

基于形状相似性比较法消除拉曼光谱的复杂背景噪声

夏良平^{1,2} 李华栋^{1,3} 尹韶云¹ 吴云峰³ 史浩飞¹ 杜春雷¹

¹ 中国科学院重庆绿色智能技术研究院微纳制造与系统集成中心, 重庆 404100
² 中国科学院光电技术研究所, 四川 成都 610209
³ 电子科技大学能源科学与工程学院, 四川 成都 610054

摘要 受到复杂环境中多种背景噪声的干扰,拉曼信号的有效信息易被噪声削弱甚至淹没。利用拉曼信息光谱与背景光谱的形状相似性,通过形状相似性比较法去除背景噪声,结合实验数据介绍该方法的去噪原理与具体步骤。对比发现,该方法所恢复的拉曼信息量大,基线平整,便于后续数据处理。

关键词 光谱学;光谱预处理;形状相似性比较;噪声去除

中图分类号 O433 文献标识码 A doi: 10.3788/AOS201333.0530003

Eliminating Complex Background Noise of Raman Spectrum Based on Configuration Similarity Comparing Method

Xia Liangping^{1,2} Li Huadong^{1,3} Yin Shaoyun¹ Wu Yunfeng³ Shi Haofei¹ Du Chunlei¹

¹ Micro-Nano Manufacturing and System Integration Center, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 404100, China

² Institute of Optical and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China

³ School of Energy Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China

Abstract The effective information of Raman spectrum is easy to be weakened or even submerged by the background noises in complex environment. The configuration similarity comparing method (CSCM), which is based on the shape similarity of the information and the background spectrum, is proposed to eliminate the background noises. Exemplified as experimental data, the CSCM principle is analyzed and the processing steps are introduced. By comparing with other two traditional methods, the CSCM recovers more Raman information and the baseline is flat, which is advantageous in further data analysis.

Key words spectroscopy; spectra preprocessing; configuration comparing; denoising

OCIS codes 140.3550; 300.6450; 290.5860

1 引 言

拉曼光谱是物质化学键或化学基团振动、转动等引起的具有指纹特性的散射谱,特异性强、可无标记、实时地获取物质分子结构信息,在食品安全、污染检测和生物医学等领域具有广泛的应用前景^[1~3]。然而,由于拉曼散射较弱,在实际测试中物质的拉曼光谱信号往往容易被机器噪声、强烈的荧

光和磷光等背景噪声所干扰^[4]。若测试对象处于多种物质的复杂体系中,则被测物质的拉曼光谱也极易与其他物质的拉曼光谱交叠甚至被覆盖。因此为了准确获得被测物质的拉曼信息,需要对实际测试的拉曼信号进行有效的光谱预处理,以去除无关信息的干扰。

常见的去除拉曼噪声干扰的光谱预处理方法可

收稿日期: 2012-12-25; 收到修改稿日期: 2013-01-24

基金项目: 中国科学院西部之光(A12K010)和重庆市科技攻关项目(cstc2012ggc0023)资助课题。

作者简介: 夏良平(1986—),男,博士研究生,主要从事表面增强拉曼传感方面的研究。E-mail: xialp@cigit.ac.cn

导师简介: 杜春雷(1962—),女,博士,研究员,博士生导师,主要从事微纳光学和表面等离子体光学等方面的研究。

E-mail: cldu@cigit.ac.cn(通信联系人,中国光学学会会员号: 463281)

以分为有背景参考和无背景参考两种情况。其中无背景噪声参考的光谱预处理方法包括:1)数字平滑滤波,常用的数字平滑滤波方法有窗口移动平均法和窗口移动二项式平滑法,可以去除因随机误差引起的信号偏离,提高信噪比,但该方法容易造成信号细节丢失^[5],这对于峰值尖锐的拉曼光谱而言是致命的;2)多项式拟合校正,该方法常用于对由荧光等引起的基线漂移进行校正,以便于后续的数据处理,但当被测物的拉曼光谱强度较弱时,与荧光光谱叠加后的信息光谱曲线与荧光背景曲线差别很小,多项式拟合容易丢失微弱的拉曼信息^[6];3)傅里叶变换(FT)与 Hadamard 变换(HT)滤波,FT 可过滤随机噪声等高频信号,HT 则可去除荧光这样的低频信号干扰,但该方法要求信号与噪声要有明显的频率差异^[7];4)小波变换,该方法通过多分辨率的特性来有效提取信号,前提是要知道噪声与信号的频率特性,再根据特性找到合适的分解方法^[8~10]。以上光谱预处理方法各有特点,但均有其局限性,对于拉曼光谱而言,在面对复杂环境背景噪声时,如在复杂基质中检测某种物质时,其他物质的拉曼光谱对于希望被探测到的物质来说也是一种背景噪声,而以上方法将难以去除这类噪声。

为了有效去除复杂环境中的背景噪声,需要借助背景拉曼光谱作为参考。在有背景参考情况下,光谱预处理方法包括直接差谱法和自适应噪声对消法。直接差谱法是将含有有效信息的光谱(简称信息光谱)直接减去背景光谱。然而,由于环境温度变化、基底增强性质不稳定等因素对拉曼光谱影响较大,实验中背景光谱和信息光谱的强度往往存在较大波动,此时差谱法只能去除部分背景噪声,且基线很不稳定,不便于后续数据处理。对于不稳定的背景噪声,自适应噪声对消法具有较好的去噪效果,但该方法要求背景噪声与有效信号的相关性不大且其均值接近零^[11],拉曼光谱与这一要求相差甚远,因此其去噪效果并不理想。

通过大量实验观察发现拉曼光谱虽然存在较大的强度波动,但其光谱曲线形状变化较小,结合拉曼光谱峰的尖锐性特征,本文基于背景拉曼光谱与信息拉曼光谱的形状相似性,提出形状比较法(CSCM)消除复杂环境下的拉曼背景噪声。

2 方法原理

拉曼光谱的测试数据为经过采样的离散数组,对于复杂的背景而言,其背景噪声包含多个部分,各

部分噪声由不同的机制产生,如仪器的随机噪声、环境温度变化引起的噪声、荧光噪声、背景物质的拉曼光谱噪声等,彼此相关性不大,因此,背景拉曼光谱可表示为

$$L(\omega) = g(\omega) + \sum_n f_i(\omega) + c_0, \quad (1)$$

式中 $g(\omega)$ 为随频率变化的非拉曼光谱的背景噪声总和,如荧光、磷光等背景噪声, $f_i(\omega)$ 为背景物质的拉曼光谱, c_0 为随机噪声和环境温度、激光功率变化等引起的拉曼强度不稳定。

图 1 中虚线给出了一个实验测试的含有大量物质的拉曼增强基底(由金属纳米结构组成,当被测物置于其上时,拉曼信号可大幅增强)的拉曼背景光谱,这 4 条曲线分别在不同时间测试同一拉曼基底不同位置获得。从这 4 条曲线可以看出,其背景拉曼光谱的曲线形状变化不大,几乎是相似的,这说明随频率变化的非拉曼光谱背景噪声 $g(\omega)$ 和背景物质的拉曼光谱噪声 $f_i(\omega)$ 与频率的相关性较大,其函数关系基本稳定,但随机噪声和环境不稳定噪声 c_0 则存在明显变化。图 1 中的实线表示在该拉曼增强基底上加入一定浓度的罗丹明 6G(R6G)溶液后所测得的信号拉曼光谱,通过该曲线与背景拉曼光谱的对比不难发现,所测得的信号拉曼光谱在曲线形状上与背景光谱整体呈相似性,只在被测物质的拉曼峰的位置存在明显不同,相当于被测物的拉曼光谱“长”在了背景光谱上。因此,可将包含被测物信息的拉曼光谱表示为

$$S(\omega) = [1 + \alpha(\omega)]L(\omega) + I(\omega) + c_1, \quad (2)$$

式中 $\alpha(\omega)$ 表示被测物拉曼光谱与背景光谱叠加的不稳定因子, $I(\omega)$ 表示被测物真实的拉曼光谱, c_1

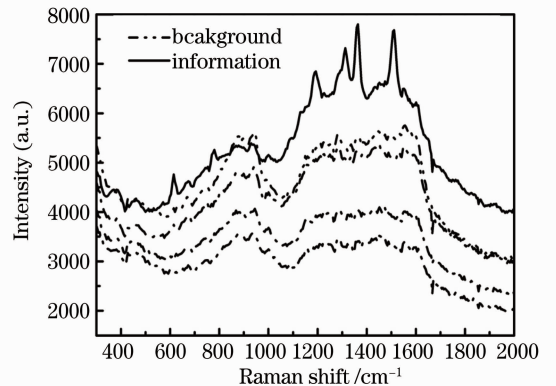


图 1 实验测得的复杂背景光谱和以 R6G 为被测物的信息拉曼光谱

Fig. 1 Experimental background spectrum of complex environment and the information Raman spectra of R6G

表示拉曼光谱强度的不稳定性,与频率无关,可通过对(2)式进行微分去掉该项。因此,将包含被测物质的信息光谱与背景光谱分别作微分后相减得到系数分光谱差值

$$C(\omega) = S(\omega)' - L(\omega)' = \alpha(\omega)L(\omega)' + \alpha(\omega)'L(\omega) + I(\omega)', \quad (3)$$

式中只有 $I(\omega)'$ 包含了有效的拉曼光谱信息,其余两项均为噪声,需要去除。由于物质的拉曼光谱峰是由物质分子键的振动、转动等所产生的能级跃迁,一般比较尖锐,散射强度在拉曼峰附近的变化非常剧烈,即在低于拉曼峰值频率的区间急剧上升,越过拉曼峰值点后急剧下降,因此(3)式中 $I(\omega)'$ 项的绝对值较大,且在拉曼峰值位置左右两侧分别呈一正一负分布规律。

根据图 1 中虚实线的比较,由于信号拉曼光谱在曲线形状上与背景光谱整体呈相似性,光谱叠加的不稳定因子 $\alpha(\omega)$ 较小,所以(3)式的第一项 $\alpha(\omega)L(\omega)'$ 的值相应较小,可通过设置合理的阈值将该项噪声去除,在去掉该项的同时也去掉了噪声项 $\alpha(\omega)'L(\omega)$ 中绝对值较小的部分。

对于 $\alpha(\omega)'L(\omega)$ 项绝对值较大的部分,由于 $\alpha(\omega)$ 与 $L(\omega)$ 均具有较大的随机性,这导致 $\alpha(\omega)'L(\omega)$ 的分布无规律可循,与 $I(\omega)'$ 项中呈左正右负的分布存在明显差异。因此,在阈值处理后的数据中,去掉不满足左正右负分布规律的部分,剩下的

部分均满足拉曼光谱分布特点,可近似认为是拉曼信息项 $I(\omega)'$ 。再对该项进行积分还原,即可获得所测物质的拉曼光谱。

3 光谱预处理步骤

结合具体的数据处理过程对所提出的基于形状相似性比较进行复杂环境下拉曼光谱预处理的操作步骤进行说明。

1) 光谱微分与差值

将图 1 中虚线所示的 4 条复杂环境下测得的背景光谱进行平均,并表示为 $L(\omega)$,实线所示的包含被测物质拉曼信息的数据标记为 $S(\omega)$,分别对 $L(\omega)$ 和 $S(\omega)$ 进行一阶微分,再利用(3)式获得如图 2(a) 实线所示的微分光谱的差值 $C(\omega)$ 。

2) 阈值设置与分段

根据第 2 节的分析可知,物质拉曼光谱一阶微分的绝对值较大,因此通过阈值比较可去除 $C(\omega)$ 中绝对值较小的背景噪声。选择 $C(\omega)$ 中没有较大绝对值且变化较为平缓的一段,如图 2(a) 中矩形方框部分所示,取其标准差 σ , σ 代表了 $C(\omega)$ 中噪声起伏的水平,因此选择阈值为 σ 。然后对 $C(\omega)$ 进行分段,将其绝对值小于阈值 σ 的部分标记为 0,值大于 σ 的部分标记为 1,值小于 $-\sigma$ 的部分标记为 -1,分段后的结果如图 2(a) 中的虚线所示。

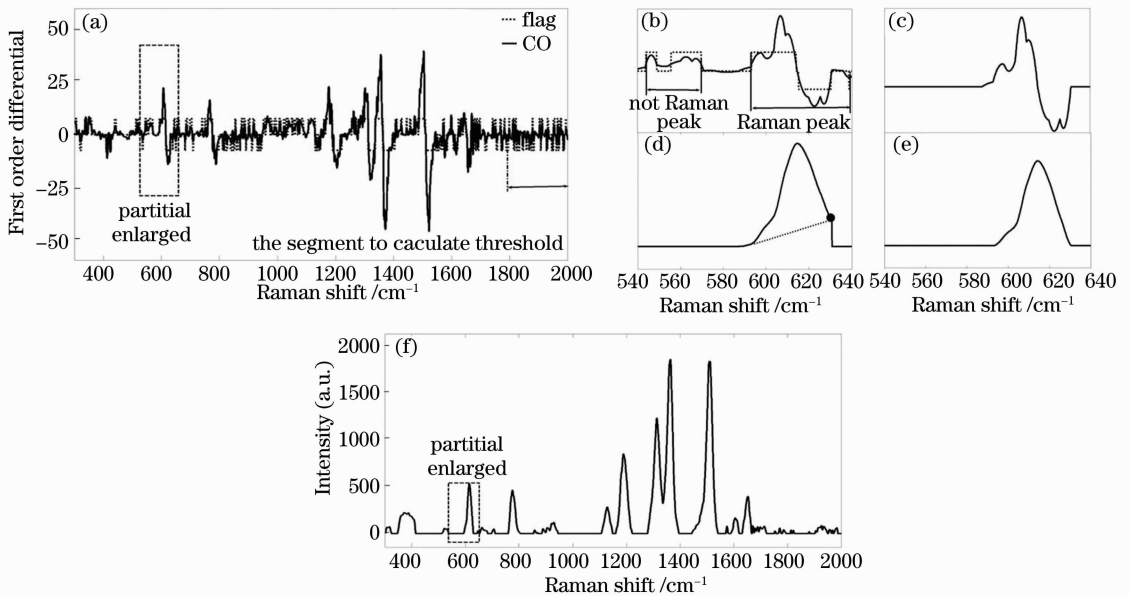


图 2 基于 CSCM 的拉曼噪声去除步骤

Fig. 2 Raman noise removing steps of CSCM

3) 噪声信息去除

根据第 2 节的分析,噪声信息的去除包含两部分,先去除 $C(\omega)$ 中绝对值较小噪声[即步骤 2) 中分段后标记为 0 的部分]。再利用 $C(\omega)$ 中绝对值较大的噪声部分拉曼光谱峰形状的特殊性去其进行去除。由于拉曼光谱峰的一阶微分严格满足左正右负分布规律,即如图 2(b) 所示,此时标记值满足 $[1, 0, -1]$ 的变化规律。因此先根据分段标记值查找所有不满足以上规律的光谱段,并将其内的 $C(\omega)$ 值置为 0,同时恢复标记值满足 $[1, 0, -1]$ 变化规律的段内的 $C(\omega)$ 值,通过以上处理即去掉了 $C(\omega)$ 中绝对值较大的噪声部分,所剩下值的包含有效拉曼信息的其中一段如图 2(c) 所示。

4) 积分还原与光谱修正

根据步骤 3) 处理后的微分数据,分别对包含有效拉曼信息的段进行分段积分,以恢复有效拉曼光谱,所获得的结果如图 2(d) 所示,该结果显示所恢复的光谱曲线存在突变点,因此需对所恢复的光谱进行修正。这里通过线性修正方式进行处理,即通过对突变点引入如图 2(d) 虚线所示的线段,再将光谱以该线段为基线整体下拉,即获得如图 2(e) 所示

的修正结果。最终经修正后获得的完整的信息拉曼光谱如图 2(f) 所示,该结果与图 1 中的原始光谱比较不难发现,预处理后的拉曼峰异常尖锐,特征明显,与原始光谱相比,绝大部分噪声已去除。

4 结果对比与讨论

为了验证基于 CSCM 的去噪方法与其他拉曼光谱预处理方法在处理复杂环境背景噪声中的效果,下面将本方法与另外两种常见的方法进行了对比。

图 3(a) 为一 R6G (购自 Sigma-Aldrich, 纯度 99%) 固体(作为 R6G 的标准拉曼光谱)和溶液在 K9 玻璃上的拉曼光谱测试结果(测试仪器: Ocean Optics 便携式拉曼光谱仪,激发波长为 785 nm,光谱分辨率为 13 cm^{-1}),固体的测试结果显示该物质具有多个特征明显的拉曼峰,而 K9 玻璃上样品的测试结果显示其背景拉曼峰与被测物质拉曼峰发生大范围重叠,且受环境温度等因素的影响,拉曼光谱强度不稳定,被测物质有效的拉曼信息几乎被背景所淹没,很难直接根据所测试的结果分辨出被测的物质。

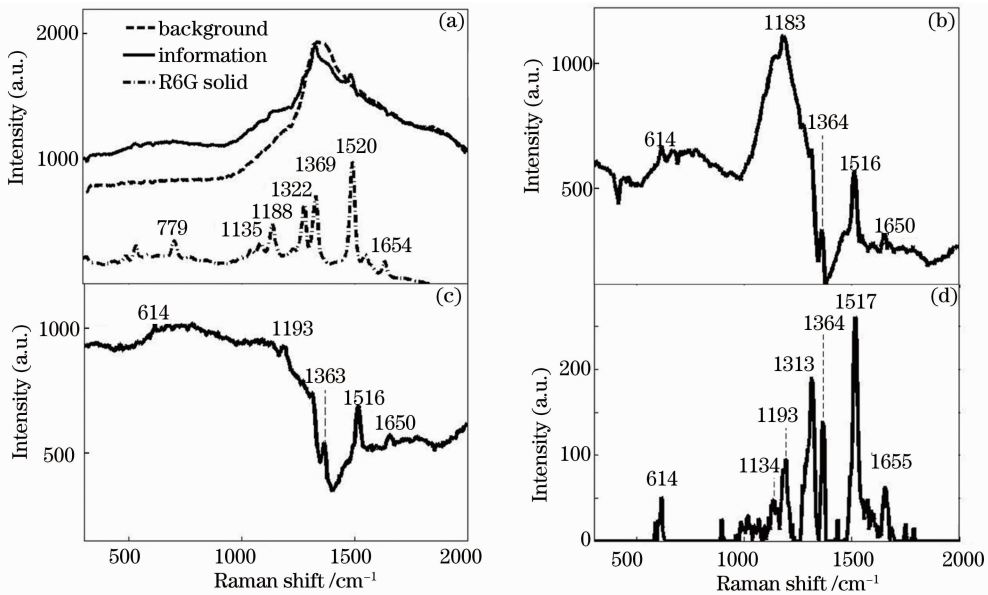


图 3 (a) 原始的背景光谱、信息拉曼光谱和 R6G 固体的拉曼谱; (b) 差谱法处理结果; (c) 自适应法处理结果; (d) CSCM 处理结果

Fig. 3 Spectra of (a) original background, information and R6G solid; obtained results of (b) subtraction method, (c) adaptive noise cancellation method; (d) proposed CSCM

采用差谱法(拉曼光谱与背景光谱直接相减)和自适应噪声对消的方法^[12]得到的拉曼光谱分别如图 3(b) 和图 3(c) 所示。两种处理方法所得到的结果只能观察到 5 个被测物的拉曼峰,且图 3(b) 中

614 cm^{-1} , 1183 cm^{-1} 的峰值不明显,与周围噪声的区别不大。图 3(c) 中 1193 cm^{-1} 处,虽然处理后的强度值较大,但其峰值宽度变宽,与 R6G 在本身的拉曼峰形状差异较大,因此容易被处理为低频噪声。

另一方面,两种预处理方法获得的光谱曲线的基线变化很大,不利于拉曼光谱数据的处理。

采用 CSCM 处理后的拉曼光谱如图 3(d)所示,与图 3(a)中固体的测试结果相比,该结果清晰地显示了 R6G 的拉曼峰值,其中有 7 个峰值的强度明显高于平均噪声,几乎涵盖了该物质的拉曼峰,且峰型尖锐,易于识别。另一方面由于采用分段积分恢复被测物的拉曼光谱,因此其基线为 0,这给光谱数据的后续处理带来了很大的方便。

在此基础上,利用 CSCM 进行了复杂背景下拉曼光谱的预处理。以滤纸加石英构成复杂背景,以丙酮和乙醇为测试试剂,以图 3 的实验仪器和参数为测试条件,获得的滤纸加石英的实验拉曼光谱如图 4(a)和图 4(b)中虚线代表的背景光谱所示,在滤纸加石英背景上分别滴入丙酮和乙醇后的拉曼光谱

如图 4(a)和图 4(b)中实线代表的的信息光谱所示。这两幅图显示背景具有较复杂的拉曼光谱,光谱范围宽,具有多个特征明显的峰值,这对测试试剂的识别与定量分析是非常不利的。对比信息光谱与背景光谱,发现其整体形状呈相似性,满足 CSCM 的适用条件,丙酮与乙醇样品的噪声去除结果分别如图 4(c)和图 4(d)所示,虚线表示以相同实验条件下测得的纯丙酮与乙醇的拉曼光谱,实线表示对图 4(a)和图 4(b)两图的处理结果,该结果显示,除了部分高频噪声,背景几乎去除,处理后的光谱与纯样品的光谱相比,对于丙酮样品所有拉曼特征峰均已恢复,且峰型、峰宽等特征一致,对于乙醇样品,除了 1100 cm^{-1} 处的拉曼峰与在形状上与背景光谱一致而无法恢复外,其余特征峰均已恢复。

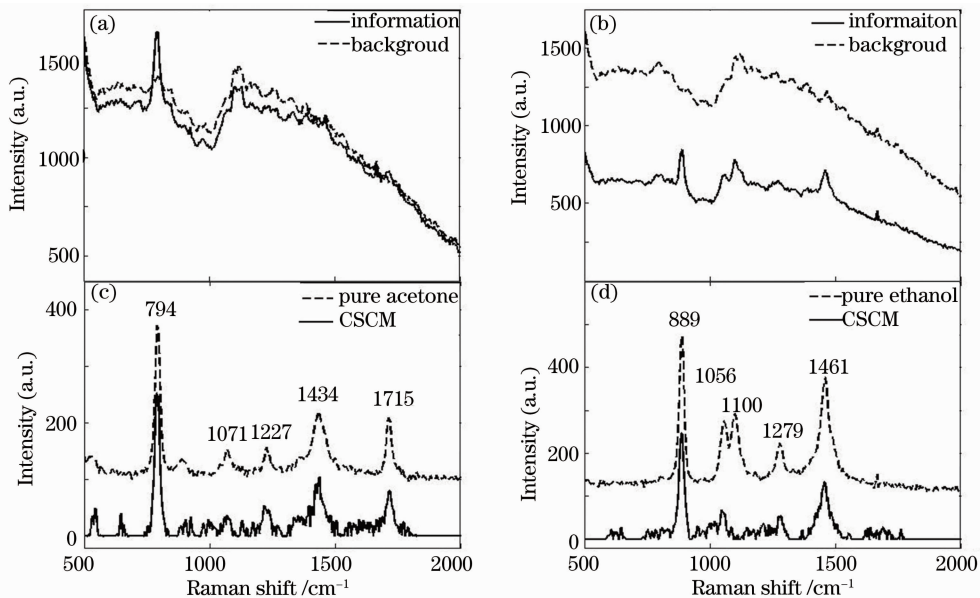


图 4 (a)、(b)复杂背景下的实验拉曼光谱; (c)、(d) CSCM 处理结果

Fig. 4 (a), (b) Experimental Raman spectra of complex background; (c), (d) CSCM result

通过以上对比不难发现,该方法对于复杂背景具有较好的光谱预处理结果,其局限性在于需先测定背景光谱,且需背景光谱的形状基本稳定。

5 结 论

提出利用拉曼信息光谱与背景光谱的形状相似性,通过微分比较的方法去除复杂环境的拉曼背景噪声,通过理论推导分析了该方法去除复杂背景噪声的原理,结合实验数据介绍了以该方法进行拉曼光谱预处理的具体步骤,将其与其他拉曼光谱预处理方法进行了比较,得出该方法在处理复杂环境背景噪声方面具有明显优势,所恢复的拉曼光谱特征

明显、提取的有效信息量大且基线平整,便于后续的数据处理与分析。

参 考 文 献

- 1 Yang Xugang, Wu Qilin. Raman Spectroscopy Analysis and Application[M]. Beijing: National Defense Industry Press, 2008. 1~3
- 2 杨序纲, 吴琪琳. 拉曼光谱学的分析与应用[M]. 北京: 国防工业出版社, 2008; 1~3
- 3 Chen Yang, Wu Songhua, Liu Zhishen. Novel method to measure sensitivity of doppler lidar using rotational Raman spectrum[J]. *Acta Optica Sinica*, 2011, **31**(s1): s100508
- 4 陈 阳, 吴松华, 刘智深. 基于转动拉曼光谱的多普勒激光雷达灵敏度测量新方法[J]. *光学学报*, 2011, **31**(s1): s100508
- 5 He Min, Dai Ye, Yu Guangjun *et al.*. Silver nanoflower fabricated by femtosecond laser pulses and its SERS effect[J].

- Chinese J. Lasers*, 2011, **38**(s1): s106005
- 何敏, 戴晔, 余光军等. 飞秒激光制备的银纳米花结构及其表面增强拉曼散射效应[J]. *中国激光*, 2011, **38**(s1): s106005
- 4 Gao Guoming, Li Xue, Qin Zongding *et al.*. New method for eliminating background fluorescence of Raman spectrum and its application[J]. *Acta Optica Sinica*, 2013, **33**(2): 0230002
- 高国明, 李雪, 覃宗定等. 消除拉曼光谱荧光背景的新方法及其应用[J]. *光学学报*, 2013, **33**(2): 0230002
- 5 Liang Yizeng, Du Yiping. *Analytical Chemometrics* [M]. Chongqing: Chongqing University Press, 2004. 91~96
- 梁逸曾, 杜一平. *分析化学计量学*[M]. 重庆: 重庆大学出版社, 2004. 91~96
- 6 Feng Xinwei, Zhu Zhongliang, Shen Mengjie *et al.*. The method of baseline drift correction of Raman spectrum and based on polynomial fitting[J]. *Computer and Applied Chemistry*, 2009, **26**(6): 759~762
- 7 Shi Yonggang, Su Bin, Tian Gaoyou. *Chemometric Methods and Realizations in Matlab*[M]. Beijing: China Petrochemical Press, 2010. 189~192
- 史永刚, 粟斌, 田高友. *化学计量学方法及 Matlab 实现*[M]. 北京: 中国石化出版社, 2010. 189~192
- 8 G. Strange, T. Nguyen. *Wavelets and Filer Banks* [M]. Wellesley: Wellesley-Cambridge Press, 1997. 35~41
- 9 Guo Jingqiu, Bu Lingbing, Wang Min *et al.*. Measurement of temperature profiles above Nanjing using Raman lidar[J]. *Chinese J. Lasers*, 2012, **39**(s1): s114009
- 郭劲秋, 卜令兵, 王敏等. 基于拉曼激光雷达的南京上空大气温度廓线观测[J]. *中国激光*, 2012, **39**(s1): s114009
- 10 Wu Fengcheng, Xie Pinhua, Li Ang *et al.*. Correction of the influence of multiple scattering on NO₂ emission flux during the pollutants source measurement by mobile differential optical absorption spectroscopy[J]. *Acta Optica Sinica*, 2011, **31**(11): 1101003
- 吴丰成, 谢品华, 李昂等. 利用 O₄ 测量去除车载差分吸收光谱测量污染源 NO₂ 排放通量计算中多次散射的影响[J]. *光学学报*, 2011, **31**(11): 1101003
- 11 Wang Xuan. *The Study of Suppressing Co-Frequency Ambient from the Measured Signal* [D]. Zhengzhou: Zhengzhou University, 2011. 14~15
- 王旋. 去除被测信号中同频背景噪声的研究[D]. 郑州: 郑州大学, 2011. 14~15
- 12 B. Widrow, J. R. Glover. Adaptive noise cancelling: principles and application[J]. *Proceedings of the IEEE*, 1975, **63**(12): 1692~1716

栏目编辑: 李文喆