

二维相关近红外光谱的植物油鉴别方法

王 哲^{1,2}, 李晨曦^{1,2*}, 钱 蕊³, 刘 蓉^{1,2}, 陈文亮^{1,2}, 徐可欣^{1,2}

1. 天津大学精密测试技术及仪器国家重点实验室, 天津 300072
2. 天津大学精密仪器与光电子工程学院, 天津 300072
3. 中国汽车技术研究中心有限公司, 天津 300300

摘 要 植物油分类鉴别方法研究在食品安全与质量监管中具有重要的研究意义及应用价值。近红外光谱方法可实现复杂成分定性与定量分析, 具有无损、快速的优势, 在植物油鉴别分类方面应用广泛。研究了基于二维相关近红外光谱的植物油分类鉴别方法: 利用傅里叶变换近红外光谱仪采集不同种类植物油的动态光谱, 对其进行二维相关分析, 得到二维相关同步谱, 同步谱的主对角元素即其自相关谱, 利用主成分分析提取自相关谱的主成分, 最后基于欧氏距离实现植物油种类鉴别。利用二维同步谱的自相关谱主成分之间的欧氏距离实现常见植物油分类判别, 提高了分类准确性和算法效率。二维相关分析以正己烷浓度为扰动因素, 采集不同浓度扰动下植物油动态光谱, 选择不同种类植物油近红外吸收光谱差异最大的 $6\ 001\sim 6\ 063\ \text{cm}^{-1}$ 波长范围计算植物油样本的二维相关谱。二维相关分析的计算过程相对于原始谱而言, 其实质就是提取不同品种的植物油随扰动因素变化的特征信息。而这些特征信息体现在大量的数据点中(二维相关同步谱矩阵), 因此需要进一步降低变量维度。二维同步谱的对角线元素, 即其自相关谱, 代表了不同波长处光谱强度随扰动因素变化的程度, 这里用自相关谱来代替二维同步谱, 大大减少了变量数量。为了进一步降低数据维度, 对各类植物油的自相关谱进行主成分分析, 将自相关谱主成分作为分类模型参数。通过计算各类植物油自相关谱主成分之间的欧式距离, 在择近原则基础上实现不同种类植物油鉴别。实验结果表明, 二维相关近红外光谱与其特征提取方法相结合可以提高植物油分类准确度, 基于自相关谱主成分之间欧式距离的分类方法也为食用油鉴别应用以及自动化处理提供了有效手段。

关键词 近红外光谱; 二维相关谱; 主成分分析; 植物油; 分类鉴别

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)10-3230-05

引 言

随着社会经济发展, 食品质量与安全越来越受到重视。植物油是每日膳食摄入的必需成分, 其质量不仅关系到人民健康, 也是食品质量与安全的重要监管对象。由于产量不同和营养价值不同, 不同种类植物油的价格差异较大。因此, 植物油种类鉴别方法具有重要的研究意义及应用价值^[1-2]。

常用的食用油鉴别和掺假检测的方法包括皂化法^[3]、电导率法^[4]、显色法^[5]等理化方法, 以及气相色谱法^[6]、液相色谱法^[7]等仪器分析方法, 这些方法均需要复杂的处理过程以及专业人员操作。近红外光谱方法具有无损、快速的优

势, 广泛应用于成分检测及过程控制领域。随着光谱仪器向着小型化、智能化方向发展, 光谱建模分析算法越来越成熟, 近红外光谱在植物油种类鉴别方面的应用也越来越广泛。Yang^[8]采集了不同植物油吸收光谱与拉曼光谱, 建立植物油快速分类模型, 利用吸收光谱实现了较高的识别率。刘翠玲^[9]结合近红外光谱技术和聚类分析方法, 实现植物油种类的快速鉴别。马啸提出了一种基于样品衰减全反射红外光谱及 KL-BP(karhunen loeve transform-back propagation)模型的分类方法, 准确率达到 90.9%。代秀迎^[10]等提出了基于主成分分析的簇类独立软模式法(soft independent modeling class analog, SIMCA)分类模型, 对菜籽油、花生油和芝麻油的识别率在 90% 以上。

收稿日期: 2019-09-09, **修订日期:** 2020-01-20

基金项目: 国家自然科学基金项目(81871396, 81971657, 81671727, 81471698, 81401454), 天津市自然科学基金(19JCYBJC29100), 天津市科技特派员计划项目, 国家重大科学仪器设备开发专项(2014YQ060773), 国家(863)高技术研究发展计划项目(2012AA022602)资助

作者简介: 王 哲, 1995 年生, 天津大学精密测试技术及仪器国家重点实验室硕士研究生 e-mail: wz1895@tju.edu.cn

* 通讯联系人 e-mail: lichenxi@tju.edu.cn

食用油成分复杂,在近红外波段谱峰重叠严重,分析模型准确度受到一定限制。对样品施加特定的微扰(浓度、温度等),诱导样品光谱信号产生动态变化,利用二维相关分析^[11-13],可分辨出吸收光谱中由于谱峰重叠而不易分辨的微小吸收特征。崔路^[14]比较了 6 种植物油的二维相关谱中自相关峰和交叉峰的位置、数目和强度,建立了植物油分类鉴别模型。

工作基于二维相关谱的植物油鉴别方法。以浓度为扰动因素,测量不同浓度扰动下植物油动态光谱,进行二维相关分析。利用主成分分析(principle component analysis, PCA)压缩自相关谱,计算其欧式距离,在择近原则基础上实现植物油鉴别。结果表明,本方法可实现六种植物油的准确分类,为植物油鉴别提供了参考。

1 实验部分

1.1 二维相关光谱

采用正己烷浓度作为动态光谱测量扰动因素,将正己烷与植物油样品以不同体积比混合,当正己烷浓度 t 在区间 $[t_{\min}, t_{\max}]$ 内变化时,测量随变量 t 的扰动而变化的光谱 $y(v, t)$,定义受到扰动诱发的动态光谱为

$$\tilde{y}(v, t) = \begin{cases} y(v, t) - \bar{y}(v) & T_{\min} \leq t \leq T_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

其中, v 为测量得到吸收光谱的波数, $\bar{y}(v)$ 是参考光谱,一般表示如式(2)

$$\bar{y}(v) = \frac{\sum_{j=1}^m y(v, t_j)}{m} \quad (2)$$

同步相关强度 $\Phi(v_1, v_2)$ 是频率 v_1 和 v_2 处动态光谱的矢量积

$$\Phi(v_1, v_2) = \frac{1}{m-1} \tilde{y}(v_1)^T \tilde{y}(v_2) \quad (3)$$

1.2 仪器与样品测量过程

光谱仪(Frontier, Perkin Elmer, USA)的测量范围 3 800~10 000 cm^{-1} ,分辨率 8 cm^{-1} ,每个样品扫描 6 次取平均。植物油的近红外吸收多为合频与倍频吸收,吸收峰一般较宽,8 cm^{-1} 的分辨率能满足需要。实验样品选择 6 种消费量较大植物油(菜籽油,大豆油,橄榄油,花生油,葵花籽油与玉米油),每种包含两个不同品牌的产品。正己烷与植物油样品按照体积分数比例 20%,40%,60%,80%,100%混合,测量得到动态光谱,并进行二维相关分析。

选择浓度做外扰,是为了使整个检测过程快捷方便。改变待测样品的浓度只需要调整溶剂(本工作选择正己烷)与待测样品的体积比,而改变温度则需要开发配合光谱仪使用的温控系统,成本高,操作复杂。

选择正己烷做扰动溶剂,是因为烷烃结构的正己烷和以脂肪酸和脂肪酸酯为主的植物油之间不发生化学反应,两者可以物理互溶,因此正己烷也是植物油提取工艺中常用的低沸点蒸馏溶剂。此外,正己烷的主要吸收峰在中红外波段,在近红外波段几乎没有吸收,因此不会影响植物油的近红外

光谱,不存在溶剂跟样品吸收峰重叠的问题。

1.3 光谱预处理与建模波段选择

光谱预处理主要目的是消除基线漂移,进一步提高光谱信噪比,从而保证二维相关分析与模型预测的准确度。根据样品光谱特征,采用一阶 15 点微分方法对测量光谱进行基线校正,利用 S-G 三次多项式 9 点平滑滤波降低噪声。测量了六种植物油吸收光谱,如图 1 所示。根据方差最大原则,选择不同样品吸收光谱方差最大波段,即 6 001~6 063 cm^{-1} 作为建模分析波段。

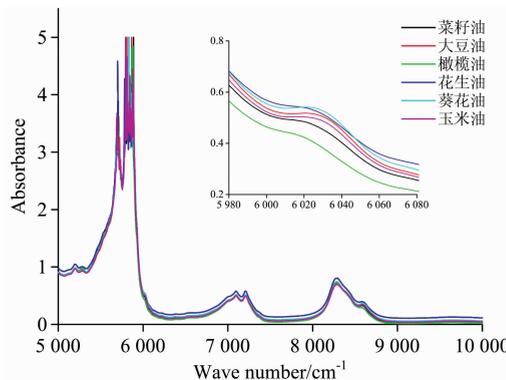


图 1 各样品的近红外光谱

Fig. 1 The infrared spectra of all samples

1.4 基于自相关谱主成分的分类方法

为了避免数据冗余造成的非线性拟合,提高分类准确率和算法效率,对自相关谱进行主成分分析,将主成分分量作为分类模型参数。以主成分之间欧式距离作为样本间相似度的度量,为植物油种类鉴别提供数值依据。

2 结果与讨论

2.1 不同种类植物油二维同步谱和自相关谱比较

比较 6 种常见植物油在 6 001~6 063 cm^{-1} 波段内的二维同步谱,结果如图 2 所示。根据二维相关谱分析结果,不同种类植物油的同步谱形状差异较大,并且特征峰数量比直接吸收谱多,分辨率更高。但二维相关谱数据量大,若将他们作为分类参数,可能会影响模型的准确性,降低运算效率。

分别比较不同种类植物油样品自相关谱,如图 3 所示。根据自相关谱特征可知,同种植物油自相关谱的变化趋势,出峰数量,出峰位置比较相似,不同种类植物油的自相关谱在出峰位置、数量以及强度上差别明显。

2.2 基于自相关谱主成分欧式距离的鉴别方法

对植物油样品的自相关谱进行主成分分析,前 8 个主成分贡献率达到 95% 以上。因此,选取前 8 个主成分来代表各个样品的同步谱,计算并比较不同种类植物油自相关谱主成分之间欧式距离,从而实现不同种类植物油分类。以 6 个不同种类植物油样品为参考样品,分别计算参考样本与其他测试样品之间的欧式距离,结果如表 1 所示,表中对角线加粗位置是同一种类植物油的欧氏距离。

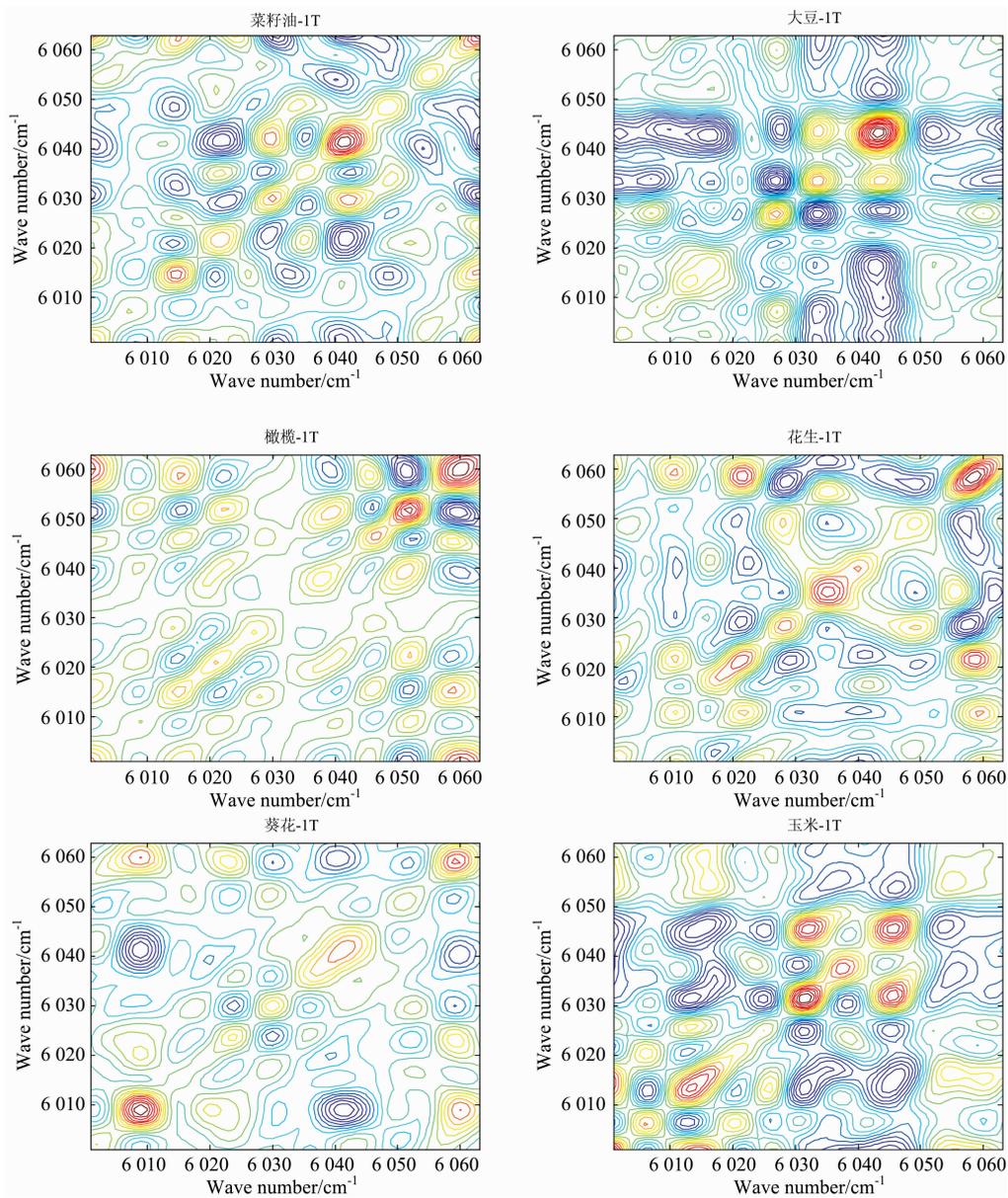
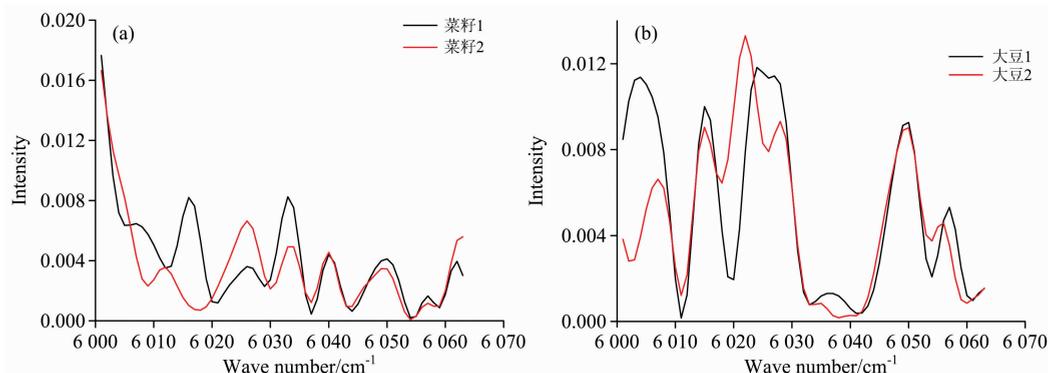


图 2 六种常见植物油的同步谱

(a): 菜籽油; (b): 大豆油; (c): 橄榄油; (d): 花生油; (e): 葵花油; (f): 玉米油

Fig. 2 2D correlation synchronization spectra of 6 kinds of common oils

(a): Rapeseed oil; (b): Soybean oil; (c): Olive oil; (d): Peanut oil; (e): Sunflower oil; (f): Corn oil



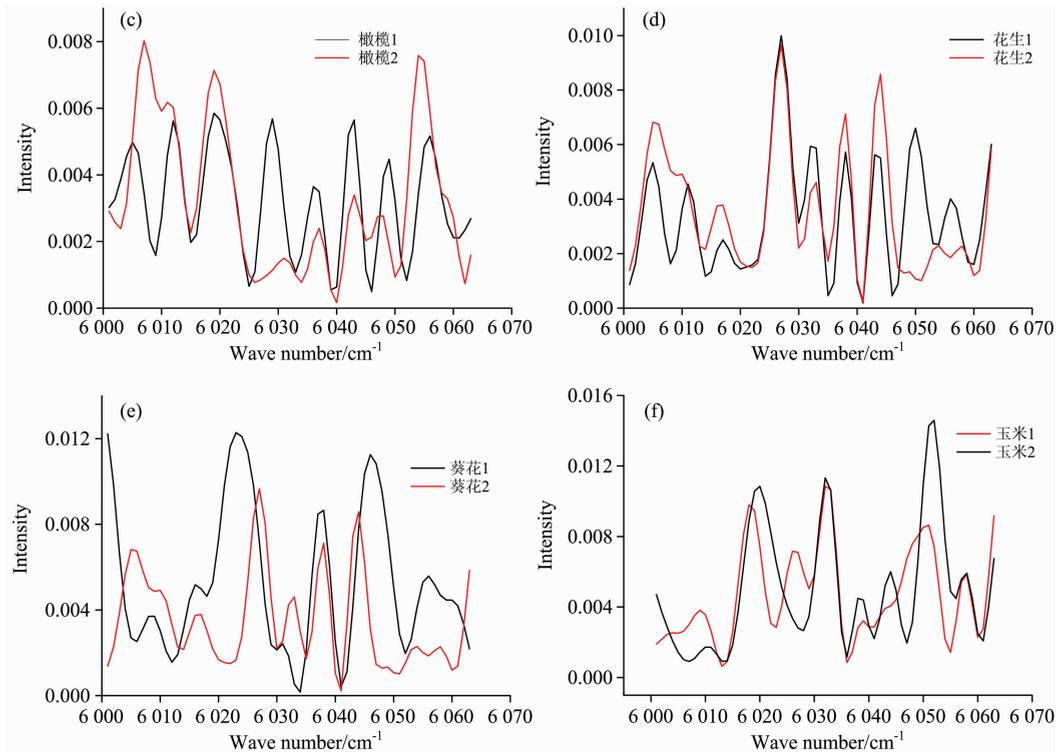


图 3 六种常见植物油的自相关谱

(a): 菜籽油; (b): 大豆油; (c): 橄榄油; (d): 花生油; (e): 葵花油; (f): 玉米油

Fig. 3 Autocorrelation spectra of 6 kinds of common oils

(a): Rapeseed oil; (b): Soybean oil; (c): Olive oil; (d): Peanut oil; (e): Sunflower oil; (f): Corn oil

表 1 欧氏距离及分类结果

Table 1 The Euclidean Distance between the reference and test sample collections and classification results

	菜籽 1	大豆 1	橄榄 1	花生 1	葵花 1	玉米 1
菜籽 2	1.944	3.402	3.207	3.259	3.340	3.488
大豆 2	4.428	2.182	4.981	3.397	3.388	4.485
橄榄 2	3.153	5.190	2.256	4.976	4.384	3.159
花生 2	3.046	3.765	3.284	2.943	3.234	3.500
葵花 2	3.119	3.618	3.465	3.826	1.699	3.060
玉米 2	3.236	3.644	4.296	3.382	2.743	2.710

根据实验结果,同种植物油的自相关谱主成分欧氏距离最小,不同种类植物油距离较大。以菜籽油为例,不同品牌菜籽油之间欧式距离为 1.944,而与其他种类植物油之间距离普遍大于 3。这也表明本研究中提出的植物油的二维相关光谱的量化方法,可以有效提取不同种类植物油的特征信息。同时,模型计算中参考样品的选择在一定程度上也会影响分类结果,如玉米油 2 同葵花 1,以及玉米油 1 的距离十分相近。差别较小的原因是,葵花油与玉米油的脂肪酸构成相近。两者的不饱和脂肪酸含量都在 85%左右,而且其中亚

油酸的含量都是 60%左右。虽然玉米油和葵花油的组分构成差别较小,实验表明本文方法仍然能够成功鉴别。

研究中虽然可以区分,但是考虑到实际应用,还需要进行进一步改进。在实际应用中,可通过扩大参考集中的样品数,采用交互验证等方法来优选参考样品集,从而提高模型分类准确性。

3 结论

研究了基于二维相关近红外光谱的植物油分类鉴别方法,利用二维相关谱中自相关谱主成分和欧氏距离判别相结合,提高了分类准确性和算法效率。以正己烷为扰动因素,测量并计算得到植物油二维相关近红外光谱,对自相关谱进行主成分分析,实现数据降维与光谱特征提取;计算其中 8 个主成分欧氏距离,基于择近原则实现不同种类植物油鉴别。将二维相关近红外光谱与其特征分析相结合,提高了植物油分类准确度,基于自相关谱主成分欧式距离的分类方法也为实际应用以及自动化处理提供了有效手段。

References

- [1] YU Qing, ZHONG Pei-pei, WANG Yuan-xing(喻 晴, 钟培培, 王远兴). *China Oils and Fats(中国油脂)*, 2018, 43(6): 81.
- [2] Dupuy N, Galtier O, Ollivier D, et al. *Analytica Chimica Acta*, 2010, 666(1-2): 23.
- [3] Spicer S S, Lillie R D. *Journal of Histochemistry & Cytochemistry*, 1959, 7(2): 123.
- [4] Alpaslan Turgut, Ismail Tavman, Sebnem Tavman. *International Journal of Food Properties*, 2009, 12(4): 741.
- [5] Guo Yujun, Liang Xiao, Bi Junmin, et al. *Food Chemistry*, 2019, 285: 450.
- [6] LI Bin, QIU Li-qun, SONG Shao-fang, et al(李 斌, 裘立群, 宋少芳, 等). *Chinese Journal of Analysis Laboratory(分析试验室)*, 2014, 33(5): 528.
- [7] WANG Chun-lei, LIU Hua-liang, MA Yong-jian, et al(王春蕾, 刘华良, 马永建, 等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2015, 43(11): 1743.
- [8] Yang Hong, Irudayaraj Joseph, Paradkar Manish M. *Food Chemistry*, 2005, 93(1): 25.
- [9] LIU Cui-ling, WEI Li-na, ZHAO Wei, et al(刘翠玲, 位丽娜, 赵 薇, 等). *China Brewaing(中国酿造)*, 2014, 33(11): 149.
- [10] DAI Xiu-ying, CHEN Bin, LIU Fei, et al(代秀迎, 陈 斌, 刘 飞, 等). *Jiangsu Agricultural Sciences(江苏农业科学)*, 2010, 4: 283.
- [11] Yang Renjie, Liu Rong, Xu Kexin. *Food Bioscience*, 2013, 2: 61.
- [12] Yang Renjie, Liu Rong, Xu Kexin, et al. *Applied Spectroscopy*, 2013, 67(12): 1363.
- [13] Noda I, Dowrey A E, Marcoli C, et al. *Applied Spectroscopy*, 2000, 54(7): 1363.
- [14] CHEN Bin, CUI Lu, LIN Zhen-xing, et al(陈 斌, 崔 路, 林振兴, 等). *Applied Chemical Industry(应用化工)*, 2016, 45(4): 784.

Research on Vegetable Oils Classification Based on Two-Dimensional Correlation Near-Infrared Spectroscopy

WANG Zhe^{1,2}, LI Chen-xi^{1,2*}, QIAN Rui³, LIU Rong^{1,2}, CHEN Wen-liang^{1,2}, XU Ke-xin^{1,2}

1. State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China

2. School of Precision Instrument and Optic Electronic Engineering, Tianjin University, Tianjin 300072, China

3. China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China

Abstract The research on classifying different kinds of vegetable oil is very important for food safety and quality supervision. The near-infrared spectroscopy could achieve a qualitative and quantitative analysis of samples with complex components. It has been widely applied to classifying different kinds of vegetable oil. The classification method based on two-dimensional correlation near-infrared spectroscopy is applied to recognize typical vegetable oil in this research. Two-dimensional correlation analysis was carried out with the concentration of n-hexane as the disturbance factor. Then the two-dimensional correlation analysis is calculated within the range of $6\ 001\sim 6\ 063\ \text{cm}^{-1}$, in which the absorption features of different kinds of vegetable oils is obvious. And these feature information is reflected in a large number of data points (two-dimensional correlation synchronization spectrum matrix), so further extraction is needed to reduce the variable dimension. The diagonal elements of the two-dimensional synchro spectrum, that is, its autocorrelation spectrum, is always positive, representing the extent to which the spectral intensity varies with the disturbance factor at different wavelengths. Taking advantage of features extracting and data dimensions reduction, the principal component analysis is adapted to extract the feature of the autocorrelation spectrum. The Euclidean distance of principal components is calculated to classify different types of vegetable oil. The experimental results indicated that the proposed method is a benefit for automatic recognition and classification of typical vegetable oil. The PCA algorithm can effectively improve the recognition efficiency and robustness of the model. It provides a new concept for the analysis and processing of food quality sensing spectroscopy.

Keywords Near-infrared spectroscopy; Two-dimensional correlation spectrum; Principal component analysis; Vegetable oil; Classification

* Corresponding author

(Received Sep. 9, 2019; accepted Jan. 20, 2020)