

文章编号: 1672-8785(2016)11-0029-07

被动红外易制毒挥发气体光谱 数据预处理技术

沈 俊 李志豪

(公安部第三研究所, 上海 200031)

摘 要: 由于信号强度低、环境复杂等因素会造成光谱杂乱和波峰不明显, 利用被动红外远程遥测技术难以准确地对易制毒化学品气体进行定性检测, 即使采用基本的数字滤波算法也无法有效解决这些问题。为此提出了一种基于卡尔曼/维纳叠加的滤波算法。通过用该算法对光谱信号进行预处理, 可以有效提高后期波峰匹配的定性检测的准确性。基于改进的算法对采集到的实验数据进行了验证。结果表明, 与其他滤波算法的处理效果相比, 本文方法可以对干扰信号进行有效过滤, 并且可以对特征峰进行准确识别。

关键词: 被动红外遥测技术; 禁毒; 光谱数据预处理; 卡尔曼滤波

中图分类号: TP312 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2016.11.006

Preprocessing of Spectra of Precursor Chemical Vapor Detected by Passive Infrared Remote Sensing

SHEN Jun, LI Zhi-hao

(The Third Research Institute of Ministry of Public Security, Shanghai 200031, China)

Abstract: Since the factors such as low signal intensity and complicated environment may cause the spectrum to be cluttered and cause the wave peak to be not obvious, it is difficult for infrared remote sensing to detect precursor chemical vapor accurately. Even the basic digital filtering algorithm is used, the problem can not be solved. For this reason, a filtering algorithm based on Kalman/Wiener superimposition is proposed. By using this algorithm to preprocess spectral signals, the qualitative detection accuracy in subsequent wave peak match can be improved effectively. The experimental data collected are verified on the basis of the improved algorithm. The result shows that compared with other filtering algorithms, the improved algorithm can filter interference signals effectively and can identify characteristic peaks accurately.

Key words: passive infrared remote sensing; drug control; spectral data preprocessing; Kalman filtering

0 引言

根据 2015 年《中国禁毒报告》, 我国目前毒品消费市场中的晶体冰毒和氯胺酮几乎都是国内生产的。广东部分地区已经成为国内非法制造冰毒、氯胺酮等合成毒品最为严重的地区。

2014 年, 全国破获制毒物品案件 549 起, 其中走私制毒物品案 29 起, 非法买卖制毒物品案 520 起, 缴获制毒物品 3.847×10^6 kg。

无论是大麻、可卡因等天然(植物)毒品还是冰毒、摇头丸等合成化学毒品, 它们的加工都

收稿日期: 2016-07-25

作者简介: 沈俊(1984-), 男, 上海人, 工程师, 学士, 主要从事安检装备技术研究。E-mail: shenjun511@126.com

<http://journal.sitp.ac.cn/hw>

INFRARED (MONTHLY)/VOL.37, No.11, Nov 2016

离不开易制毒化学品。从某种意义上来说,没有易制毒化学品就没有毒品。而其中醋酸酐、乙醚、丙酮、氯化亚砷、三氯甲烷和石油醚等在非法制造毒品中作为基本原料起着化学合成、精制等作用。

作为一种有效的大气气体探测与识别方法,红外光谱遥测技术已被广泛应用于灰卤辐射检测、有毒气体识别、军用化学战剂探测等领域。相关研究表明,大多数易制毒化学品在制作过程中会产生大量挥发气体,并在 $700 \sim 1400 \text{ cm}^{-1}$ 的红外大气窗口中显现出非常明显的特征光谱。因此,红外光谱遥测技术已经成为远距离探测非法制毒窝点的最为有效的方法之一。

1 红外光谱遥测技术的原理

红外吸收光谱分析方法的主要原理是,利用物质分子对红外辐射的吸收,使得分子由基态振动和转动能级跃迁到激发态,从而获得红外吸收光谱。红外吸收光谱被记录成红外光的百分透射比(吸光度)与波数或波长之间的关系曲线。由于不同分子的振动能级和转动能级不同,能级间的能量差值不同,各种物质对红外光的吸收波长也必然不同。所以根据物质的红外吸收波长就可对物质进行定性分析。

如图 1 所示,我们可以利用一个简单的三层



图 1 被动红外遥测过程

模型来描述被动红外遥测的基本特性。这里,每层都被认为含有均匀的物理和化学特性。由天空或地表散发出的背景辐射通过蒸汽云及其与光谱仪之间的大气层传播。背景辐射透过所有层所产生的辐射特性变化将会被红外光谱遥测仪接收。当含有化学活性物质的蒸汽云团进入被动红外遥测仪的视野时,后者将记录背景辐射透过蒸汽云团后所产生的相应辐射特性变化,从而推测出蒸汽云所包含的化学物质成分。

如图 2 所示,当红外辐射被遥测仪捕捉到后,首先将会进入一个光学系统受到相应处理,然后通过半透半反膜被分为两路:一路被直接传给 CCD 图像传感器进行成像;另一路依次通过光学干涉系统和 MCT 检测器处理后,与经 CCD 图像传感器处理后的数据一起由 DSP 信号处理器统一生成一幅图像,并反馈给 PC 机或屏幕进行成像。图 3 所示为整体仪器的外观造型。

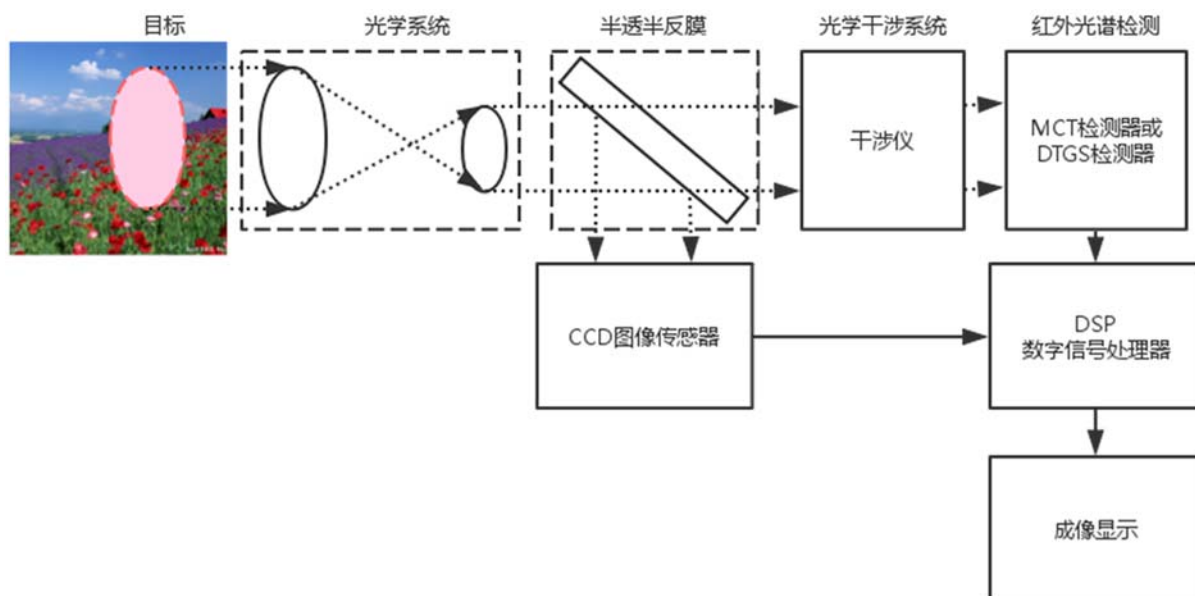


图 2 被动红外光谱遥测仪的结构示意图



图 3 被动红外光谱遥测仪的实物图

2 光谱数据预处理

基于双光束干涉的傅里叶红外光谱技术通过干涉条纹的傅里叶变换获得的光谱, 可用于红外波段 ($700 \sim 1300 \text{ cm}^{-1}$) 中具有特征吸收光谱的气体分子的测量^[1]。为了有效提高原始光谱数据的信噪比, 剥离背景干扰噪声信号, 进一步提高信号质量, 同时提高被测光谱数据与标准光谱样本库比对的准确度, 光谱数据预处理显得尤为重要。

2.1 基本数字滤波算法

作为一种最常用的分析信号处理方法, 数字平滑滤波算法的实质主要是去除由随机误差引起的分析信号偏离。常用的平滑方法包括移动平均值法、Savitsky-Golay 法 (S-G 法) 和中位数平滑法 (中值滤波) 三种。

2.1.1 窗口移动平均滤波算法

数字平滑滤波算法的具体实现是, 将原始数据分割成若干个相邻的或者递进重叠的等宽度窗口, 然后由每个窗口计算出平均值或拟合值, 从而构成处理后的平滑曲线。这个窗口的大小被定义为滤波宽度。随着窗口的不断移动, 即可得到经滤波后的一些列信号值。设滤波值为

$$y_k^* = \frac{1}{2m+1} \sum_{j=-m}^{j=m} y_{k+j} \quad (1)$$

式中, k 为原始数据点的序号; $2m+1$ 为窗口 (滤波带宽) 的大小; m 为调节滤波带宽的参数。

窗口移动平均滤波算法具有实现简单、计算速度快等优点。但在实际使用中, 如果选择的窗口参数值过大, 比如大于 5 点以上, 那么就很容易导致数据结构产生扭曲, 从而引起数据失真。

2.1.2 S-G 滤波

为了解决窗口移动平均滤波给原始数据都赋予一样的权重从而使数据容易发生扭曲的问题, S-G 滤波通过给数据赋以不同权重来进行滤波。由于数据平滑基于高次多项式来实现, 所以 S-G 滤波也被称为卷积平滑。

假设窗口内的数据点可用一个 $k-1$ 次多项式进行拟合, 即

$$y_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_{k-1} i^{k-1} \quad (2)$$

于是就有 n 个这样的方程, 从而构成 k 元线性方程组。用矩阵表示为

$$Y_{(2m+1) \times 1} = X_{(2m+1) \times k} \cdot A_{k \times 1} + E_{(2m+1) \times 1} \quad (3)$$

A 的最小二乘解为

$$\hat{A} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (4)$$

Y 的模型预测值 (滤波值) 为

$$\hat{Y} = X \cdot A = X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Y = B \cdot Y \quad (5)$$

$$B = X \cdot (X^T \cdot X)^{-1} \cdot X^T \quad (6)$$

通过确定滤波带宽 n 和多项式的最高次幂 ($k-1$), 即可得到相应的滤波公式。

设 $n=5$, $k=3$, 得到 5 点 2 次滤波模型:

$$\hat{Y}_{-2} = \frac{1}{35}(31y_{-2} + 9y_{-1} - 3y_0 - 5y_1 + 3y_2) \quad (7)$$

$$\hat{Y}_{-1} = \frac{1}{35}(9y_{-2} + 13y_{-1} + 12y_0 + 6y_1 - 5y_2) \quad (8)$$

$$\hat{Y}_0 = \frac{1}{35}(-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2) \quad (9)$$

$$\hat{Y}_1 = \frac{1}{35}(-5y_{-2} + 6y_{-1} + 12y_0 + 13y_1 + 9y_2) \quad (10)$$

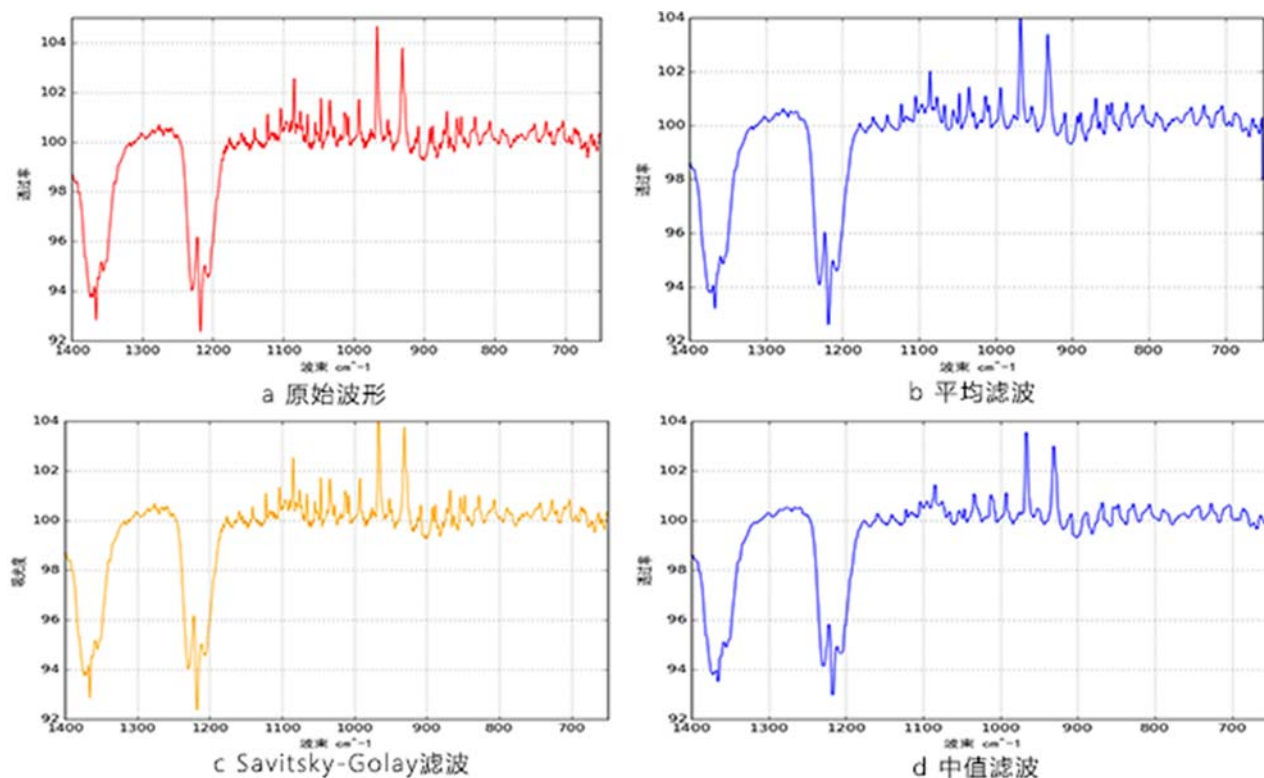


图 4 针对丙酮 - 透过率的三种滤波算法对比情况

$$\hat{Y}_2 = \frac{1}{35}(3y_{-2} - 5y_{-1} - 3y_0 + 9y_1 + 31y_2) \quad (11)$$

式中, $\hat{Y}_{-2}$ 、 $\hat{Y}_{-1}$ 、 $\hat{Y}_1$ 和 $\hat{Y}_2$ 为当前窗口两端的滤波值; $\hat{Y}_0$ 为窗口中心点的滤波值。

2.1.3 中位数平滑法

中位数平滑法 (中值滤波) 是一种基于排序统计理论的可有效抑制噪声的非线性信号处理技术。它对脉冲干扰级椒盐噪声的抑制效果较好, 在抑制随机噪声的同时能够有效保护边缘少受模糊。使用这种方法得出的结果受外界值的影响较小, 其计算过程与窗口移动平均值算法一致:

$$y_k^* = \text{median}(y_{k+j}), j = -m, -m+1, \dots, m-1, m \quad (12)$$

图 4 所示为通过软件模拟对以上三种滤波算法进行验证的效果对比情况。其中, 图 4(a) 所示为原始波形, 图 4(b) 所示为经平均滑动滤波处理后的波形, 图 4(c) 所示为经 S-G 滤波处理后的波形, 图 4(d) 所示为经中值滤波处理后的波形。由此可见, 平均滑动平均滤波、S-G 滤波

和中值滤波算法虽然在不同程度上对原始波形数据滤波有所优化, 但对较大的干扰信号只能起到一定程度的抑制作用, 且对波形平滑的效果存在局限性。基于这三种滤波算法, 通过微分方法进行搜峰, 物质识别的准确率约为 86%。

2.2 基于卡尔曼滤波的改进方法

谱图检索非常重视最强峰, 它是该化合物的最显著特征。根据萨特勒的谱线索引方法, 以谱图中的最强峰为抓手, 将全范围分成若干小区域, 然后按规则将各峰填入小区, 并查表加以核对。基于这种方法, 我们在保留最强峰的前提下, 尽可能地抑制波形上叠加的其他干扰信号, 以减少由于干扰信号给谱图检索带来的误差。

卡尔曼滤波器^[2]是一种“最优化自回归数据处理算法”。它可以写成一个方程, 但它经常被理解为“预测”和“更新”两个不同阶段。在预测阶段, 基于以前的时间状态在当前时间产生一个估计的状态。这一预测状态估计也被称为先验状态估计。它是一个当前的估计状态, 而不是当前的观测状态。在更新阶段, 将当前的先

验状态估计与当前观测状态相结合, 以完善的的状态进行估计。这种改进的估计被称为后验状态估计。

卡尔曼滤波的五个基本公式如下:

(1) 预测 (先验) 状态估计

$$\hat{X}_{k|k-1} = F_k \hat{X}_{k-1|k-1} + B_k U_k \quad (13)$$

式中, $\hat{X}_{k|k-1}$ 为预测上一个状态的结果; $\hat{X}_{k-1|k-1}$ 为上一个状态的最优结果; U_k 为状态控制量; 当不需要控制量时, U_k 可设为 0。

(2) 预测 (先验) 估计协方差

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \quad (14)$$

式中, P 为协方差; $P_{k|k-1}$ 为 $\hat{X}_{k|k-1}$ 的协方差; F_k^T 为 F_k 的转置矩阵; Q_k 为系统过程的协方差。

(3) 最优卡尔曼增益

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (15)$$

通过结合预测值和测量值, 可以得到现在状态 (k) 的最优化估算值 $\hat{X}_{k|k}$, 由此可以推导出卡尔曼增益。

(4) 更新 (后验) 状态估计

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \tilde{y}_k \quad (16)$$

k 状态下的最优估算值为 $X(k|k)$ 。

(5) 更新 (后验) 估计协方差

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \quad (17)$$

式中, I 为 1 的矩阵, 单模型、单测量中 $I=1$ 。

由图 5 可见, 经卡尔曼滤波后, 整个波形数据有了较大的优化, 波形上的干扰信号得到了明显抑制, 但还是会存在一些较为明显的毛刺信号。由此我们基于卡尔曼滤波, 在对波形数据进行卡尔曼滤波之前, 采用维纳滤波对其进行叠加滤波处理。

维纳滤波^[3]是一种用于从噪声中提取信号的过滤 (或滤波) 方法。这种线性滤波问题可以

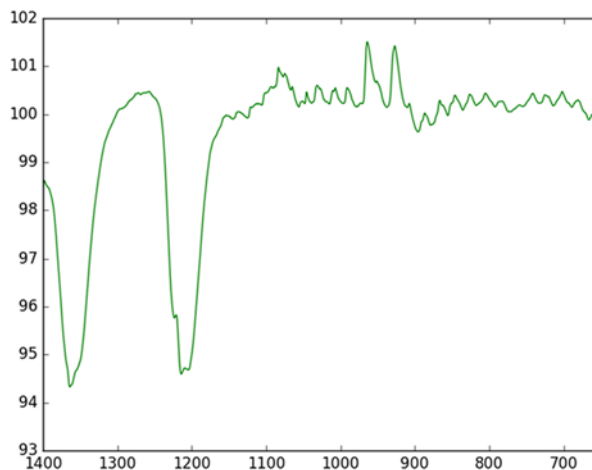


图 5 经卡尔曼滤波后的丙酮 - 透过率波形

看做是一种估计问题或者线性估计问题。

如果一个线性系统的单位样本响应为 $h(n)$, 并且输入一个随机信号 $x(n)$:

$$x(n) = s(n) + v(n) \quad (18)$$

则输出 $y(n)$ 为

$$y(n) = \sum_m h(m)x(n-m) \quad (19)$$

式中, $x(n)$ 为信号; $v(n)$ 为噪声。图 6 所示为维纳滤波器的输入 - 输出关系。



图 6 维纳滤波器的输入 - 输出关系

$x(n)$ 通过线性系统 $h(n)$ 后得到的 $y(n)$ 尽量接近于 $s(n)$, 因此 $y(n)$ 称为 $s(n)$ 的估计值, 用 $\hat{s}(n)$ 表示。

我们将噪声估计设定为输入数据的局部方差平均值。图 7 所示为用软件模拟出的经维纳滤波、改进型卡尔曼 (卡尔曼/维纳) 滤波后的波形图形。从中可以看出, 经过卡尔曼/维纳叠加滤波后, 在保留最大特征峰的前提下, 整体波形数据在平滑度上比之前其他算法的滤波效果有了明显提升。

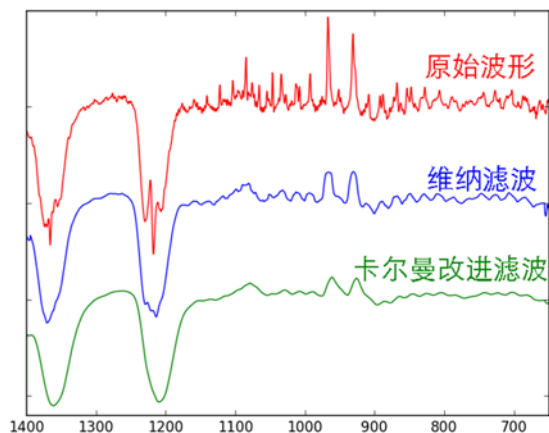


图 7 采用卡尔曼/维纳叠加滤波方法得到的丙酮-透过率波形

2.3 基线校准

受仪器背景、样品粒度以及其他因素的影响,近红外分析中常常会出现基线漂移和倾斜现象。通过基线校正可以有效地消除这些影响。

我们采用基于非对称最小二乘法的基线校正^[4]。非对称最小二乘法的优化目标如下:

$$S = \sum_i w_i (y_i - z_i)^2 + \lambda \sum_i (\Delta^2 z_i)^2 \quad (20)$$

式中, w_i 为权重因子; Δ 为二阶差分算子; λ 为正则化参数; $\sum_i w_i (y_i - z_i)^2$ 为拟合函数与原始数据的非对称拟合程度; $\lambda \sum_i (\Delta^2 z_i)^2$ 用于保证拟合函数的光滑性。

基于 Python 软件模拟对基线校正算法进行验证:

```
def baseline_als(y, lam, p):
    L = len(y)
    D = sparse.csc_matrix(np.diff(np.eye(L), 2))
    w = np.ones(L)
    for i in xrange(10):
        W = sparse.spdiags(w, 0, L, L)
        Z = W + lam * D.dot(D.transpose())
        z = sparse.linalg.spsolve(Z, w*y)
        w = p * (y > z) + (1-p) * (y < z)
    return z
```

图 8 和图 9 为运行效果图。

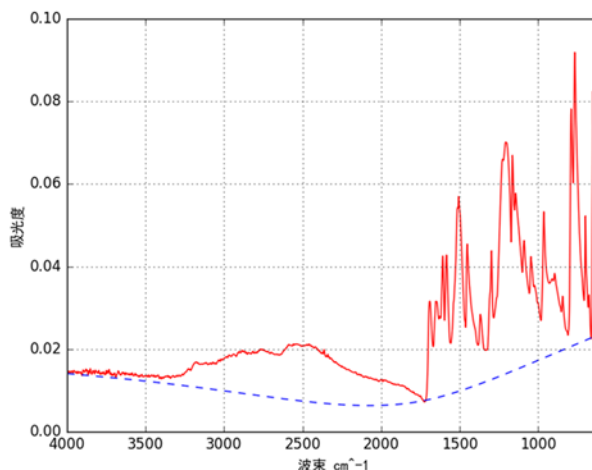


图 8 经计算得出的基底位置上的乙酰邻氨基苯甲酸 100 - 吸光度

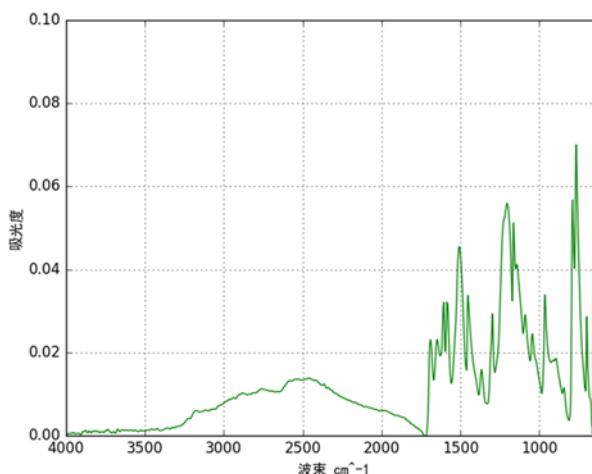


图 9 扣除波形基底后的乙酰邻氨基苯甲酸 100 - 吸光度

2.4 谱峰定位

作为一种非常有用的信号处理技术,信号求导(微分法)可以消除信号的背景,确定谱峰位置,改善谱峰分辨。信号的一阶和二阶导数分别为

$$\left(\frac{\partial y_i}{\partial i} \right)_{i=0} = a_1 \quad (21)$$

$$\left(\frac{\partial^2 y_i}{\partial i^2} \right)_{i=0} = 2! \cdot a_2 \quad (22)$$

则利用 7 点三次多项式求得窗口中心所对应的一阶和二阶导数的计算公式为

$$\left(\frac{\partial y}{\partial i} \right)_{i=0} = \frac{1}{252} (22y_{-3} - 67y_{-2} - 58y_{-1} + 58y_1 + 67y_2 - 2y_3) \quad (23)$$

一阶导数等于 0 的地方即为原始波形的峰值。

$$\left(\frac{\partial^2 y}{\partial i^2}\right)_{i=0} = \frac{1}{252}(15y_{-3} - 9y_{-1} - 12y_0 - 9y_1 + 15y_3) \times 2 \quad (24)$$

二阶导数光谱的谱峰方向与原谱相反, 其最小值为原始波形的峰值。通过二阶导数可以搜索到原始波形的半肩峰、隐藏峰等用常规方法很难搜到的峰位。

在利用谱峰定位算法得到当前谱峰有哪几个主要峰后, 需要将其谱峰与样本库中的信息进行比对, 这样才能确认物质的最终成分。根据 Webbook^[5] 化学数据库所提供的红外吸收光谱, 我们通过比对可以找到相应的物质。图 10、图 11 和图 12 所示分别为未经数据预处理的、标准库中的以及经数据预处理后的氯仿-透射率。

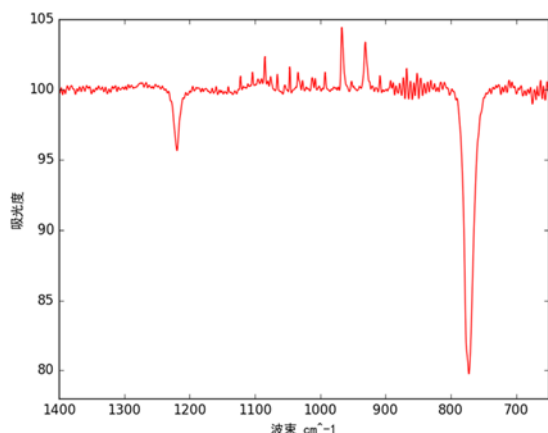
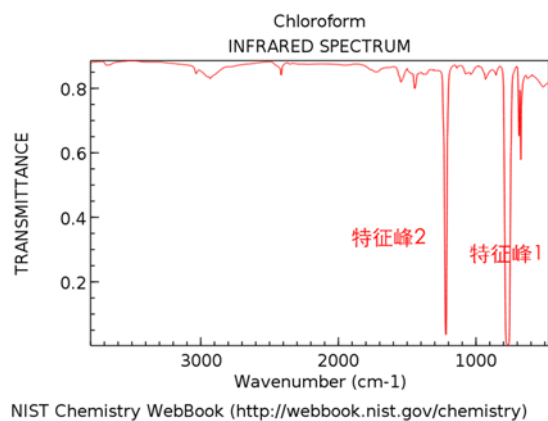


图 10 未经数据预处理的氯仿-透射率



NIST Chemistry WebBook (<http://webbook.nist.gov/chemistry>)

图 11 标准库中的氯仿-透射率

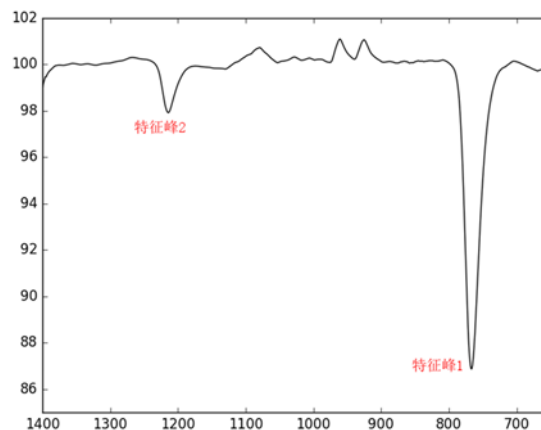


图 12 经数据预处理后的氯仿-透射率

3 结束语

通过对红外光谱数据进行一系列预处理, 可以将谱图检索对物质定性的准确率提升至 93%。因此, 我们认为在红外光谱遥测设备中使用这种光谱数据预处理方法是行之有效的。

随着光谱数据预处理技术的不断发展, 被动红外光谱遥测技术对在易制毒化学品制作过程中产生的挥发气体的识别效率和准确性会有较大提高。如果此项技术能够服务于公安一线, 那么将可从源头上有效打击国内毒品生产窝点, 同时对非法制造冰毒、氯胺酮等合成毒品的管控和缉查工作也能起到很好的助力作用。

参考文献

- [1] 张记龙, 聂宏斌, 王志斌, 等. 化学战剂红外光谱遥测技术现状及发展趋势 [J]. *中北大学学报*, 2008, **29**(3): 265-271.
- [2] Kalman R E. A New Approach to Linear Filtering and Prediction Problems [J]. *Transactions of the ASME-Journal of Basic Engineering*, 1960, **82**(1): 35-45.
- [3] Izzetoglu M, Devaraj A, Bunce S, et al. Motion Artifact Cancellation in NIR Spectroscopy Using Wiener Filtering [J]. *IEEE Trans Biomed Eng*, 2005, **52**(5): 934-938.
- [4] Peng J T, Peng S L, Jiang A, et al. Asymmetric Least Squares for Multiple Spectra Baseline Correction [J]. *Analytica Chimica Acta*, 2010, **683**(1): 63-68.
- [5] NIST: NIST Chemistry WebBook [DB/OL]. <http://webbook.nist.gov/chemistry>, 2011.