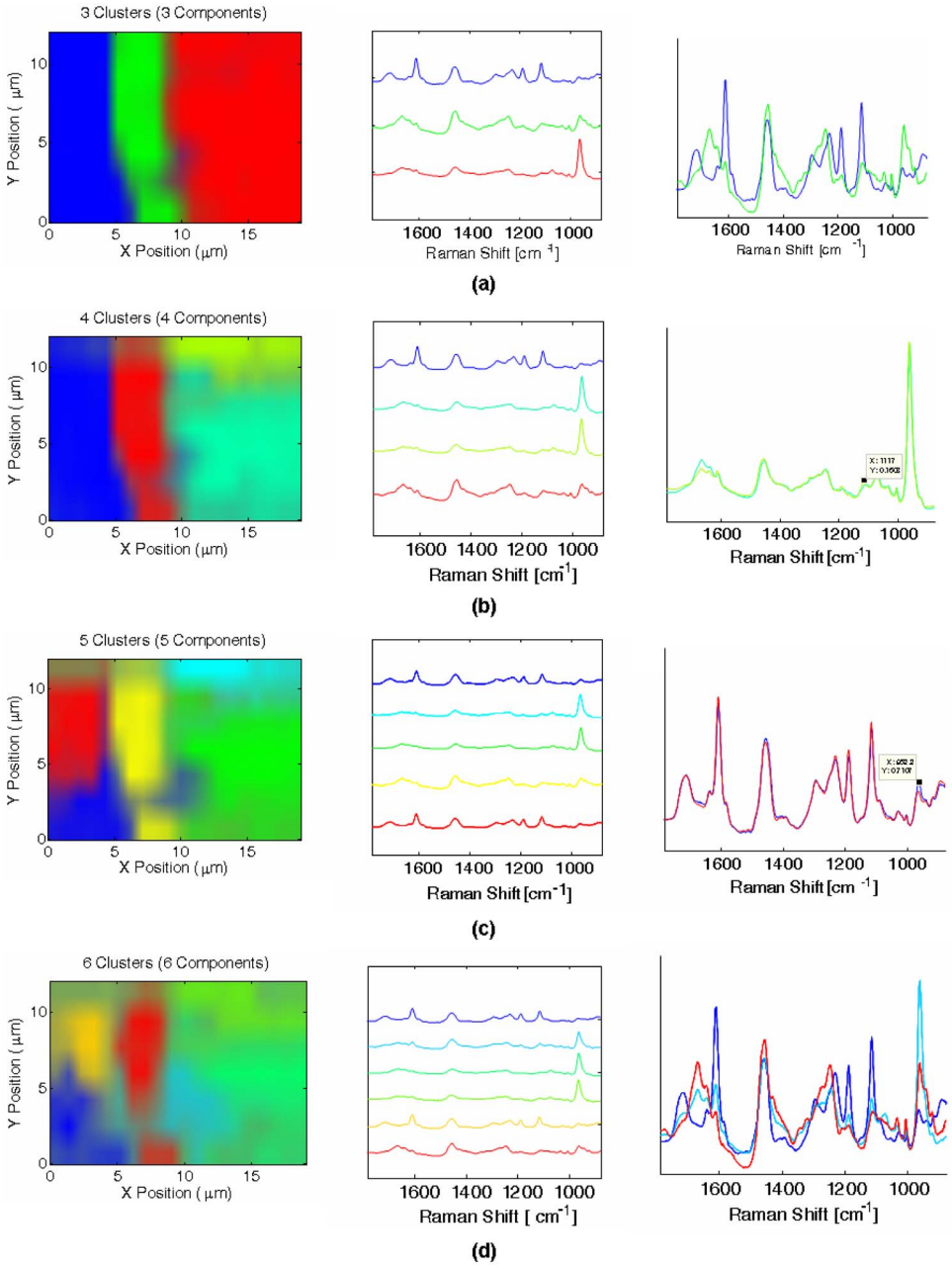

Fig. 6 The clustered images show the division of the specimen into chemically distinct regions.

dent spectral results. It is thus seen that principal component scores extract the spectral information in independent sets in decreasing order of their importance. The latter components are perceived to be either the minor changes in the chemical structure or noise, and their scores are therefore not considered here in extracting information. Other methods such as clustering will be considered for further detailed analysis.

The spectral components detected from the PC analysis are not spectra of pure components but rather linear combinations of pure spectra. The scores of the PCA represent the distribution of linearly independent components that are either complementary or supplementary linear combinations of pure components. Hence, rather than the set of Raman shift values, the principal components can be used as the set of basis vectors to represent the hyperspectral dataset. Correspondingly, the scores of these principal components are their contributions at every pixel and can be used in place of intensity values in fuzzy c-means clustering analysis.

3.3 *Fuzzy C-Means Clustering*

Cluster analysis separates the group of spectra into clusters with clear similarities within each cluster and distinctions between the clusters. Figure 6(a) shows the first image where the data has been split into three clusters. The regions are identified by observation of the cluster centroid spectra shown in the same figure. This is a low level of resolving the data, where it can be seen that the specimen is divided into three

major components: the adhesive, the interface, and the dentin. Different regions have different spectra clusters. The demineralized collagen matrix region is obtained directly without having to subtract out the CH_2 peak of the adhesive. From the comparison of the interface spectrum with the adhesive spectrum [shown in Fig. 6(a)], the BisGMA monomer concentration is readily observed to be lesser in the interface region as compared to the other SB adhesive components (mostly HEMA) concentration.

To observe further detail in the specimen, the number of clusters is increased successively and the details are shown in Fig. 6(b). In addition to the details observed in three clusters, a separate region where there is more mineral concentration and correspondingly, a lesser concentration of adhesive penetration, can be seen. This can be observed from the comparison of the spectra of the two new clusters obtained [Fig. 6(b)]. The new cluster region can be correlated with score 1 in Fig. 5(a), where the region with higher mineral concentration corresponds closely with the cluster that has a lesser concentration of the adhesive. Also, it is observed that adhesive penetration is not consistent with dentin demineralization. Comparison of Fig. 6(c) to the partially demineralized zone observed in score 2 [Fig. 5(b)] shows that the partially demineralized zone to the right of the interface beyond 12 μm shows almost no adhesive infiltration.

From Fig. 6(c), in addition to the earlier clusters formed, the adhesive can be seen to be split into two regions in which the lower region has a slightly lesser concentration. The centroid of the new cluster is observed to have a slightly lower BisGMA content compared to the entire adhesive cluster. Using the clustering method, the chemically and structurally altered regions in the adhesive layer are clearly seen. The alterations are almost impossible to see using univariate methods. This information is very important for understanding the quality and durability of the adhesive/dentin interface.

From Fig. 6(d), in addition to the information obtained from five clusters, we can now see still greater detail in the interface region. The new cluster formed indicates the penetration of the adhesive through the interface. This cluster is observed to be particularly rich in BisGMA monomer, compared to the hybrid layer. The BisGMA monomer has therefore not penetrated very well into the interface, but has spread around it. These results match the univariate image shown in Fig. 3(c). These results agree well with the previous results.[11] The new cluster centroid also indicates the presence of layers with lesser concentration of adhesive that has penetrated through the demineralized dentin matrix.

Further clustering results in the formation of duplicate clusters, and inclusion of more principal components is not found to produce significant changes in the clusters. Hence, the number of clusters as well as factors is fixed at six. A minimum number of clusters divide the specimen into regions of different chemical makeup, while further clusters tend to reveal differences in the concentrations of the chemical constituents present within basic clusters.

In this study, both univariate and multivariate methods were used to extract information from 2-D Raman imaging data. Univariate methods bring out the distribution of functional groups that are taken to represent particular constituents in the specimen. For example, to study mineral distribution, adhesive resins, BisGMA monomer, and collagen, 961 cm^{-1} (P-O), 1453 cm^{-1} (CH_2), 1113 cm^{-1} (C-O-C), and 1667 cm^{-1} (amide) are used, respectively. However, the peaks represented by these groups sometimes overlap and are sometimes inadequate to separate out chemically different regions. Apart from this, it is difficult to separate out regions at successively greater resolutions. For example, it becomes difficult to separate out the zone of partially demineralized dentin from fully demineralized dentin.

These results can be obtained without such *a-priori* knowledge and spectral manipulation by using multivariate methods, namely, PCA and fuzzy c-means clustering as demonstrated in this work. From the score images of the principal components, the distributions of partially demineralized dentin, adhesive, and mineral distribution, as well as collagen fiber arrangement, were observed. The fuzzy c-means clustered images of the specimen indicated results like the unequal distribution of the two monomers—HEMA and BisGMA—and the mismatch between the extent of demineralization and adhesive penetration. While five clusters showed the penetration of the adhesive into the collagen fiber network, six clusters further split the specimen into regions that show the difference in monomer concentrations and penetration over the specimen. Since the wet bonding technique was used, relative hydrophobic BisGMA could not penetrate as well as hydrophilic HEMA, as observed from the relevant cluster centroids in all the cluster sets. Similar results were observed using univariate methods. However, while using pseudo color representation of fuzzy clustered scores as demonstrated in this work, all the zones can be observed on the same image, rendering the interpretation easy. Also, separation of chemical spectra distinct to each chemical is inherent in the algorithm, and digital subtraction of spectra is not required.

4 Conclusion

The thickness of the adhesive-dentin hybrid layer can be observed directly from the clustered images. According to the clustering results, the specimen in this study is seen to be composed of four broad zones: 1. adhesive zone with SB adhesive; 2. lower layer preceding the interface containing higher concentration of BisGMA compared to the interface, and also independently penetrated through the hybrid layer; 3. interface zone where the adhesive and collagen fibers have intermingled (it has relatively lower BisGMA penetration); and 4. zone where BisGMA monomer has also managed to penetrate a little independent of the interface. In addition, a demineralized layer with lesser adhesive penetration compared to the interface portion, and partially demineralized dentin with almost no penetration of adhesive, were also differentiated. As compared to the univariate method, not only do the results image adhesive penetration more accurately, but they also enhance the confidence in the results based on the fact that the imaging represents whole spectra and not intensities at individual Raman shifts. For the first time, the chemical and structural altered regions in the adhesive/dentin interface were clearly observed.

In summary, the application of multivariate methods to analyze the adhesive/dentin interface provides a relatively unsupervised technique to bring out the chemical detail in the adhesive/dentin specimen compared to traditional univariate

methods. The principal component analysis defines a set of basis spectra for the specimen. This obviates the necessity to worry about overlapping peaks and other such problems, since each principal component represents one or two chemical constituents in the specimen that are uncorrelated with the others. By trial and error, six principal components are found to be significant in this study. Their scores form the input for clustering. In the current study, an analysis of six centroid spectra give significant information about the spatial distribution of chemical components in the specimen. Principal component analysis serves to get rid of noise and greatly reduce the size of the basis needed to describe the spectral dataset. The specific implementation of PCA results as an input to fuzzy c-means clustering is somewhat novel. When input into the clustering algorithm, the scores of the principal components split the specimen into regions characteristic of their chemical composition. The cluster centroids represent spectra characteristic of the cluster; the cluster represents the region as a whole, as opposed to studying individual pixels or intensities. In other words, the spectra are classified in a systematic manner so that the study of a relatively small number of spectra directly gives information about the distribution of various components in the specimen, as opposed to univariate methods that involve a more trial-and-error-like procedure along with a necessity for *a-priori* knowledge of the specimen.

Acknowledgments

This investigation was supported by U.S. Public Health Service (USPHS) research grants DE 015281 (Wang) and DE 15735 (Wang) from the National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland. The authors gratefully acknowledge 3M ESPE, Dental Products Division, for donating the dentin adhesive products used in this study.

References

1. C. J. Collins, R. W. Bryant, and K. L. V. Hodge, "A clinical evaluation of posterior composite resin restorations: 8-year findings," *J. Dent.* **26**, 311–317 (1998).
2. E. A. Kidd and D. Beighton, "Prediction of secondary caries around tooth-colored restorations: a clinical and microbiological study," *J. Dent. Res.* **75**, 1942–1946 (1996).
3. E. A. Kidd, F. Toffenetti, and I. A. Mjor, "Secondary caries," *Int. Dent. J.* **42**, 127–138 (1992).
4. M. Hashimoto, H. Ohno, M. Kaga, K. Endo, H. Sano, and H. Oguchi, "Resin-tooth adhesive interfaces after long-term function," *Am. J. Dent.* **12**, 211–215 (2001).
5. G. W. Marshall, S. J. Marshall, J. H. Kinney, and M. Balooch, "The dentin substrate: structure and properties related to bonding," *J. Dent.* **25**, 441–458 (1997).
6. J. Kanca, 3rd, "Effect of resin primer solvents and surface wetness on resin composite bond strength to dentin," *Am. J. Dent.* **5**, 213–215 (1992).
7. P. Spencer, Y. Wang, J. L. Katz, and A. Misra, "Physicochemical interactions at the dentin/adhesive interface using FTIR chemical imaging," *J. Biomed. Opt.* **10**, 031104 (2005).
8. P. Tramini, B. Bonnet, R. Sabatier, and L. Maury, "A method of age estimation using Raman microspectrometry imaging of the human dentin," *Forensic Sci. Int.* **118**, 1–9 (2001).
9. Y. Wang and P. Spencer, "Analysis of acid-treated dentin smear debris and smear layers using confocal Raman microspectroscopy," *J. Biomed. Mater. Res.* **60**, 300–308 (2002).
10. Y. Wang and P. Spencer, "Hybridization efficiency of the adhesive dentin interface with wet bonding," *J. Dent. Res.* **82**, 141–145 (2003).
11. Y. Wang, P. Spencer, and X. M. Yao, "Micro-Raman imaging analysis of monomer/mineral distribution in intertubular region of adhesive/dentin interfaces," *J. Biomed. Opt.* **11**, 024005 (2006).
12. P. Spencer, Y. Wang, M. P. Walker, D. M. Wieliczka, and J. R. Swafford, "Interfacial chemistry of the dentin/adhesive bond," *J. Dent. Res.* **79**, 1458–1463 (2000).
13. P. Geladi and H. Grahn, *Multivariate Image Analysis*, John Wiley and Sons, New York (1996).
14. M. Otto, *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, John Wiley and Sons, New York (1999).
15. C. A. Hayden and M. D. Morris, "Effects of sampling parameters on principal components analysis of Raman line images," *Appl. Spectrosc.* **50**, 708–714 (1996).
16. J. R. Mansfield, M. G. Sowa, J. R. Payette, B. Abdulrauf, M. F. Stranc, and H. H. Mantsch, "Tissue viability by multispectral near infrared imaging: a fuzzy c-means clustering analysis," *IEEE Trans. Med. Imaging* **17**, 1011–1018 (1998).
17. P. Lasch, W. Haensch, D. Naumann, and M. Diem, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis," *Biochim. Biophys. Acta* **1688**, 176–186 (2004).
18. K. E. Shafer-Peltier, A. S. Haka, J. T. Motz, M. Fitzmaurice, R. R. Dasari, and M. S. Feld, "Model-based biological Raman spectral imaging," *J. Cell. Biochem.* **Suppl. 39**, 125–137 (2002).
19. C. Krafft, K. Thummler, S. B. Sobottka, G. Schackert, and R. Salzer, "Classification of malignant gliomas by infrared spectroscopy and linear discriminant analysis," *Biopolymers* **82**, 301–305 (2006).
20. S. Albyrak and F. Amasyali, "Fuzzy c-means clustering on medical diagnostic systems," *J. Microbiol.* **144**, 1157–1170 (1998).
21. Y. Wang and P. Spencer, "Quantifying adhesive penetration in adhesive/dentin interface using confocal Raman microspectroscopy," *J. Biomed. Mater. Res.* **59**, 46–55 (2002).
22. J. Kanca, 3rd, "Improving bond strength through acid etching of dentin and bonding to wet dentin surfaces," *J. Am. Dent. Assoc.* **123**, 35–43 (1992).
23. Z. Ye and G. Auner, "Principal component analysis approach for biomedical sample identification," *IEEE Intl. Conf. Syst. Man Cyber.* **2**, 1348–1353 (2004).
24. E. R. Malinowski, *Factor Analysis in Chemistry*, John Wiley and Sons, New York (1991).
25. R. Goodacre, E. M. Timmins, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell, and P. J. Rooney, "Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks," *Microbiology* **144**, 1157–1170 (1998).
26. Y. P. Wang, Y. Wang, and P. Spencer, "Fuzzy clustering of Raman spectral imaging data with a wavelet-based noise-reduction approach," *Appl. Spectrosc.* **60**, 826–832 (2006).