Morphology-Based Automated Baseline Removal for Raman Spectra of Artistic Pigments

ROSANNA PEREZ-PUEYO,* MARIA JOSE SONEIRA, and SERGIO RUIZ-MORENO

Signal Theory and Communications Department, ETSETB, Universitat Politècnica de Catalunya, C/ Sor Eulalia de Anzizu s/n, D5, Campus Nord, 08034 Barcelona, Spain

The interpretation of a Raman spectrum is based on the identification of its characteristic molecular bands. However, the assignment of the vibrational modes is often compromised by the presence in the spectrum of an intense fluorescence background that covers the measured spectra. Several techniques have been employed to minimize the presence of this fluorescence in order to resolve and analyze Raman spectra. In this paper a new automated method for fluorescence subtraction is described, based on morphology operations. This method is compared with the most commonly used polynomial fitting methods. Results indicate that the proposed automated method is efficient in fluorescence subtraction and retains the line shapes and positions of the Raman bands in the spectra.

Index Headings: Pigment identification; Art conservation; Raman spectroscopy; Fluorescence; Mathematical morphology; Background subtraction; Spectral processing.

INTRODUCTION

Raman spectroscopy is an extremely valuable tool for research in a wide range of fields, and numerous publications in recent years have attested to its popularity, success, and range of applications.[1–3] This technique has proven capable of identifying pigments,[4–9] mainly because it enables *in situ*, nondestructive analysis directly on the unprepared sample or surface of the work of art and provides identification. However, the efficiency of the Raman effect is low because only a small number of incident photons are inelastically diffused. Therefore, the detection of the Raman signal can be limited by the different noise sources commonly encountered in Raman spectroscopy: shot noise, instrument noise, and background noise.[10]

The most common source of background light encountered in Raman spectroscopy is the fluorescence that can be intrinsic to a pictorial sample due to the pigment itself and/or the binding media used in the paint layer and/or the varnishes used to enhance the work of art. This fluorescence background is usually some order of magnitude higher than the Raman signal intensity and it can mask the Raman bands, compromising the final result of the analysis. Therefore, the presence of fluorescence in a Raman spectrum must be reduced to facilitate the effective observation of the Raman signal in the measured spectrum. This is particularly important in the application of Raman spectroscopy to the identification of pigments in a work of art.

There is no single strategy for reducing fluorescence background in Raman spectra. Several instrumental and mathematical methods have been proposed to eliminate fluorescence baseline to enhance the Raman signal.[11–19] The non-instrumental methods (mathematical treatments) are usually employed because they do not require the upgrading of existing instrumentation. One simple and effective method consists of fitting the background using a polynomial function and then subtracting it from the spectrum (baseline subtraction). In this technique, the user manually optimizes the subtraction of fluorescence. Thus, the procedure is neither automatic nor systematic. It can take a long time since the user must treat each spectrum individually to find the best approximation. To overcome these limitations, an automatic method for the elimination of the fluorescence background is proposed. The method is based on mathematical morphology techniques, which are commonly used in image analysis and processing to simplify the images and preserve the main characteristics of the objects' shapes. In a Raman spectrum, the bands are the main features (objects) to consider in order to identify materials. Therefore, the proposed method fits best the baseline of the spectrum not considering the Raman peaks, and subtracting only the intense background without manual intervention. The Raman spectrum is now prepared for subsequent signal processing to identify the analyzed material.

The aim of this paper is to investigate the power of mathematical morphology to remove the background of Raman spectra without modifying the Raman bands. For this purpose, the paper is organized as follows. First of all, a brief introduction describing the principles of mathematical morphology, the rationale of the proposed method for the baseline removal, and the experimental system for measuring the spectra is given. Second, a discussion concerning the results of applying this method to Raman spectra of pigments is provided. Furthermore, a comparison of distortions introduced by the proposed method with the polynomial approximation is made.

THEORY

Morphology-Based Operations. Mathematical morphology is a nonlinear signal processing technique based on classical set theory.[20] It can be used in digital image processing in which the components of an image are considered as sets, subsets, or parts of a space. The first studies were undertaken by Minkowski,[21] continued by Hadwiger,[22] and reformulated by Matheron[23] and Serra.[24] Mathematical morphology is currently used in different research fields because it requires only to work with sets of data and their properties.[25–27] It is particularly useful in fields in which shape is the most important characteristic, so it can be applied to the processing of one-dimensional signals such as spectra by considering the functions as sets that represent the signals.

All morphological operations are the result of one or more operations (union, intersection, etc.) between two sets X and Y, which both pertain to a Z space. One of the two sets, for instance Y, is named the structuring element. To operate with the other set, X, the structuring element is shifted through the Z space, transforming the original X into another set. The aim of
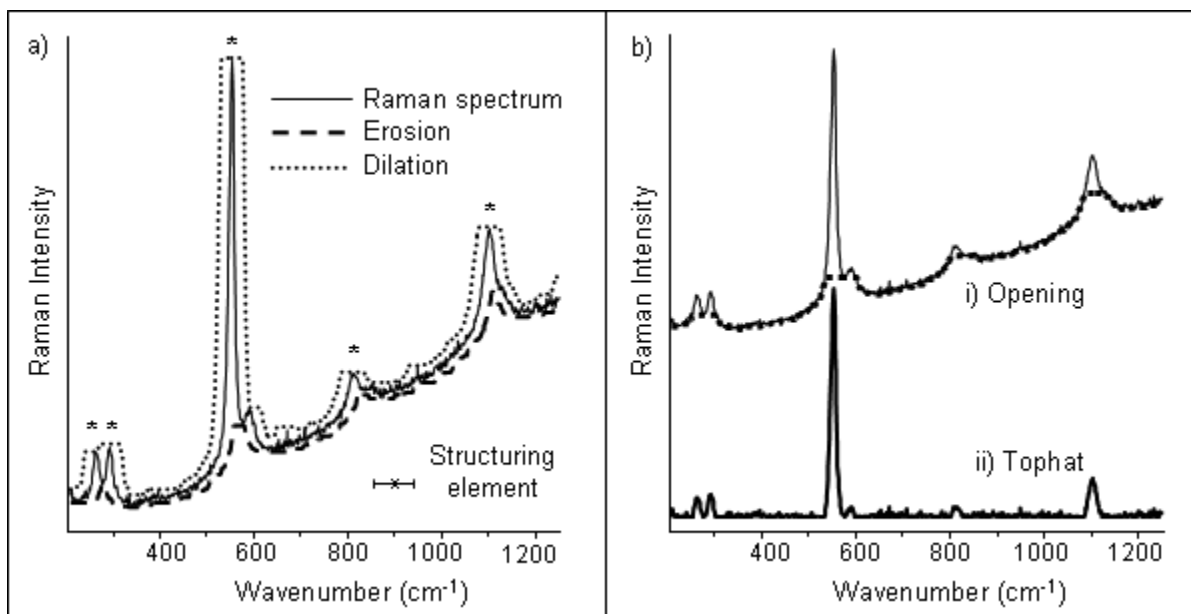
FIG. 1. (**a**) Erosion and dilation and (**b**) opening (spectrum i, dashed line) and top-hat (spectrum ii) of the spectrum corresponding to an ultramarine blue pigment with representative peaks located at 258 cm^{-1}, 286 cm^{-1}, 548 cm^{-1}, 820 cm^{-1}, and 1096 cm^{-1}.

this transformation is to search for geometric structures in the original X using the structuring element, Y, whose shape is chosen *a priori* according to the "morphology" of the set to be transformed and the special structures (forms) to be extracted. The two most basic operations in mathematical morphology are erosion and dilation. These operators are the basis of any other morphological transformation.

To adapt the above operator definitions to one-dimensional signal processing, the erosion, $\varepsilon_Y(f)$, of a one-dimensional function $f(x)$ by a structuring element Y is defined as the minimum value of the function in the window, as determined by the structuring element, centered at x. The expression for the erosion of f by Y is described mathematically as:

$$\varepsilon_Y(f)(x) = \min_{s \in Y} f(x + s) \qquad (1)$$

The dilation, $\delta_Y(f)$, is the dual operator of the erosion. Thus, the dilation of a one-dimensional function f by a structuring element Y is defined as the maximum value of the function in the window determined by the structuring element, centered at x. The expression for the dilation of f by Y is described mathematically as:

$$\delta_Y(f)(x) = \max_{s \in Y} f(x + s) \qquad (2)$$

Figure 1a shows the erosion and dilation of a Raman spectrum (that is, a one-dimensional function) by a plane structuring element that is a window of 41 points centered at the processed point. This spectrum was measured from an oil painting, corresponding to the ultramarine blue pigment ($[(Na_{10}Al_6Si_6O_4S_2)_x]$), and was excited using red radiation from a HeNe laser. The erosion of the function results in a narrowing of the peaks. Peaks that are narrower than the structuring element disappear completely and the minima of the function expand. Dilation of the function leads to the opposite effects.

Generally, dilation and erosion are transformations with no inverse, so, there is no way to recover the original signal from its erosion or dilation. However, it is possible to approximate the original signal by using a combination of these two basic operators. In other words, dilation can be used with the same structuring element as erosion to try to retrieve the original set. This combination of operators—erosion followed by dilation with the same structuring element—is another morphological operator called the opening operator. The same happens when a dilation is followed by an erosion. This composition is called the closing operator. Opening and closing are the basic operations of morphological noise removal. Opening removes small features, while closing removes small holes. These techniques can also be used to find specific shapes in an image. Opening can be used to find things into which a specific structuring element can fit (edges, corners, etc.). The opening of a function f by a structuring element Y is obtained by the erosion of f by Y, followed by the dilation of the resulting function by Y, and is described mathematically as

$$\gamma_Y(f) = \delta_Y[\varepsilon_Y(f)] \qquad (3)$$

Opening smoothes the function nonlinearly. It removes the positive peaks that are narrower than the structuring element. Therefore, the opening always takes values lower than the original function. Figure 1b (spectrum i, dashed line) shows the opening of the signal that is represented in Fig. 1a by a structuring element of 41 points (44.4 cm^{-1}).

Another morphological operation is the top-hat transformation. The top-hat by opening is the remainder of the original signal and its opening, that is,

$$\rho(f) = f - \gamma(f) \qquad (4)$$

This operation allows one to determine which structures have been eliminated by the opening. The top-hat by opening (now top-hat) of the signal represented in Fig. 1a is shown in Fig. 1b (spectrum ii). It must be pointed out that the size of the plane structuring element has been chosen to show how the top-hat operator works. Taking into account this result, it is
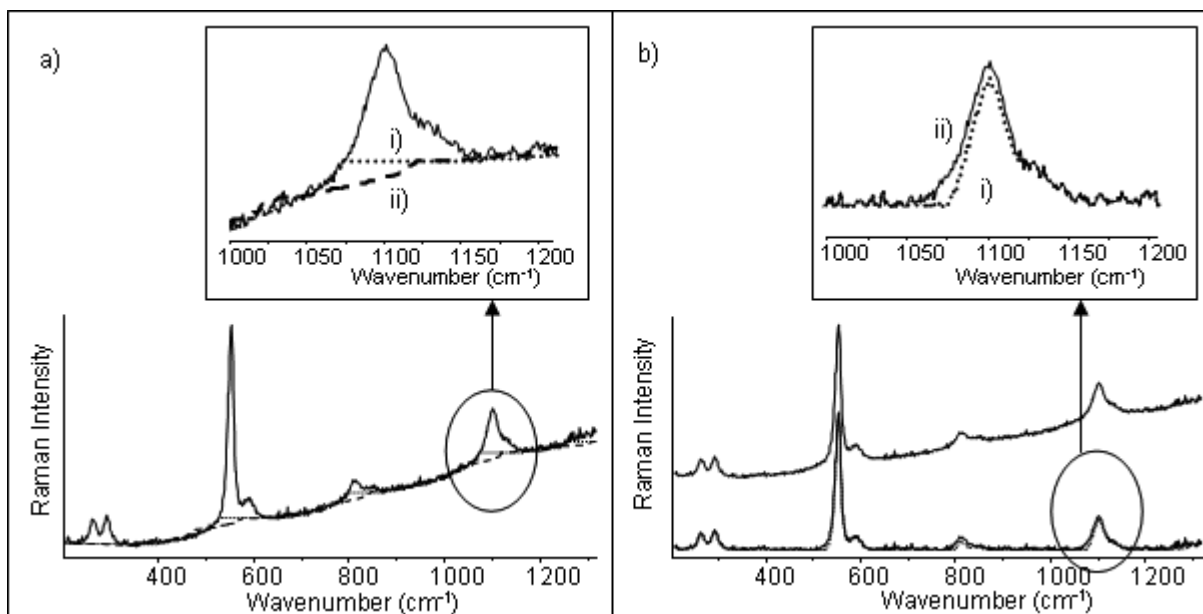
FIG. 2. (a) Opening (spectrum i) and new approximation (spectrum ii) of the spectrum in Fig. 1 and (b) spectrum filtered by the top-hat (spectrum i) and by the proposed method (spectrum ii), with the insets showing an enlargement of the Raman band at 1096 cm^{-1}.

reasonable to suppose that, with an optimum size of the structuring element, the opening could approximate the baseline. The top-hat could be used to obtain a baseline-free spectrum. On the basis of this idea, a morphological filter is proposed in this paper to eliminate the fluorescence baseline in Raman spectra.

Proposed Baseline Removal Method. The purpose of this work was the use of mathematical morphology for the automatic removal of the baseline of the Raman spectra of works of art. The main objective was to eliminate the background of Raman spectra, which is mainly caused by fluorescence in the analyzed sample. Hence, fluorescence-free spectra could be obtained without modifying the shape or position of the Raman bands in order to identify the material (i.e., pigments).

As mentioned in the previous section, an opening is a morphological transformation that removes peaks of width equal to or smaller than the structuring element size and always takes values that are lower than those of the signal. With a suitable structuring element, the opening of a Raman spectrum will remove all Raman bands narrower than the structuring element size. The result obtained upon applying the opening operation can be approximated to the baseline of the Raman spectrum and the top-hat transformation is the background-free final spectrum. As the Raman bands in a spectrum can have different widths that are not known *a priori*, the first step is to optimize the size of the structuring element so as not to lose any band contributions. The size, in terms of the number of data points of the structuring element, must be at least equal to the number of data points of the widest Raman band of the spectrum. The following method for obtaining the optimum size of the structuring element does not involve peak recognition, i.e., the Raman bands do not have to be explicitly located in advance.

First, the opening of the spectrum is calculated using a three-point structuring element. Second, the size of the structuring element is increased by two points and the opening of the spectrum is recalculated, and so on. When three consecutive openings are equal, the method is stopped and the optimum size of the structuring element is the smallest of the three used to calculate these openings. In other words, the opening of the Raman spectrum is iteratively applied, starting from a minimum size of the structuring element (three points), until three consecutive openings are the same. This iterative procedure was performed using a home-made algorithm.

If this method is applied to the spectrum in Fig. 1, an optimum size of the structuring element of 89 points (97.7 cm^{-1}) is obtained. Nevertheless, the opening of the Raman spectrum with the optimized structuring element may show residual bands predominantly in those spectral regions that show non-resolved multiple bands. Consequently, the top-hat of the spectrum can result in a baseline-free spectrum with some slightly distorted bands in its shape. To overcome this problem, a new combination of erosion and dilation of the opening is proposed to obtain a better approximation of the spectrum baseline in these spectral ranges. This consists of the average of dilation and erosion of the opening of the spectrum. The expression of this approximation is described mathematically as

$$\gamma'(f) = \frac{\delta[\gamma_Y(f)] + \varepsilon[\gamma_Y(f)]}{2} \qquad (5)$$

This new approximation of the baseline is closer to baseline than the opening in the regions where there are Raman bands, while the opening is closer to baseline than the new approximation in regions where there are no bands (see Fig. 2a). Therefore, we establish that the best approach to baseline will be the minimum value between the opening and the new approximation to the baseline at each point. That is,

$$\gamma_{opt}(f) = \min[\gamma'(f), \gamma(f)] \qquad (6)$$

The baseline is optimally removed by applying the top-hat transformation, using the proposed correction for the opening.

Thereby, starting from a Raman spectrum with baseline f, the spectrum with no baseline $\rho_{opt}(f)$ is obtained as:

$$\rho_{opt}(f) = f - \gamma_{opt}(f) \qquad (7)$$

Figure 2b shows the spectrum of Fig. 1 before and after filtration with top-hat (spectrum i) and with the proposed technique (approximated top-hat) (spectrum ii). As can be seen, the approximated top-hat avoids the slight distortion in the shape of the bands that is produced by top-hat. In both cases, the positions of the peaks remain unchanged after filtering.

As discussed previously, the optimum length of the structuring element is determined by the width of the widest band and/or overlapping bands of the spectrum. Thus, for a spectrum with wide and/or overlapping peaks, a structuring element of larger length will be obtained. If the baseline has a similar width, it could be confused with a band and therefore would not be efficiently eliminated. This is a weakness of the proposed method, although it is a very unlikely situation.

MEASUREMENT SYSTEM: EXPERIMENTAL SETUP

Raman spectra were collected by using a commercial device (model INDURAM) by Horiba Jobin Yvon provided with *Labspec* software. Three different laser excitation frequencies were employed for spectral acquisition: a HeNe laser (632 nm), an Ar laser (514 nm), and a near-infrared (NIR) diode laser (785 nm), each of which provided approximately 4 mW of laser power at the sample. The most suitable laser excitation frequency was selected in order to obtain the best signal-to-noise ratio in the spectrum from the sample. The light from the laser was guided by an optical fiber to the probe head and directed to the sample. The scattered light was collected and filtered by the same probe head. It was then guided by optical fiber to the monochromator and detected by a thermoelectrically cooled charge-coupled device (CCD) detector. An 1800 lines/mm dispersive grating was used for the measurements. All spectra were acquired in a spectral range suitable for measuring the most representative bands for each pigment. Acquisition times were on the order of 100 s. This Raman instrumentation with an optical fiber enabled direct measurements on the works of art.

RESULTS AND DISCUSSION

Analysis of Experimental Spectra. To evaluate the performance of the proposed method, the technique was applied to Raman spectra of artistic pigments from different works of art. Below, the results obtained with some experimental spectra, showing different shapes of baseline, are shown.

Figure 3a shows the Raman spectrum of cadmium yellow pigment (ZnS·CdS) before and after filtering. The size of the structuring element was 109 points (59.4 cm^{-1}). The spectrum was measured on an oil painting with a laser wavelength of 514 nm. The bands arise from a flat baseline and their positions remained unchanged after filtering. Another example of baseline correction is shown in Fig. 3b. The Raman spectrum corresponds to the vermillion pigment (HgS). The spectrum was measured on an oil painting using a laser wavelength of 632 nm. The size of the structuring element was 35 points (40.1 cm^{-1}). It can be seen that the baseline was completely removed

and the bands still remained in their positions. So, the filtered spectrum is ready for further processing if necessary.

The Raman spectrum of lead tin yellow pigment (Pb$_2$SnO$_4$) is shown in Fig. 3c. The spectrum was measured on an oil painting with a laser wavelength of 514 nm. The spectrum baseline was quite different from that of the previous spectra. The spectrum shows a sharp slope in the region of the lower frequencies probably due to the notch filter edge. For this reason, the filtered spectrum showed a little band that is not observable in the original spectrum. Nevertheless, the background has been completely removed with a structuring element of 23 points (15.1 cm^{-1}) and no shift in the band's position has been noticed after baseline subtraction.

Figure 3d shows the last example of baseline correction. The Raman spectrum was acquired from a sample of commercial standard red vermillion French substitute pigment manufactured by Sennelier (N°675, PR4, β-naphthol). This is an azo group pigment. Although it was directly measured on pure pigment using an excitation wavelength of 785 nm, this spectrum showed baseline itself because the pigment has intrinsic fluorescence. It was chosen because the baseline is completely different from the other examples. The proposed (filter) technique performed the baseline correction with a structuring element of 41 points (44 cm^{-1}) and without shifting the bands.

The figures enable a qualitative visual inspection of the distortion imposed by the given (filter) method on the baseline-removed spectra. Despite slight distortion, the bands are clearly visible. Furthermore, in a molecular identification, the information of interest is in the band position. The filter eliminated the fluorescence baseline without shifting this position. Thus, the spectra are ready for further numerical processing if necessary.

The demonstration of the proposed method on experimental spectra shows its effectiveness. The technique was then compared with the polynomial fitting method using the root mean square error to demonstrate its potential for complete automation.

Comparison with the Conventional Polynomial Approach. The polynomial approach is the most popular method in Raman spectroscopy for subtracting the baseline from a spectrum. This method consists of finding the polynomial that best fits the baseline. Many commercial software provide this application. A comparison between the polynomial approach and the proposed filter has been made. The degree of the polynomial was selected *a priori*, taking into account the shape of the baseline of the processed spectrum. When the collected spectra exhibit large variability in the baseline slope and shapes, it might be difficult to carry out an empirical parameter selection.

The Raman spectrum of ternary yellow pigment (lead-antimony-tin triple oxide pigment, Pb$_2$(Sn$_x$Sb$_{2-x}$)O$_{7-x/2}$), is shown in Fig. 4. This spectrum was measured from an oil painting using an excitation wavelength of 632 nm. The size of the structuring element was 41 points (46.8 cm^{-1}). The figure shows how the proposed method performs on the spectrum relative to the polynomial method used (fifth order). The result shows the effectiveness of the proposed method.

The root mean square error (RMSE) was used to compare the results of the proposed technique and the polynomial approach. To this end, the RMSE between the ideal Raman spectrum of a pure pigment sample (a commercial standard) and either of the fitting methods has been calculated. The ideal
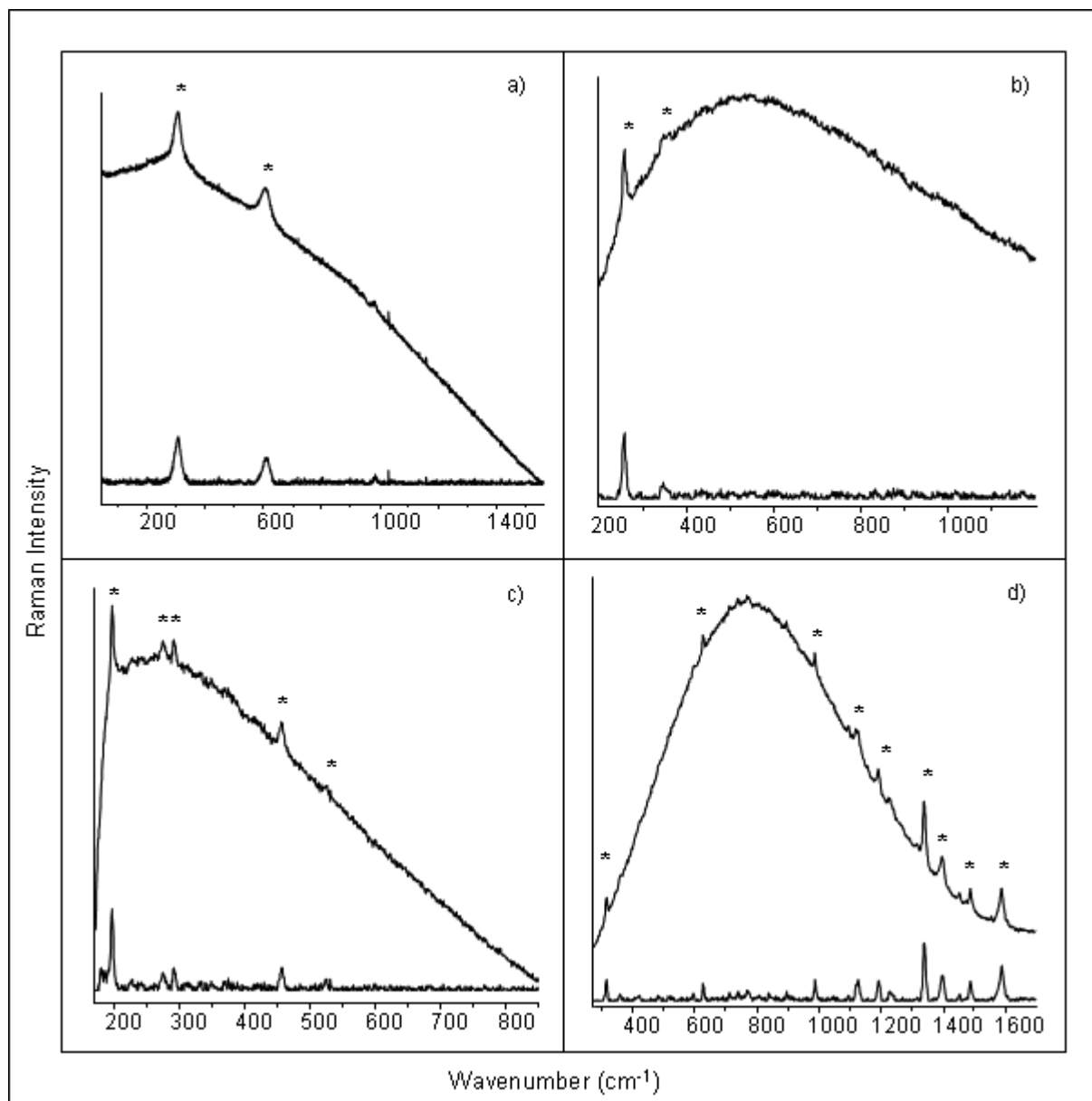
FIG. 3. Examples of Raman spectra before and after filtering. (**a**) Cadmium yellow pigment with representative peaks located at 304 cm^{-1} and 609 cm^{-1}, (**b**) vermillion pigment with representative peaks located at 253 cm^{-1} and 344 cm^{-1}, (**c**) lead tin yellow with representative peaks located at 196 cm^{-1}, 275–291 cm^{-1}, 457 cm^{-1}, and 525 cm^{-1}, and (**d**) red vermillion French substitute pigment with representative peaks located at 312 cm^{-1}, 625 cm^{-1}, 988 cm^{-1}, 1122 cm^{-1}, 1191 cm^{-1}, 1337 cm^{-1}, 1396 cm^{-1}, 1486 cm^{-1}, and 1588 cm^{-1}.

Raman spectrum was measured under optimum conditions. The values of RMSE obtained for the proposed morphologic filter were 0.0703 for the cadmium yellow pigment, 0.0967 for the vermillion pigment, 0.06 for the lead-tin yellow pigment, and 0.0395 for the ternary yellow pigment and for the polynomial approximation were 0.1290, 0.4748, 0.402, and 0.1041, respectively. These last values correspond to the best polynomial approximation for each spectrum, that is, the polynomial degree was five for the spectra of cadmium yellow and ternary yellow pigments and nine for the spectra of vermillion and lead-tin yellow pigments. As can be seen, the filter gave better RMSE values than the polynomial approximation. Furthermore, the polynomial approximation requires user intervention and involves critical subjectivity, while the proposed filter only works with the measured spectrum,

regardless of the user. The baseline has a strong influence on the polynomial approximation as the polynomial degree must be selected according to the shape of the baseline to subtract, while the proposed technique works independently of the baseline.

CONCLUSION

We have proposed and experimentally demonstrated a novel, simple scheme for fluorescence rejection. The basic principle is to extract the bands from the baseline. The morphological filter that we designed automatically removes the fluorescence baseline and avoids user intervention. From a practical perspective, the proposed method produces results that are within the general range of manual methods and has the potential for complete automation. The specific reasons for
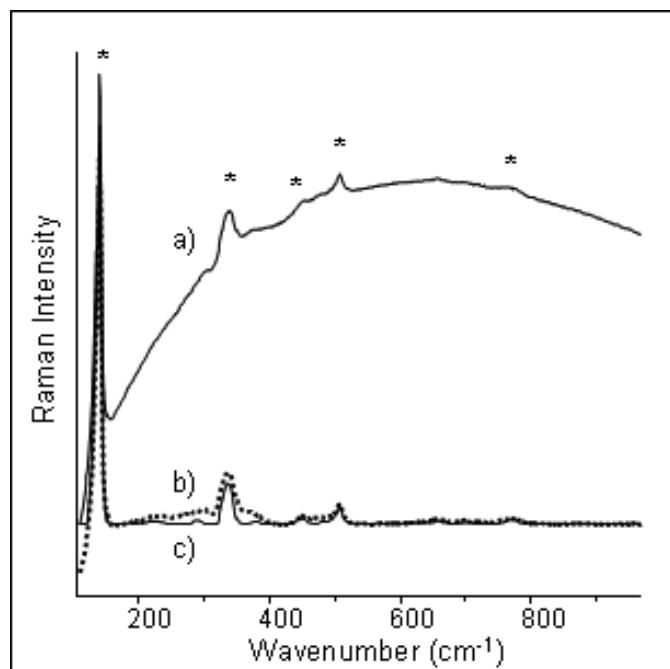
<small>Fig. 4. Raman spectrum of the ternary yellow pigment with representative peaks located at 140 cm^{-1}, 336 cm^{-1}, 452 cm^{-1}, 509 cm^{-1}, and 770 cm^{-1}, (**a**) before filtering, (**b**) filtered with the fifth-order polynomial method, and (**c**) filtered with the proposed method.</small>

quality of the original Raman signals. Moreover, although the proposed method has been proved with Raman spectra from pigments used in works of art, it can be used to remove the baseline from spectra obtained from other materials.

eliminating the baseline depend on the application. These reasons may range from relatively simple requirements, such as presentation, to more demanding requirements, such as quantification or preparation for further numerical processing. Baseline correction may be just an intermediate step within a large scheme for a complete automatic identification so that automation of this step can be very important. The proposed technique is a "self-contained" method, which avoids user intervention.

The proposed method does not require knowledge of the baseline shape, so it can be used to automate fluorescence subtraction from measured spectra. The algorithm is computationally efficient, conceptually simple, and easy to implement. It is an alternative to other methods that use complicated mathematical data treatment and take advantage of the spatial characteristics of spectrum data. This method does not need automatic peak recognition. As the peaks are extracted, they do not have to be explicitly located in advance. However, it must be pointed out that the Raman bands should be distinguishable in the spectral range, also with strong fluorescence, so that the proposed method works efficiently. This condition is necessary for any mathematical method. Of course, the quality of the reconstructed fluorescence-free spectrum is linked to the

1. P. Matousek, E. R. C. Draper, A. E. Goodship, I. P. Clark, K. L. Ronayne, and A. W. Parker, Appl. Spectrosc. **60**, 758 (2006).
2. M. F. Escoriza, J. M. VanBriesen, S. Stewart, and J. Maier, Appl. Spectrosc. **61**, 812 (2007).
3. Y. Liu, K. Chao, M. S. Kin, D. Tuschel, O. Olkhovyk, and R. J. Priore, Appl. Spectrosc. **63**, 477 (2009).
4. F. Schulte, K.-W. Brzezinka, K. Lutzenberger, H. Stege, and U. Panne, J. Raman Spectrosc. **39**, 1445 (2008).
5. L. Burgio, R. J. H. Clark, V. S. F. Muralha, and T. Stanley, J. Raman Spectrosc. **39**, 1482 (2008).
6. D. de Waal, J. Raman Spectrosc. **38**, 956 (2007).
7. S. E. Jorge Villar, H. G. M. Edwards, J. Medina, and F. Rull Perez, J. Raman Spectrosc. **37**, 974 (2006).
8. R. Pérez-Pueyo, M. J. Soneira, and S. Ruiz-Moreno, J. Raman Spectrosc. **35**, 808 (2004).
9. M. Castanys, R. Pérez-Pueyo, M. J. Soneira, and S. Ruiz-Moreno, J. Raman Spectrosc. **37**, 1003 (2006).
10. M. J. Pelletier, *Analytical Applications of Raman Spectroscopy* (Blackwell Science, New York, 1999).
11. I. Osticioli, A. Zoppi, and E. M. Castellucci, J. Raman Spectrosc. **37**, 1078 (2006).
12. Y. Oshima, Y. Komachi, Ch. Furihata, H. Tashiro, and H. Sato, Appl. Spectrosc. **60**, 964 (2006).
13. I. Osticioli, A. Zoppi, and E. M. Castellucci, Appl. Spectrosc. **61**, 839 (2007).
14. Ch. A. Lieber and A. Mahadevan-Jansen, Appl. Spectrosc. **57**, 1363 (2003).
15. A. Cao, A. K. Pandya, G. K. Serhatkulu, V. M. Naik, R. Naik, G. W. Auner, R. Rabah, and D. C. Freeman, J. Raman Spectrosc. **38**, 1199 (2007).
16. Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, and J. Hu, Chemom. Intell. Lab. Syst. **85**, 94 (2007).
17. G. Schulze, A. Jirasek, M. M. L. Yu, A. Lim, R. F. B. Turner, and M. W. Blades, Appl. Spectrosc. **59**, 545 (2005).
18. N. N. Brandt, O. O. Brovko, A. Y. Chikishev, and O. D. Paraschuk, Appl. Spectrosc. **60**, 288 (2006).
19. F. Rosi, M. Paolantoni, C. Clementi, B. Doherty, C. Miliani, B. G. Brunetti, and A. Sgamellotti, J. Raman Spectrosc., DOI 10.1002/jrs.2447.
20. A. A. Fraenkel, Y. Bar-Hillel, and A. Levy, *Foundations of Set Theory,* A. Heyting, H. J. Keisler, A. Mostowsky, A. Robinson, and P. Suppes, Eds. (Elsevier Science Publishers B.V., New York, 1984), 2nd rev.
21. H. Minkowski, Math. Ann. **57**, 447 (1903).
22. H. Hadwiger, *Vorslesunger über Inhalt, Oberfläche und isoperimetrie* (Springer Verlag, New York, 1957).
23. G. Matheron, *Random Sets and Integral Geometry* (Wiley, New York, 1975).
24. J. Serra, *Image Analysis and Mathematical Morphology. Vol 2: Theoretical Advances* (Academic Press, New York, 1988).
25. Z. Yu-qian, G. Wei-hua, C. Zhen-cheng, T. Jing-tian, and L. Ling-yun, *Proceedings of 27th Annual International Conference of the Engineering in Medicine and Biology Society,* 6492 (2005).
26. D. Ze-Feng, Y. Zhou-Ping, and X. You-Lun, IEEE Signal Process. Lett. **14**, 31 (2007).
27. C. Hao-Teng, L. Chih-Hong, and P. Tun-Wen, J. Mol. Recognit. **21**, 431 (2008).