

基于光谱融合的油漆分类方法比较

古锬山, 王继芬*

中国人民公安大学侦查学院, 北京 100038

摘要 为了实现对油漆物证的快速、无损以及准确分类, 实验收集了犯罪现场常见的 5 个油漆品牌共计 50 个油漆样本的红外原始光谱数据和导数光谱数据, 结合光谱融合技术, 建立了基于 KNN、SVM 以及逐步判别分析的油漆分类模型。实验结果表明, 3 种分类模型对于融合光谱的识别率要高于单一光谱; KNN 以及 SVM 分类模型对于其中的 3 种油漆样本识别率高, 但对其余 2 种样本的分类效果并不好, 而逐步判别分析模型对 5 种油漆样本的各种光谱数据识别率均高于 KNN 和 SVM 模型, 其中采用逐步判别分析中的 Smallest F ratio 判别模型对一阶导数光谱和三阶导数光谱融合数据的训练集和测试集实现了完全识别。本文方法的检验效率高, 定性能力强, 满足公安机关对于相关物证的快速检验要求, 为刑事技术人员快速识别油漆物证提供了一种有效的手段。

关键词 光谱学; 油漆; 光谱融合; 化学计量学

中图分类号 O433

文献标志码 A

doi: 10.3788/LOP202158.2230002

Comparison of Paint Classification Methods Based on Spectral Fusion

Gu Kunshan, Wang Jifen*

School of Investigation, People's Public Security University of China, Beijing 100038, China

Abstract To achieve fast, non-destructive, and accurate classification of paint evidence, the infrared raw and derivative spectral data of fifty paint samples from five common crime scenes were collected. The paint classification models based on KNN, SVM, and stepwise discriminant analysis were created using spectral fusion technology. The experimental results show that the three classification models for the fusion spectrum have a higher recognition rate than a single spectrum. The recognition rate of the KNN and SVM classification model for three paint samples is high, but the classification effect for the remaining two samples is not good. The recognition rate of the stepwise discriminant analysis completely model for all kinds of spectral data of five paint samples is higher than that of the KNN and SVM models. To achieve 100% recognition of the training and test sets, the Smallest F ratio discriminant model of stepwise discriminant analysis identifies the first derivative and third derivative spectral fusion data. This method has high efficiency and strong qualitative ability, and it meets the requirements for rapid inspection of relevant material evidence by public security organs. It also provides criminal technicians a quick way to identify paint evidence.

Key words spectroscopy; paint; spectral fusion; stoichiometry

OCIS codes 300.6170; 300.6300; 300.6340

1 引言

油漆是一种能起保护装饰、绝缘、降低材料腐蚀速度、耐高温等作用的化学混合涂料, 油漆主要是由

主要成膜物质、次要成膜物质以及辅助成膜物质构成。油漆在家庭以及社会的各个方面都具有广泛的应用, 因此在各种类型的案件现场中经常会遗留下微量的油漆碎片, 这使得油漆物证的快速分析成为

收稿日期: 2020-12-27; 修回日期: 2021-01-23; 录用日期: 2021-02-04

基金项目: 中国人民公安大学国家安全高精尖学科高端论文项目(2020GDLW037)

通信作者: *wangjifen58@126.com

犯罪现场物证分析的重中之重,例如:在各类入室盗窃案件中,嫌疑人使用工具撬门锁时,门锁上的油漆可能会脱落到作案工具上;在交通事故现场,车辆之间的碰撞以及汽车与其他交通工具或行人碰撞都会导致油漆物证的转移或脱落。检验现场遗留物是否为油漆物证以及遗留的油漆物证与比对样品是否相同,可为案件分析提供重要的线索证据,缩小侦查范围。

各鉴定机构对于油漆物证的鉴定大多是基于自身的实验经验、实验设备等条件而采取不同的方法进行检验鉴定。雷平^[1]利用扫描电镜/能谱法对油漆物证进行检验,实验结果表明该方法可以得到油漆的层数、厚度、颗粒程度,以及每层油漆的元素成分。孙其然等^[2]利用裂解气相色谱-质谱法(Py-GC/MS)将油漆瞬间热解成小分子碎片,然后根据这些碎片对油漆的结构进行分析,实验结果表明该方法对油漆的解析能力强且能够得到油漆的高分子单体结构。另外,常用的油漆物证鉴定方法还有热分析法^[3]、原子发射光谱法^[4]等。虽然上述方法在灵敏度、检测速度等方面表现优秀,但是这些方法或耗时较长,或需要昂贵的仪器,无法将其运用于现场检材的快速筛查,影响办案效率;且有些方法还会消耗待测样品,不适用于微量油漆物证的检验。红外光谱法具有分析速度快,不损坏样品,可提供检测物的丰富结构信息并且特征性强,不损坏样品等优点,与微量物证对仪器分析的要求完美匹配^[5-7],已被广泛用于油漆等微量物证的检验鉴定^[8-9],但当需要比对的样本数较多时,各红外吸收峰交叉混合,分析结果受检测者主观性干扰较大,且耗时费力、准确率低^[10]。

光谱融合可以认为是将多个光谱信息进行整合,即将不同光谱之间的优势进行互补,从而获得高质量且相关性更高的光谱。利用光谱融合技术能够克服单一光谱信息缺陷并将光谱信息进行整合,从而提高模型识别的准确度^[11]。李亚惠等^[12]建立了一种黑果枸杞识别的模型,并比较了该模型对于黑果枸杞近、中红外光谱融合和单一光谱的识别率,结果表明该模型对于黑果枸杞融合后的数据识别率可以达到 100%,能够达到快速准确识别黑果枸杞相关信息的目的;Feng 等^[13]将拉曼光谱和红外光谱进行融合并与偏最小二乘法相结合,构建了定量分析土壤中多环芳烃的模型,实验结果表明该模型对于土壤中葱的决定系数为 0.9514,均方根误差为 0.9143 mg/g,对土壤中荧葱的决定系数为 0.9609,

均方根误差为 0.8614 mg/g,与单一的红外光谱以及拉曼光谱相比,融合光谱对于土壤中多环芳烃的预测性能更佳。

为了实现油漆物证的快速有效分类,本研究以油漆样本的原始红外光谱数据以及其一阶、二阶和三阶导数红外光谱数据为基础,利用光谱融合技术并结合数学建模的方法,对比了 K 近邻分析、支持向量机、逐步判别分析方法对油漆物证进行区分的准确性,为现场油漆物证的快速检验提供一定的参考。

2 实验部分

2.1 实验样本

收集不同类型、不同品牌的油漆样本共 50 个,包含 7 个聚氨酯漆样本、9 个不饱和树脂漆样本、8 个硝基漆样本、19 个光固化涂料和 7 个水性涂料,分别编号为 Y1、Y2、Y3、Y4 和 Y5。

2.2 主要仪器及设备

所用的仪器主要包括 Nicolet 5700 型傅里叶变换红外光谱仪(美国 Thermo Fisher Scientific 公司)和 KBr 分束器(美国 Thermo Fisher Scientific 公司)。扫描次数设为 32 次,光谱分辨率为 4 cm^{-1} ,收集 $1800\sim 500\text{ cm}^{-1}$ 波数范围的谱图,并且每个样本采集 3 次,最终取平均值作为该样本的最终谱图数据^[14]。

2.3 实验条件及预处理

在温度为 $(25\pm 3)\text{ }^\circ\text{C}$,相对湿度为 80% 的实验条件下,对谱图进行自动基线校准^[15]、峰面积归一化^[16]、多元散射校正^[17]后,分别用 S-G 算法平滑^[18]得到原始、一阶、二阶、三阶导数光谱图。

3 实验建模

3.1 K 近邻分析

K 近邻分析(KNN)是机器学习工具箱中最常见、最直接的方法之一。KNN 概念的简单性使其成为不同应用领域的最佳分类工具。作为模式识别中基于实例的学习方法,KNN 可以根据特征空间中最接近的训练示例对研究案例的每个元素进行排序^[19]。其过程主要是对一个给定的原始数据集 m 进行训练,当输入新的数据 α 时,在原始数据集 m 中提取出与 α 相近的 N 条近邻信息,假设在这 N 条信息中,属于 Ω_1 类型的有 S_1 个,属于 Ω_2 类型的有 S_2 个,以此类推,属于 Ω_i 类型的有 S_i 个。存在 $S_i \in N(i=1,2,\dots,t)$,任取 $S_x \in N(x=1,2,\dots,t)$

都有 $S_i \geq S_x$ 时,此时 α 属于 Ω_i 类型。对本实验来说,利用 KNN 模型对给定油漆样本数据集 m 进行训练,当输入需要识别的油漆数据时,会在数据集中给出 N 条近邻信息,在这些信息里数量最多的类型即为输入油漆样本的类别。

3.2 支持向量机

支持向量机(SVM)是一种基于统计学习理论、VC 维数理论的二分类模型。SVM 能够有效地处理回归、分类、判别分析等问题,因此被广泛应用于预测和综合评价,在小样本、高维的模式识别问题上有着出色的表现^[20]。SVM 的主要原理是在高维数据中构建一个超平面,通过该超平面将不同类型的数据区分开,最终转化为求凸二次规划问题的最优解^[21]。对于线性数据,通过下列变换可以得到线性 SVM 的分类决策函数表达式。假设训练样本集为 $(\mathbf{x}_t, y_t), t=1, 2, 3, \dots, l, \mathbf{x} \in \mathbf{R}^n, y \in \{-1, +1\}$, 并假设训练集是线性的,存在超平面 $(\mathbf{w} \cdot \mathbf{x}) + b = 0$, 使得 $y_t [(\mathbf{w} \cdot \mathbf{x}_t) + b] \geq 1$, 根据计算的分类间隔得到的最优超平面问题的约束式为

$$\min \varphi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2, \quad (1)$$

式中: $\varphi(\mathbf{w})$ 表示超平面; \mathbf{w} 为法向量, 决定超平面的方向。

为求解上述约束式, 引入拉格朗日函数

$$L = \frac{1}{2} \|\mathbf{w}\|^2 - \alpha_i \{ [y_i (\mathbf{w} \cdot \mathbf{x}_i) + b] - 1 \}, \quad (2)$$

式中: α_i 为大于 0 的 Lagrange 乘子。将任意一个支持向量和 KKT 条件 $\alpha_i \{ [y_i (\mathbf{w} \cdot \mathbf{x}_i) + b] - 1 \} = 0$ 结合可以求出 $b = y_i - (\mathbf{w} \cdot \mathbf{x}_i)$ 。最后将 \mathbf{w}, b 代入 $(\mathbf{w} \cdot \mathbf{x}_i) + b = 0$ 得到线性 SVM 的分类决策函数表达式:

$$f(x) = \text{sign}(\mathbf{w} \cdot \mathbf{x}_i + b). \quad (3)$$

然而, 光谱数据大多属于非线性数据, 对于此类数据, 需要通过核函数进行非线性变换, 将原本非线性的数据转化为某个空间维度上的线性分类, 通过一系列变换可以得到非线性 SVM 分类决策函数的表达式

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i \beta_i k(\mathbf{x}_i, \mathbf{x}_j) + b^* \right], \quad (4)$$

式中: $\beta_i \in \{0, 1\}$ 表示样本类别; $k(\mathbf{x}_i, \mathbf{x}_j)$ 为所选的核函数。

3.3 逐步判别分析

逐步判别分析是统计学上常用的判别方法, 其模型采用输入一个变量又输出一个变量的动态调节

方法, 即每次输入变量后都会进行检验, 将判别能力强的变量引入判别函数, 而将判别能力弱的变量及时输出, 使得最终的判别函数既简单又具备强的判别能力^[22]。

根据判别准则的不同, 逐步判别分析有以下 5 种方法, 分别是 Wilks' Lambda、Unexplained variance、Mahalanobis distance、Smallest F ratio 以及 Rao's V。Wilks' Lambda 法是在所有的输入变量中寻找使模型 Wilks' Lambda 统计量值最小的变量; Unexplained variance 法是在所有未知变量中比较出使得各组之间的未解释方差值总和最小的变量; Mahalanobis distance 法是在变量中寻找使得各类间的马氏距离最大的变量; Smallest F ratio 法是在输入的变量中寻找使得各类间的最小 F 值最大的变量; Rao's V 法是在输入的变量中寻找出使得模型 V 值增幅最大的变量。

4 结果与讨论

4.1 光谱分析和光谱融合

50 个油漆样本的原始光谱、一阶导数光谱、二阶导数光谱以及三阶导数光谱如图 1 所示。观察图 1(a) 的原始光谱得知: 在波数为 3000 cm^{-1} 左右处有一个左高右低的尖峰, 在波数为 1720 cm^{-1} 处均有一个较强的尖峰, 在波数为 1250 cm^{-1} 处均存在一个强吸收的尖峰, 在波数为 2800 cm^{-1} 左右有一个左高右低的中强峰, 在波数 1100 cm^{-1} 处有一个左低右高的尖峰, 而在波数 715 cm^{-1} 处有一个左高右低的尖峰, 在波数为 1500 cm^{-1} 左右有一个较弱的双峰。观察图 1 可知, 油漆样本的 4 种红外谱图出峰位置基本一致, 但各谱图的峰高、峰强有所差异, 难以通过直接观察进行区分。

由图 1 可知, $1800 \sim 500 \text{ cm}^{-1}$ 以及 $3000 \sim 2700 \text{ cm}^{-1}$ 波数区域的峰形比较明显, 特征性强, 包含了能够反映样本之间差异的重要信息。按层次分类, 光谱融合的方法可分为初级、中级以及高级 3 个层次。初级层次的光谱融合是将光谱数据经过一系列预处理后进行简单叠加, 以此达到丰富光谱信息的目的; 中级层次的光谱融合是通过提取光谱数据中的特征信息后进行叠加整合, 从而实现精简数据、提高分析准确度的目的; 高级层次的光谱融合是要将每种数据进行回归分析并将结果进行组合, 从而得到最优融合模式^[23]。本文采用初级层次的光谱融合方法对油漆样本的红外光谱进行融合。分别将 50 个油漆样本的红外光谱波数为 $1800 \sim 500 \text{ cm}^{-1}$

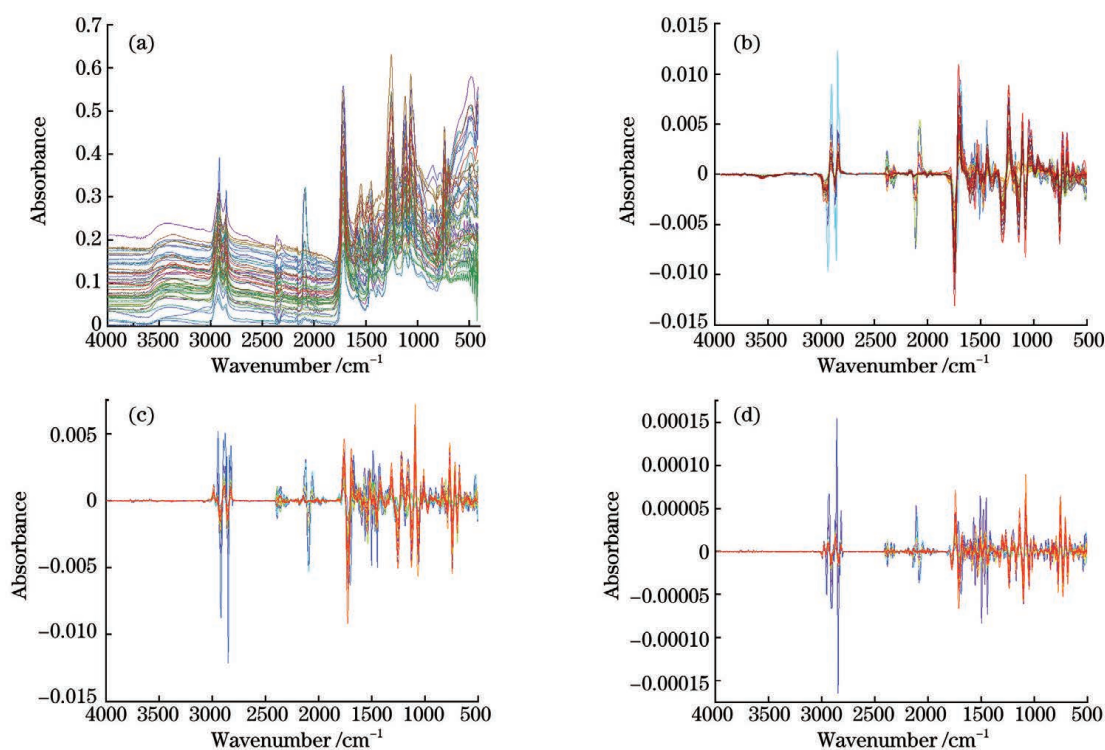


图 1 50 个油漆样本的 4 种红外谱图。(a)原始光谱;(b)一阶导数光谱;(c)二阶导数光谱;(d)三阶导数光谱

Fig. 1 Four infrared spectra of 50 paint samples. (a) Original spectra; (b) first derivative spectra; (c) second derivative spectra; (d) third derivative spectra

以及 $3000 \sim 2700 \text{ cm}^{-1}$ 区域的原始谱图与一阶导数谱图、原始谱图与二阶导数谱图、原始谱图与三阶导数谱图、一阶导数谱图与二阶导数谱图、一阶导数谱图与三阶导数谱图以及二阶导数谱图与三阶导数谱图首尾串行叠加,得到 6 种光谱融合数据集,分别用 G1、G2、G3、G4、G5 以及 G6 表示上述融合光谱数据。

4.2 KNN 分析结果

KNN 算法的基本构成元素有 K 值大小、KNN 分类器的距离以及对样本进行分类判别的规则^[24]。对于所建立的 KNN 判别模型,上述 3 个基本元素里 K 值大小的选择是 KNN 判别模型识别率的最主要影响因素。选取的 K 值较小会使得模型变得复杂,导致各类数据发生拟合的概率增加;相反,选取的 K 值较大就会使得模型变得简单,导致各类样本的误判率增加。本实验采用交叉验证法来选择最优 K 值,该 KNN 模型 K 值的选择错误率结果如图 2 所示。观察图 2 可以发现,模型的分类错误率随着 K 值的增加而不断变化,在 K 值大于 10 以后模型的错误率稳定为 0.688,当 $K=1$ 时模型的错误率最低,只有 0.152,因此本实验中 KNN 模型的 K 值为 1。在 $K=1$ 的条件下分别构建原始光谱、导数光谱以及融合光谱的 KNN 模型。该模型对于

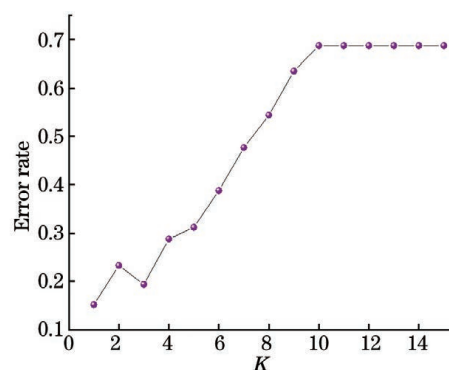


图 2 K 值选择对错误率的影响

Fig. 2 Effect of K value selection on error rate

不同样本的分类结果如表 1 所示。

由表 1 可知,当 $K=1$ 时 KNN 分类模型对 Y2 样本实现了全部分类,平均分类准确率达到 100%;对于 Y3 和 Y4 样本的平均分类准确率也较高,分别为 98.8% 和 82.5%;但对于 Y1 和 Y5 样本来说,其平均分类准确率不理想,只有 45.7% 和 27.3%。分析认为,KNN 模型对于 Y1 和 Y5 样本识别率低可能是受样本不均匀的影响,即有的样本少,有的样本多,少的样本就有可能被误判到多的样本中。在本实验中,不同油漆样本的数量不一致,如 Y4 样本有 19 个,Y1 和 Y5 样本数量最少,只有 7 个,这就使得 KNN 模型在预测样本时偏向于 Y4 样本,从而导致

表 1 $K=1$ 时各光谱数据对油漆样本的分类识别率

Table 1 Classification and recognition rate of paint samples for each spectral data when $K=1$

Spectral type	Accuracy /%				
	Y1	Y2	Y3	Y4	Y5
Original spectra (OG)	14	100	88	84	14
1st derivative spectra (FD)	57	100	100	74	14
2nd derivative spectra (SD)	57	100	100	89	43
3rd derivative spectra (TD)	57	100	100	74	29
G1	29	100	100	79	14
G2	29	100	100	89	29
G3	43	100	100	84	29
G4	71	100	100	89	43
G5	57	100	100	79	29
G6	43	100	100	84	29
Average accuracy	45.7	100	98.8	82.5	27.3

误判数量增加。而 Y2 及 Y3 样本虽然分别只有 9 个和 8 个,但 KNN 模型对于这两类样本的识别率仍比较理想。分析认为,Y2 及 Y3 样本虽然数量少,但其具备与 Y4 样本不同的特征多,且其特征均比较容易被 KNN 模型识别,这使得这两种样本虽然数量和 Y1 及 Y5 样本差不多,但识别率远高于 Y1 及 Y5 样本。

观察图 3 可知,光谱融合后对于油漆样本的总体分类识别率要比未融合更加稳定,且光谱融合后的平均识别率要比未融合的高 1.3%,说明利用光谱融合后的数据进行 KNN 模型识别的准确率以及稳定性相比于原始数据都要更加出色。G4(一阶导数光谱与二阶导数光谱融合)的总体分类识别率最高,达到 84%,远高于原始数据的 68%。由此可见,光谱融合后的数据更加适合于 KNN 模型的判别。

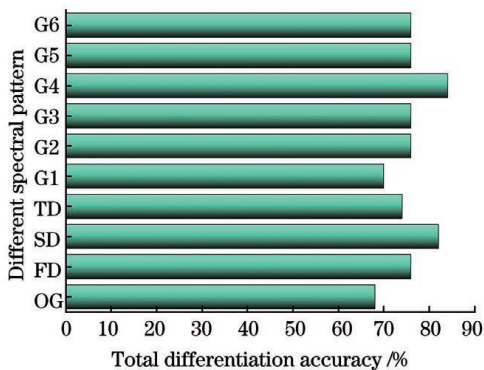


图 3 5 种油漆样本在 10 种光谱数据模型下的总体分类识别率

Fig. 3 Overall classification and recognition rate of five paint samples under 10 spectral data models

4.3 SVM 分类结果

SVM 中常用的核函数有多项式 (polynomial)、径向基函数 (RBF)、拉普拉斯 (Laplacian)、Sigmoid 等^[25],在预实验中比较了上述 4 种核函数对于油漆样本的最小分类误差,实验结果如图 4 所示。

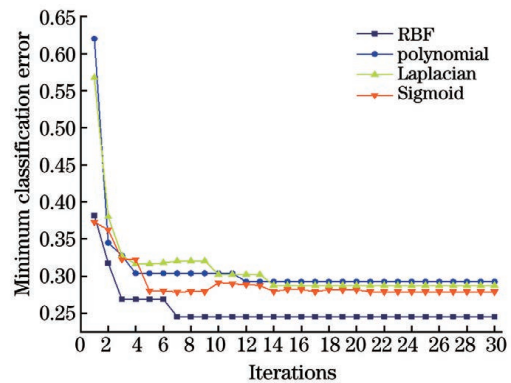


图 4 4 种核函数的最小分类误差统计

Fig. 4 Statistical results of minimum classification errors of four kernel functions

由图 4 可知:当迭代数较少时,各核函数的最小分类误差值随着迭代次数的改变而发生变化;当迭代次数逐渐增大时,各核函数的最小分类误差值逐渐趋于稳定。以多项式作为核函数时,最小分类误差值最大,为 0.2925;以 RBF 作为核函数时,最小分类误差值最小,为 0.24533。因此,本实验采用 RBF 作为 SVM 的核函数,SVM 的其余参数均为 MATLAB R2019a 软件默认参数(γ 值为 $1/k$, k 为样本类型数,本实验 $k=5$,即 γ 值为 0.2;惩罚系数 C 和回归精确度均为 0.1),RBF 的表达式为

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\xi^2}\right), \quad (5)$$

式中: x_j 为 RBF 核函数的中央; ξ 为 RBF 核函数中与宽度有关的参数, 该参数将会控制 RBF 核函数沿半径或直径的作用范围。当 $\|x_i - x_j\|$ 无限接近于 0 时, RBF 核函数的取值为 1; 当 $\|x_i - x_j\|$ 远大于 0 时, RBF 核函数的取值为 0。由于 RBF 核函数的函数类型与高斯分布非常类似, 因此 RBF 核函数又被称为高斯核函数^[26]。

对 5 种油漆样本的各类光谱数据集构建以 RBF 为核函数的 SVM 模型进行分类识别, 该模型对各类样本的识别率如图 5 所示, SVM 对各类样本的平均分类识别率如图 6 所示。

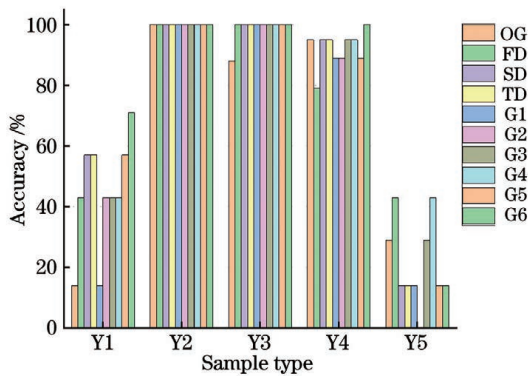


图 5 SVM 对不同光谱数据集的油漆样本识别率

Fig. 5 Recognition rate of paint samples from different spectral data sets by SVM

观察图 5 可知, 基于光谱融合的 SVM 分类模型对油漆样本的总体识别率要优于基于导数光谱的 SVM 分类模型, 基于导数光谱的 SVM 分类模型对油漆样本的总体识别率又优于基于原始数据的 SVM 分类模型。光谱融合能让各数据进行相互补充, 让每个

表 2 5 种判别分析模型在不同光谱数据下的训练集识别率以及测试集的识别率

Table 2 The recognition rate of training sets and test sets of five discriminant analysis models under different spectral data

Spectral type	Accuracy / %									
	Wilks' Lambda		Unexplained variance		Mahalanobis distance		Smallest F ratio		Rao's V	
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
OG	94	78	94	88	94	88	94	86	92	76
FD	90	84	94	90	94	86	100	90	92	72
SD	96	90	96	90	100	96	100	90	90	84
TD	98	90	98	88	100	96	98	94	98	86
G1	94	88	92	86	100	98	100	90	92	72
G2	96	90	96	90	100	96	100	90	90	84
G3	98	88	96	88	100	96	100	92	98	86
G4	90	86	100	90	100	96	100	94	96	82
G5	96	84	96	88	98	94	100	100	88	80
G6	92	84	96	88	100	98	100	96	96	86

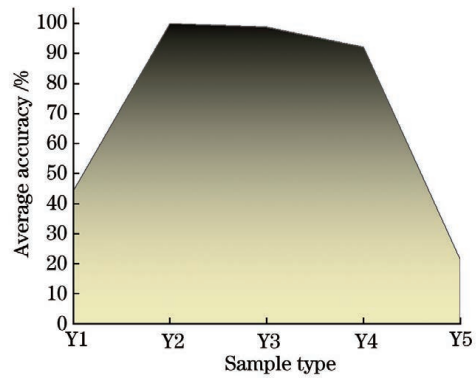


图 6 SVM 对各类样本的平均分类识别率

Fig. 6 The average classification and recognition rate of all kinds of samples by SVM

样本所具备的特征更加突出, 使得 SVM 对各样本的识别率更加稳定。其中 G6(二阶导数光谱与三阶导数光谱融合)的总体分类识别率最高, 可以达到 84%。

由图 6 可知: 无论是哪种光谱数据, 该模型对于 Y2、Y3 和 Y4 样本的分类效果都非常好, 对 Y2 样本更是实现了完全识别, 对 Y3 和 Y4 样本的识别率也高达 98.8% 和 92.1%; 但是对于 Y1 和 Y5 样本的识别率并不是特别好。这可能是由于 Y1 和 Y5 样本的数据维度过高, 以 RBF 核函数的 SVM 模型对于解决高维映射问题的能力有所欠缺, 导致该模型难以区分出 Y1 以及 Y5 样本。

4.4 逐步判别分析结果

按照训练集和测试集数量 4:1 的比例, 随机在每种油漆样本中选择两个样本作为测试集, 对于测试集的识别率采用交叉验证的方法进行验证, 其余样本用作训练集。以光谱数据集为基础, 分别构建 5 种判别分析模型, 5 种判别分析模型在不同光谱数据下的训练集识别率以及测试集的识别率如表 2 所示。

从表 2 可以看到,5 种判别模型对于训练集的认识率非常高,普遍可以达到 90% 以上,对于测试集的认识率大部分都能超过 80%。使用 Mahalanobis distance 判别方法的判别模型可以实现对于二阶导数光谱、三阶导数光谱、G1、G2、G3、G4 以及 G6 光谱和使用 Smallest F ratio 判别方法的判别模型对于一阶导数光谱、二阶导数光谱、G1、G2、G3、G4、G5 以及 G6 光谱的训练集的完全识别。使用 Smallest F ratio 判别方法的判别模型对于 G5

光谱更是可以实现对训练集以及测试集的完全识别。

下面将进一步研究 G5 光谱的 Smallest F ratio 判别模型。将 Fisher 函数作为判别函数的系数,采用 Box's M 工具统计量进行描述以及各种油漆样本的协方差矩阵的检验,最后用 F 检验的概率来决定变量是否能够加入最终的判别函数^[27]。表 3 为 G5 光谱 Smallest F ratio 判别模型的函数特征表。

表 3 判别函数特征表

Table 3 Discriminant function characteristic table

Function	Eigenvalue	Canonical correlation	Test of function	Wilks' Lambda	P
F_1	90.092	0.994	1 through 4	0.000	0.000
F_2	14.707	0.968	2 through 4	0.002	0.000
F_3	11.233	0.958	3 through 4	0.028	0.000
F_4	1.876	0.808	4	0.348	0.001

特征值越大表明该函数在样本区分中所扮演的角色越重要。从表 4 可以看到,4 个函数的特征值均大于 1,且特征值从大到小排列,即对于判别过程函数所起的作用排序依次为函数 F_1 、 F_2 、 F_3 、 F_4 。典型相关性表征各函数之间的相互关联程度,典型相关性的值越大,表明各判别函数的相关性强,样本区分更明显。Wilks' Lambda 是组内平方和与总平方和的比值,Wilks' Lambda 值越小,表明判别函数对于模型的影响越大。 P 表征函数的显著性, P 小于 0.01,表明判别函数显著相关。在 G5 光谱的 Smallest F ratio 判别模型中构建了 3 个判别函数,这三个判别函数的 Wilks' Lambda 值分别为 0.000、0.002 和 0.028,表明这两个判别函数对于模型的判别能力有很大的影响,且 3 个判别函数的显著性均 < 0.001 ,表明这 3 个判别函数对模型的分类影响特别显著。因此,本实验采用 F_1 、 F_2 以及 F_3 作为判别函数,得到 G5 光谱 Smallest F ratio 判别模型的判别分析结果,如图 7 所示。

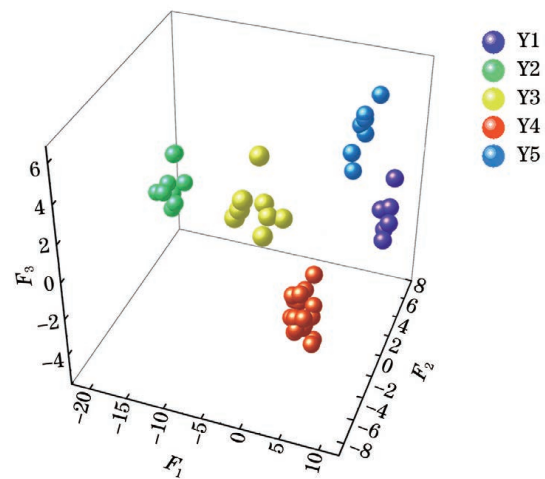


图 7 判别分析图

Fig. 7 Discriminant analysis diagram

图 7 中 5 种油漆样本不仅实现了完全分离,且各个样本之间的距离也比较大,表明该方法非常适合于这 5 种油漆样本的识别。图 7 中函数 F_1 的判别式为

$$F_1 = -2547.981x_1 - 167.642x_2 + 2071.819x_3 - 2792.776x_4 - 7390.720x_5 - 1212002.728x_6 - 532430.551x_7 + 681237.011x_8 + 1694922.030x_9 + 1270538.183x_{10} + 2192463.818x_{11} - 270195.798x_{12} + 1067707.828x_{13} - 1526931.596x_{14} + 1315651.006x_{15} + 683220.136x_{16} - 837916.401x_{17} + 735471.510x_{18} - 1.083, \quad (6)$$

函数 F_2 的判别式为

$$F_2 = -3571.076x_1 - 4088.391x_2 - 5939.488x_3 + 3294.154x_4 + 5085.985x_5 + 520910.627x_6 + 210687.362x_7 + 79283.307x_8 + 14479.933x_9 + 679309.029x_{10} - 1089493.665x_{11} + 778667.043x_{12} + 49107.119x_{13} -$$

$$568698.443x_{14} - 18951.234x_{15} - 230020.703x_{16} + 314075.805x_{17} + 304998.394x_{18} - 0.68, \quad (7)$$

函数 F_3 的判别式为

$$F_3 = -1697.638x_1 - 2722.098x_2 + 456.042x_3 + 7291.474x_4 + 664.318x_5 + 89611.168x_6 + 787210.259x_7 - 88601.745x_8 - 96553.122x_9 + 1155939.507x_{10} - 2936294.492x_{11} + 1138041.987x_{12} + 544878.911x_{13} - 270584.248x_{14} - 11584.230x_{15} + 2325.651x_{16} + 186690.175x_{17} + 148908.898x_{18} + 0.38. \quad (8)$$

对于未知类型的油漆样本,只需要在以上 3 个判别函数中输入相应的值,就会在判别分析图上显示出其位置以及该样本所属的类型。

5 结 论

本实验利用红外谱图的原始光谱和导数光谱并结合光谱融合技术与 KNN 模型、SVM 模型以及逐步判别分析对油漆样本进行分类研究,并对比了三种模型对于油漆样本的识别率。结果表明,三种模型对于融合光谱的识别率更高更稳定,原因是光谱融合能够使各光谱数据相互补充,从而提高识别准确率。其中无论是 KNN 模型还是 SVM 模型都能实现对 Y2 样本的完全识别,对 Y3、Y4 样本区分准确率也非常高,但对于 Y1 和 Y5 样本的识别率却不理想,因此 KNN 模型和 SVM 模型适合于 Y2、Y3 以及 Y4 样本的识别。对于逐步判别分析来说,其总体识别率均要比 KNN 和 SVM 模型高,其中 G5 光谱 Smallest F ratio 判别模型更是实现了对训练集和测试集的完全判别。该方法操作简单,可视性强,能够在短时间内完成对现场油漆物证的准确无损识别,满足公安机关对于现场微量油漆物证检测的相关要求。本实验所采集的样本数量以及种类不多,下一步将收集更多的样本数据并与更多的其他化学计量学方法进行比较,遴选出对各油漆样本识别率最优的模型,为法庭科学中油漆物证的鉴定提供参考。

参 考 文 献

- [1] Lei P. The identification method of paint evidences in traffic accidents using SEM/EDS [J]. Journal of Chinese Electron Microscopy Society, 2014, 33(4): 373-376.
雷平. 交通事故中油漆物证的扫描电镜/能谱鉴定方法[J]. 电子显微学报, 2014, 33(4): 373-376.
- [2] Sun Q R, Luo Y W, Xu C. Comparative analysis of red automobile paints by pyrolysis gas chromatography-mass spectrometry and infrared spectrometry[J]. Chinese Journal of Forensic Sciences, 2013(5): 51-54.
孙其然, 罗仪文, 徐彻. 裂解气相色谱-质谱法和红外光谱法对红色汽车油漆的比较分析[J]. 中国司法鉴定, 2013(5): 51-54.
- [3] Zheng S J, Yao Z D, Xi J H, et al. Study on identification of paint by thermal analysis[J]. Chemical World, 1996(2): 77-79.
郑尚菊, 姚中栋, 奚建华, 等. 热分析法鉴别油漆的研究[J]. 化学世界, 1996(2): 77-79.
- [4] Xi J H, Yao Z D. Identification of paints by atomic emission spectrometry [J]. Chemical World, 1997(4): 190-192.
奚建华, 姚中栋. 用原子发射光谱法鉴别油漆的研究[J]. 化学世界, 1997(4): 190-192.
- [5] Yu J, Zhang Y, Pang S Y, et al. Application of IR in the field of evidence identification [J]. Spectroscopy and Spectral Analysis, 2016, 36(9): 2807-2811.
余静, 张云, 庞松颖, 等. 红外光谱技术在物证鉴定中的应用[J]. 光谱学与光谱分析, 2016, 36(9): 2807-2811.
- [6] Martyna A, Michalska A, Zadora G. Interpretation of FTIR spectra of polymers and Raman spectra of car paints by means of likelihood ratio approach supported by wavelet transform for reducing data dimensionality[J]. Analytical and Bioanalytical Chemistry, 2015, 407(12): 3357-3376.
- [7] Grunert T, Stephan R, Ehling-Schulz M, et al. Fourier transform infrared spectroscopy enables rapid differentiation of fresh and frozen/thawed chicken [J]. Food Control, 2016, 60: 361-364.
- [8] Ferreira K B, Oliveira A G G, Gonçalves A S, et al. Evaluation of hyperspectral imaging visible/near infrared spectroscopy as a forensic tool for automotive paint distinction [J]. Forensic Chemistry, 2017, 5: 46-52.
- [9] Wang Y, Xu J, Cheng Y, et al. Application of IR differential spectrum technology into testing the mixed paint of micro-quantity [J]. Forensic Science and Technology, 2020, 45(2): 197-200.
王延, 徐俊, 程昱, 等. 红外差谱技术用于微量混合油漆的比对检验[J]. 刑事技术, 2020, 45(2): 197-200.
- [10] Zięba-Palus J, Trzcińska B M. Application of infrared and Raman spectroscopy in paint trace examination [J]. Journal of Forensic Sciences, 2013,

- 58(5): 1359-1363.
- [11] Yang Q L, Deng X J, Sun X D, et al. Application and research progress of spectral data fusion technology in food testing [J]. Science and Technology of Food Industry, 2020, 41(18): 324-329.
杨巧玲, 邓晓军, 孙晓东, 等. 光谱数据融合技术在食品检测中的应用研究进展[J]. 食品工业科技, 2020, 41(18): 324-329.
- [12] Li Y H, Li Y X, Tan W L, et al. Rapid identification and evaluation of *Lycium ruthenicum* Murr. by near-infrared and Fourier transform infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(12): 3878-3883.
李亚惠, 李艳肖, 谭伟龙, 等. 基于近、中红外光谱法融合判定黑果枸杞产地及品质信息[J]. 光谱学与光谱分析, 2020, 40(12): 3878-3883.
- [13] Feng Y Z, Zhang T L, Zhao M J, et al. Raman-infrared spectral fusion combined with partial least squares (PLS) for quantitative analysis of polycyclic aromatic hydrocarbons in soil[J]. Analytical Methods, 2020, 12(9): 1203-1211.
- [14] Deiss L, Margenot A J, Culman S W, et al. Optimizing acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy of soils[J]. Soil Science Society of America Journal, 2020, 84(3): 930-948.
- [15] Zhang F, Tang X J, Tong A X, et al. Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method [J]. Spectroscopy Letters, 2020, 53(3): 222-233.
- [16] He X L, Wang J F, Li Q S, et al. Identification of vehicle bumper debris based on multi-layer perception-Fisher discriminant and infrared spectroscopy[J]. China Measurement & Test, 2019, 45(5): 74-78, 92.
何欣龙, 王继芬, 李青山, 等. 基于多层感知器-Fisher 判别分析的车用保险杠红外光谱鉴别[J]. 中国测试, 2019, 45(5): 74-78, 92.
- [17] Rouchi M B, Khorrami M K, Garmarudi A B, et al. Application of infrared spectroscopy as Process Analytics Technology (PAT) approach in biodiesel production process utilizing Multivariate Curve Resolution Alternative Least Square (MCR-ALS)[J]. Spectrochimica Acta: Molecular and Biomolecular Spectroscopy, 2019, 213: 347-353.
- [18] Sivagami, Vaishali A, Ramakrishnan R, et al. Weather prediction model using savitzky-golay and Kalman filters[J]. Procedia Computer Science, 2019, 165: 449-455.
- [19] He X L, Chen L B, Wang J F, et al. Raman spectroscopy analysis of plastic steel window based on K nearest neighbors algorithm[J]. Laser & Optoelectronics Progress, 2018, 55(5): 053001.
何欣龙, 陈利波, 王继芬, 等. 基于 K 近邻算法的塑钢窗拉曼光谱分析[J]. 激光与光电子学进展, 2018, 55(5): 053001.
- [20] Li D, Jin Y Y, Tong Y, et al. Intelligent detection and defect classification of infusion bags based on support vector machine[J]. Laser & Optoelectronics Progress, 2019, 56(13): 131502.
李丹, 金媛媛, 童艳, 等. 基于支持向量机的输液袋智能检测与缺陷分类[J]. 激光与光电子学进展, 2019, 56(13): 131502.
- [21] Chen Y, Yan X, Zhang X, et al. Quantitative analysis of surface enhanced Raman spectroscopy of polycyclic aromatic hydrocarbons based on support vector machine algorithm [J]. Chinese Journal of Lasers, 2019, 46(3): 0311005.
陈阳, 严霞, 张旭, 等. 基于支持向量机算法的多环芳烃表面增强拉曼光谱的定量分析[J]. 中国激光, 2019, 46(3): 0311005.
- [22] Revelou P K, Pappa C, Kakouri E, et al. Discrimination of botanical origin of olive oil from selected Greek cultivars by SPME-GC-MS and ATR-FTIR spectroscopy combined with chemometrics[J]. Journal of the Science of Food and Agriculture, 2021, 101(7): 2994-3002.
- [23] Shen Y, Chen X P, Yuan Y B, et al. Infrared and visible image fusion based on significant matrix and neural network [J]. Laser & Optoelectronics Progress, 2020, 57(20): 201007.
沈瑜, 陈小朋, 苑玉彬, 等. 基于显著矩阵与神经网络的红外与可见光图像融合[J]. 激光与光电子学进展, 2020, 57(20): 201007.
- [24] He Y, Wang J F. Rapid nondestructive identification of wood lacquer using Raman spectroscopy based on characteristic-band-Fisher-K nearest neighbor[J]. Laser & Optoelectronics Progress, 2020, 57(1): 013001.
何亚, 王继芬. 基于特征波段-Fisher-K 近邻的木器漆拉曼光谱的快速无损鉴别[J]. 激光与光电子学进展, 2020, 57(1): 013001.
- [25] Ma Y J, Ma Y T, Chen J H. Vehicle recognition based on multi-layer features of convolutional neural network and support vector machine [J]. Laser & Optoelectronics Progress, 2019, 56(14): 141001.
马永杰, 马芸婷, 陈佳辉. 结合卷积神经网络多层特征和支持向量机的车辆识别[J]. 激光与光电子学进展, 2019, 56(14): 141001.
- [26] Zhang Z Y, Qin X P, Li Y F. Recognition method of waste non-ferrous metal fragments based on machine

vision[J]. Laser & Optoelectronics Progress, 2020, 57(16): 161017.

章振原, 秦训鹏, 李轶峰. 基于机器视觉的废旧有色金属碎料识别方法[J]. 激光与光电子学进展, 2020, 57(16): 161017.

[27] He X L, Liu W H, Wang J F. Examination the front

bumpers based on infrared spectroscopy and multivariate statistical[J]. The Journal of Light Scattering, 2018, 30(1): 70-76.

何欣龙, 刘文浩, 王继芬. 红外光谱结合多元统计学检验汽车前保险杠[J]. 光散射学报, 2018, 30(1): 70-76.