

用于光谱分类的光谱峰谷沿匹配算法^{*}

陈 扬^{1)†} 贾丽萍²⁾ 张太宁¹⁾ 郭 澎¹⁾ 王湘晖¹⁾ 常胜江¹⁾

1) 南开大学现代光学研究所, 天津 300071)

2) 石家庄经济学院, 石家庄 050031)

(2009 年 3 月 9 日收到, 2009 年 4 月 29 日收到修改稿)

针对超短脉冲激光与气体相互作用产生的非线性荧光光谱的特点, 提出了一种新的光谱去噪、特征提取、特征匹配的有效方法. 通过对三种气体不同浓度的荧光光谱进行小波压缩去噪, 得到 5 层 664 个小波系数, 重构后获得除噪效果满意而特征信息保持不变的光谱. 对重构后的光谱进行光谱峰/谷沿相似度的分析, 构造了三种气体的特征谱. 在此基础上, 经对三种气体光谱进行特征峰/谷沿的匹配分析, 提出一种利用同种物质不同浓度下的光谱总体相似度来构造标准特征谱新方法. 通过与标准特征谱的匹配分析得到分类结果. 实验结果表明, 该算法的分类结果受浓度变化影响小, 对低至 0.05% 浓度条件下的全部 27 条光谱均获得 100% 的分类识别结果.

关键词: 非线性荧光光谱, 特征匹配, 光谱峰谷沿

PACC: 4230S, 3350D

1. 引 言

随着工业的发展, 大气污染问题已经成为人们关注的焦点. 为更好地检测和治理污染, 有必要对大气状况进行实时监测. 传统的湿式化学技术以及后续发展起来的气相色谱法^[1]、质谱和色谱联合技术^[2]等是以吸气取样后的实验分析为基础, 不具备实时和连续监测能力. 近些年, 基于差分光学吸收光谱技术、傅里叶变换红外光谱技术以及差分吸收激光雷达技术^[3,4]等光学和光谱学方法以其大范围、高灵敏度、多组分的在线检测能力而成为大气环境监测的主要技术手段. 但由于这些技术需要瞄准光源安装探测系统或使用角反射器, 因此只能在固定的场合使用.

近几年, 超短脉冲激光与大气相互作用产生荧光光谱^[5-7]现象引起了人们的重视. 由于自聚焦效应, 超短脉冲激光在空气中传输时会形成等离子体通道, 即成“丝”现象出现. 丝状脉冲光场内的高密度能量使生化污染物分子发生非线性电离和解离, 产生具有分子光谱特性的荧光光谱^[8-10]. 这些光谱可以作为分子指纹模式加以识别, 从而得到大气中污

染气体的成分和浓度. 同已有的光学或光谱学环境监测方法相比, 超短脉冲激光技术具有三个明显的优势: 1) 丝状光场是一个稳定的脉冲光场, 污染气体分子释放的荧光光谱具有很好的稳定性和可重复性; 2) 超连续谱分布(加上展宽的三次谐波光谱后)光谱范围为 250 nm 至几个微米, 的“丝”状脉冲光场可同时满足所有污染气体激发荧光的需要, 因此使用单个激光器就可满足多种气体监测要求; 3) 可以把激光光源与遥感探测系统集成为一体形成车载、船载或机载系统, 可实现对任意地点的大气环境进行实时监测. 因此它的机动性很强, 潜在的应用领域更广. 目前, 该方法在环境监测中应用的主要困难是缺乏光谱分析的有效方法. 由于非线性荧光光谱的产生涉及的非线性效应较多, 且同时测量多种污染气体, 这些污染成分的荧光光谱相互影响, 产生交叠, 出现光谱强度与气体浓度之间复杂的非线性关系, 采用传统的光谱分析方法分析上述光谱存在很大难度. 解决这一问题的有效方法是研究和探索新型的光谱分析技术.

传统光谱分类方法, 依据光谱形态可以有导数光谱分析、波峰/波谷总数、波峰/波谷宽度、波峰/波谷斜率等各种分析方法^[11,12]. 由于受噪声干扰, 飞秒

^{*} 国家重点基础研究发展计划(973 计划)批准号: 2007CB310403); 天津市自然科学基金重点项目(批准号: 07JCZDJC05500); 国家自然科学基金(批准号: 60772105, 10704043); 河北省科技厅科技支撑项目(批准号: 08213517D)资助的课题.

[†] 通讯联系人. E-mail: ych@mail.nankai.edu.cn

脉冲激光与大气相互作用产生的荧光光谱存在一些不可忽略的尖峰噪声,导致一些原本较平坦的区域由于噪声干扰出现随机波动的伪特征;同时,随着浓度的变化,各个峰变化趋势不尽相同,其叠加谱的变化规律更为复杂.传统的光谱分析方法对以上因素较为敏感,例如对偶然出现的尖峰噪声,导数光谱分析、波峰/波谷总数分析、波峰/波谷斜率分析等都将给出错误的信息,采用传统的光谱分析方法分析上述荧光光谱难度较大.

本课题组对复杂非线性荧光光谱数据已进行了较为深入的研究^[10,13-17].针对超短脉冲激光与大气相互作用产生的荧光光谱的具体特点,本文提出了一种新的光谱去噪、特征提取、特征匹配的有效方法.我们以空气中混杂的各种浓度的乙烯、*N*-丁烷和1-丁烯气体为分类对象,通过对三种气体不同浓度的荧光光谱进行小波压缩去噪,得到5层664个小波系数,重构后可以获得除噪效果满意而特征信息保持不变的光谱.在此基础上,经对三种气体光谱进行特征峰/谷沿的匹配分析,我们提出一种利用同种物质不同浓度下的光谱总体相似度来构造标准特征谱的新方法,通过与标准特征谱的匹配分析得到分类结果.计算机模拟结果表明,该算法的分类结果受浓度变化影响小,在低至0.05%浓度条件的所有光谱均获得100%的分类识别结果.

2. 特征峰匹配的特征提取与分类算法

2.1. 小波压缩去噪

小波压缩去噪的理论基础是小波变换的多分辨率分析.含有噪声的信号经过小波变换的多分辨率

逐级分解,得到其小波变换后的离散细节和离散逼近. Mallat 等证明,噪声的离散细节信号的幅度随着小波变换尺度的增长而不断减小,但信号的小波变换系数与尺度的关系则不同.利用光谱噪声和小波变换的特点,选择一门限对小波变换后各尺度离散细节进行门限处理,绝对值较小的点取为零,对非零值重新索引后数据点数即可大大压缩.将离散逼近信号和处理后离散细节经小波逆变换重构信号,即可以达到去噪的目的^[18-20].

2.2. 光谱峰谷沿匹配的特征提取算法

在荧光光谱特征提取方面,考虑到飞秒脉冲激光与大气相互作用产生的荧光光谱的特点后,我们提出一种基于特征峰/谷匹配算法的特征提取算法.

2.2.1. 光谱峰谷沿的确定

采样光谱是离散的,我们用其一阶差分来确定光谱的极大值点和极小值点.某一点处的差分由负值变为正值,则该点为极小值点,同理可确定极大值点.极小值到极大值之间的一段为上升沿,标注为“+”,极大值到极小值之间的一段为下降沿,标注为“-”.忽略光谱上升下降的具体数值,将其抽象为峰谷变化的趋势,有利于克服各个峰谷绝对强度随浓度变化不同带来的分析问题.

2.2.2. 光谱的匹配图

将待比较的两条光谱进行如上处理之后,分别投影到直角坐标系中的 x - y 平面内,根据已标注的两条谱线的“+”“-”区域,我们可以将坐标平面分割成许多矩形块,这些矩形块依据两个边乘积的正负被赋予正负号.矩形块为“+”表示其对应的两段光谱同为上升沿或同为下降沿,如图1所示的阴影部分,为“-”表示一段为上升沿另一段为下降沿.其

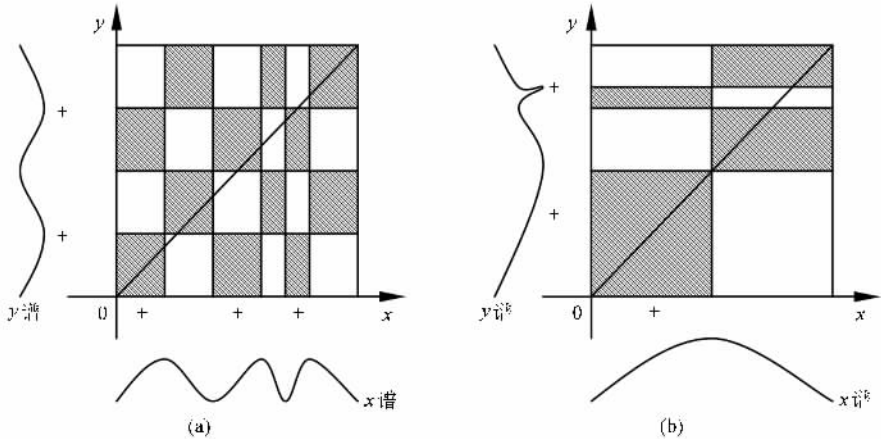


图1 光谱匹配图示例

中投影于 x 轴的光谱简称为 x 谱,也称为标准谱.投影于 y 轴的光谱简称为 y 谱,也称为待匹配谱.在匹配图中,我们只关心 $y = x$ 这条线附近的情况,因为这条线上的点对应两条谱线相同的波长位置.

2.2.3. 峰谷沿的相似度

我们定义峰谷沿的相似度

$$s_{xy}(x_1, x_2, y_1, y_2) = \begin{cases} \frac{\sqrt{2}(\min(x_2, y_2) - \max(x_1, y_1))}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}, \\ \text{flag}(x_1, x_2, y_1, y_2) = "+", \\ 0, \text{flag}(x_1, x_2, y_1, y_2) = "-", \end{cases} \quad (1)$$

其中 x_1, x_2, y_1, y_2 为图 1 中矩形块的四个顶点坐标.当我们以 x 谱为标准谱,分析 y 谱与 x 谱的相

似度时将上式左边简写为 $s_y(x_1, x_2)$. (1) 式表示了 x 谱(x_1, x_2)段与 y 谱(y_1, y_2)段的相似度,从图 1 中直观的看,就是矩形块(x_1, x_2, y_1, y_2)所夹 $y = x$ 线段长度与矩形块对角线长度之比.只有当两段光谱同为上升沿或下降沿时才有正值,否则相似度为 0.

2.2.4. 对尖峰噪声的优化

通常光谱中不可避免地包含着噪声,导致光谱中存在不可忽略的较窄的噪声尖峰.采用峰谷沿匹配算法能很好地解决噪声对分类的影响.考虑图 1 (b) 所示情形, y 谱为 x 谱叠加尖峰后的光谱.在光谱的匹配图上出现了 x 谱(x_1, x_2)段对应 y 谱(y_1, y_2)(y_3, y_4)两段的情况.我们通过定义一个整体相似度将这两段谱的相似度合并,整体相似度定义为

$$\tilde{s}_y(x_1, x_2) = \frac{s_{xy}(x_1, x_2, y_1, y_2) \cdot (y_2 - y_1) + s_{xy}(x_1, x_2, y_3, y_4) \cdot (y_4 - y_3)}{y_4 - y_1}. \quad (2)$$

此时(x_1, x_2)段的相似度会有一定程度的下降,但仍保持较大正值,不同于完全失配下相似度为 0 的情况.(2) 式满足加法结合律.如果出现一段 x 谱对应多段 y 谱的情况,利用(2)式同样能得到整体相似度.

2.2.5. 平移系数与拉伸系数

由于光谱的采样是离散的,采样点的细微偏差

可能引起采样点幅度值较大的变化,在极值点附近还可能引起极值点定位的偏差.另外受仪器精度的影响,采样所得的光谱可能会有整体细微的漂移或拉伸.为了把光谱本身的差别和仪器采样误差区分开,需要对光谱匹配情况进行更细致地描述,我们定义峰谷沿平移系数为

$$\text{tran}_{xy}(x_1, x_2, y_1, y_2) = \frac{(y_1 + y_2) - (x_1 + x_2)}{2\max(x_2 - x_1, y_2 - y_1) - |(y_2 - y_1) - (x_2 - x_1)|}. \quad (3)$$

峰谷沿平移系数表示两段相似度大于 0 的谱沿相对的位移,取值范围为 $[-1, 1]$,当谱沿没有相对位移时平移系数为 0;当相对位移达到最大时平移系数为 1 或 -1.平移系数为正值表示 y 谱相对 x 谱有向右的位移,负值表示有向左的位移.从匹配图上直观地看,平移系数表示矩形块中心到 $y = x$ 的距离与矩形块对角线半长度的比值.

峰谷沿的拉伸系数定义为

$$\text{deform}_{xy}(x_1, x_2, y_1, y_2) = \frac{y_2 - y_1}{x_2 - x_1}. \quad (4)$$

峰谷沿的拉伸系数表示相似度大于 0 的两段谱沿在坐标轴投影的长度之比.当拉伸系数大于 1 时 y 谱相对于 x 谱是拉伸的,当系数小于 1 时 y 谱相对于 x 谱是压缩的.从匹配图上直观地看,拉伸系数表示矩形块 y 轴长度与 x 轴长度之比.

2.2.6. 修正的峰谷沿相似度

当峰谷沿相似度不为 1 时,匹配的两段谱有相对位移或拉伸.通常采样过程会引起细微的漂移与拉伸,与有较大漂移与拉伸的不匹配有本质的区别.为区分这种差别,我们修正峰谷沿的相似度公式使其与漂移、拉伸量的平方相关.

$$\begin{aligned} \text{sim}_y(x_1, x_2) &= \tilde{s}_y(x_1, x_2) \cdot (1 - \text{tran}^2) \\ &\times \left(1 - \left(1 - \min\left(\text{deform}, \frac{1}{\text{deform}}\right) \right)^2 \right), \end{aligned} \quad (5)$$

其中 tran , deform 为(4)(5)式中 x_1, x_2 所对应矩形块的平移系数与拉伸系数,该矩形块为完成(2)式所做尖峰噪声优化后的大矩形块.通过(5)式的修正,当平移、拉伸较小时,相似度被修正的量很小,几乎可以忽略;当平移、拉伸较为严重时,原本未修正时谱

线的相似度就比较小,而修正后相似度被进一步减小,使我们更加容易区分出不匹配的情况.当待匹配光谱段仅出现尖峰噪声时,由于合并后整体峰谷沿的位置还是匹配的,大矩形块对应的平移系数和拉伸系数都与完全匹配时相近,修正后的相似度几乎没有改变而保持较大值.为了便于分析,我们将按峰谷沿分组表示的相似度改写成按光谱点逐点表示的相似度,即

$$\text{sims}_{xy}(x) = \begin{cases} \text{sims}_y(x_i, x_{i+1}), & x \in (x_i, x_{i+1}), \\ \frac{1}{2}(\text{sims}_y(x_{i-1}, x_i) + \text{sims}_y(x_i, x_{i+1})), & x = x_i, \\ 0, & \text{其他}. \end{cases} \quad (6)$$

一个光谱点落在某段峰谷沿范围内,则赋予该点相似度为对应峰谷沿的相似度;如果一个光谱点为标准谱的极大值点或极小值点,则该点同时落在前后两段光谱沿范围内,我们赋予该点相似度为前后两段峰谷沿相似度的平均值.将(6)式求得的各个谱点的相似度值顺序排列组成相似度矢量 sims_{xy} , 矢量维数为光谱点的数量.

2.2.7. 光谱特征峰的提取与标准特征峰谱的自适应构造法

为了有效提取光谱的特征,我们将同一物质不同浓度的光谱进行两两间的匹配分析,方法如下:不同浓度的谱 $z_1, z_2, \dots, z_n$, 依次将各浓度下的光谱 z_a 作为标准谱,其余浓度作为待匹配谱分别与标准谱进行(1)至(6)式的计算得到相似度矢量 $\text{sims}_{z_a z_1} \dots \text{sims}_{z_a z_{a-1}}, \text{sims}_{z_a z_{a+1}} \dots \text{sims}_{z_a z_n}$, 将上述矢量取平均后得到平均相似度矢量

$$\text{sims}'_{z_a} = \frac{1}{n-1} \sum_{\substack{i=1 \\ i \neq a}}^n \text{sims}_{z_a z_i}, \quad (7)$$

它表示了以 z_a 作为标准谱时与其他浓度谱的总体

匹配情况,对平均相似度矢量 sims'_{z_a} 进行阈值处理

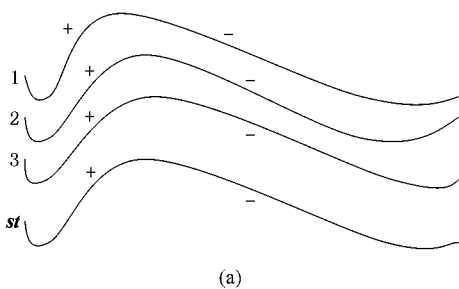
$$\text{sims}_{z_a}(z) = \begin{cases} \text{sims}'_{z_a}(z), & \text{sims}'_{z_a}(z) \geq \xi, \\ 0, & \text{其他}. \end{cases} \quad (8)$$

得到 z_a 作为标准谱与其他浓度谱的总体相似度矢量 sims_{z_a} .对于非特征峰的光谱部分,由于噪声的随机性,不同浓度光谱作为标准谱与其他谱进行匹配时相似度都有大有小,但平均值较小,通过阈值作用将被置为0.对于特征峰的光谱部分,由于总体相似度值大于阈值,相似度被保留下来.

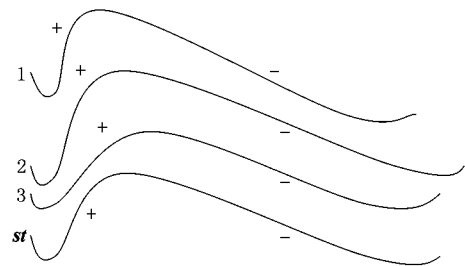
基于以上分析,我们提出一种利用同种物质不同浓度下的光谱总体相似度来构造标准特征谱新方法.具体方法如下.

每个特征峰/谷在 $\text{sims}_{z_1} \dots \text{sims}_{z_n}$ 中对应一段相似度不为0光谱段,以不同谱 $z_1, z_2, \dots, z_n$ 为标准谱时这个峰谷沿的总体相似度不同,我们选取总体相似度最大的光谱段代表特征峰谷沿,特征峰谷沿的端点为该光谱段对应沿的端点,峰沿的上升(下降)趋势标注与谱线对应沿一致.当相邻的特征峰谷沿对应不同光谱段时,边界点可能会发生重叠或空白.我们处理的原则是:重叠部分划归为相似度高的峰谷沿,空白部分划归为相似度低的峰谷沿.

以三条谱线为例,图2模拟了两种不同情况下构造的特征谱 st .图2(a)中光谱段2作为标准谱时,其上升沿所对应的总体相似度 $\text{sims}_2(1) \times \text{sims}_2$ (1)代表谱2上升沿各个点的总体相似度,它们都是相等的,因此放在一起简化表示,较其他两条光谱作为标准谱时在该处的总体相似度 $\text{sims}_1(1), \text{sims}_3(1)$ 大,因此我们将谱线2的上升沿作为标准谱的上升沿.同理我们将谱线1下降沿作为标准谱的下降沿.谱线2的上升沿与谱线1的下降沿有坐标重合的部分,在这些重合点处 $\text{sims}_2(1) > \text{sims}_1(2)$,因此我们将这些点划归为上升沿部分.图2(b)为构造标准谱



(a)



(b)

图2 构造标准谱示例

的另外一种情况,上升沿取为与光谱 2 相同,下降沿取为与光谱 3 相同,空白部分划归为下降沿.这样的处理保证了标准谱中既能完整地包含高相似度的峰谷沿,也能对低相似度的峰谷沿进行调整.实际操作中我们处理是相似度矢量,因此在构造标准谱时,除了标记上升沿和下降沿外,我们在相似度矢量为 0 的点处将标准谱标记为 0,表示在这些点特征峰不明显或无法肯定是否存在.经过这些步骤后,我们可以得到维数为光谱点数的标准特征峰谱 st ,其各点值标记为 0“+”或“-”.

2.3. 基于构造标准特征谱的分类算法

当对未知光谱 z 进行分类时,如何提取标准谱的特征点是很关键的.以两种气体为例:首先对两种光谱按照 2.2.7 节的方法构造标准谱 st_x 和 st_y ,将这两种光谱的标准特征谱进行一次相似度分析,此过程只需将(6)式的变量 x, y 用 st_x 和 st_y 代替,并将(1)式的结果代入即可得到相似度矢量 sims_{st_x, st_y} .标准谱的特殊之处在于有许多点的值为“0”,处理时按照“-”来处理,因为它们同为不匹配的情况.

对于多种气体的光谱的情况,依次考虑各个谱的特征.以 st_{s_1} 为例,我们将 st_{s_1} 分别与剩下的 $n-1$ 种气体标准谱进行相似度分析得到相似度 $\text{sims}_{st_{s_1}, st_{s_2}}, \dots, \text{sims}_{st_{s_1}, st_{s_n}}$. st_{s_1} 谱与 st_{s_n} 谱的相似度分析在特征点处有几种情况: st_{s_1} 有明显的特征而 st_{s_n} 没有明显特征,即 st_{s_n} 谱在这些点为 0,此时求得这些点的相似度为 0; st_{s_1} 与 st_{s_n} 都有明显特征但特征相反,此时相似度很小; st_{s_1} 与 st_{s_n} 都有明显特征且特征相近,此时相似度较大.我们赋予不同特征点不同的权值,第一种情况 0.5,第二种情况 1,第三种情况 0,用以区分它们对分类贡献的大小.同样方法,我们可以得到 st_{s_1} 与余下 $n-1$ 种气体谱的所有 $n-1$ 组权值,将得到的权值相加作为特征点的权重.我们依次对 $st_{s_1}, \dots, st_{s_n}$ 进行这样的运算,得到 n 个气体谱的特征点权重.

以三种气体为例,表 1 和图 3 显示了各个光谱的标准谱示例和计算出的特征点权重.表格中的“-”表示该点不是该标准谱的特征点,因为其特征谱在该点处为 0.通过图和表我们看到,大的权重对谱的分类是非常重要的,如 st_1 谱中的 1 和 2 处,如

果待分类谱在 1 处匹配该上升沿,2 处匹配该下降沿,则为我们的分类提供相当大的帮助,因为待分类谱与 st_2 谱和 st_3 谱中这两处是不匹配的,我们相信该谱为 st_1 类的可能性很大,而 1,2 两处对于 st_2 和 st_3 的分类作用则没有那么大,因为如果匹配了 st_2, st_3 的特征,我们仅能得出不是 st_1 的结论,并不能判断是 st_2 或 st_3 中的哪一类.

表 1 标准谱示例中计算得到的分类特征点权值

	1	2	3	4	5	6	7	8	9	10
st_1	2	2	-	-	1.5	1.5	0.5	0.5	0	0
st_2	1	1	1	1	-	-	-	-	0	0
st_3	1	1	-	-	1.5	1.5	0.5	0.5	0	0

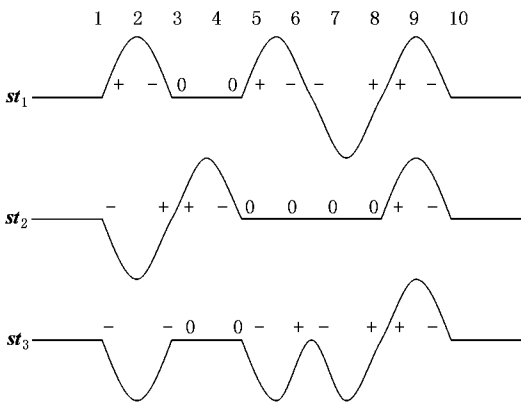


图 3 三种气体的标准谱示例

求得特征点权重后,我们将待分类谱 z 与各特征谱 $st_{s_1}, \dots, st_{s_n}$ 分别进行加权的相似度分析

$$\overline{\text{sims}}_{st_n, z} = \frac{\sum_{i=1}^d w_{ni} \text{sims}_{st_n, z}(i)}{\sum_{i=1}^d w_{ni}}, \tag{9}$$

式中 d 为光谱维数, w_{ni} 表示标准谱 st_{s_n} 第 i 点的权重,如果该点不是特征点则权重为 0. $\text{sims}_{st_n, z}(i)$ 表示 z 谱与 st_{s_n} 谱的相似度.在理想情况下,当 z 谱与 st_{s_n} 在特征点处加权相似度 $\overline{\text{sims}}_{st_n, z}$ 完全匹配时为 1,不匹配时较小.在分类过程中,考虑到特征点众多,可以对权重取阈值操作,仅保留特征明显对分类很有帮助的点.权重小的点处与其他谱特征相似,代表性不强且容易受到干扰,可弃之不用.

基于构造标准特征谱的分类方法,不但给出了分类结果,还给出了与各标准谱匹配的相似度,得出

更加明确的结论.在与已知谱相似度都较低时,该方法有助于发现新物质和确定新物质的特征.

3. 实验及其结果分析

3.1. 非线性荧光光谱的获得

获取气体荧光光谱的实验装置参考我们前期的工作^[7-12],飞秒激光脉冲入射到气体中时由于自聚焦效应,在空气中形成等离子体通道,即成“丝”现象出现,丝状脉冲光场内的高密度能量使生化污染物分子发生非线性电离和解离,产生具有分子光谱特性的荧光光谱.经透镜会聚后被分光光度计(Acton Spectrapro2150)和 CCD 探测,得到混合气体的非线性荧光光谱.实验中一个输入脉冲激光能量为 10 mJ、脉宽为 50 fs,重复频率为 10 Hz.我们特别选取三种光谱十分接近的掺杂气体,分别为乙烯(Ethylene),1-丁烯(1-Butene),N-丁烷(N-Butane).这三种气体分别与空气混合的体积比:乙烯为 25%,12.5%,6.25%,3.13%,1.56%,0.78%,0.39%,

0.20%,0.10%,0.049%(10 个数据);1-丁烯为 25%,12.5%,6.25%,3.13%,1.56%,0.78%,0.39%,0.20%,0.10%,0.05%(10 个数据);N-丁烷为 12.5%,1.56%,0.78%,0.39%,0.20%,0.098%,0.049%(7 个数据),总共 27 条光谱数据.

3.2. 光谱的小波压缩去噪

原始光谱中含有大量噪声,有必要进行去噪处理.我们采用小波去噪的方法,将信号作五层分解,采用硬阈值函数对各层系数进行处理,细节中绝对值较小的点置为零.我们前期的工作已经验证了对于该类复杂荧光光谱,采用双正交的 bior3.9 小波的滤波及重构光谱均方误差最小,且特征峰保留完好.本文依然采用 bior3.9 小波进行压缩去噪.最终光谱数据由 3979 个点压缩为 664 个系数,重构后光谱的噪声大大降低,信号特征更为明显.

3.3. 相似度矢量

我们以乙烯气体浓度为 0.39% 和 0.78% 的两条谱为例分析一下它们的相似度.我们将 0.78% 乙

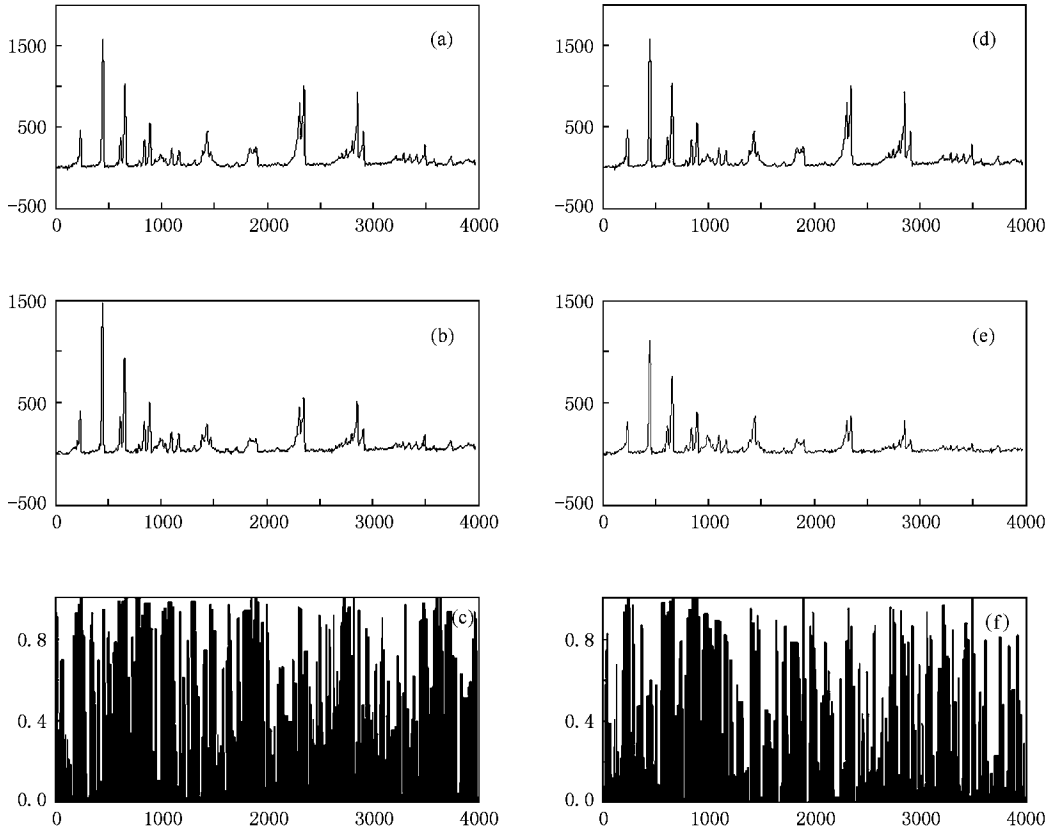


图 4 同种气体的相似度与不同气体的相似度 (a)0.78% 乙烯光谱 (b)0.39% 乙烯光谱 (c)0.78% 与 0.39% 乙烯光谱相似度矢量 (d)0.78% 乙烯光谱 (e)0.39% 1-丁烯光谱 (f)0.78% 乙烯光谱与 0.39% 1-丁烯光谱相似度矢量

烯光谱作为标准谱,0.39% 乙烯光谱作为待匹配光谱进行(1)至(6)式的运算.图 4(a)(b)(c)显示了采用小波去噪处理后两气体光谱和它们的相似度矢量.从图中我们可以看出两种浓度的气体光谱在光谱强大的地方相似度很高,而在强度小的地方相似度则有高有低.

我们再以 1-丁烯浓度为 0.39% 和乙烯浓度 0.78% 的两条谱为例分析相似度.我们仍然将 0.78% 乙烯光谱作为标准谱,将 0.39% 的 1-丁烯光谱作为待匹配谱.图 4(d)(e)(f)显示了两气体光谱和它们的相似度矢量.从图中我们可以看出这两种不同气体的光谱在强度较大处特征峰处相似度依然很高,不论直接观察或与图 4(a)(b)(c)的相似度进行比较我们都难以明显区分 1-丁烯和乙烯两种气体谱的差别.因为乙烯与 1-丁烯两种有机气体的电离能很接近,主要化学键也接近,两种气体的主

要官能团也都为 $C=C$ 双键,因此它们的谱很相似;且在很低浓度情况下,光谱中还包含大量空气中各主要气体成分的光谱,因此区分这两种谱是有一定困难的.

3.4. 构造标准特征谱

根据 2.2.7 节的方法,我们利用同种物质不同浓度的光谱构造标准特征谱,平均相似度阈值 ξ 取 0.5.图 5 显示了 *N*-丁烷与 1-丁烯气体谱与构造的标准特征谱.左图为 *N*-丁烷,我们只选取 *N*-丁烷 7 条光谱中的 2 条为例.图 5(a)(b)光谱浓度为 0.78% 0.05%.对应的(c)图为构造的标准谱,忽略了幅度信息,构造标准谱的“+”;“-”;“0”用斜率 1,-1,0 的线段来表示.(d)(e)(f)为 1-丁烯各浓度光谱与构造标准谱,其中的光谱选取 3.13%,0.05% 为代表.

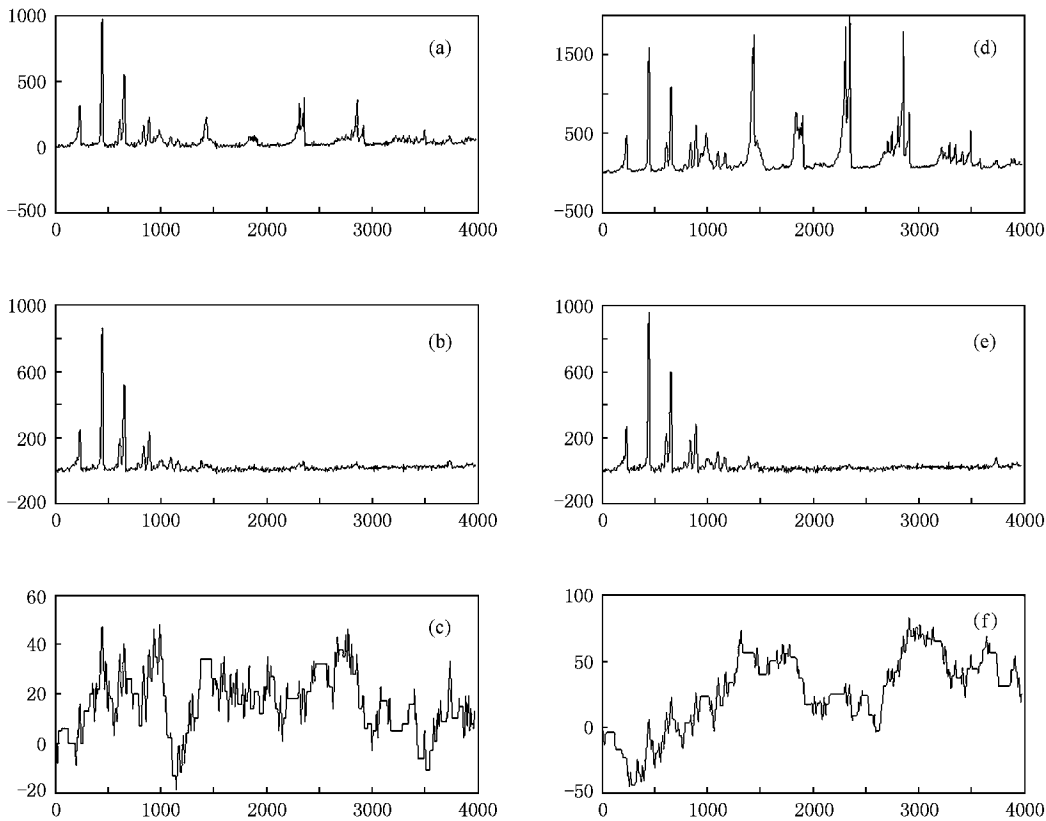


图 5 光谱与其构造标准谱 (a)0.78% *N*-丁烷光谱 (b)0.05% *N*-丁烷光谱 (c)*N*-丁烷构造标准谱 (d)3.13% 1-丁烯光谱 (e)0.05% 1-丁烯光谱 (f)1-丁烯构造标准光谱

3.5. 基于构造标准特征谱的分类算法

采用 2.3 节的方法计算三种气体谱的分类特征

点权重,图 4(a)(b)(c)显示了三种气体谱的分类特征点权重.我们以乙烯光谱权重为 2 的点 3638—3653 为例,直观分析一下对应光谱的特点.图 4(d),

(e)(f)显示了浓度分别为 1.56% 与 0.10% 的 *N*-丁烷、1-丁烯、乙烯气体谱在点 3638—3653 附近的

光谱。
从图 6 中可以看出 ,在乙烯特征点权重为 2 的

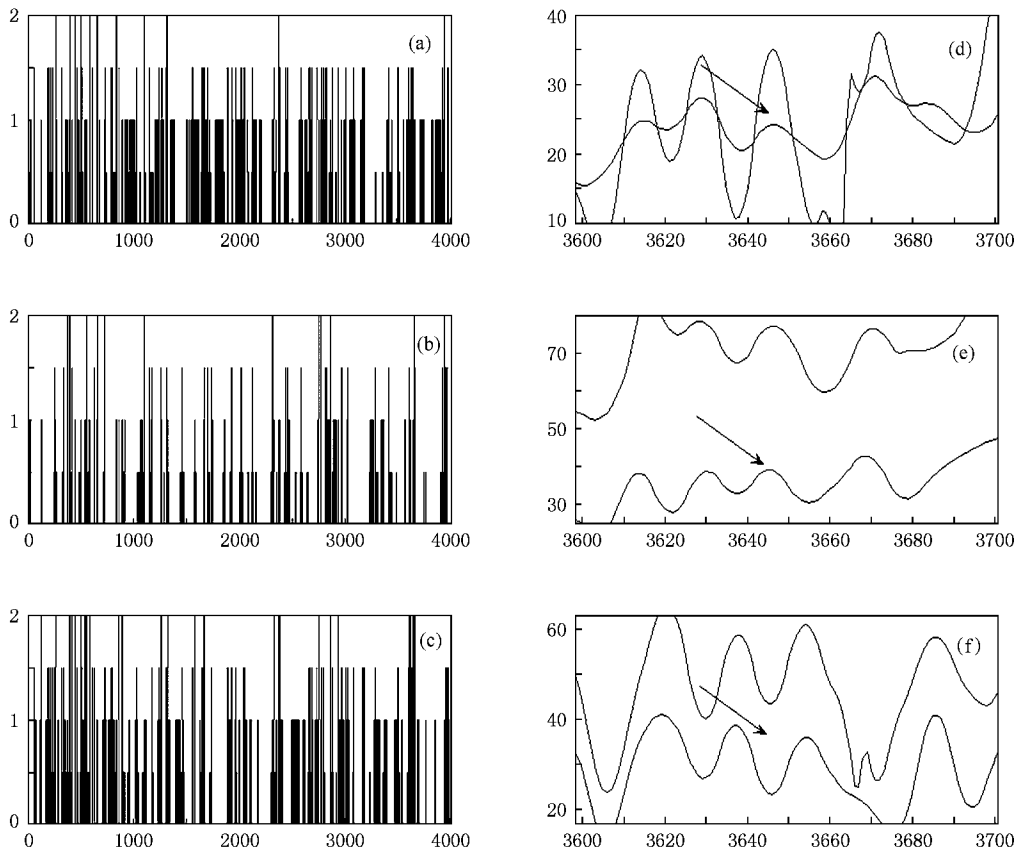


图 6 气体谱的分类特征点权重与三种谱在乙烯特征点 3638—3653 附近的光谱 (a) *N*-丁烷特征点权重 (b) 1-丁烯特征点权重 (c) 乙烯特征点权重 (d) *N*-丁烷示例光谱 (e) 1-丁烯示例光谱 (f) 乙烯示例光谱

地方 ,乙烯谱与 *N*-丁烷、1-丁烯谱有着明显不同的特征 ,*N*-丁烷、1-丁烯谱在这些点处为明显的峰而乙烯为明显的谷 ,为我们区分乙烯与其他两种谱提供了很大的帮助 .从左图中可以看出 ,乙烯有 20 多处特征权重为 2 的谱段 ,*N*-丁烷与 1-丁烯也有 10 多处 ,特征权重 1.5 的地方更多 .这些谱段都对应明显的特征峰 ,有助于我们鉴别污染气体 .

我们轮流用三种气体 27 条光谱中的 1 条作为测试光谱 ,其余 26 条用来进行 2.2.7 节与 2.3 节中标准谱的构造与特征权重值的求解 ,用这条测试光谱与构造的三种气体标准谱进行带权重的匹配 .权重只取 2 和 1.5 的部分 ,其余部分置为 0 .表 2 列出了 27 条光谱的相似程度 .

表中编号 1—7 为 *N*-丁烷浓度 12.5%—0.05% ,8—17 为 1-丁烯浓度 25%—0.05% ,18—27

为乙烯浓度 25%—0.05% . bta 表示 *N*-丁烷 ,bte 表示 1-丁烯 ,eth 表示乙烯 .加权相似度未进行归一化 .

从表中我们可以看出 ,对 27 条光谱取相似度最大匹配谱为分类结果 ,正确率为 100% .这三种有机物光谱十分相似 ,特别是 *N*-丁烷与 1-丁烯这两种物质 ,它们在化学结构差别仅为一个碳—碳双键与一个碳—碳键 ,在浓度很低的情况下 ,它们混合空气谱的相似度更加接近 .而本方法能很好地区分这两种差别细微的物质 ,而且能计算物质与物质之间谱相似度的差异 .在气体浓度很低的时候 ,光谱淹没在很大的噪声中 ,而本实验中 7 ,17 ,27 号谱所对应浓度分别为 0.05% 的 *N*-丁烷、1-丁烯、乙烯谱 ,与其他浓度的谱相比 ,基于相似度分析的分类结果没有显著地变差 ,提示我们本方法有分析识别更低浓度气体的能力 .

表 2 光谱的相似度分析

光谱编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14
bta 相似度	0.67	0.64	0.69	0.72	0.63	0.68	0.64	0.49	0.46	0.37	0.20	0.31	0.31	0.35
bte 相似度	0.45	0.37	0.40	0.49	0.46	0.39	0.30	0.54	0.59	0.57	0.75	0.69	0.73	0.71
eth 相似度	-0.02	-0.08	-0.02	-0.12	0.05	-0.05	-0.03	-0.09	0.02	0.07	0.12	0.15	0.09	0.18
光谱编号	15	16	17	18	19	20	21	22	23	24	25	26	27	
bta 相似度	0.38	0.38	0.47	-0.02	0.00	-0.08	0.03	0.00	-0.08	0.01	-0.09	-0.09	-0.06	
bte 相似度	0.64	0.51	0.57	-0.01	-0.02	0.11	0.25	0.27	0.22	0.28	0.02	0.01	0.03	
eth 相似度	0.12	0.11	-0.06	0.53	0.66	0.52	0.64	0.73	0.74	0.69	0.74	0.67	0.63	

4. 结 论

本文提出了一种新的光谱去噪、体征提取及分类的有效方法.研究了在空气中混杂的各种浓度的乙烯、*N*-丁烷和 1-丁烯气体的分类算法 ,通过对三种气体不同浓度的荧光光谱进行小波压缩去噪 ,得到 5 层 664 个小波系数 ,重构后可以获得除噪效果

满意而特征信息保持不变的光谱 .在此基础上 ,经对三种气体光谱进行特征峰/谷沿的匹配分析 ,我们提出一种利用同种物质不同浓度下的光谱总体相似度来构造标准特征谱新方法 ,通过与标准特征谱的匹配分析得到分类结果 ,该算法可用于解决其他类型复杂光谱的分类问题 ,特别对低浓度气体的分类有很好的鲁棒性 .

[1] Shao G R ,Jiao D M 1996 *Zhejiang Chemical* **27** 37 (in Chinese) [邵光荣、焦东苗 1996 浙江化工 **27** 37]

[2] Zhao Z ,Wang R Z ,Sun T H 1999 *Chin . Space Sci . and Tech .* **4** 46 (in Chinese) [赵忠、王荣宗、孙天辉 1999 中国空间科学技术 **4** 46]

[3] Edner H ,Fredriksson K ,Sunesson A 1987 *Appl . Opt .* **26** 4330

[4] Zhang G X ,Zhao Y F ,Zhang Y C ,Zhao P T 2008 *Acta Phys . Sin .* **57** 7390 (in Chinese) [张改霞、赵曰峰、张寅超、赵培涛 2008 物理学报 **57** 7390]

[5] Hu X Y ,Zhong F C ,Deng J 2001 *Acta Opt . Sin .* **21** 641 (in Chinese) [胡雪原、钟方川、邓建 2001 光学学报 **21** 641]

[6] Zhu S T ,Zhang J L ,Shen W D 1994 *Chinese Journal of Lasers* **21** 678 (in Chinese) [朱蔚通、章介伦、沈文达等 1994 中国激光 **21** 678]

[7] Feng Z H ,Fu X Q ,Zhang L F ,Xu H W ,Wen S C 2008 *Acta Phys . Sin .* **57** 2253 (in Chinese) [冯则胡、傅喜泉等 2008 物理学报 **57** 2253]

[8] Cao S Y ,Zhang Z G ,Chai L ,Wang Q Y ,Yang J J ,Zhu X N 2007 *Acta Phys . Sin .* **56** 2765 (in Chinese) [曹士英、张志刚、柴路、王清月、杨建军、朱晓农 2007 物理学报 **56** 2765]

[9] Yang S Z ,Chen D Y 2008 *Chin . Phys . B* **17** 817

[10] Shen J Y ,Han Y Z ,Chang S J 2004 *Acta Opt . Sin .* **24** 1000 (in Chinese) [申金媛、韩应哲、常胜江 2004 光学学报 **24** 1000]

[11] Liu Y ,Song C Y ,Luo X S ,Lu J ,Ni X W 2007 *Chin . Phys .* **16** 1300

[12] Su H J ,Du P J ,Sheng Y H 2008 *Application Research of Computer* **25** 390 (in Chinese) [苏红军、杜培军、盛业华 2008 计算机应用研究 **25** 390]

[13] Han Y Z ,Li S M ,Zhang Y X 2005 *J . Optoelectronics Laser* **16** 710 (in Chinese) [韩应哲、李素梅、张延 2005 光电子激光 **16** 710]

[14] Han Y Z ,Zhang Y X ,Chang S J 2005 *J . Optoelectronics Laser* **16** 718 (in Chinese) [韩应哲、张延、常胜江 2005 光电子激光 **16** 718]

[15] Li S M ,Han Y Z ,Zhang Y X 2006 *Acta Opt . Sin .* **26** 147 (in Chinese) [李素梅、韩应哲、张延 2006 光学学报 **26** 147]

[16] Wang K ,Zhang H ,Chang S J ,Shen J Y 2007 *Acta Phys . Sin .* **56** 3613 (in Chinese) [王凯、张会、常胜江、申金媛 2007 物理学报 **56** 3613]

[17] Wang J H ,Zheng Z ,Zhang Y X 2007 *J . Optoelectronics Laser* **18** 741 (in Chinese) [王建华、郑震、张延 2007 光电子激光 **18** 741]

[18] Yuan L 2008 *Chin . Phys . B* **17** 170

[19] Ding G ,Zhong S S ,Li Y 2008 *Chin . Phys . B* **17** 1998

[20] Sun Y K 2005 *Wavelet Analysis and Application* (Beijing :China Machine Press) p17 (in Chinese) [孙延奎 2005 小波分析及其应用(北京 :机械工业出版社)第 17 页]

A classification method for nonlinear fluorescent spectra based on edges matching^{*}

Chen Yang^{1)†} Jia Li-Ping²⁾ Zhang Tai-Ning¹⁾ Guo Peng¹⁾ Wang Xiang-Hui¹⁾ Chang Sheng-Jiang¹⁾

¹⁾ *Institute Of Modern Optics ,Key laboratory of Opto-electronic Information Science and Technology*

of Ministry of Education ,Nankai University ,Tianjin 300071 ,China)

²⁾ *Shijiazhuang University of Economics ,Shijiazhuang 050031 ,China)*

(Received 9 March 2009 ; revised manuscript received 29 April 2009)

Abstract

A pre-process of feature extraction and classification approach based on spectrum edges matching is proposed to analyze the complicated nonlinear fluorescence spectra emitted by the interaction between femto-second(fs) laser and the impurities in air. The spectra data is denoised and compressed from 3979 points to 664 points using wavelet(WT) transform. By similarity analysis we create the characteristic spectra of 3 kinds of gases and the weights for classification. A new method of classification is proposed based on the edges matching and the comparison with characteristic spectra. Compared with existing methods ,our method can not only get 100% classification accuracy ,but also gives the characteristic position and the matching degree. The analysis of the matching degree shows that our method works well at low concentrations and has a potential application of identifying gases of lower concentration.

Keywords : nonlinear fluorescence spectrum , feature matching , edges of spectrum

PACC : 4230S , 3350D

^{*} Project supported by the National Basic Research Program of China(Grant No. 2007CB310403) ,the Key Program of the Natural Science Foundation of Tianjin ,China (Grant No. 07JCZDJC05500) ,the National Natural Science Foundation of China (Grant Nos. 60772105 ,10704043) and the Support Program of Science and Technology Department of Hebei Province(Grant No. 08213517D).

[†] Corresponding author. E-mail : ych@mail.nankai.edu.cn