

酥梨货架期的高光谱成像无损检测模型研究

李 雄, 刘燕德*, 欧阳爱国, 孙旭东, 姜小刚, 胡 军, 欧阳玉平

华东交通大学机电与车辆工程学院, 江西 南昌 330013

摘 要 水果新鲜度是反映水果是否新鲜、饱满的重要品质指标, 为了探讨水果不同货架期的预测和判别方法, 以酥梨为研究对象, 利用高光谱成像技术, 结合偏最小二乘判别法(PLS-DA)和偏最小二乘支持向量机(LS-SVM)算法对酥梨货架期进行判别。由光源、成像光谱仪、电控位移平台和计算机等构成的高光谱成像装置采集样品光谱, 装置光源采用额定功率为 200 W 四个溴钨灯泡成梯形结构设计, 光谱范围为 1 000~2 500 nm, 分辨率率为 10 nm。选取优质酥梨 30 个, 货架期设置为 1, 5 和 10 d, 对 30 个样品完成 3 次光谱图像的采集, 并矫正原始图像。实验结果表明: 基于图像的酥梨货架期定性分析时, 对不同货架期样品的原始图像进行 PCA 压缩, 得到三种不同货架期的权重系数数据, PC1 图像提取特征波长点为 1 280, 1 390, 1 800, 1 880 和 2 300 nm, 以特征图像的平均灰度值作为自变量且以货架期作为因变量建立定性判别模型, 建模集 68 个, 预测集 22 个。最小二乘支持向量机以 RBF 为核函数时, 预测集中样品的误判个数为 1, 误判率为 4.5%。而当采用 lin 核函数时, 样品的误判个数为 0, 误判率为 0。PLS-DA 定性分析时 RMSEC 为 1.24, R_c 为 0.93。RMSEP 为 1, R_p 为 0.96, 预测集误判率为 0。特征图像对酥梨货架期判别 LS-SVM 中的 lin 核函数所建立的模型结果较好, 优于 RBF 核函数的建模效果, 也优于 PLS-DA 判别模型。ENVI 软件提取实验样品光谱后建立 LS-SVM 和 PLS-DA 判别模型, LS-SVM 利用 RBF 和 lin 核函数误判率分别为 4.5% 和 0。与 RBF 核函数相比, lin 核函数所建立的模型预测酥梨货架期的效果更好。PLS-DA 方法主成分因子数为 12, RMSEC 和 RMSEP 分别为 0.48 和 0.78, R_c 和 R_p 分别为 0.99 和 0.97, 建模集与预测集的误判率均为零。LS-SVM 中的 lin 核函数所建立的模型结果较好, 依然优于 PLS 所建立的检测模型。酥梨的光谱信息结合 LS-SVM 可以实现对酥梨货架期的检测和判别。基于图像建立酥梨的货架期预测模型与光谱相比, 都实现了酥梨货架期的判别, 而特征图像法, 选择区域较少流失部分信息, 计算量小, 建模结果相对略差。酥梨货架期的高光谱成像检测模型研究为消费者正确评价水果新鲜度提供了理论指导, 也为后期水果货架期检测仪器的开发提供了技术支持。

关键词 高光谱成像; 货架期; 特征图像; 最小二乘判别; 偏最小二乘支持向量机

中图分类号: S24 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)08-2578-06

引 言

水果新鲜度是反映水果是否新鲜、饱满的重要品质指标, 随着人们生活品质的提高和科技进步, 水果成为生活中必不可少的一部分, 交通的便利让人足不出户就能享受世界各地的水果, 人们对水果的新鲜度提出了新的要求。一些长途运输且不易保鲜的水果, 需要提供一定的储存条件延长其保质期, 例如梨, 黄桃、荔枝等水果在常温下放置几天或分类、

筛选、加工运输到达市场的过程表面似乎没有变化, 但内部品质却已发生改变, 影响口感^[1-3]。实际生活中, 水果加工企业或者消费者在购买水果时仅仅通过眼睛和手感去判断水果的新鲜度, 该方法主观性强、准确率低, 消费者和水果加工企业容易购买到大量的内部腐烂的不新鲜水果, 造成较大的经济损失, 如何准确、快速、无损地检测水果的新鲜度显得尤为重要。

目前, 高光谱和近红外光谱对水果品质检测较为成熟, 尤其是近红外漫透射或漫反射技术可以较好地检测水果的内

收稿日期: 2018-07-27, 修订日期: 2018-11-20

基金项目: 国家自然科学基金项目(31760344), 南方山地果园智能化管理技术与装备协同创新中心(赣教高字[2014]60号), 江西省优势科技创新团队(20153BCB24002)资助

作者简介: 李 雄, 1992年生, 华东交通大学机电与车辆工程学院助教 e-mail: 1598657020@qq.com

* 通讯联系人 e-mail: jxliuyd@163.com

部品质。Manuela Zude 等^[4]利用近红外光谱技术检测苹果果实硬度、可溶性固形物并预测货架期,果实的分类正确率达 93%,从而准确预测水果的保质期。Sinelli 等^[5]有效利用了近红外光谱技术对包装的意大利乳清干酪储存期间发生的变化实现了监测。但光谱技术对水果新鲜度及货架期评价方面的报道极少,高光谱成像技术具有“图谱合一”的特点,能够有效识别水果损伤、结痂等外部品质,从而对水果货架期做出预测^[6-8]。本研究以酥梨为研究对象,利用高光谱成像技术对其进行货架期的判别,为食品加工企业和消费者在购买水果时对水果新鲜度的判断提供参考依据。

1 实验部分

1.1 材料

实验所用的酥梨购买于南昌市某水果店,样品选取新鲜、表面无损伤、果型大小均匀优质酥梨 30 个,对 30 个酥梨样品完成 3 种不同货架期的光谱图像的采集。如图 1 所示为酥梨的部分样品示意图,图 2 为酥梨样品的 RGB 图像。



图 1 酥梨样品

Fig. 1 The sample of pear

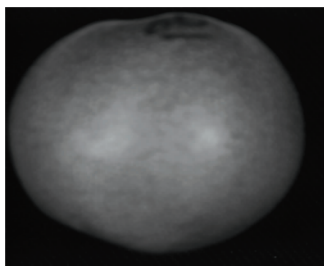


图 2 样品 RGB 图像

Fig. 2 The RGB imaging of sample

1.2 高光谱成像系统

用于采集实验样品图像的高光谱成像装置如图 3 所示,硬件主要包括光源、成像光谱仪、电控位移平台和计算机等。其中光源采用四个溴钨灯泡成梯形结构设计,其额定功率为 200 W,输入电压为 12 V,由稳压电源供电,光谱范围为 1 000~2 500 nm,光谱分辨率为 10 nm,相机像素为 320×256;为了使样品在采集图像时候不受环境光的影响,整个成像系统集成在一个密闭柜中。

1.3 图像采集

图像采集之前为了消除基线漂移对数据采集的影响,成像系统预热 30 min。图像采集中,照射在样品表面的光经反

射后携带着的光谱信息被相机镜头所捕获,得到一维的影响以及光谱信息,当所有的数据均被计算机软件记录后最终形成一个包含了影像信息和光谱信息的三维数据立方体。所采集的图像模块既包含有特定像素下的光谱信息也具有特定波长下的图像信息,数据模块如图 4 所示。

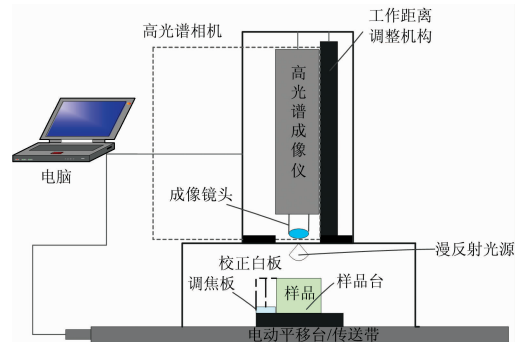


图 3 高光谱成像装置

Fig. 3 Hyperspectral imaging device

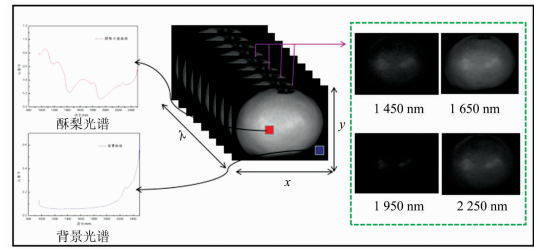


图 4 酥梨高光谱数据模块

Fig. 4 Crisp pear hyperspectral data block

由于采集样品的高光谱图像会受到光源强度在各波段下分布不均匀以及 CCD(charge coupled device)相机中暗电流的影响,在对数据进行分析处理之前需要对所获得的高光谱图像进行黑白标定,再利用软件中分析工具选项对所有采集的实验样品的原始高光谱图像进行矫正,其矫正的公式如(1)所示。

$$R_{\lambda} = \frac{I_{\lambda} - H_{\lambda}}{B_{\lambda} - H_{\lambda}} \quad (1)$$

式(1)中, R_{λ} 为标定后的数据, H_{λ} 为全黑数据, B_{λ} 为全白数据, I_{λ} 为原始数据。

1.4 数据处理

偏最小二乘法通过最小化误差的平方并找到一组数据的最佳匹配函数,常用于对化学组分与光谱数据之间关系的回归分析。支持向量机通过非线性内积核函数将线性不可分的低维空间数据映射到一个线性可分的高维特征空间,在这个特征空间中进行回归拟合,与 SVM(support vector machine)相比,LS-SVM 具有更快的求解速度和高精度等优势^[9-10]。

2 结果与讨论

2.1 酥梨特征图像的获取

利用酥梨的图像和光谱分别对其货架期进行建模并对比

哪种建模结果更好。基于图像处理酥梨货架期预测方法为：利用特征图像的平均灰度值建立与酥梨货架期的对应关系并建模。但每一个样品的原始图像中包含了 288 幅图像，要对 30 个样品中每一个样品的 288 幅图像进行图像处理，不仅计算复杂且工作量大还有可能获得较差的建模效果。因而需要利用相关的化学计量学方法对原始图像进行分析处理，最终获得最能代表实验样品品质信息的若干幅光谱图像，即特征图像。研究中采用的化学计量学方法为主成分分析法 (PCA)，主成分分析法是高光谱图像处理的一项常用技术，具有增强光谱图像的信息含量、隔离噪声和降低数据维数等优势。原理为通过多个波段的线性变换使得原始数据映射到一个新的坐标系，以使数据的差异性达到最大。

利用 PCA 分析方法分别对第 1 天、第 5 天和第 10 天货架期中实验样品的若干图像进行处理分析，主成分分析后选择了不同货架期的前 5 个主成分图像，即每一幅原始图像经压缩后得到 5 幅主成分图像，该 5 个图像能够清晰的反映出酥梨样品的基本信息。然后从 5 幅图像中选取质量最佳的图像，通过提取 PC 图像中的权重系数数据可以得到各个波长点对该 PC 图像贡献率的大小，进而可以得到最佳波长点即特征波长点，特征波长点所对应的 RGB 图像即为特征图像。对不同货架期中样品的原始图像进行 PCA 压缩后，最终得到三种不同货架期的权重系数数据。如图 5 所示分别为 3 个货架期下的原始高光谱图像，从左到右依次为一个样品的第 1 天、第 5 天、第 10 天的高光谱原始图像，其中在进行 RGB 图像运算时，R、G 和 B 的取值分别为 985、1 205 和 1 469 nm。

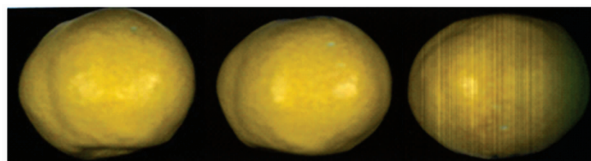


图 5 不同货架期下的原始高光谱图像

Fig. 5 The raw hyperspectral image of different shelf life

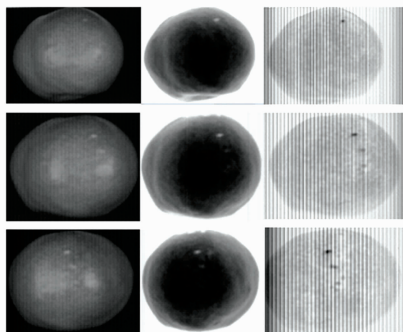


图 6 不同货架期下的 PC 图

Fig. 6 The PC imaging of different shelf life

通过观察，其中 3 幅 PC 图能更清晰和更准确地反映出酥梨的真实信息，图 6 为实验样品在不同货架期下进行主成份压缩后各自所得的 3 幅 PC 图，图中从上到下依次为某一

个样品第 1 天、第 5 天和第 10 天的样品 PC 图，在 3 幅 PC 图像中，第 1 幅 PC 图像最清晰，可以准确的反映出样品的全部信息，适合样本的分割和信息提取，PC2 图像中样品中心与边缘颜色差异较大，而第 3 幅 PC 图像则太模糊。因此，选取 3 个不同货架期的 PC1 图像来进行权重系数的提取，图 7 所示为从样品不同货架期的 PC1 图像中提取的权重系数数据所绘制的 3 条权重系数曲线图，三条权重曲线的变化规律基本一致，通常认为权重系数曲线中的峰和谷所对应的波长点就是特征波长点，选取了 5 个特征波长点如图中箭头标记处所示，对应的特征波长分别为：1 280、1 390、1 800、1 880 和 2 300 nm。在获取特征波长点之后，根据特征波长点可得到特征图像。

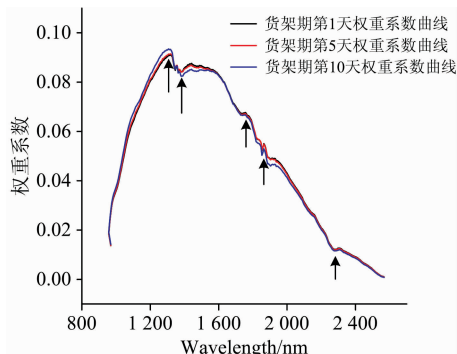


图 7 不同货架期下的权重系数图

Fig. 7 The weight coefficient of different shelf life

如图 8 所示为根据特征波长点所截取的某个实验样品三个货架期的特征图像，从上而下依次为货架期第 1 天、第 5 天和第 10 天。

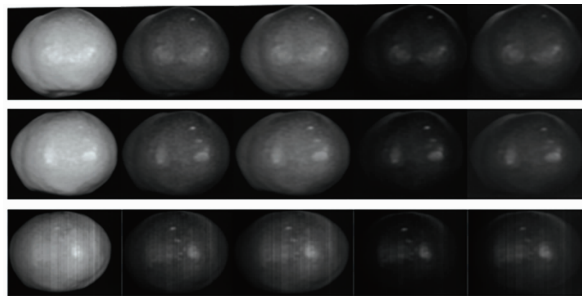


图 8 3 种货架期下的特征图像

Fig. 8 The feature imaging of 3 shelf life

2.2 酥梨特征图像的处理

根据特征波长点获取每一个实验样品的特征图像后，需要对其进行平均灰度值的计算，利用 matlab 对图像进行读取，然后根据样品边缘与除样品之外区域的差异选取合适的阈值，阈值选取后对图像进行二值化处理，使得样品区域的灰度值保持不变，其他区域的灰度值变为零，然后计算实验样品区域的所有像素点和像素值并求取总和，最后利用总的灰度值与总的像素点进行比值处理得到实验样品的平均灰度值。对所有的特征图像进行处理后，以每一个实验样品的五幅特征图像的平均灰度值作为自变量且以货架期作为因变

量,通过 unscrambler 软件对其进行定性模型的建立。

2.3 基于图像和光谱的酥梨货架期定性分析模型

2.3.1 基于图像的酥梨货架期定性分析模型

利用最小二乘支持向量机 (LS-SVM) 和偏最小二乘法 (PLS) 分别对图像数据建立模型。实验数据一共 90 个,其中 68 个用于建模集,剩余的 22 个用于模型的预测。如表 1 所示为根据 LS-SVM 中不同的核函数对图像数据处理所得到的结果,从表 1 可知当最小二乘支持向量机以 RBF 为核函数时,预测集中样品的误判个数为 1,误判率为 4.5%。而当采用 lin 核函数时,样品的误判个数为 0,误判率为 0。

表 1 基于图像处理的不同核函数下所建模型结果

Table 1 Model results based on different kernel functions of image processing

LS-SVM	误判样品数	误判率/%	RMSEP	R_p
RBF 核函数	1	4.5	0.76	0.98
lin 核函数	0	0	0.98	0.97

通过 PLS-DA 对实验样品定性分析的结果如图 9 所示,从图 9 可知 RMSEC 为 1.24, R_c 为 0.93。RMSEP 为 1, R_p 为 0.96, 其中截距 B 为 24.7。图中的阈值 T_1 和 T_2 分别为 3 和 7.5。当预测值小于 3 的时候判定其储藏期为 1 d, 当预测值介于 3 和 7.5 之间则判定其储藏期为 5 d, 预测值大于 7.5 时判定货架期为 10 d, 以此作为判定预测集样品是否被误判的标准。图中红色点代表预测集, 蓝色代表建模集, 可以看出红色的点全部预测正确, 误判率为 0, 而建模集中有 3 个点发生误判。

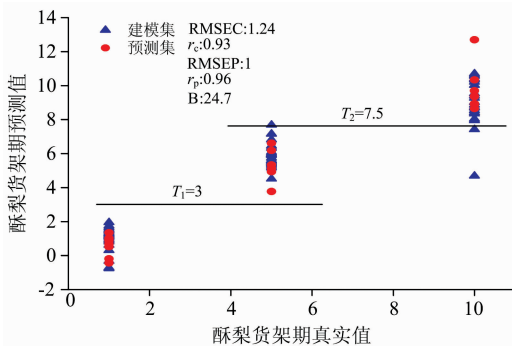


图 9 建模集货架期与预测集货架期关系图

Fig. 9 The relationship graph of modeling set shelf life and predicting set shelf life

2.3.2 基于光谱的酥梨货架期定性分析模型

为了对比光谱和图像对酥梨货架期预测的结果好坏,在对图像进行处理和分析外同时从样品的原始图像中提取样品的光谱,利用 ENVI 软件对实验样品的光谱进行提取,在提取的过程中每个样品表面选择的像素点均为 30 个。

利用所有样品的光谱并结合 LS-SVM 和 PLS-DA 算法分别建立模型,寻找最佳的建模结果。表 2 所示为 LS-SVM 的建模结果,利用 RBF 和 lin 核函数分别对 22 个预测集样品进行预测,其误判的样品个数分别为 1 和 0,误判率分别

为 4.5% 和 0。与 RBF 核函数相比,利用 lin 核函数所建立的模型预测酥梨货架期的结果更好。

表 2 基于光谱处理的不同核函数所建立的模型结果

Table 2 Model result based on different kernel function of spectral processing

LS-SVM	误判样品个数	误判率/%	RMSEP	R_p
RBF 核函数	1	4.5	0.71	0.98
lin 核函数	0	0	0.65	0.98

图 10 为利用光谱数据结合 PLS 方法所建立酥梨货架期模型的主成分因子数决定图,红色曲线代表建模集均方根误差随着主成分因子数增加而变化的曲线,蓝色曲线代表预测集均方根误差随着主成分因子数增加而变化的曲线,从图中可知随着主成分因子数的增加建模集和预测集的均方根误差呈现的趋势为先降低后保持大体不变,建模集曲线虽然在主成分因子数为 15 的地方有所反弹但是随之又降低,预测集曲线在主成分因子数为 9 之后基本保持水平且均方根误差保持基本不变,结合两条均方根误差曲线的变化最终得到酥梨货架期主成分因子数为 12。图 11 为模型预测结果图,实验

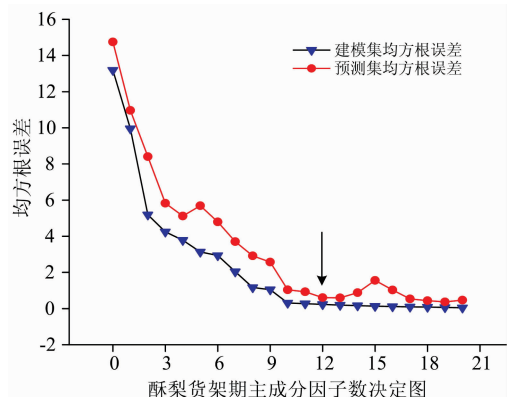


图 10 酥梨货架期主成分因子数决定图

Fig. 10 The number principal component factor decision diagram of pear shelf

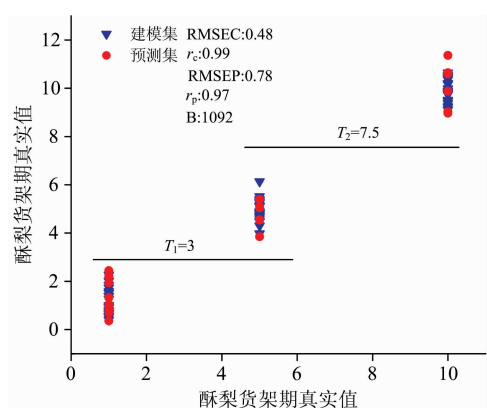


图 11 酥梨货架期预测值与真实值之间的关系

Fig. 11 The relationship of pear shelf life predicted value and true value

样品一共 90 个, 其中 68 个用于建模集进行模型的建立, 剩余的 22 个用于预测集检测模型的预测结果, 从图中可知, RMSEC 和 RMSEP 分别为 0.48 和 0.78, R_c 和 R_p 分别为 0.99 和 0.97, 其中截距 B 为 10.92, 阈值 T_1 和 T_2 分别为 3 和 7.5, 判断一个样品是否被误判的标准与前文基于图像建立判别模型的标准相同, 从图中可以看出建模集与预测集均在阈值范围内, 因此建模集与预测集的误判率均为零。

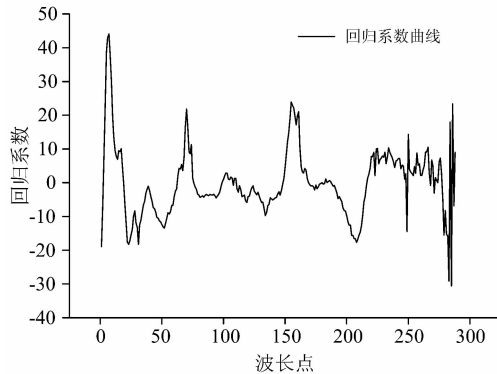


图 12 酥梨货架期回归系数图

Fig. 12 The regression coefficient map of pear shelf

图 12 所示为模型的回归系数图, 全谱中的各个变量均在酥梨货架期判别中发挥了作用, 在不同的峰值处有对应的波长点, 而波长点分别对应 1 030, 1 368, 1 863, 2 145 和 2 410 nm 处波长, 此处提供信息较多, 对变量货架期预测贡献较大。将采集到的酥梨能量谱代入式(2)中, 经计算得到判别值 \hat{y} , 当 \hat{y} 值小于阈值 3 时被判定为 1 d 即货架期为 1 d, \hat{y} 值介于阈值 3 和 7.5 之间时被判定为 5 d, 而当 \hat{y} 值大于 7.5 时则被判为 10 d。

3 结 论

通过高光谱成像技术基于图像和光谱建立酥梨的货架期预测模型, 对比发现光谱并结合最小二乘支持向量机中的 lin 核函数所建立的模型结果最优, 其预测集相关系数和预测集均方根误差分别为 0.98 和 0.65, 误判率为零。基于图像所建立的模型与之相比略差, 原因可能是特征图像选择相对较少导致某些重要信息的流失, 同时在对图像处理的过程中阈值的选择是否合理也是导致建模结果较差的一个重要原因。本实验利用特征图像的平均灰度值作为模型的自变量计算量小、效率高, 实现了酥梨货架期的判别。两种途径结合两种算法对酥梨的货架期进行了预测, 都取得了较好的实验效果, 该实验可为实际生产加工中对梨的品质鉴定提供参考和依据。

References

- [1] SONG Xue-jian, WANG Hong-jiang, ZHANG Dong-jie, et al(宋雪健, 王洪江, 张东杰, 等). Nondestructive Testing(无损检测), 2017, 39(10): 71.
- [2] SUN Hai-xin, YAN Mei-ling, LU Rong-rong(孙海新, 闫美伶, 陆容容). Food Technology(食品科技), 2015, (5): 334.
- [3] WEN Shu-xian, LI Shao-wen, JIN Xiu, et al(温淑贤, 李绍稳, 金秀, 等). Journal of Computer Science(计算机科学), 2017, 44(S1): 216.
- [4] Manuela Zude, Bernd Herold, Jean-Michel Roger, et al. Journal of Food Engineering, 2005, 77(2): 254.
- [5] Sinelli N, Barzaghi S, Giardina C, et al. Journal of Near Infrared Spectroscopy, 2006, 13(5): 293.
- [6] SUN Jun, JIN Xia-ming, MAO Han-ping, et al(孙俊, 金夏明, 毛罕平, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2014, 30(10): 167.
- [7] HUANG Shuang-ping, QI Long, MA Xu, et al(黄双萍, 齐龙, 马旭, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(1): 212.
- [8] LIU Ya, Mu Hetar · Migiti, CAO Peng-cheng, et al(刘亚, 木合塔尔·米吉提, 曹鹏程, 等). Agricultural Technology and Equipment(农业科技与装备), 2016, (5): 50.
- [9] LIU Xue, LI Ya-mei, LIU Jiao, et al(刘雪, 李亚妹, 刘娇, 等). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2015, 46(10): 328.
- [10] LIU Yan-de, CHENG Meng-jie, HAO Yong, et al(刘燕德, 程梦杰, 郝勇). Journal of East China Jiaotong University(华东交通大学学报), 2018, 35(4): 1.

Study on Non-Destructive Testing Model of Hyperspectral Imaging for Shelf Life of Crisp Pear

LI Xiong, LIU Yan-de*, OUYANG Ai-guo, SUN Xu-dong, JIANG Xiao-gang, HU Jun, OUYANG Yu-ping
School of Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract Fruit freshness is an important quality index reflecting whether the fruit is fresh and full. In order to explore the prediction and discrimination methods of different shelf life of fruits, this paper takes the pear as the research object, and uses hyperspectral imaging technology combined with partial least squares discrimination (PLS), DA and partial least squares support vector machine (LS-SVM) algorithm to distinguish the shelf life of pears. The spectrum of the sample is collected by a high-spectrum imaging device consisting of a light source, an imaging spectrometer, an electronically controlled displacement platform, and a computer. The device light source is designed with a ladder power of 200 W four bromine tungsten bulbs, and the spectral range is 1 000~2 500 nm. 10 nm. The material was selected from 30 high-quality pears, and the shelf life was set to 1 day, 5 days and 10 days. Three spectral images were acquired for 30 samples and the original image was corrected. The experimental results show that the image-based analysis of the shelf life of the pears is carried out by PCA compression of the original images of different shelf life samples, and the weight coefficient data of three different shelf periods are obtained. The wavelength points of PC1 image extraction are 1 280, 1 390 and 1 800 nm. 1 880 and 2 300 nm, with the average gray value of the feature image as the independent variable and the shelf life as the dependent variable to establish a qualitative discriminant model, 68 modeling sets and 22 prediction sets. When the least squares support vector machine uses RBF as the kernel function, the number of misjudgments in the predicted concentrated samples is 1, and the false positive rate is 4.5%. When the lin kernel function is used, the number of misjudgments of the sample is 0, and the false positive rate is 0. The RMSEC for PLS-DA qualitative analysis was 1.24, which was 0.93. The RMSEP is 1, which is 0.96, and the prediction set false positive rate is zero. The characteristic image has better model for the lin kernel function in the LS-SVM of the shelf life of the pear, which is better than the modeling effect of the RBF kernel function and better than the PLS-DA discriminant model. The LS-SVM and PLS-DA discriminant models were established by ENVI software to extract the spectra of the experimental samples. The false positive rates of RB-SVM using RBF and lin kernel functions were 4.5% and 0, respectively. Compared with the RBF kernel function, the model established by the lin kernel function predicts the shelf life of the pears better. The PLS-DA method has a principal component factor of 12, RMSEC and RMSEP of 0.48 and 0.78, respectively, and 0.99 and 0.97, respectively. The false positive rate of the modeling set and the prediction set are both zero. The model established by the lin kernel function in LS-SVM is better than the detection model established by PLS. The spectral information of the pears combined with LS-SVM can realize the detection and discrimination of the shelf life of the pears. Compared with the spectrum, the shelf life prediction model based on the image was used to distinguish the shelf life of the pear, while the feature image method, the selected area was less lost part of the information, the calculation amount was small, and the modeling result was relatively poor. The research on the hyperspectral imaging detection model of the shelf life of the pear provides theoretical guidance for consumers to correctly evaluate the freshness of the fruit, and also provides technical support for the development of the fruit shelf detection instrument in the later stage.

Keywords Hyperspectral imaging technique; Shelf life; Feature image; Least squares discriminant; Partial least squares support vector machine

(Received Jul. 27, 2018; accepted Nov. 20, 2018)

* Corresponding author