Article

Applied Spectroscopy

Applied Spectroscopy
2017, Vol. 71(8) 1947–1953

DOI: 10.1177/0003702817700656
journals.sagepub.com/home/asp
SAGE

An Automated Algorithm of Peak Recognition Based on Continuous Wavelet Transformation and Local Signal-to-Noise Ratio

Fang Qian, Yihui Wu, and Peng Hao

Abstract

Raman peaks carry valuable information about constituent chemical bonds. Therefore, peak recognition is a very essential part of spectral analysis. The fully automated peak recognition is convenient in practical application. A fully automated Raman peaks recognition algorithm based on continuous wavelet transformation and local signal-to-noise ratio (LSNR) is proposed. This algorithm extracts feature points through continuous wavelet transformation and recognizes peaks through LSNR. This algorithm also can be used to eliminate spike, noise, and baseline. Both simulated and experimental data are used to evaluate the performance of the CWT-LSNR algorithm compared with the other two algorithms. The results show that CWT-LSNR gives better accuracy and has the advantage of easy use.

Keywords

Peak recognition, spectral preprocessing, wavelet transformation, signal-to-noise ratio (SNR)

Date received: 23 November 2016; accepted: 21 February 2017

Introduction

Raman spectroscopy provides detailed information and is used in chemical, biochemical, environmental, and food sciences. The spectral peaks carry valuable information about constituent chemical bonds.[1–4] It is necessary to recognize the peaks for further analysis. In recent years, different peak recognition algorithms have been used in the field of chromatogram and some of them are described in detail.[5–10] The continuous wavelet transformation algorithm (CWT)[11–14] is widely used and is a method based on finding the camber line in the wavelet coefficient matrix, and moreover, length threshold and signal-to-noise ratio (SNR) threshold of the camber line should be prepared. One general disadvantage of this method is the requirement to optimize parameters, which can be very time consuming. Vivó-Truyols uses high-order derivatives to assess the number of underlying compounds under a given peak cluster (DW).[15] High-order derivatives belong to high-pass filter, so that this algorithm may enhance the noise and Raman signal may submerge in the noise. Therefore, a multitude of algorithms are unavailable to the Raman spectrum.

A new fully automated algorithm for Raman peak recognition is proposed in this study. This algorithm is based on continuous wavelet transformation and local signal to noise ratio (LSNR). Through the wavelet transformation, the feature points in the original spectrum signal, such as peak, valley, starting point, ending point, and inflection point, are classified. In using the LSNR, peaks can be recognized.

Method

The proposed algorithm can recognize the spectral peaks automatically by means of CWT and LSNR, namely CWT-LSNR. The algorithm is divided into four steps:

Step 1: *Baseline removal*. Preprocessing is required to eliminate effects of unwanted signals such as fluorescence, cosmic rays, laser power fluctuations, and detector noise. A fully automated algorithm based on wavelet feature points

State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China

Corresponding author:
Yihui Wu, State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China.
Email: yihuiwuciomp@126.com

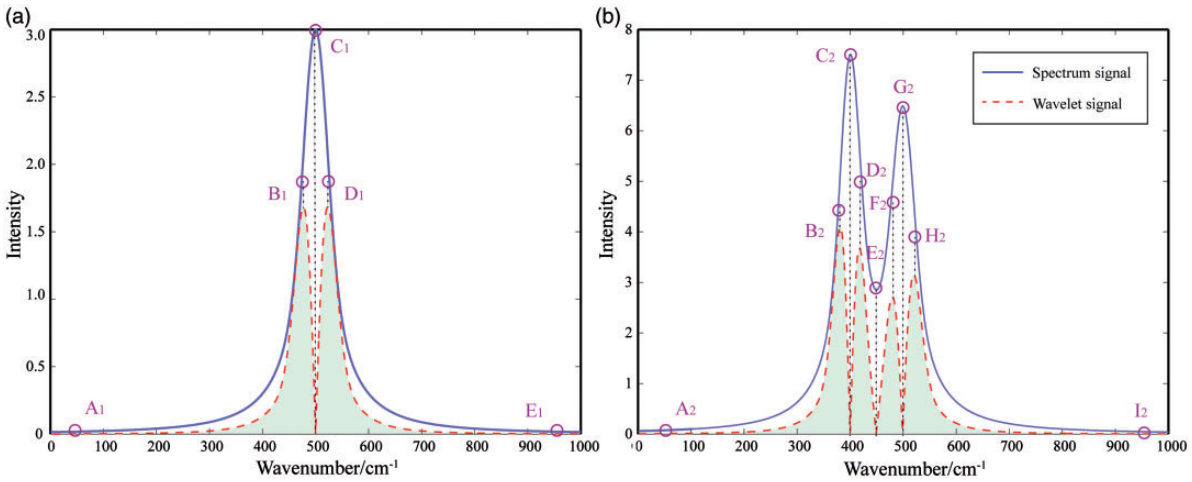

Figure 1. Simulated spectrum signal. (a) Single peak; (b) overlap peak.

Table 1. The correspondence between wavelet extreme points and the spectral feature points.

Peak style	Single peak	Overlap peak	Spectrum signal	Wavelet signal	Spectral first derivative
Starting point	A1	A2	Minimum	Minimum	×
Left inflection point	B1	B2,F2	Rising edge	Maximum	×
Peak	C1	C2,G2	Maximum	Minimum	v
Right inflection point	D1	D2,H2	Trailing edge	Maximum	×
Valley	None	E2	Minimum	Minimum	v
Ending point	E1	I2	Minimum	Minimum	×

Table 2. Simulated pure spectrum signal.

Peak position (cm^{-1})	645	1090	1535	2130	2200	2270
Peak height	3	6	9	6	6	6
FWHM	25	30	35	40	45	50

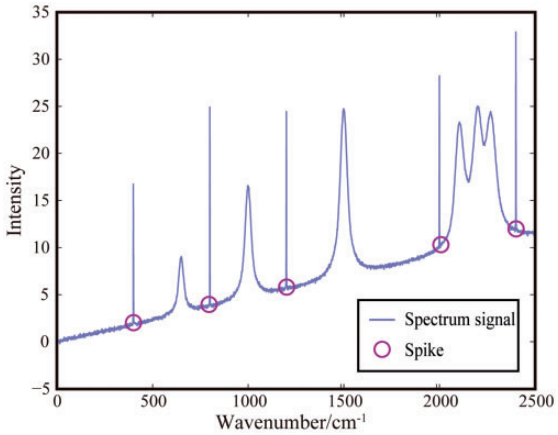

Figure 2. Simulated spectrum signal with spikes, noise, and baseline,

and segment interpolation (AWFPSI) is used to remove the baseline.

Step 2: *Spike removal and noise attenuation*. Spike removal is a very important part of preprocessing.[16,17] It is necessary to remove the spikes of the raw signals to increase the accuracy of the peak recognition. Generally, the spike is sharper compared with the genuine Raman band and its full width at half-maximum (FWHM) does not exceed 2 cm^{-1}.

A method based on local extreme values is proposed in Step 2. Three-point zero-order Savitzky–Golay (SG)[18] filter is used to reduce the noise and retains all important spectral bands. The SG filter is a moving window based on local polynomial fitting procedure. The size of the moving window is three and the polynomial order is zero in this method. Because the SG filter is a non-linear, weighted, smoothing function, it is guaranteed that high-frequency

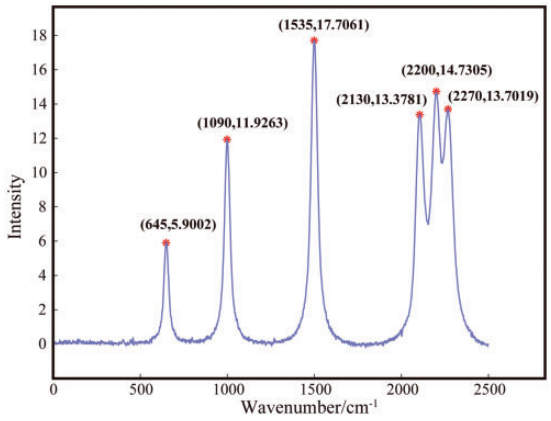

Figure 3. Peak recognition.

noise is well suppressed. After filtering, the singularities are detected through residual error. The median of such a local window of intensity values at each singularity is calculated to cover the singularity value. Other intensity values keep constant. Preliminary spike removal spectra $S_1(n)$ is obtained as follows

$$S_1(n) = \begin{cases} S(n), & res < 3 \times \sigma \\ S'(n), & res \geq 3 \times \sigma \end{cases} \quad (1)$$

$$res = S(n) - S'(n)$$
$$\sigma = 0.7431 \times iqr(res) \quad (2)$$

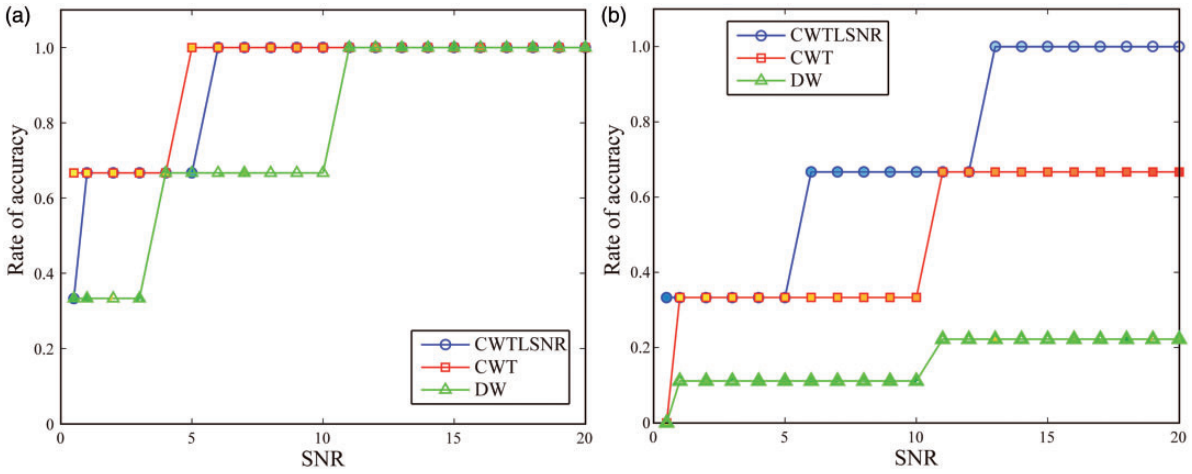

Figure 4. Rate of peak recognition accuracy for three algorithms. (a) Single peak; (b) overlap peak.

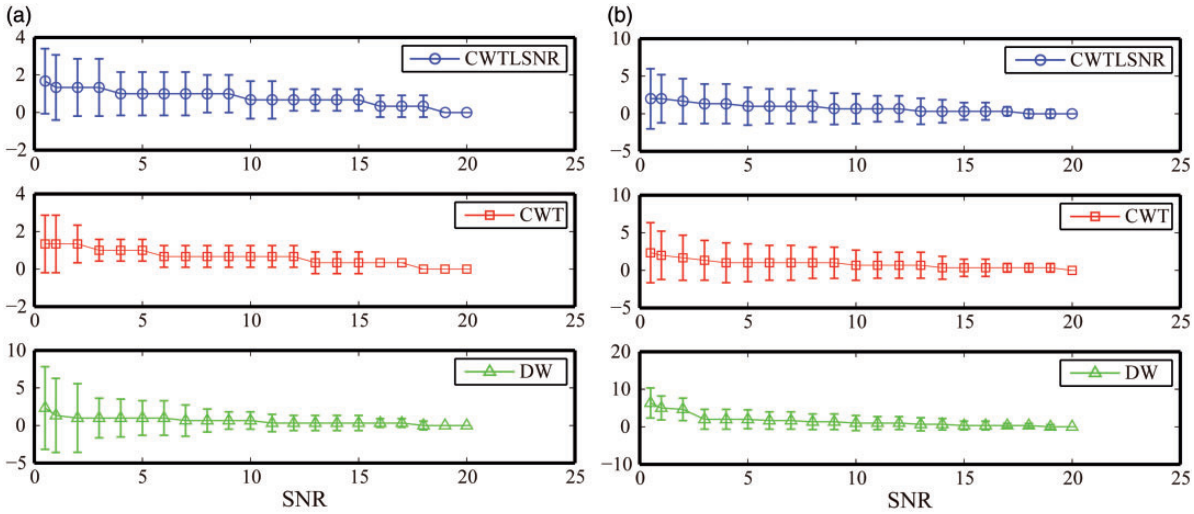

Figure 5. Error bar of peak position for three algorithms. (a) Single peak; (b) overlap peak.

Table 3. Comparison of peak evaluated error of three algorithms.

Method	Single peak		Overlap peak	
Equalizing value	Error average	Error standard deviation	Error average	Error standard deviation
CWT-LSNR	0.6508	0.6133	0.7937	1.8843
CWT	0.6349	0.5505	0.8571	2.4961
DW	0.7937	1.7909	1.6032	2.9051

Table 4. Experimental conditions used to evaluate the CWT-LSNR algorithm.

Spectral type	Description
Raman spectrum	Raman spectrometer Optical source: 785 nm laser Resolution: 3 nm Integration time: 100 ms–10 s Confocal microscope Raman spectrometer Optical source: 488 nm laser Resolution: 1 nm Integration time: 10 s

where $S(n)$ is original spectrum signal, $S'(n)$ is smoothed signals by SG filtering, *res* is the residual error, and *iqr* is a Matlab function to calculate the quartile. The residual error threshold σ is calculated with reference to the literature.[19,20]

The local maximum values are extracted from $S_1(n)$. The assumed coordinate of a certain maximum value is (x_{max}, y_{max}). Then, two symmetry points in the nearest neighborhood, whose values are $y_{max}/2$, are extracted. The FWHM of this maximum value can be calculated through the deviation of the two symmetry point abscissa. If the FWHM is less than or equal to 2 cm^{-1}, this point can be considered as a spike. The spike needs to be removed and the intensity value of the spike should be replaced by the median of its four nearest neighborhoods.

Step 3: *Computing continuous wavelet transformation*. The CWT of signal $s(t)$ at scale a ($a \in R$) and translational value b ($b \in R$) can be expressed by the following integral:

$$s_w(a,b) = \frac{1}{|a|}\int_{-\infty}^{+\infty} s(t)\psi\left(\frac{t-b}{a}\right)\mathrm{d}t \tag{3}$$

where $\psi(t)$ is a continuous function in both the time domain and frequency domain called mother wavelet. When the signals are decomposed with CWT, we can get some local extreme values of wavelet coefficients. These extreme values correspond to the feature points in the spectrum signal, including peak, valley, starting point, ending point, and inflection point. There are many wavelet families available in the literature such as Harris, Daubechies, Biorthgonal, Ceiflets, and Symlets, and different wavelet families have different mother wavelets. A series of functions that can be obtained by the scale and translation of the mother wavelets are wavelet basic functions.

Using the wavelet basic function, frequency-like information from the spectrum signals can be extracted. In order to extract the extreme values, the wavelet basis function should have first order vanishing moment, such as Harris, Daubechies, and Biorthgonal. If the wavelet basis function is odd symmetry, the corresponding wavelet filter coefficient has linear phase. This characteristic makes the extreme value points unbiased in any scale, but Daubechies does not have symmetry properties. Harris and Bior1.1 have bad localization properties. Considering the exact reconstruction, Bior1.3 is the wavelet basis function in this algorithm.

Simulated spectrum signal with a single peak and its wavelet transformation signal are shown in Figure 1a. Simulated spectrum signal with overlap peaks and the wavelet transformation signal are shown in Figure 1b.

The wavelet extreme points corresponding to the feature points in the spectrum signal are shown in Figure 1.

Table 1 shows the peak, valley, starting point, and ending points as minimum values in the wavelet curve. The first derivatives of peak and valley in the spectrum signal go through zero point. By comparing these characteristics, we can identify the different feature points.

Step 4: *Computing the local signal-to-noise ratio*. The second order difference (step is k) of the feature point p in the spectrum signal is defined as follows

$$\Delta(\Delta S_k) = S_{p-k} - 2S_p + S_{p+k} \tag{4}$$

where S is the spectrum signal and its length is n. S_p is the spectrum signal at point p.

The LSNR (step is k) is defined as follows,

$$LSNR = \frac{|S_{p-k} - 2S_p + S_{p+k}|}{Noise_p} \tag{5}$$

$Noise_p$ is the energy of noise at point p. It can be written as:

$$Noise_p = S_p - S_p^{smooth} \tag{6}$$

S_p^{smooth} is the smoothed signal using SG filter.

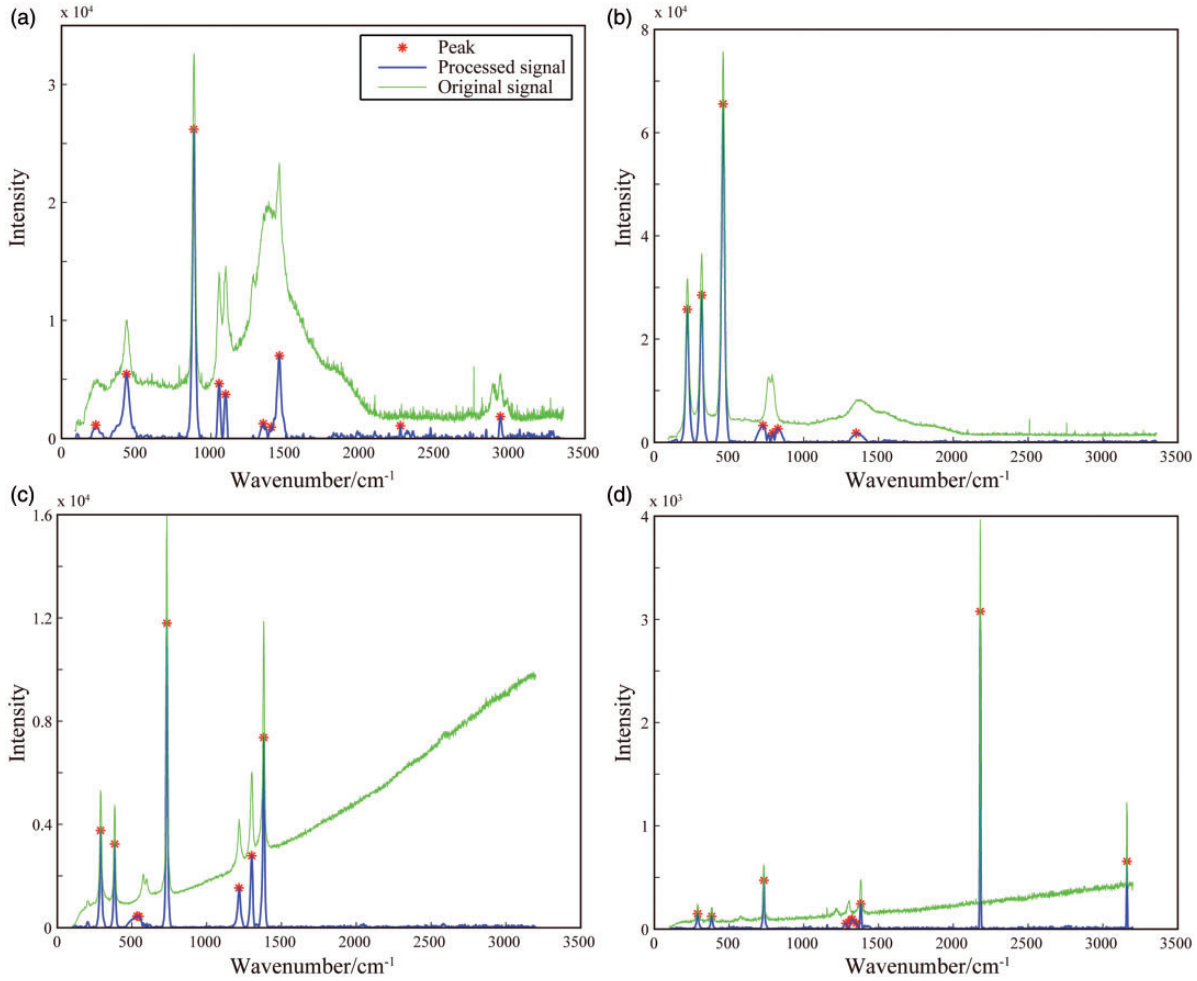

Figure 6. Peak recognition of (a) alcohol; (b) carbon tetrachloride; (c) polytetrafluoroethylene; (d) methanol.

If the scale of step k matches the width of the peak, Eq. 5 can be defined as:

$$LSNR_p = \frac{2|S_p|}{Noise_p} \tag{7}$$

The peak SNR (PSNR) at point p can be defined as:

$$PSNR_p = \frac{|S_p|}{Noise_p} \tag{8}$$

Comparing Eq. 7 with Eq. 8, the PSNR is twice as big as LSNR.

By combining the characteristics of wavelet feature points and the relationship between LSNR and PSNR, the spectral peak can be recognized accurately.

Experimental Results and Analysis

Both simulated and experimental data are used to evaluate the performance of the CWT-LSNR algorithm.

Simulated Signals for Comparison

Mathematically, the simulated spectrum signal $s(t)$ can be expressed as follows:

$$s(t) = x(t) + k(t) + n(t) + b(t) \tag{9}$$

where $s(t)$ is the simulated spectrum signal, $x(t)$ is the pure spectrum signal, $k(t)$ is the simulated spike, $n(t)$ is the noise, and $b(t)$ is the baseline.

Three single peaks and three overlap peaks comprise the simulated spectrum signal $s(t)$: each peak varied in intensity. The parameters are shown in Table 2.

Spikes are added in 400, 800, 1200, 2000, and 2400 cm^{-1}, their peak height is 9, 15, 10, 12, and 15, respectively. Figure 2 shows the simulated spectrum signal with spikes, noise, and baseline.

Figure 3 shows the result of peak recognition using the proposed CWT-LSNR algorithm. Through preprocessing, the spikes, noise, and baseline are removed and the peaks can be recognized accurately.

Table 5. The peak position and height of the Raman spectrum.

Substance	Peak position (cm^{-1})	Height (cm^{-1})
Alcohol	220, 435, 885, 1054, 1099, 1350, 1404, 1457, 2267, 2936	836, 5440, 26 195, 4631, 3724, 1246, 945, 6987, 1065, 1849
Carbon tetrachloride	220, 317, 460, 762, 789, 826, 1350	26 344, 29 253, 66 980, 1663, 1806, 2575, 1787
Polytetrafluoroethylene	290, 383, 533, 733, 1216, 1300, 1380	3890, 3345, 440, 12 480, 1565, 2824, 7665
Methanol	290, 383, 733, 1283, 1323, 1380, 2180, 3161	147, 121, 470, 54, 94, 242, 3076, 655

A range of SNR factors are multiplied by the Gaussian white noise created to evaluate the ability of the algorithms to perform peak recognition in different noise levels. The factors from the set {0.5, 1, 2, 3, 4, . . . ,18, 19, 20} are used for low and high noise levels, respectively.

Figure 4 is a diagram of peak recognition accuracy in different noise environment. The accuracy rises as the SNR increases. Continuous wavelet transformation and LSNR algorithm ranks second in terms of the single peak recognition. When the SNR reaches 6 dB, the accuracy of CWT-LSNR will arrive at 100%. The accuracy of CWT-LSNR is higher than CWT and DW algorithms in the overlap peak recognition. When the SNR reaches 13 dB, the accuracy of CWT-LSNR will arrive at 100%.

Figure 5a shows the average and the standard deviation of the single peak evaluated error for the various SNR. Figure 5b shows the average and the standard deviation of the overlap peak evaluated error for the various SNR. The circles, blocks, and triangles in Figure 5 represent the average of the error between the ideal peak position and estimated peak position. The horizontal lines in Figure 5 represent the standard deviation of the error between the ideal peak position and estimated peak position. The error bar in Figure 5 shows the average and the standard deviation decrease as the SNR increases.

Table 3 shows the mean value of the error average and error standard deviation of CWT-LSNR, CWT, and DW in Figure 5. In the single peak, the mean error of CWT-LSNR is lower than DW. The mean error of CWT-LSNR is comparable to CWT and has advantage of ease of use. In the overlap peak, the mean error of CWT-LSNR is the lowest.

Experimental Signals for Comparison

Experimental signals from Raman spectrum are used to show the applicability of the CWT-LSNR algorithm in actual databases. Information about the experimental condition is shown in Table 4.

The different chemical substances, such as alcohol and carbon tetrachloride, are measured using the laser of 785 nm wavelength for excitation by Raman spectrometer. Polytetrafluoroethylene and methanol are measured using the laser of 488 nm wavelength for excitation by confocal microscope Raman spectrometer. The results of peak recognition are shown in Figure 6.

The recognition ability of the CWT-LSNR algorithm is demonstrated by the Raman spectrum with four different chemical substances and two different integration time. Figure 6 shows CWT-LSNR could remove the spike, baseline and abate noise. Continuous wavelet transformation and LSNR not only recognizes the peaks accurately, but also provides the peak position and height, as shown in Table 5.

Conclusion

As mentioned previously, peak recognition is a very important part of spectral analysis. A fully automated Raman peaks recognition algorithm based on CWT-LSNR is proposed in this study. Both simulated and experimental data are used to evaluate and compare the performance of the CWT-LSNR algorithm. The results show that DW has low accuracy in low noise ranges, but CWT-LSNR proves highly accurate. Peaks with SNR down to approximately one and below can be recognized by CWT-LSNR. This algorithm is also available if the spikes, baseline and noise are in existence. Moreover, the CWT-LSNR is a fully automated peak recognition algorithm whereas CWT requires optimizing parameters. When processing large data sets, a fully automated algorithm such as CWT-LSNR would be desirable as it is not necessary to set any parameters. Thus, the CWT-LSNR algorithm is more suitable for peak recognition for Raman spectrum and has advantage of ease of use.

Conflict of Interest

The authors report there are no conflicts of interest.

Funding

This work was funded by the 863 Program (2015AA042402).

References

1. B. Singh, R. Gautam, S. Kumar, B.N. Vinay Kumar, U. Nongthomba, et al. "Application of Vibrational Microspectroscopy to Biology and Medicine". Curr. Sci. 2012. 102(2): 232–244.
2. T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, J. Popp. "How to Pre-Process Raman Spectra for Reliable and Stable Models?" Anal. Chim Acta. 2011. 704(1–2): 47–56.
3. R. Gautam, S. Vanga, F. Ariese, S. Umapathy. "Review of Multidimensional Data Processing Approaches for Raman and Infrared Spectroscopy". EPJ Techniques and Instrumentation. 2015. 2(8): 1–10.

4. M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, et al. "Molecular Pathology via IR and Raman Spectral Imaging". J. Biophotonics. 2013. 6(11–12): 855–886.
5. G. Reich. "Recognizing Chromatographic Peaks with Pattern Recognition Methods: Part 2. Evaluation of Different Distance Measures". Anal. Chim. Acta. 1987. 201: 171–183.
6. G. Reich. "Recognizing Chromatographic Peaks with Pattern Recognition Methods III. Application of the Algorithm for Peak Recognition in Trace Analysis". Chromatographia. 1987. 24(1): 659–665.
7. G. Reich. "Recognizing Chromatographic Peaks with Pattern Recognition Methods: Part 1. Development of a k-Nearest-Neighbour Technique". Anal. Chim. Acta. 1987. 201: 153–170.
8. F. Janssens, J.P. Francois. "Evaluation of Three Zero-area Digital Filters for Peak Recognition and Interference Detection in Automated Spectral Data Analysis". Anal. Chem. 1991. 63(4): 320–331.
9. H. Wätzig. "Peak Recognition Technique by a Computer Program Copying the Human Judgment". Chromatographia. 1992. 33(5–6): 218–224.
10. A. Likar, T. Vidmar. "A Peak Search Method based on Spectrum Convolution". J. Phys. D: Appl. Phys. 2003. 36(15): 1903.
11. G. Cooper, M. Kubik, K. Kubik. "Wavelet Based Raman Spectra Comparison". Chemometrics and Intelligent Laboratory Systems. 2011. 107(1): 65–68.
12. A. Wee, D.B. Grayden, Y. Zhu, K. Petkovic-Duran, D. Smith. "A Continuous Wavelet Transform Algorithm for Peak Detection". Electrophoresis. 2008. 29(20): 4215–4225.
13. 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.
14. Z. Cai, J. Wu. "An Automatic Peak Detection Algorithm for Raman Spectroscopy based on Wavelet Transform". Proc. SPIE 8200. 2011 International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Processing Technology, 82000E (29 November 2011); doi:10.1117/12. 904862.
15. G. Vivó-Truyols, J.R. Torres-Lapasió, A.M. Van Nederkassel, Y. Vander Heyden, D.L. Massart. "Automatic Program for Peak Detection and Deconvolution of Multi-Overlapped Chromatographic Signals: Part I: Peak Detection". J. Chromatogr. A. 2005. 1096(1–2): 133–145.
16. 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.
17. F. Ehrentreich, L. Sümmchen. "Spike Removal and Denoising of Raman Spectra by Wavelet Transform Methods". Anal. Chem. 2001. 73(17): 4364–4373.
18. H.G. Schulze, R.B. Foist, A. Ivanov, R.F. Turner. "Fully Automated High-Performance Signal-to-Noise Ratio Enhancement Based on an Iterative Three-Point Zero-Order Savitzky–Golay Filter". Appl. Spectrosc. 2008. 62(10): 1160–1166.
19. D. Zhang, K.N. Jallad, D. Ben-Amotz. "Stripping of Cosmic Spike Spectral Artifacts Using a New Upper-bound Spectrum Algorithm". Appl. Spectrosc. 2001. 55(11): 1523–1531.
20. 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.