Article

Applied Spectroscopy
2018, Vol. 72(12) 1752–1763

DOI: 10.1177/0003702818789695
journals.sagepub.com/home/asp
SAGE

A Novel Pre-Processing Algorithm Based on the Wavelet Transform for Raman Spectrum

Yang Xi, Yuee Li, Zhizhen Duan, and Yang Lu

Abstract

Noise and fluorescent background are two major problems for acquiring Raman spectra from samples, which blur Raman spectra and make Raman detection or imaging difficult. In this paper, a novel algorithm based on wavelet transform that contains denoising and baseline correction is presented to automatically extract Raman signals. For the denoising section, the improved conventional-scale correlation denoising method is proposed. The baseline correction section, which is performed after denoising, basically consists of five aspects: (1) detection of the peak position; (2) approximate second derivative calculation based on continuous wavelet transform is performed using the Haar wavelet function to find peaks and background areas; (3) the threshold is estimated from the peak intensive area for identification of peaks; (4) correction of endpoints, spectral peaks, and peak position; and (5) determine the endpoints of the peak after subtracting the background. We tested this algorithm for simulated and experimental Raman spectra, and a satisfactory denoising effect and a good capability to correct background are observed. It is noteworthy that this algorithm requires few human interventions, which enables automatic denoising and background removal.

Keywords

Raman spectroscopy, wavelet transform, denoising, baseline correction, derivative spectra

Date received: 29 March 2018; accepted: 25 June 2018

Introduction

In recent years, Raman spectroscopy has been applied to many different scientific branches, including biology, chemistry, and materials science, due to its capability to provide information about the physical and chemical characteristics of materials.[1,2] As a nondestructive technique, Raman spectroscopy has also attracted much attention in biological imaging wherein different biological components could be distinguished due to the differences in the resonance nature. Notably, the relative intensities of peaks in Raman spectra are related to the molecular concentration, which makes Raman spectroscopy an attractive tool for mapping biological objects.[3]

The Raman spectrum generally has a low signal-to-noise ratio (SNR) and high magnitude of background, which arises mainly from laboratory equipment, environment, and autofluorescence. In fact, while acquiring the Raman spectrum, the cosmic rays or the alpha rays and gamma rays emitted from the environment also illuminate the detector and form a series of very sharp peaks, usually no more than 2 cm^{-1} in width, called the spike.[4–7] The spikes should be removed before denoising and removing the background. In this paper, we use the previous method to remove the spikes.[4]

Noise suppression is an important operation in data processing, but today there is no standard denoising strategy. With the development of the wavelet theory, the wavelet transform has been increasingly applied and even become a hot topic, as one of the main methods of signal denoising.[8,9] Wavelet denoising can be categorized into three major groups based on the denoising principle. The first group is the denoising method based on the wavelet modulus maxima proposed by Mallat and Hwang.[10] This method has heavy computation and contains uncertainty with the use of modulus maxima to reconstruct the signal. The second group is the wavelet threshold shrinkage denoising method originally proposed by Donoho et al.[11–13] The main idea is to reconstruct the signal by introducing a threshold

School of Information Science & Engineering, Lanzhou University, China

Corresponding author:
Yuee Li, Tianshui South Road, 222 Lanzhou, 730000 China.
Email: liyuee@lzu.edu.cn

to process coefficients. The calculation of this method is simple, but the denoising effect mainly depends on threshold determination. There are two well-known threshold methods, named the hard threshold method and the soft threshold method. The hard threshold function is prone to fluctuation and the pseudo-Gibbs effect in the process of reconstructing the signal, which can keep the details of the original signal and will make the signal distorted while. In addition, the soft threshold function is born with constant error, which means that the reconstructed signal is excessively smooth and there is a constant error between the processed and original signals. The third group is the scale correlation denoising method in the wavelet domain proposed by Xu et al.[14] This method determines the type of wavelet coefficients (signal of interest or noise) by calculating the wavelet coefficient correlation of adjacent scales, and then chooses coefficients to reconstruct the signal; this method requires moderate computation. However, coefficient processing with only a simple reservation will lead to a serious pseudo-Gibbs effect that is similar to that of the hard threshold method.

We propose an improved algorithm based on the denoising method of the third group. The soft threshold method is used to achieve adaptive shrinkage of wavelet coefficients corresponding to the signal of interest. Compared with the conventional-scale correlation denoising method and the soft threshold method, the denoising effect is significantly improved.

In addition, many of spectra possess problems of a varying baseline drift, which distorts the peak height and complicates automatic analysis. Therefore, it is necessary to develop an intelligent and robust background correction method for online application. Three popular methods, namely polynomial fitting,[15–17] derivatives[18,19] and wavelet transformation,[20–22] were applied for baseline correction. However, each of them had their shortcomings: the manual polynomial fitting was not so effective and it overly depends on the user's experience.[16] The performance of the automatic polynomial fitting was poor in a low SNR environment.[17] The traditional derivatives were easily affected by local loud noise, which is incorrectly identified as a spectral signal.[19] The wavelet transform method divides the signal into two parts: approximation and detail. The approximation is roughly regarded as a background and set to zero for reconstructing the signal.[20] In this way, it might lose some useful information and cause distortion of the spectra; meanwhile, it is difficult to choose the scale of decomposition.

In our algorithm, we draw on the advantages of the above three baseline correction methods and propose a novel algorithm. Meanwhile, we apply this algorithm to the simulated spectra with different background curves, noise standard deviations, and signal intensities. Finally, by testing the real Raman spectra, we demonstrate the practicability of our algorithm for the application of most Raman spectra.

Method

Mathematical Model of the Raman Spectrum

The peak of the Raman spectrum is Lorentz linear in nature. However, due to the influence of the instrument and the sample characteristics, the measured Raman spectrum is usually the convolution of the Lorentz function and the Gaussian function, that is, the Voigt function.[23] Since a Voigt profile has no analytical form, for analysis, the peak is modeled as a Gaussian function (Eq. 1).

$$f(x) = a \cdot e^{\left(\frac{-0.5(x-c)^2}{\sigma^2}\right)} \tag{1}$$

where a is the intensity controller and c and σ represent the position and standard deviation of the Gaussian function: the parameters are shown in Table I. We use four types of mathematical functions to simulate the background: sigmoidal, exponential, Gaussian, and fifth polynomial. The simulated Raman spectra are shown in Fig. 1.

Basics

The stationary wavelet transform is on the basis of the orthogonal wavelet, and it is a redundant wavelet transform with shift invariance. Down-sampling is not needed for each decomposition as the transformed high- and low-frequency components will have the same size as the input signal.[24]

In general, a wavelets function that resembles the signal or attributes can yield better signal and noise separation as well as sparsity, and we found that the Coiflet wavelet family is better suited for Raman spectra compared to other wavelet families. In the denoising section, we use the coif3 wavelet to perform stationary wavelet transform and the first four detail scales were selected for processing.

For a spectrum peak (with noise and background), the peak detection limit (PDL) is defined as $\frac{h_{peak}}{\sigma}$, where h_{peak}

Table I. Gaussian function parameters used for the simulated Raman spectra.

c	σ	a
350	10	0.7
550	12	1.4
800	10	0.5
840	8	0.9
880	9	0.7
1200	14	0.6
1500	10	0.4
1550	12	0.3
1800	10	0.5
2000	8	0.3

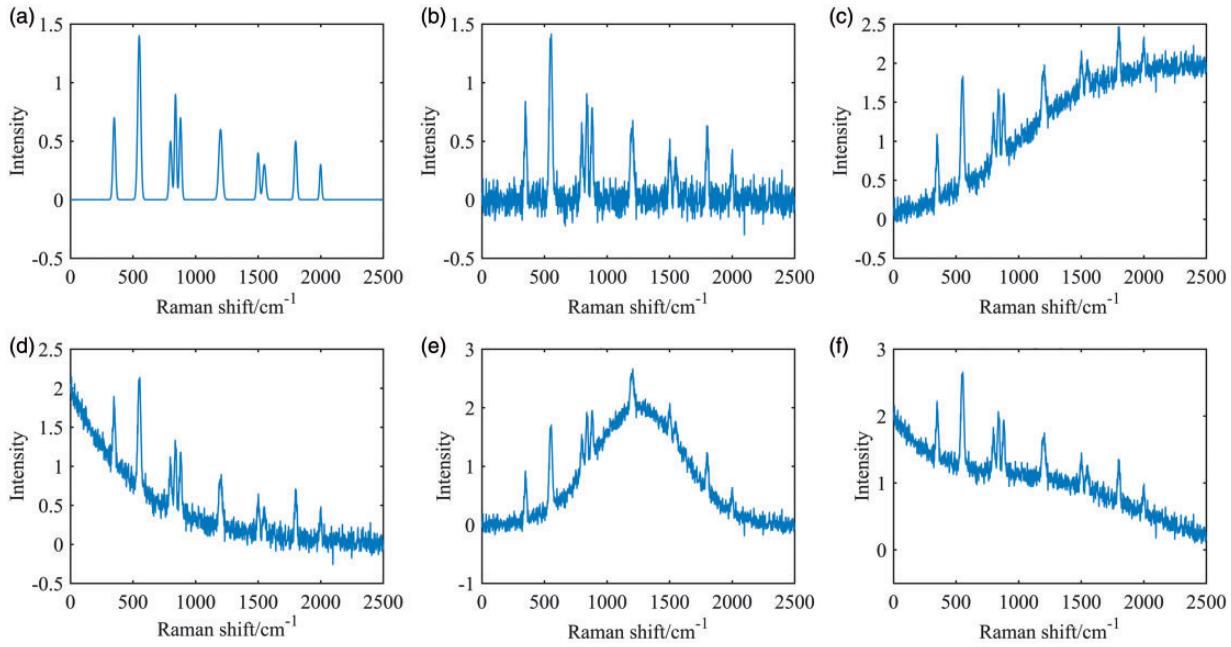

Figure 1. (a) Simulated spectrum without noise and background. (b) Spectrum after adding white noise. (c)–(f) Spectrum after adding sigmoidal, exponential, Gaussian, and fifth polynomial background, respectively.

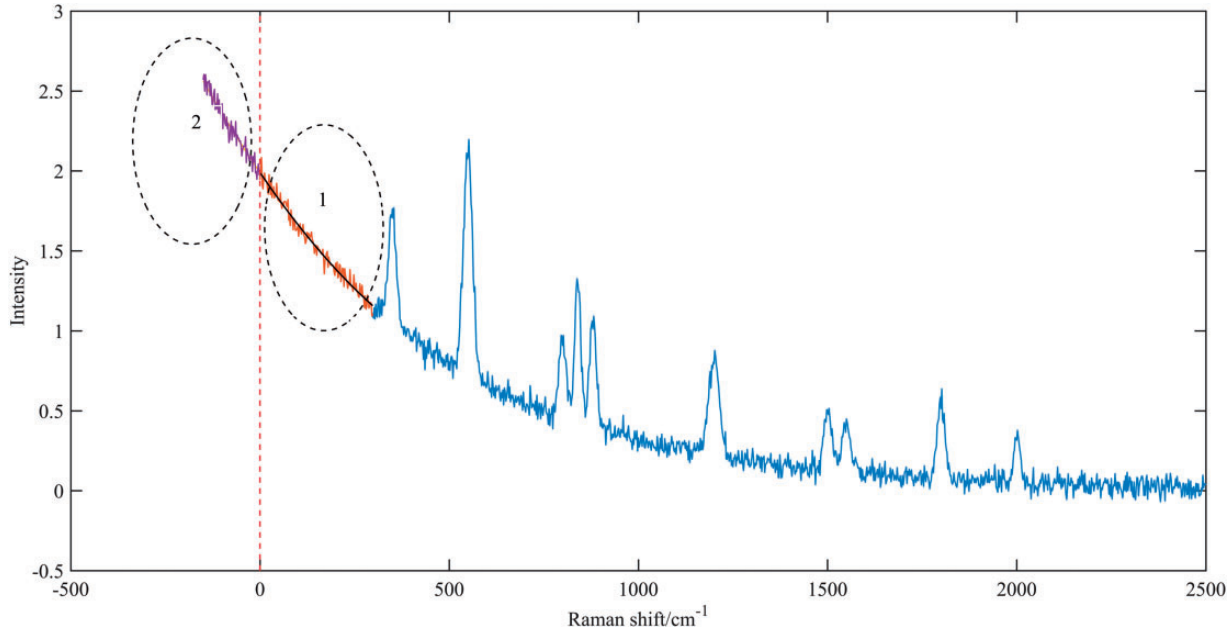

Figure 2. The simulated Raman spectrum after extension of the left end.

represents the height of the pure peak (without noise and background) and σ represents the noise standard deviation of the peak. It is generally believed that the Raman peak is not detectable while the PDL is lower than 3; here, we choose the PDL as low as this critical value for testing the efficiency of our algorithm. It should be noticed that all Raman spectra need to be extended before the stationary wavelet transform and the continuous wavelet transform are applied to avoid the boundary effect that will affect the accuracy of the experiment. Notably, a gentle extension is preferred because a steep one may be misidentified as a Raman peak. As shown in Fig. 2, the simulated spectrum with a wavenumber range of 0–2500 was extended. Firstly, the second-order polynomial fitting is

performed at the dot-line circled area 1. Then, based on the above fitting results, the cubic spline interpolation is used to extrapolate and the estimated noise is added, as shown in dot-line circled area 2. Here, the noise standard deviation of the high-pass filtered signal is used for estimating the noise of the original signal (see Eq. 2). This is because the high-frequency component is dominated by the noise and the interference of the signal is small, and the noise standard deviation obtained in this method is more accurate.

$$\sigma = \frac{\sigma'}{\|g\|} \tag{2}$$

where σ represents the noise standard deviation of the original signal, σ' represents the noise standard deviation after high-pass filtering, and $\|g\|$ represents the modulus of the response vector of the high-pass filter and is easy to obtain for a specific wavelet function. Here, we adopt the wavelet to estimate σ, since the wavelet belongs to the high-pass filter.

An Improved Denoising Algorithm

In the conventional-scale correlation denoising method, the normalized correlation of wavelet coefficients is calculated using the following formula:

$$NCor_j(i) = Cor_j(i) \cdot \sqrt{\frac{\sum_{n=1}^{N} |W_j(i)|^2}{\sum_{n=1}^{N} |Cor_j(i)|^2}} \tag{3}$$

where $Cor_j(i) = W_j(i) \cdot W_{j+1}(i)$, $W_j(i)$ represents the ith wavelet coefficient on the j-scale and N indicates the length of the j-scale. This method actually selects the wavelet coefficients one by one using the hard threshold method. Although the detailed information of the signal can be preserved, a serious pseudo-Gibbs effect may occur, which makes the following analysis difficult. Therefore, it is necessary to propose a way to improve it. Combining the advantages of conventional-scale correlation denoising method and soft threshold method, we propose an improved algorithm, where each wavelet coefficient owns an adaptive threshold and the soft threshold method is used to process each coefficient.

We define $\left|\frac{NCor_j(i)}{W_j(i)}\right|$ as the degree of signal (DOS). According to the scale correlation denoising theory, with the increase of the possibility of the wavelet coefficient corresponding to the signal of interest, the correlation becomes stronger and the DOS becomes larger. We hope that the wavelet coefficient corresponding to the signal of interest has a small threshold to keep the edge information as much as possible, while the wavelet coefficient corresponding to the noise has a relatively large threshold to remove the noise. The universal threshold is widely used as a threshold selection method (Eq. 4).[13]

$$thr_j = \sigma_j \cdot \sqrt{2\ln(N)} \tag{4}$$

where σ_j is noise standard variance of the j-scale and N is the length of the signal. We substitute the DOS into the following formula, and the threshold corresponding to each wavelet coefficient is obtained (Eq. 5).

$$thr_j(i) = thr_j \cdot \left(\frac{1}{DOS}\right)^t = thr_j \cdot \left|\frac{W_j(i)}{NCor_j(i)}\right|^t \tag{5}$$

where $thr_j(i)$ is the threshold corresponding to the ith wavelet coefficient on the j-scale and t is a modulation factor. For the coefficients corresponding to the signal of interest, with the increase of t, $\frac{1}{DOS}$ becomes smaller, the pseudo-Gibbs effect is more likely to occur and, with the decreases of t, $\frac{1}{DOS}$ becomes larger and the signal will be excessively smooth. We found that the denoising effect is optimal when t is selected as 3; the selection of t is discussed in the Supplemental Material. Here we combine the universal threshold with the DOS so that the wavelet coefficients corresponding to the signal of interest own different thresholds and the modulation factor t selection is supported by a large number of experiments and theories.

When the correlation operation is performed, the correlation of wavelet coefficients would be affected because the position of the wavelet coefficients on adjacent scales may not correspond exactly. Therefore, instead of taking the value of a specific point, we take the average over a region and the width of the region is conservatively chosen as 3.

Acquisition of Noise Standard Deviation, σ_j

The original signal is decomposed by wavelets and the fourth detail scale was selected for analysis because the wavelet coefficient corresponding to the noise is small enough and the wavelet coefficient corresponding to the signal of interest is relatively large on this scale. Here, a window with fixed width is selected to find an area in the fourth detail scale, while ensuring the sum of the absolute value of the wavelet coefficients in this area is the smallest and recording this area, then the above steps are repeated to find the second and third areas, while ensuring the three areas do not intersect with each other. Then we merge these three areas and map them to the first three detail scales, and the noise standard deviation of each scale is approximately obtained by the sample estimation method. As for the fourth detail scale, we use med[abs(w_4)]/0.6745 (med means taking the median) to estimate the noise standard deviation, where w_4 represents wavelet coefficients of the fourth detail scale.[13] The width of the window selected here is 60 and the width of the areas ($60 \times 3 = 180$) selected is about one-sixth of the simulated spectra length.

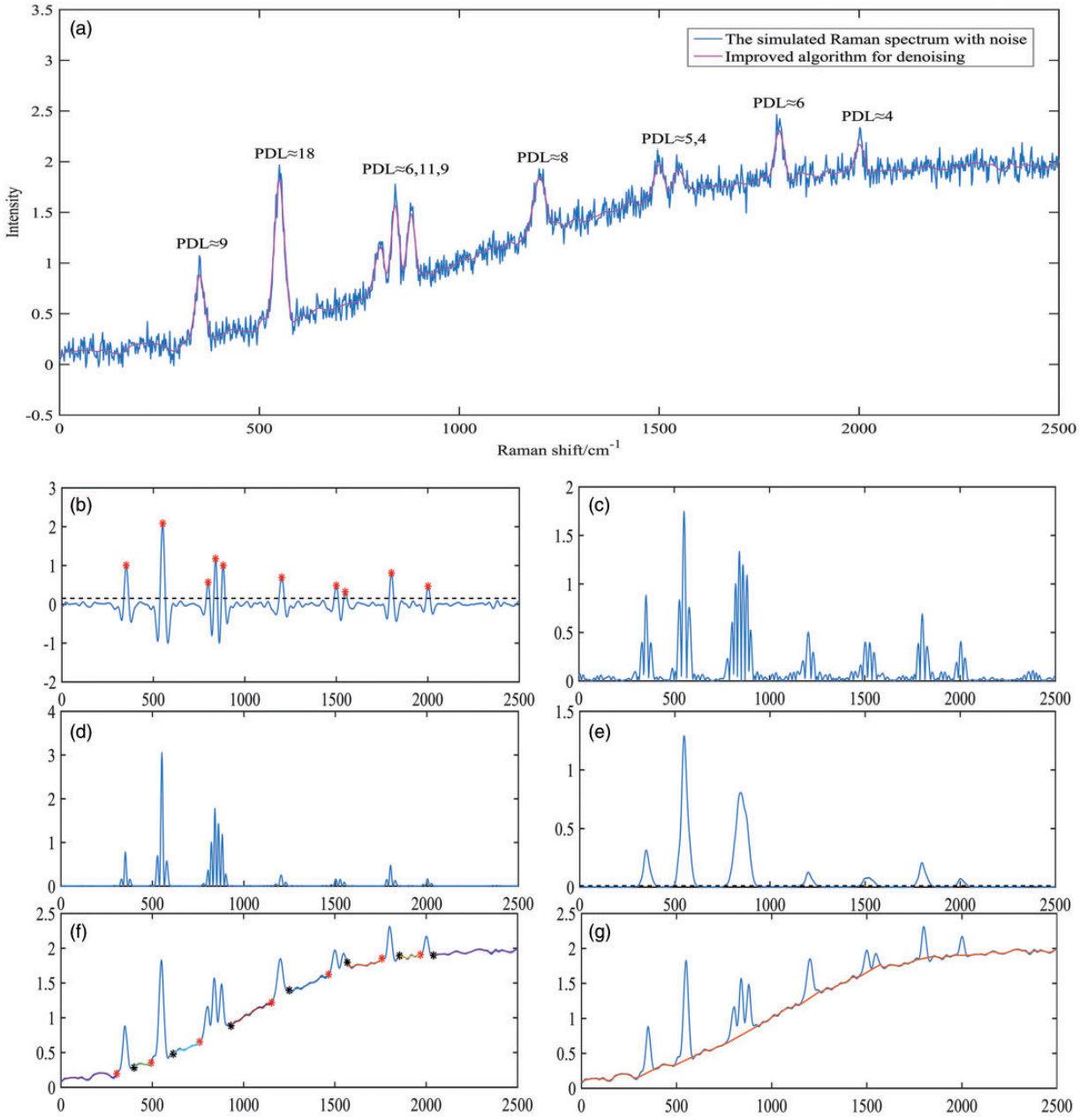

Figure 3. (a) The improved algorithm is used to denoise the simulated Raman spectrum and each peak with a different peak detection limit (PDL). (b)–(g) The steps of the baseline correction algorithm.

Baseline Correction Algorithm

The improved denoising algorithm is used to denoise the simulate Raman spectrum with sigmoidal background and Gaussian noise, of which the noise standard deviation is 0.08, as shown in Fig. 3a. Then the baseline correction test is done and the steps are as shown in Figs. 3b–g.

1. Detect the peak position through continuous wavelet transform with the Mexican Hat wavelet as the mother wavelet.[25] To reduce the amount of calculation, we only analyze the fifth scale, which generally meets the requirements of experimental accuracy (Fig. 3b). Here we use the method used in the above denoising section to estimate the noise standard deviation σ and choose $3 \times \sigma$ as the threshold, conservatively. The peak position will be corrected later.

   This part also involves the selection of the peak intensive area (PIA): Firstly, select a fixed window with a width of about one-third of the spectral length, then use it to find an area in the coefficients of Fig. 3b, while ensuring the number of peak positions in this area is the largest.

2. Based on continuous wavelet transform, use the Haar wavelet function for approximate second derivative calculation and take the absolute value of it (Fig. 3c).[26,27] The larger scale of decomposition can avoid the impact of noise while broadening the peak when using this method; we choose a Gaussian function to study the effect of decomposition scale on the peak width, as shown in Figs. 4a and 4b. When we choose the decomposition scale of 10; its impact on the peak width can be ignored.
3. In order to make the difference between the coefficients corresponding to the signal of interest and the coefficients corresponding to the noise larger, the second derivative is simply squared (Fig. 3d).
4. In order to facilitate the analysis, the coefficient of each point is obtained by averaging that of the adjacent points, which is equivalent to taking a moving average. The width of the moving average window cannot be too large or small, and no in-depth study on it has been done here. The width is chosen as 5 and the number of iterations is 10. Due to multiple iterative averaging, the coefficients of the non-PIA gradually approach zero, so we rely on the PIA to select the threshold. The threshold of the PIA is obtained by the method used in step 1. The starting endpoint and the ending endpoint of the peak are found based on the relationship between the endpoints and the threshold (Fig. 3e). The correction of the endpoints is mentioned later. In order to make it easier to understand for the reader, we use the Gaussian peak as an example in the Supplemental Material (Fig. S2).
5. The endpoints of the identified peak are marked, while the iteration averaging fit is applied to the non-peak regions and they form the new fitting background (Fig. 3f). The iterative process here is the same as that of step 4: when the number of iterations is large, the corrected non-peak regions will fluctuate severely. Conversely, when the number of iterations is small, the corrected non-peak regions will be smooth. We found that this choice basically meets the baseline correction requirements in practical application.
6. For the peak, the background is obtained by simply connecting the two endpoints, where the endpoints are the new endpoints that have been processed in step 5 (Fig. 3g). The correction of the spectral peaks is mentioned later.

Corrections for Endpoints, Spectral Peaks, and Peak Position

It is notable that some endpoints found in step 4 of the baseline correction algorithm belong to pseudo-endpoints. For example, unpaired endpoints (the starting endpoint and the ending endpoint of the peak are paired endpoints) at the left or right of the spectrum need to be removed. If there is no peak position (identified in step 1) between paired endpoints, these paired endpoints also need to be ignored.

For spectra with a low PDL, the non-peak regions after denoising may fluctuate violently and may be mistakenly identified as a signal of interest. Here we define PDL_{peak} for correction (Eq. 6).

$$PDL_{peak}(i) = \frac{2 \cdot m_i - l_i - r_i}{\sigma} \tag{6}$$

where σ represents the estimated noise standard deviation of the original signal, m_i represents the maximum value of the ith-identified peak, and l_i and r_i, respectively, represent the values corresponding to the starting endpoint and the ending endpoint of the ith-identified peak. As mentioned above, the spectrum peak is not detectable while the PDL is lower than 3; considering that peak may have severe height differences, it is more scientific to use PDL_{peak} $(2 \times 3 = 6)$ for comparison. We conservatively estimate that when PDL_{peak} is less than 5, the identified peak is considered to be caused by noise and removed.

When the identified peak is wider and there is a large height difference between the paired endpoints, the processed peak appears to be defective, as shown in Figs. 4c and 4d. In this case, we propose a modified method: we subtract the background of each identified peak, and if a negative value arises at a certain point (the value of this point is the minimum negative value) and its absolute value is greater than the threshold, this point is connected to the endpoints on both sides to form a new background (as shown in Fig. 4c, marked as 1, the triangle marked point), then the above work is repeated until the iteration condition is not satisfied (the place marked as 2 has not meet the iterative conditions). Then the background correction for the peak is finished. Here three times the noise standard deviation was selected as the threshold, where the noise standard deviation is calculated from the corrected non-peak regions. Figure 4e is the result after the correction for Fig. 4d.

Then, the peak positions identified in step 1 are corrected considering that the identified peaks should cover related peak positions.

After the baseline correction, we make some modifications to eliminate errors caused by the former processing. We calculate the noise standard deviation of the corrected non-peak regions and set it as the threshold to correct the endpoints of the peak. There are two reasons for doing this: (1) In step 4 of the baseline correction algorithm, the threshold selection is conservative and the peak broadens. (2) The corrected spectrum fluctuates in the non-peak regions, so we choose the noise standard deviation as the threshold to make the identification of endpoints more accurate, without mistaking the non-peak regions as the peaks. In addition, we can also determine the form of the spectrum peaks (single peak or overlapping peaks) based on the number of identified peak positions in the peak.

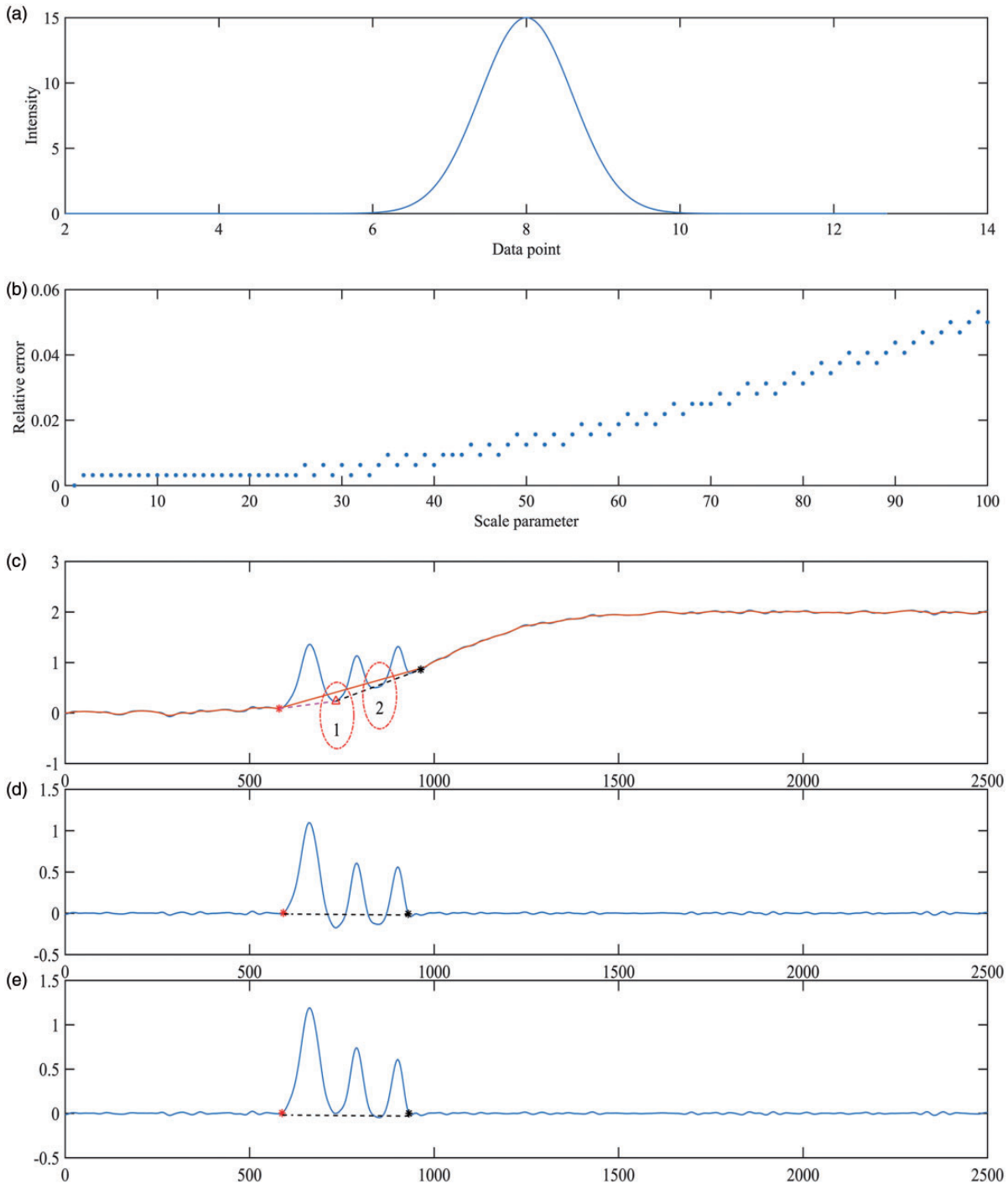

Figure 4. (a) Gaussian function used for testing. (b) The influence of scale parameters on peak width. (c) Unprocessed peak. (d) The peak after removing the background. (e) The result after peak correction.

Results and Discussion

Evaluation of the Denoising Effect

In order to test our algorithm, three of Matlab's built-in signal (shown in supplemental material) blocks, bumps, heavy sine, and the simulated Raman signal with sigmoidal background, were used for denoising tests. Here, the SNR and root mean square error (RMSE) are used to evaluate the denoising effect of signals with different noise standard deviation; the definition is given by Eq. 7 and Eq. 8. In this case, $x(n)$ represents the original pure signal, $y(n)$ denotes the denoised signal, and N is the length of the signal. The denoising effect becomes better with the increase of the SNR and the decrease of the RMSE. Three denoising

methods, the soft threshold method (S), conventional-scale correlation denoising method (C), and improved algorithm (I), are used to denoise three standard signals and the simulated Raman signal with different noise standard deviations utilizing Matlab, and the results are shown in Tables II–IV. We gave different noise standard deviations for each test signal, and we found that our algorithm has a big advantage even if each peak owns a PDL as low as 3. As the value of the PDL continues to decrease, the denoising effect will gradually deteriorate. To visualize this difference, the bumps signal is selected for analysis, as shown in Fig. 5, of which the noise standard deviation is 0.3. It can be seen that the improved algorithm effectively suppresses the pseudo-Gibbs effect (parts marked 1, 2, and 3) while preserving the details of the signal (parts marked 4), and the denoising effect is better than the conventional-scale correlation denoising method

Table II. Signal-to-noise ratio with different denoising methods (DMs) for denoising three standard signals with different noise standard deviations (NSDs).

DM / NSD	Blocks			Bumps			Heavy sine		
	S	C	I	S	C	I	S	C	I
0.1	25.82	33.41	33.35	23.38	29.85	30.65	35.47	37.96	38.86
0.2	21.97	27.13	28.35	19.98	25.64	25.45	30.69	30.1	32.94
0.3	18.6	24.38	25.43	16.97	23.26	23.6	28.82	27.47	30.72
0.4	17.92	22.1	24.58	16.3	22.07	22.38	27.53	27.18	28.43
0.5	16.52	20.72	22.07	14.81	20.06	20.6	26.18	26.15	26.22
0.6	15.81	18.05	20.67	13.87	17.94	18.48	23.45	23.88	23.79
0.7	15.65	17.54	19.66	13.83	17.7	18.1	23.68	23.68	24.18

Table III. Root mean square error with different denoising methods (DMs) for denoising three standard signals with different noise standard deviation (NSDs).

DM / NSD	Blocks			Bumps			Heavy sine		
	S	C	I	S	C	I	S	C	I
0.1	0.1519	0.0634	0.0639	0.122	0.0579	0.0528	0.052	0.039	0.0352
0.2	0.2366	0.1306	0.1136	0.1804	0.094	0.0961	0.0901	0.0965	0.0695
0.3	0.3489	0.1794	0.159	0.2551	0.1236	0.1189	0.1117	0.1305	0.0898
0.4	0.3775	0.233	0.1753	0.2755	0.1418	0.1369	0.1297	0.135	0.1169
0.5	0.4436	0.2734	0.2339	0.3272	0.1787	0.168	0.1514	0.152	0.1508
0.6	0.481	0.3718	0.275	0.3647	0.2281	0.2144	0.2006	0.199	0.1994
0.7	0.49	0.3941	0.3089	0.3663	0.2345	0.224	0.2037	0.2037	0.1907

Table IV. Signal-to-noise ratio (SNR) and root mean square error (RMSE) with different denoising methods (DMs) for denoising simulated Raman signals with different noise standard deviations (NSDs).

DM / NSD	SNR			RMSE		
	S	C	I	S	C	I
0.03	34.4432	40.88	41.9795	0.0271	0.0129	0.0114
0.06	30.2838	34.5735	35.8116	0.0437	0.0266	0.0231
0.08	28.8612	33.0298	34.7849	0.0514	0.0318	0.026
0.1	28.1945	29.9099	32.5417	0.0555	0.0456	0.0337
0.12	27.6486	29.265	31.5953	0.0591	0.0491	0.0375

and the soft threshold method.

$$\mathrm{SNR} = 10\lg\left(\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}[x(n)-y(n)]^2}\right) \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{N}[x(n)-y(n)]^2}{N}} \tag{8}$$

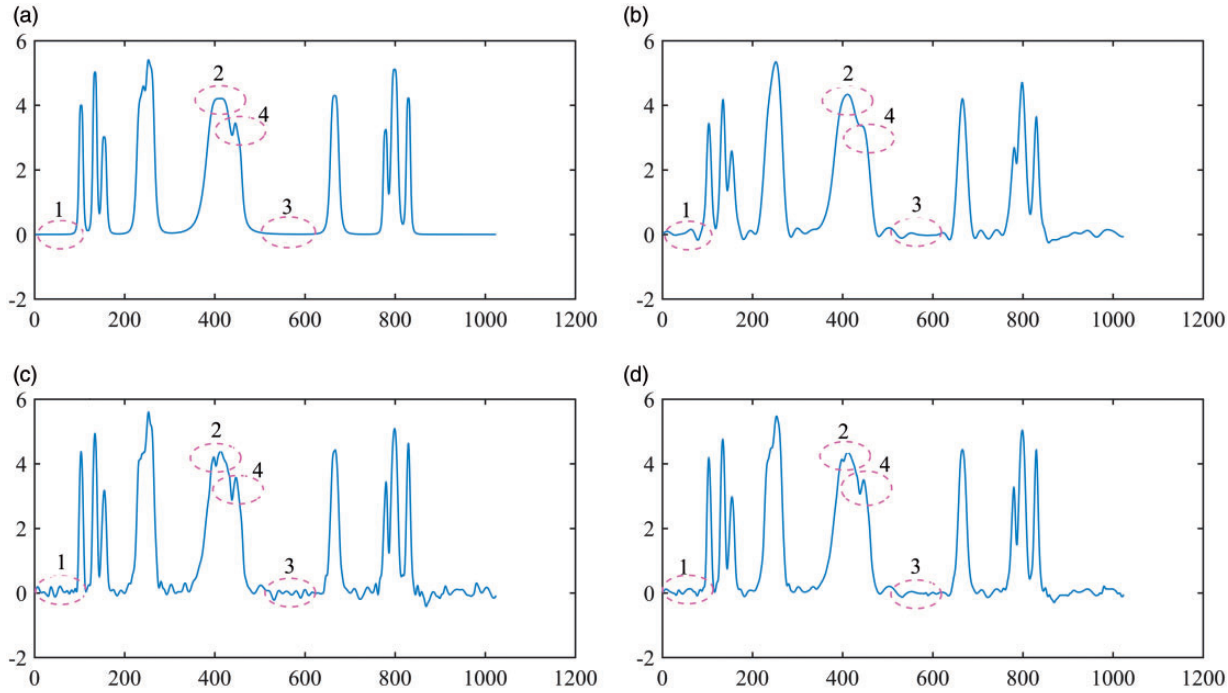

Figure 5. Three denoising methods are used to denoise the bumps with a noise standard deviation of 0.3: (a) pure bumps; (b) soft threshold method; (c) conventional-scale correlation denoising method; and (d) improved algorithm.

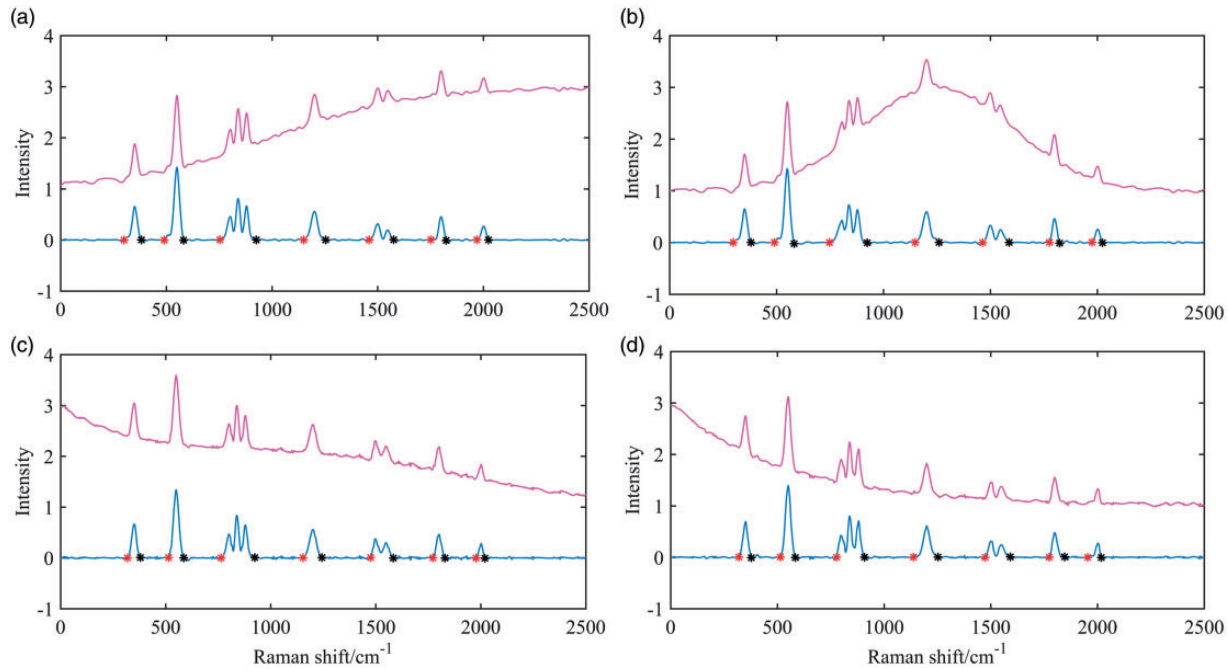

Figure 6. Baseline correction for simulated spectra with (a) sigmoidal background, (b) Gaussian background, (c) exponential background, and (d) fifth polynomial background.

Test for Simulated and Experimental Spectra

We denoise the simulated Raman spectrum with four background and noise standard deviations of 0.08, and a baseline correction test of the denoised spectrum is performed. Figure 6a is the final result of the above baseline correction algorithm, and Figs. 6b–d show the baseline correction results of the three other types of backgrounds. Moreover, we selected two representative spectral signals from the Raman spectra collected in our laboratory for

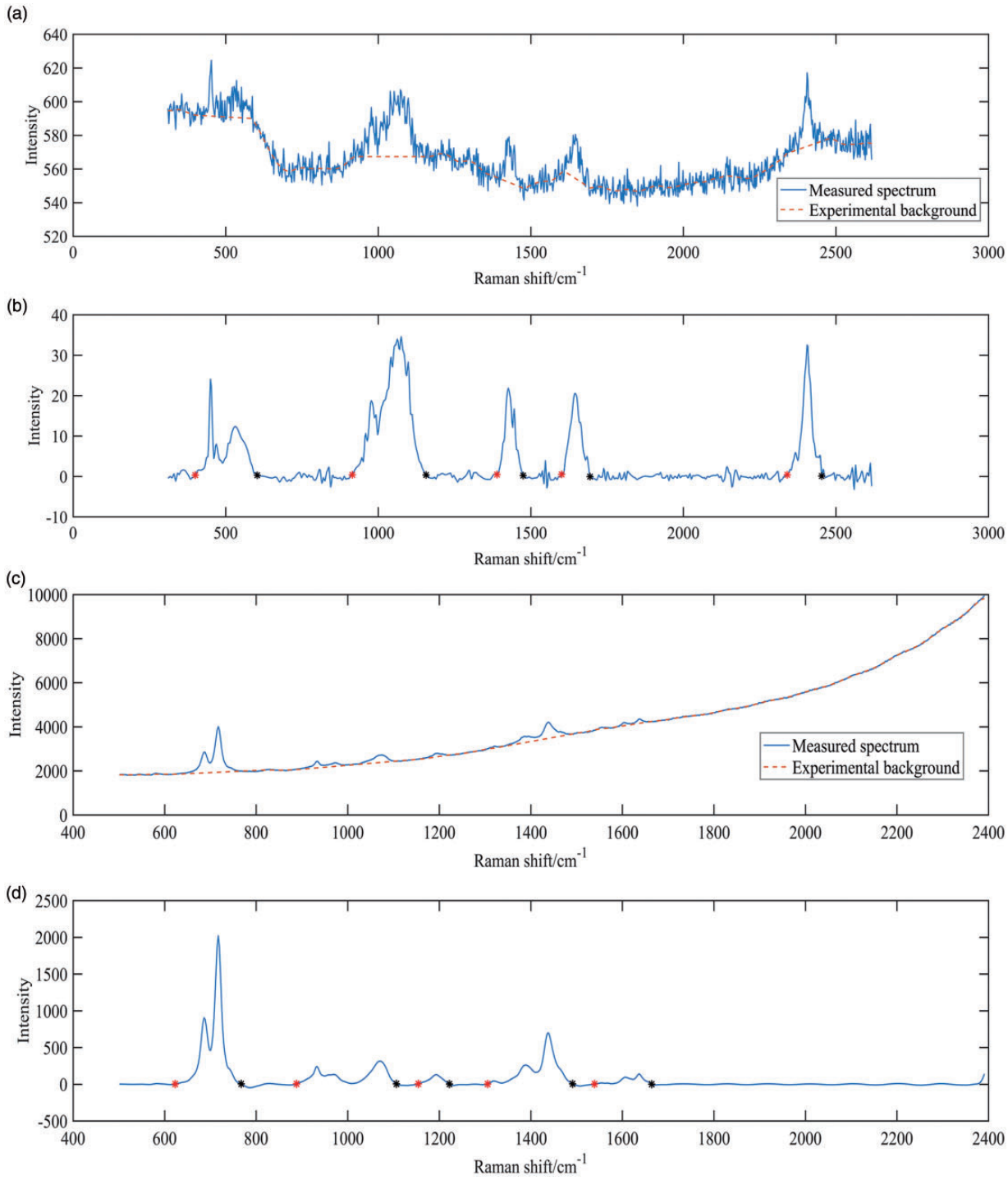

Figure 7. (a) The measured Raman spectrum of the skin cancer tissue. (b) Raman spectrum of the skin cancer tissue after denoising and background removal. (c) The measured Raman spectrum of the crystal violet. (d) Raman spectrum of the crystal violet after denoising and background removal.

processing, as shown in Figs. 7a and 7c, where one is the case of the low SNR (the PDL of each peak is in the range of 2–6) and the other is the case of the low signal baseline ratio (SBR), defined as the height of the peak (before adding baseline or noise) relative to the baseline amplitude; here, the SBR of each peak is in the range of 0.0125–0.25. These two rough signals are chosen to prove the validity of our proposed algorithm. The spectra after processing are shown in Figs. 7b and 7d. All Raman peaks of interest are preserved with little background after applying our algorithm, which is very helpful for Raman spectra analysis and Raman imaging reconstruction in complex environments. The experimental Raman spectra here were acquired from a Witec Confocal Raman System with 532 nm laser excitation; the accumulation time is 1 s and the laser power is 5 mW.

Conclusion

In summary, this algorithm incorporates the functions of denoising and baseline correction, as well as the identification of the spectrum peak and the recognition of single peaks and overlapping peaks. This novel algorithm was used to test various simulated and real Raman spectra, and all spectra are processed completely and precisely. We believe that the new pre-processing algorithm proposed in this paper can be used as a powerful tool for Raman spectra processing. In addition, the algorithm can be applied for denoising and baseline correction of Raman spectra with few human interventions. Notably, it also has great potential applications for processing other spectral signals and chromatograms.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant Nos. 61405083), the Fundamental Research Funds for the Central Universities (lzujbky-2018-129), and the Natural Science Foundation of Gansu Province of China (Grant Nos. 17JR5RA197).

Supplemental Material

All supplemental material mentioned in the text, including parameter *t* settings, Figs. S1 and S2, and three of Matlab's built-in signals, is available in the online version of the journal.

ORCID iD

Yuee Li http://orcid.org/0000-0001-5430-1587

References

1. B. Plowman, S.J. Ippolito, V. Bansal. "Gold Nanospikes Formed Through a Simple Electrochemical Route with High Electrocatalytic and Surface Enhanced Raman Scattering Activity". Chem. Commun. 2009. 45(33): 5039–5041.
2. A.F. Chrimes, K. Khoshmanesh, P.R. Stoddart. "Active Control of Silver Nanoparticles Spacing Using Dielectrophoresis for Surface-Enhanced Raman Scattering". Anal. Chem. 2012. 84(9): 4029–4035.
3. H.F. Wang, Y. Fu, P. Zickmund. "Coherent Anti-Stokes Raman Scattering Imaging of Axonal Myelin in Live Spinal Tissues". Biophys. J. 2005. 89(1): 581–591.
4. D. Zhang, K.N. Jallad, D.B. Amotz. "Stripping of Cosmic Spike Spectral Artifacts Using a New Upper-Bound Spectrum Algorithm". Appl. Spectrosc. 2001. 55(11): 1523–1531.
5. G.R. Phillips, J.M. Harris. "Polynomial Filters for Data Sets with Outlying or Missing Observations: Application to Charge-Coupled-Device-Detected Raman Spectra Contaminated by Cosmic Rays". Anal. Chem. 1990. 62(21): 2351–2357.
6. W. Hill, D. Rogalla. "Spike-Correction of Weak Signals from Charge-Coupled Devices and Its Application to Raman Spectroscopy". Anal. Chem. 1992. 64(21): 2575–2579.
7. F. Ehrentreich, L. Summchen. "Spike Removal and Denoising of Raman Spectra by Wavelet Transform Methods". Anal. Chem. 2001. 73(11): 4364–4373.
8. D. Giaouris, J.W. Finch, O.C. Ferreira. "Wavelet Denoising for Electric Drives". IEEE Trans. Ind. Electron. 2008. 55(2): 543–550.
9. F. Ehrentreich. "Wavelet Transform Applications in Analytical Chemistry". Anal. Bioanal. Chem. 2002. 372(1): 115–121.
10. S. Mallat, W.L. Hwang. "Singularity Detection and Processing with Wavelets". IEEE Trans. Inf. Theory. 1992. 38(2): 617–643.
11. D.L. Donoho, I.M. Johnstone, G. Kerkyacharian. "Wavelet Shrinkage: Asymptopia?" J.R. Statist. Soc. B. 1995. 57(2): 301–369.
12. D.L. Donoho. "De-Noising by Soft-Thresholding". IEEE Trans. Inf. Theory. 1995. 41(3): 613–627.
13. D.L. Donoho, I.M. Johnstone. "Ideal Spatial Adaptation by Wavelet Shrinkage". Biometrika. 1994. 81(3): 425–455.
14. Y. Xu, J.B. Weaver, D.M. Healy. "Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique". IEEE Trans Image Process. 1994. 3(6): 747–758.
15. C.A. Lieber, A.M. Jansen. "Automated Method for Subtraction of Fluorescence from Biological Raman Spectra". Appl. Spectrosc. 2003. 57(11): 1363–1367.
16. A. Jirasek, G. Schulze, M.L. Yu. "Accuracy and Precision of Manual Baseline Determination". Appl. Spectrosc. 2004. 58(12): 1488–1499.
17. J. Zhao, H. Lui, D.I. Mclean. "Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy". Appl. Spectrosc. 2007. 61(11): 1225–1232.
18. M.N. Leger, A.G. Ryder. "Comparison of Derivative Preprocessing and Automated Polynomial Baseline Correction Method for Classification and Quantification of Narcotics in Solid Mixtures". Appl. Spectrosc. 2006. 60(2): 182–193.
19. A. O'Grady, A.C. Dennis, D. Denvir. "Quantitative Raman Spectroscopy of Highly Fluorescent Samples Using Pseudosecond Derivatives and Multivariate Analysis". Anal. Chem. 2001. 73(9): 2058–2065.
20. Y. Hu, T. Jiang, A. Shen. "A Background Elimination Method Based on Wavelet Transform for Raman Spectra". Chemom. Intell. Lab. Syst. 2007. 85(1): 94–101.
21. L. Shao, P.R. Griffiths. "Automatic Baseline Correction by Wavelet Transform for Quantitative Open-Path Fourier Transform Infrared Spectroscopy". Environ. Sci. Technol. 2007. 41(20): 7054–7059.
22. P.M. Ramos, I. Ruisanchez. "Noise and Background Removal in Raman Spectra of Ancient Pigments Using Wavelet Transform". J. Raman Spectrosc. 2005. 36(9): 848–856.
23. A. Das, S. Pisana, B. Chakraborty. "Monitoring Dopants by Raman Scattering in an Electrochemically Top-Gated Graphene Transistor". Nat. Nanotechnol. 2008. 3(4): 210:215.

24. G.P. Nason, B.W. Silverman. "The Stationary Wavelet Transform and Some Statistical Applications". In: A. Antoniadis, G. Oppenheim, editors. Wavelets and Statistics. New York: Springer, 1995, pp.281–299.
25. P. Du, W.A. Kibbe, S.M. Lin. "Improved Peak Detection in Mass Spectrum by Incorporating Continuous Wavelet Transform-Based Pattern Matching". Bioinformatics. 2006. 22(17): 2059–2065.
26. X. Shao, A.K. Leung, F. Chau. "Wavelet: A New Trend in Chemistry". Acc. Chem. Res. 2003. 36(4): 276–283.
27. X. Shao, C. Pang, Q. Su. "A Novel Method to Calculate the Approximate Derivative Photoacoustic Spectrum Using Continuous Wavelet Transform". Fresenius' J. Anal. Chem. 2000. 367(6): 525–529.