Proceedings of the International MultiConference of Engineers and Computer Scientists 2019
IMECS 2019, March 13-15, 2019, Hong Kong

Effective and Efficient Baseline Correction Algorithm for Raman Spectra

Yung-Sheng Chen and Yu-Ching Hsu *

Abstract—**Baseline shifts caused by a variety of factors depending on the type of spectroscopy are common degradation problems in Raman spectra. An effective and efficient method is presented for the baseline correction of Raman spectra. A kernel smoothing function is developed for the estimation of baseline, and a negative signal filter is designed for the removal of negative impulse responses caused from the spectroscopy. Results on the computer and instrument implementation confirm the feasibility of the proposed method.**

Keywords: Spectroscopy, Raman, Spectrometers, spectroscopic instrumentation

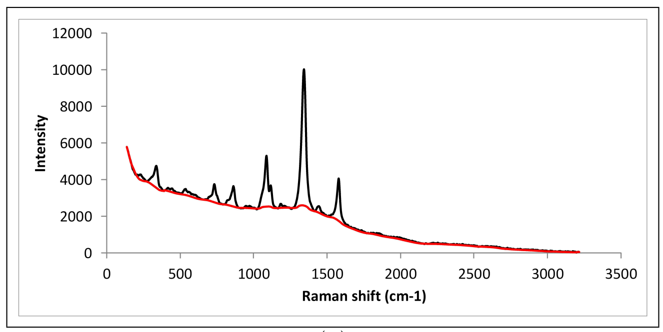
1 Introduction

Raman spectroscopy is a well-established but significant technique for revealing the molecular fingerprints based on its vibrational information. Due to the label-free and non-invasive property as well as convenient application, Raman spectroscopy has been widely applied for material analysis [1], biological investigation [2], medical diagnosis [3], and so on. Baseline is a common degradation problem in Raman spectra and usually results from the spurious background signal or instrument fluctuation. The problem of baseline will severely result in the difficulty of quantitative or qualitative analysis of Raman spectra. To overcome such a problem, baseline correction is a necessary step before performing the Raman spectra analysis and usually classified into three main types of methods, i.e., smoothing spline fitting, wavelet decomposition and integration, and least squares error modelling (or divided into physically and mathematically motivated approaches). The merits and disadvantages of those methods have been investigated and can be found in [1, 2, 4]. In recent, due to the progress of micro Raman spectroscopy, e.g., Micro Raman Identify (MRI) spectrometer [5], users usually need a friendly interface to fast correct the baseline with only a few parameters easily controlled. To achieve such a goal, an effective and efficient baseline correction algorithm is presented in this paper.

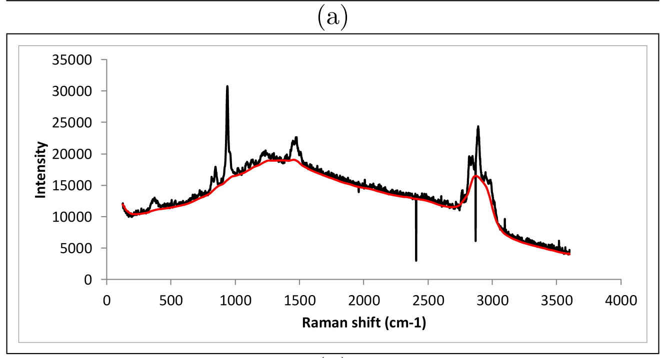
*Y.-S. Chen and Y.-C. Hsu are with the Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan, ROC. (phone: 886-3-4638800; fax: 886-3-4639355; e-mail: eeyschen@saturn.yzu.edu.tw).

2 Proposed Approach

All the samples of Raman spectra used in this study are from the MRI spectrometer [5]. Two samples are illustrated in Fig. 1, where the baseline (red colour) is estimated by the proposed method. Note that the negative impulse responses result from the MRI spectrometer can also be filtered out by the proposed method as illustrated in Fig. 1(b). Their final baseline correction results are shown respectively in Fig. 2. The details of the presented algorithm will be described as follows.



(a)



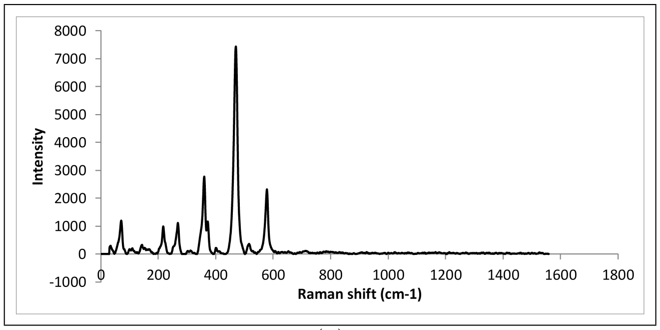
(b)

Figure 1: Illustrations of the measured Raman spectrum and the baseline (red colour) estimated by the proposed algorithm. (a) Sample-1: Measured Raman spectrum including the baseline. (b) Sample-2: Measured Raman spectrum including not only the baseline but also the negative impulse responses.

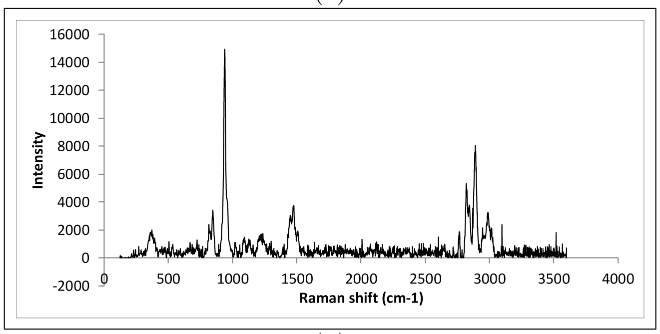
2.1 Baseline removal modeling

Let $f_{\texttt{raw}}$ be the raw observed Raman spectra data having N data points. Generally, the raw data $f_{\texttt{raw}}$ is composed of the true signal (s), the baseline (b), as well as the noise

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(a)



(b)

Figure 2: Results of baseline correction by the proposed algorithm. (a) Baseline correction result for the Raman spectrum given in Fig. 1(a). (b) Baseline correction result for the Raman spectrum given in Fig. 1(b).

(n). It can thus be modelled as

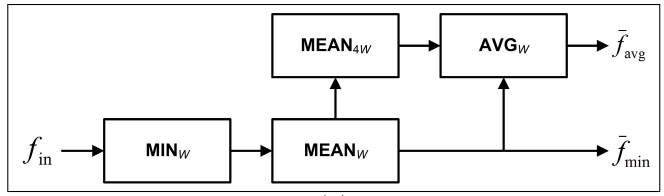
$$f_{\texttt{raw}} = s + b + n \tag{1}$$

Based on the quality of nowadays technology, the noise term n can be ignored since it has a little effect compared to the baseline. Therefore, the signal s can be estimated by subtracting the baseline b from the raw data $f_{\texttt{raw}}$. In other words, the estimation of baseline is the main task for extracting the signal s. Let $f_{\texttt{baseline}}$ be the estimated baseline. Then the correct signal $f_{\texttt{correct}}$ can be expressed as

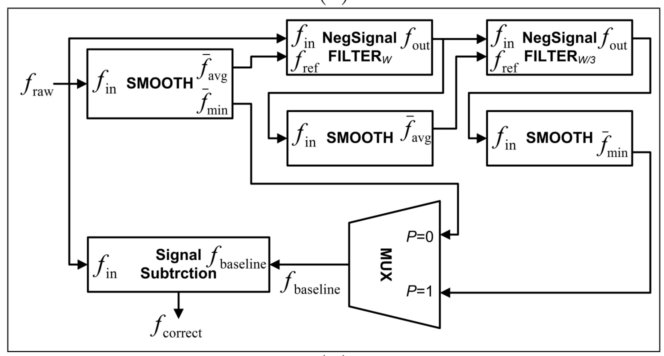
$$f_{\texttt{correct}} = f_{\texttt{raw}} - f_{\texttt{baseline}} \tag{2}$$

2.2 Smooth function

From the characteristic of Raman spectra data, the baseline can be regarded as a stable component from the human visual inspection, which is used to carry the signal. Therefore, the key function of estimating the baseline is first developed in this study for finding the local minima of the raw data and smoothing them as the **SMOOTH** function depicted in Fig. 3(a). Here $f_{\texttt{in}}$ is the input signal, whereas the $\overline{f}_{\texttt{min}}$ and $\overline{f}_{\texttt{avg}}$ are the resultant local minima and averaged information. Assume the i-th data point is to be processed, the sliding window is defined as $[i-W, i+W]$, where W is a positive integer and controls the window size. The functional blocks (**MIN**, **MEAN**, and **AVG**) shown in Fig. 3(a) are formulated respectively as follows.



(a)



(b)

Figure 3: Signal flow of the proposed method. (a) **SMOOTH** function block diagram with **MIN**, **MEAN** and **AVG** operations for finding useful local stationary information. (b) Flowchart of removing the negative impulse responses, extracting the baseline, and obtaining the baseline correction result.

$$\text{MIN}_W(i) = \min_{\forall j \in [i-W, i+W]} f_{\texttt{in}}(j) \tag{3}$$

$$\text{MEAN}_W(i) = \left(\sum_{\forall j \in [i-W, i+W]} \text{MIN}_W(j) \right) / (2W+1) \tag{4}$$

$$\text{MEAN}_{4W}(i) = \left(\sum_{\forall j \in [i-4W, i+4W]} \text{MEAN}_W(j) \right) / (8W+1) \tag{5}$$

$$\text{AVG}_W(i) = \left(\text{MEAN}_W(i) + \text{MEAN}_{4W}(i) \right)/2 \tag{6}$$

Here $\mathbf{MIN}_W$ computes the fundamental local minima for the input signal and after the processing of $\mathbf{MEAN}_W$, the smoothed local minima $\overline{f}_{\texttt{min}}$ can be obtained. However, in our study on the MRI spectrometer [5], some negative impulse responses (possibly resulting from the sensor device or the integrated mechanism) may occur during the signal acquisition as the Sample-2 shown in Fig. 1(b). To filter out such an unwanted negative signal, it is reasonable to provide a more smoother local minima as a reference for comparison. A more wider $\mathbf{MEAN}_{4W}$ function is thus performed on the resultant signal from $\mathbf{MEAN}_W$ and then combined with $\mathbf{MEAN}_W$ by $\mathbf{AVG}_W$ to obtain the useful reference signal $\overline{f}_{\texttt{avg}}$. As long as the mean input signal is less than the reference, it can be replaced by the mean signal, otherwise the original input signal is remained.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

2.3 Negative signal removal

The function block of $(\textbf{NegSignal Filter})_W$ in Fig. 3(b) performs such a negative signal removal, which can be formulated as below.

$$f_{\texttt{out}}(i) = \begin{cases} \overline{f}_{\texttt{ref}}(i), & \overline{f}_{\texttt{in}}(i) < \overline{f}_{\texttt{ref}}(i) \\ f_{\texttt{in}}(i), & \texttt{otherwise} \end{cases} \quad (7)$$

Here $\overline{f}_{\texttt{in}}$ represents the resultant mean signal from the input signal $f_{\texttt{in}}$, and can be computed as the formula in Eq. (4) with W windows size. $\overline{f}_{\texttt{ref}}$ is obtained by the same way. Since the negative impulse responses may be large or small as illustrated in Fig. 1(b), $(\textbf{NegSignal Filter})_W$ cascading (**SMOOTH**) and $(\textbf{NegSignal Filter})_{W/3}$ cascading (**SMOOTH**) are adopted respectively for the removal of large and small negative signal as depicted in Fig. 3(b). It can be regarded as a two-stage negative signal filter. Here the **SMOOTH** extracts continuously the useful $\overline{f}_{\texttt{min}}$ and $\overline{f}_{\texttt{avg}}$ from the filtered signal obtained by the **NegSignal Filter**. In this case, the final baseline is estimated and output $\overline{f}_{\texttt{min}}$ located at the middle right of the signal flow given in Fig. 3(b). Note here that if there is not any negative impulse responses such a two-stage negative signal filter would not be performed, the baseline can be directly output from the $\overline{f}_{\texttt{min}}$ of the first **SMOOTH** function block.

2.4 Baseline correction algorithm

The whole signal flow of the proposed algorithm is depicted in Fig. 3(b). Let P be a binary parameter to control a multiplexer (**MUX**), which outputs the estimate baseline $f_{\texttt{baseline}}$ that comes from the two-stage negative signal filter ($P = 1$) or the first **SMOOTH** function block ($P = 0$). The baseline correction result $f_{\texttt{correct}}$ can be further obtained by the **Signal Subtraction** function block. For the given Raman spectra data shown in Fig. 1, the estimated baselines (red colour) and their baseline correction results shown in Fig. 2 demonstrate the feasibility of the proposed algorithm.

3 Result and evaluation

The algorithm is implemented in Microsoft® Visual Studio® C++ 2013 and run on a laptop with Intel® Core™ 2 Duo 1.8 GHz CPU and 4G RAM. In addition to the baseline correction results for Sample-1 and Sample-2 shown in Fig. 2, the results of other three samples are also given in Fig. 4 for further evaluations. Fig. 4(a) shows our method can estimate effectively the baseline for a background spectrum. Fig. 4(b) shows the result of a material measured under such a background. Fig. 4(c) shows the baseline estimation result for a spectrum with some negative impulse responses corrupted, and its baseline correction result is shown in Fig. 4(d). The performances reported in Table 1 are evaluated by using data size, W, P, as well as execution time (with milliseconds or ms), where parameters W and P can be selected by user. As a result, our experimental results have confirmed the effectiveness and efficiency of the proposed baseline correction algorithm for Raman spectra.

Table 1: Performances of the five samples from the Micro Raman Identify (MRI) spectrometer [5] using the proposed algorithm.

Samples	**data size**	W	P	**execution time (ms)**
Sample-1	1559	15	0	5
Sample-2	1979	15	1	15
Sample-3	1010	15	0	3
Sample-4	1014	15	0	3
Sample-5	1973	31	1	23

4 Conclusion

In the field of Raman spectrometer, it is of great importance to correct the baseline before performing the quantitative and qualitative analysis for a raw Raman spectrum. The presented algorithm can effectively not only correct the baseline but also remove the negative impulse responses. The execution time with milliseconds performs the efficiency of the proposed method. This algorithm has been implemented in the instrument software of Micro Raman Identify (MRI) spectrometer [5] and thus confirms its feasibility in the engineering and laboratory application.

Acknowledgment

This work was supported in part by the Protrustech (Taiwan), which is a professional instrument company, working on Raman spectroscopy, Photoluminescence, Photoreflectance, and so on. In addition, the attendance fee of this conference is supported by the Ministry of Science and Technology, Taiwan, Republic of China, under the grant number MOST 107-2221-E-155-054.

References

[1] M. Koch, C. Suhr, B. Roth, and M. Meinhardt-Wollweber, "Iterative morphological and mollifier-based baseline correction for raman spectra," *Journal of Raman Spectroscopy*, vol. 48, pp. 336–342, 2017.

[2] S. Guo, T. Bocklitz, and J. Popp, "Optimization of raman-spectrum baseline correction in biological application," *Analyst*, vol. 141, pp. 2396–2404, 2016.

[3] P. Kumar, T. Bhattacharjee, M. Pandey, A. Hole, A. Ingle, and C. M. Krishna, "Raman spectroscopy in experimental oral carcinogenesis: investigation of abnormal changes in control tissues," *Journal of Raman Spectroscopy*, vol. 47, pp. 1318–1326, 2016.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

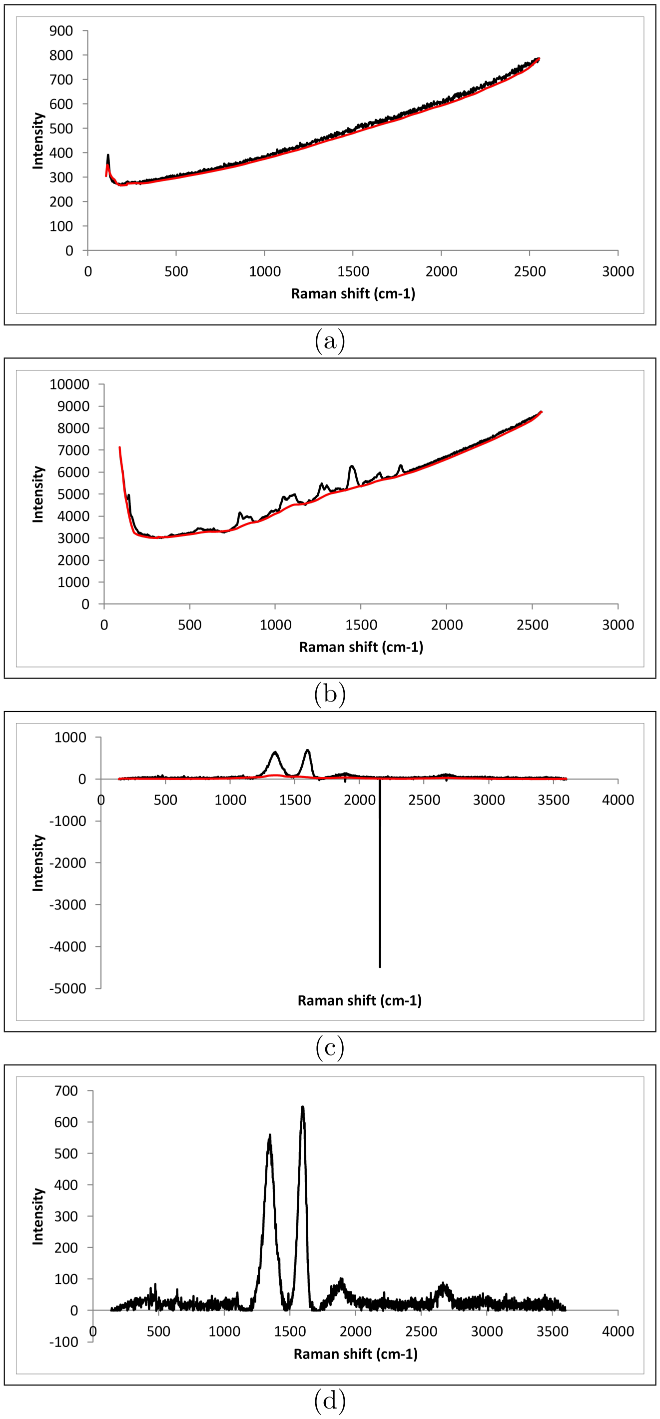


Figure 4: Results of the other samples for baseline correction. (a) The estimated baseline fits well for a background spectrum. (b) Result of a material measured under such a background in (a). (c) The estimated baseline for a spectrum with some negative impulse responses corrupted. (d) The final baseline correction result for the spectrum in (c).

[4] J. Liu, J. Sun, X. Huang, G. Li, and B. Liu, "Goldindec: a novel algorithm for raman spectrum baseline correction," *Applied Spectroscopy*, vol. 69, pp. 834–842, 2015.

[5] Raman Spectroscopy, "Protrustech CO., LTD, Taiwan," http://www.protrustech.com/en/pro_index.php.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)