

基于临近比较的快速拉曼光谱基线校正方法

许英杰^{1*}, 范贤光¹, 林智乐¹, 王昕¹, 左勇²

(1. 厦门大学 航空航天学院, 福建 厦门 361005; 2. 北京长城计量测试技术研究所国防科技工业第一计量测试研究中心, 北京 100095)

摘要: 拉曼光谱成像技术是基于拉曼散射效应所开发的一项现代检测技术, 在现代生产、科学研究过程中使用非常广泛。拉曼光谱信号受荧光效应和仪器等方面的影响, 往往会产生基线漂移, 严重影响对信号特征的进一步提取。因此, 必须对拉曼光谱信号进行基线校正。传统的基线校正方法, 只针对单一光谱信号, 计算量较大, 在处理由大量拉曼信号组成的成像数据时, 耗时较长且效果不佳。该文提出一种基于临近比较的快速基线校正方法, 根据在相同背景下采集的光谱之间的相关性, 实现快速基线校正, 提高了拉曼成像数据的处理速度。

关键词: 拉曼光谱; 基线校正; 临近比较; 快速成像

中图分类号: O657.37 文献标识码: A 文章编号: 1004-4957(2017)08-1047-04

A Fast Baseline Correction Method for Raman Spectroscopy Based on Comparison

XU Ying-jie^{1*}, FAN Xian-guang¹, LIN Zhi-le¹, WANG Xin¹, ZUO Yong²

(1. School of Aerospace Engineering, Xiamen University, Xiamen 361005, China;
2. Changcheng Institute of Metrology & Measurement, Beijing 100095, China)

Abstract: Raman imaging is a kind of modern testing technology based on Raman scattering, which has been widely applied in manufacturing and scientific research. However, due to the fluorescent effect and instruments drift, the baseline shift could easily occur, which has a strong impact on the feature extraction of the Raman signals. Therefore, the baseline correction is necessary and inevitable in the signal processing of Raman spectra. The traditional baseline correction methods can only correct the Raman spectrum one by one, so once a large amount of Raman imaging signals have to be processed, the processing time is too long to accept. In this paper, the baseline correction base on comparison which considers the correlation between the Raman spectra and the same background is proposed to realize the fast baseline correction and improve the processing speed of Raman imaging data.

Key words: Raman spectroscopy; baseline correction; comparison; fast imaging

拉曼光谱(Raman spectroscopy)又称拉曼效应,由印度科学家C. V. Raman发现并命名^[1-2]。拉曼光谱是一种基于拉曼散射效应的散射光谱。在早期发展阶段,拉曼光谱采用聚焦的太阳光、汞弧灯等作为光源,限制了其发展。上世纪60年代随着激光技术的兴起,以及光电信号转换器件的发展给拉曼光谱带来新的转机。受益于激光器、单色仪和弱光信号探测技术的不断进步,如今的拉曼光谱不仅能够借助于特征拉曼频率分辨微量混合物之中的各种化学成分信息,而且可以给出其中各种成分的空间分布信息,其空间分辨率已经接近光的衍射极限,这就是拉曼成像技术^[3-6]。通过设定测量范围、步长以及测量条件后,由软件控制在样品表面一定范围内,逐点采集光谱,然后利用拉曼特征峰的强度变化,构建出该特征峰对应化学成分在样品上的空间分布图,即为拉曼光谱图像。然而,仪器所得到的拉曼光谱并不是理想拉曼光谱,无法直接提取特征峰信息。其原因是,受荧光效应和仪器等方面的影响,所测得的拉曼光谱存在不同程度的基线漂移^[7-8]。因此,为消除基线漂移对拉曼成像的影响,必须对光谱数据进行基线校正。通过基线校正,拉曼光谱数据特征峰的提取和被测物质的识别将变得更加明晰。现有的基线校正方法通常只考虑单一光谱信号的特点,通过循环迭代的方法逼近基线^[9-13]。

收稿日期: 2017-04-17; 修回日期: 2017-06-22

* 通讯作者: 许英杰, 工程师, 研究方向: 测试技术与仪器, Tel: 0592-2182687, E-mail: xuyingjie@xmu.edu.cn

然而,在处理拉曼成像光谱信号时,大量相似的拉曼信号在较短时间内被检测出,如果仍采用传统方法,重复性的计算会极大地浪费校正时间。

针对上述问题,本文提出基于临近比较的基线校正方法,实现对大量拉曼光谱信号的快速基线校正。考虑到对同一样品成像时,各个拉曼光谱信号之间的基线漂移程度和类型相近,因此本方法利用光谱漂移之间的相关性进行基线校正,从而节省校正时的运算时间,极大提高基线校正的效率,进而提高整个样品的成像速度。

1 算法原理

常规基线校正策略主要是对多组光谱信号中的每一条基线分别进行基线校正,未考虑基线之间的相关性,本文提出一种快速基线校正算法,主要流程为读取一批光谱数据,计算其平均光谱,然后使用传统基线校正方法对平均拉曼光谱进行基线校正。进而利用各基线之间的相关性,通过对该基线进行缩放、平移、旋转等操作,获得每个光谱信号的基线。算法流程如图 1 所示。

具体原理如下:

(1) 获得所有拉曼光谱的平均谱 S_0 , 然后利用传统方法对其进行基线校正, 获得平均基线 B_0 。本文采用分段线性拟合方法^[14]。

(2) 计算每个光谱 S_i 与平均光谱 S_0 的比例系数 $k_i (i=1, 2, \dots, n)$, 如式 1 所示:

$$k_i = \frac{S_i \times \sqrt{\frac{1}{n} \sum_{j=1}^n (a_{ji} - \bar{a}_i)^2}}{S_0 \times \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{a}_0)^2}} \quad (1)$$

其中, a_{ji} 表示 S_i 中第 j 个数据, \bar{a}_i 表示这一列的列向量平均值, a_i 表示 S_0 中第 i 个数据, \bar{a}_0 表示 S_0 的列向量平均值。利用式(1)分别计算每一组光谱与平均光谱的比例系数 k_i , 然后使用系数 k_i 对 B_0 进行比例缩放处理, 从而获得 B_i 。

(3) 对每条基线 B_i 进行平移, 计算 B_i 与光谱 S_i 间的距离最小值 d_{\min} , 移动 B_i 使其与光谱 S_i 有 1 个相切点 (a, b) 。

$$d_{\min} = \min\{|a_{ji} - b_{ji}|\} \quad (2)$$

其中, b_{j_0} 为基线 B_i 的第 j 个点。

(4) 以相切点 (a, b) 为旋转中心, 左右两边分别进行旋转调整, 旋转一定角度直至左右两边基线与原始拉曼光谱曲线的距离最小值 $d_{\min} = 0$, 从而获得该条光谱的基线。

$$\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x(t) - a \\ y(t) - b \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix} \quad (3)$$

其中, $y(t), x(t)$ 表示当前基线 B_i 上的每个数据的横纵坐标, θ 表示旋转的角度, $u(t), v(t)$ 分别表示旋转后新基线上每一点的横纵坐标。

(5) 分别从所有光谱信号 S_i 中减去上一步所获得的基线 B_i , 实现对所有信号的基线校正。

2 仿真分析

为了验证本算法的可行性, 首先对其进行仿真验证。仿真所用的拉曼光谱采用多个高斯函数^[15]模拟拉曼特征峰。仿真信号如图 2 曲线 a 所示, 曲线 b 为基线校正结果。

利用仿真程序生成 100 组基线形状相近, 拉曼信号各不相同的仿真信号, 利用本文方法对其进行

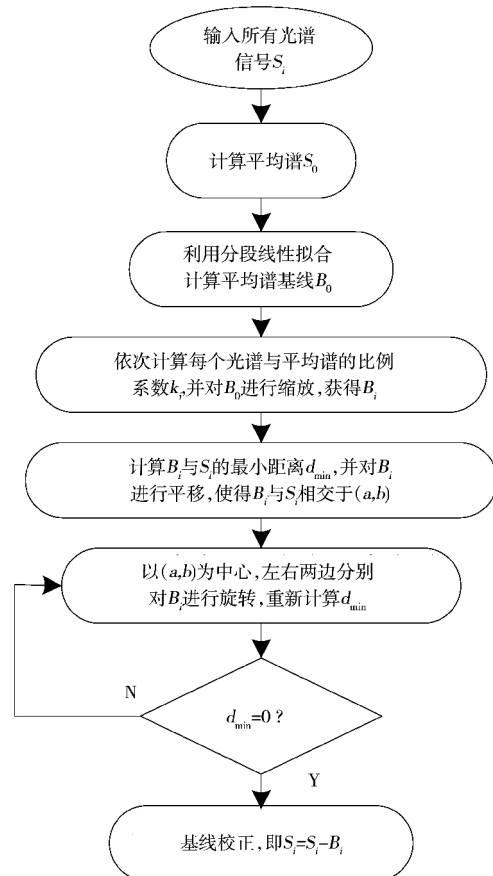


图 1 基于临近比较的快速拉曼光谱基线校正方法流程图
Fig. 1 Flow chart of the fast baseline correction method based on comparison

快速基线校正, 并计算每一组通过快速基线校正算法得到的基线与仿真程序添加的基线之间的均方误差 (Mean square error, MSE), 其结果如图 3 所示。MSE 的计算公式如式 (4) 所示。

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (b_{i,t} - b_{i_0,t})^2 \quad (4)$$

其中, $b_{i,t}$ 和 $b_{i_0,t}$ 分别为本文算法得到的基线和给定基线的第 t 个点。

从图 3 可看出, 由于上述 100 组仿真光谱均由 3 个高斯函数叠加而成, 而高斯函数的向量参数由范围内的随机数决定, 所以有几组的 MSE 值略大。但从散点图可以直观看出, 所有仿真结果的均方误差最大值为 0.56, 最小值为 0.13, 平均值在 0.2 左右, 说明本文提出的快速基线校正方法在处理仿真拉曼光谱信号时, 能够很好地完成拟合, 并保证较小的 MSE 值。

3 实验

在仿真分析的基础上, 为进一步验证本方法的可行性, 对真实实验数据 (泰诺和替硝唑) 的拉曼光谱进行基线校正。实验设备为日本 NANO-PHOTON 公司的高速高分辨率拉曼成像 RAMAN-11 系统。进行数据处理的计算机配置参数为: Intel (R) Core (TM) i5-2450M CPU @ 2.5GHz, 装载系统 Windows 7 旗舰版 64 位。

为显示清晰, 两种药品均只采用 11 组数据对算法进行验证 (图 4)。在算法用时方面, 对 11 组泰诺的实验光谱数据运行快速基线校正方法在上述设备环境下用时 0.080 015 s, 而运行多项式拟合算法用时 0.465 975 s。对 11 组替硝唑的实验光谱数据运行快速基线校正方法用时 0.001 664 s, 而运行多项式拟合算法用时 0.390 645 s。从需要运行的时间对比可以看出, 对 11 组实验数据进行基线校正时, 现有方法花费时间较多, 而快速基线校正方法可以大量节省时间, 实现快速校正的效果。实际进行拉曼成像时, 拉曼光谱信号的组数远高于 11 组 (取决于成像范围, 大多数拉曼成像图所包含的拉曼信号组数均在 2 000 以上), 显著节省了校正时间。

在拟合效果方面, 本方法由于并未对每个光谱进行循环迭代, 通过平移旋转获得的基线很难保证在每个数据点上均小于原始光谱, 因此扣除基线后可能出现负数。但负数一般位于无拉曼谱峰的位置, 因此不会影响拉曼光谱的波形, 只需将负数直接置零即可获得与传统方法相近的校正结果。

4 结论

本文提出了一种基于临近比较的快速光谱基线校正方法, 针对同一拉曼成像图不同组拉曼信号之间基线相近的特点, 提出通过临近比较的方法实现快速基线校正。首先, 用传统方法计算平均谱的基线, 并以此为基准, 通过比例缩放、平移和旋转操作, 分别获得每个拉曼信号的基线。在仿真和实验中可看出, 本文方法能够很好地实现大量拉曼信号的快速基线校正, 可在大量节省校正时间的前提下, 得到与传统方法相当的校正效果。

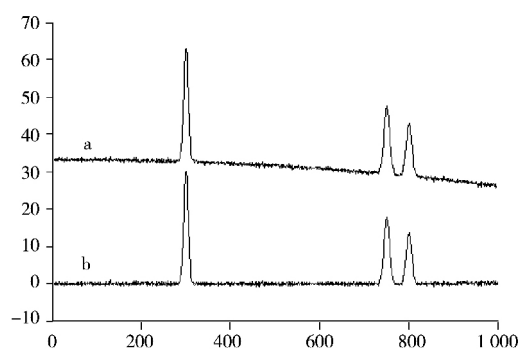


图 2 仿真拉曼光谱信号及其校正结果
Fig. 2 Simulate Raman signal (a) and the baseline correction result (b)

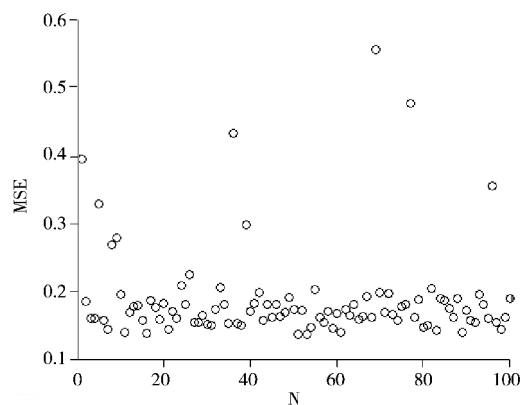


图 3 100 组仿真数据基线校正 MSE 分布图
Fig. 3 MSE for the baseline correction results of the 100 simulated Raman data

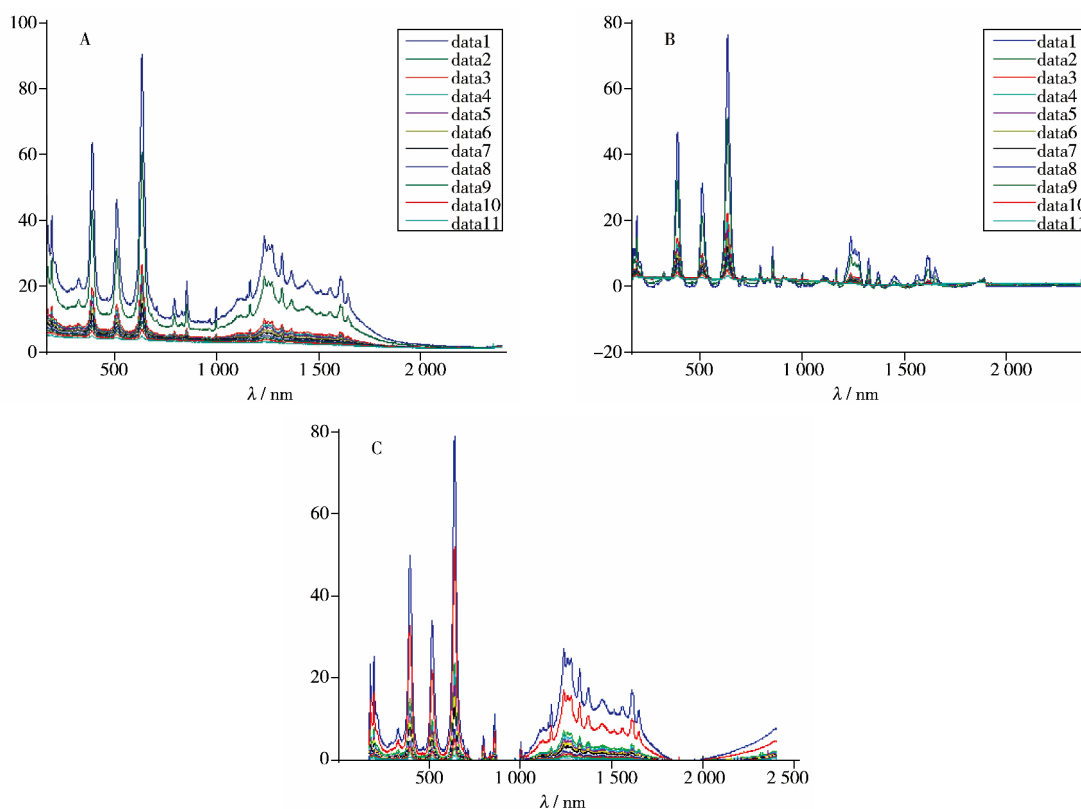


图 4 泰诺的拉曼光谱信号(A)及本文方法(B)和传统多项式拟合(C)基线校正结果

Fig. 4 Raman signals of tylenol (A) and the correction results of the proposed method (B) and polynomial fitting (C)

参考文献:

- [1] Chalmers J M, Griffiths Peter R. *Handbook of Vibrational Spectroscopy (vol. 1-5)*. Wiley: Netherlands, **2003**.
- [2] Wu Z K. *Spectro. Spectral Anal.* (吴征铠. 光谱学与光谱分析), **1983**, 3(2): 65-71.
- [3] Tian G H, Chen Y J, Feng Q M. *Chem. Eng.* (田国辉, 陈亚杰, 冯清茂. 化学工程师), **2008**, 22(1): 34-36.
- [4] Chuan K H, Hiro H, Shinsuke S. *Chem. Commun.*, **2011**, 33(47): 9423-9425.
- [5] Ismail H B, Havva T T, Hüseyin E G. *RSC Adv.*, **2015**, 70(5): 56606-56624.
- [6] Wang L, Guo S X, Dai Y Z, Yang L B, Liu G K. *Chin. J. Anal. Chem.* (王磊, 郭淑霞, 戴吟臻, 杨良保, 刘国坤. 分析化学), **2015**, 43(1): 33-39.
- [7] McCreer R L. *Raman Spectroscopy for Chemical Analysis*. Wiley-Interscience: Columbus, **2000**.
- [8] Feng X W, Zhu Z L, Shen M J. *Comput. Appl. Chem.*, **2009**, 26(6): 759-762.
- [9] Jagtiani A V, Sawant R, Carletta J. *Meas. Sci. Technol.*, **2008**, 19(6): 1-15.
- [10] O'Grady A, Dennis A C, Denvir D. *Anal. Chem.*, **2001**, 73: 2058-2065.
- [11] Ehrentreich F, Summchen L. *Anal. Chem.*, **2001**, 17(73): 4364-4373.
- [12] Wang X, Fan X G, He H. *Meas. Sci. Technol.*, **2015**, 26(11): 115503.
- [13] Feng X W, Zhu Z L, Shen M J, Cong P S. *Comput. Appl. Chem.* (冯昕韡, 朱仲良, 沈梦洁, 丛培盛. 计算机与应用化学), **2009**, 26(6): 759-762.
- [14] Qin Z J, Tao Z H, Liu J X, Wang G W. *Spectro. Spectral Anal.* (覃赵军, 陶站华, 刘军贤, 王桂文. 光谱学与光谱分析), **2013**, 2: 383-386.
- [15] Li J R, Dai L K. *Spectro. Spectral Anal.* (李津蓉, 戴连奎. 光谱学与光谱分析), **2012**, 32(3): 594-597.