

基于高光谱成像技术的李果实成熟度判别

李丽丽¹, 王斌², 张学豪³, 张淑娟³

(1. 山西农业大学信息科学与工程学院, 山西太谷 030801) (2. 山西农业大学信息学院, 山西太谷 030801)

(3. 山西农业大学工学院, 山西太谷 030801)

摘要: 为了探究快速识别不同成熟度李果实的有效方法, 本研究以李果实作为研究对象, 基于高光谱成像技术对不同成熟度的李果实(未熟、半熟、成熟、过熟)样本的光谱信息(420~1000 nm)进行采集, 对采集样本的光谱信息进行平滑处理(Smoothing)与标准正态变量校正(SNV)相结合的方法预处理光谱数据, 分别以预处理后的全光谱(FS)数据和采用主成分分析(PCA)法提取主成分、采用连续投影算法(SPA)提取特征波长作为输入变量, 建立偏最小二乘法(PLS)模型, 比较不同判别模型的准确性。结果表明, FS-PLS建立的模型判别准确率最高, 综合准确率达到91.88%; 但考虑实验计算量及复杂程度来说, SPA-PLS建立的模型判别准确率最优, 综合准确率达到91.25%。该研究为李果实成熟度的判别检测提供了新的理论基础。

关键词: 李果实; 成熟度; 高光谱; 判别; 无损

文章编号: 1673-9078(2019)06-258-263

DOI: 10.13982/j.mfst.1673-9078.2019.6.034

Discrimination of Plum Fruit Maturity Based on Hyperspectral Imaging Technology

LI Li-li¹, WANG Bin², ZHANG Xue-hao³, ZHANG Shu-juan³

(1. College of Information Science and Engineering, Shanxi Agricultural University, Taigu 030801, China)

(2. College of Information, Shanxi Agricultural University, Taigu 030801, China)

(3. College of Engineering, Shanxi Agricultural University, Taigu 030801, China)

Abstract: In order to investigate an effective method for rapid discrimination of plum fruits at different maturity levels, plum fruits were used as the research object of this study, and spectral information (420~1000 nm) of plum fruits at different maturity levels (unripe, mid-ripe, ripe, over-ripe) was collected using hyperspectral imaging technology. A combination of smoothing and standard normalized variate (SNV) was applied to pretreat spectral data. Then, the partial least squares method (PLS) model was established for comparing the accuracy of different discriminant models, through using the pre-treated full spectrum (FS) data, the principal components extracted by the principal component analysis (PCA) method, and the feature wavelengths extracted by the successive projection algorithm (SPA) technique as input variables. The results revealed that the model established by FS-PLS had the highest discriminative accuracy and the accuracy rate reached 91.88%. However, considering the amount and complexity of the experimental calculations, the accuracy rate of the model established by SPA-PLS was the best with the comprehensive accuracy reaching 91.25%. This study provides a new theoretical basis for discriminative detection of plums maturity.

Key words: plum; maturity; hyperspectral imaging; discrimination; non destructive

李果实跟其他大部分水果一样, 都是一种典型的呼吸跃变型水果。水果采收期是影响水果品质的主要时期, 果实的成熟度是评价果实品质的重要参数。随着科技的发展, 农产品进出口市场竞争越来越激烈,

收稿日期: 2018-09-14

基金项目: 国家自然科学基金资助项目(31271973); 山西省自然科学基金资助项目(2012011030-3); 山西农业大学青年科技创新项目(2016005)

作者简介: 李丽丽(1988-), 女, 讲师, 硕士, 研究方向: 农业信息化技术

通讯作者: 张淑娟(1963-), 女, 教授, 研究方向: 农产品智能化检测与分级装备

确定合适的采摘成熟度是保证水果品质的关键。因此, 迫切需要对水果质量检测方法及设备的研发, 为水果分选, 分级提供科学依据。

高光谱成像检测技术不仅聚集光学、力学、电磁学、声学及电子鼻等方法的优点, 而且具有光谱信息、图像信息以及波段的多样化。国内外诸多学者采用该技术在农副产品无损检测方面获得硕果累累^[1-8]。例如, 黄双萍^[9]等基于高光谱成像对水稻穗瘟病害程度进行分级, 提出光谱词袋模型分析法并采用支持向量机分类算法建立穗瘟病害程度分级模型, 识别精度达到94.72%。孙静涛^[10]等采用高光谱技术结合特征波长

筛选的方法,同时对哈密瓜的可溶性固形物含量、硬度及成熟度进行了无损检测研究结果显示, MSC-CARS-SPA 方法所建立的可溶性固形物和硬度 SVM 预测模型最优, CARS-PCA-SVM 模型的判别结果与全光谱 SNV-SVM 模型相同,其校正集和预测集判别正确率分别为 95%和 94%。P. Rajkumar 等^[11]利用高光谱成像技术(400~1000 nm)对处在 20、25 及 30 ℃ 香蕉品质及成熟度进行了研究,通过多元线性回归对香蕉的可溶性固形物、水分及硬度进行了识别,决定系数分别为 0.85、0.87 及 0.91。Nayeli Velez Rivera^[12]等利用近红外高光谱图像(650~1100 nm)和机器学习对芒果机械损伤进行了检测,结果表明,对损伤七天的样本可以有效的区分出来,并且使用了不同的方法,其中最优方法为 K-NN(最近邻分类算法)达到 97.95%。

本研究利用高光谱成像技术对不同成熟度的李果实样本的光谱信息进行采集,对采集样本的光谱信息进行平滑处理(Smoothing)与标准正态变量校正(SNV)相结合的方法预处理光谱数据,分别以预处理后的全光谱(FS)数据和采用主成分分析(PCA)法提取主成分、采用连续投影算法(SPA)提取特征波长作为输入变量,建立偏最小二乘法(PLS)模型并进行预测。

1 材料和方法

1.1 实验样本

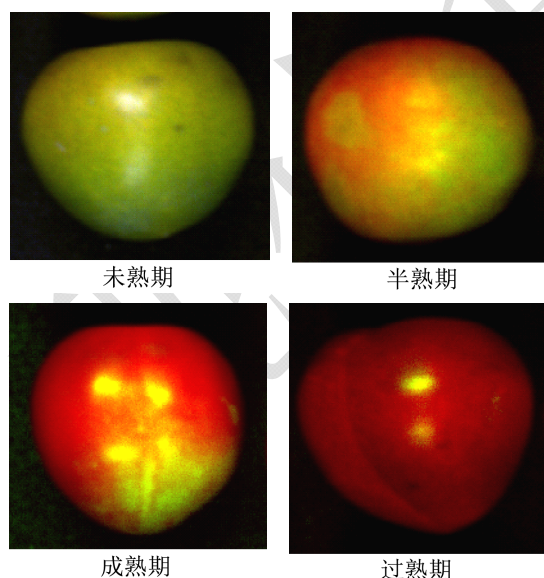


图1 不同成熟度的样本图

Fig.1 Different Maturity of plums

实验材料李果实样本于 2017 年 6 月~8 月份采摘至山西太谷侯城乡。分别采摘并从中挑选出过熟样本 80 个、成熟样本 80 个、半熟样本 80 个、未熟样本 80

个(共计 320 个样本)。利用 Kennard-Stone 算法^[12,13]将各类样本按照 3:1 的比例随机分成校正集和预测集,并选出 20 个未熟样本、20 个半熟样本、20 个成熟样本、20 个过熟样本作为预测集,其余 240 个校正集样本为 60 个未熟样本、60 个半熟样本、60 个成熟样本、60 个过熟样本。4 类样本各假定一个值作为成熟度判别的依据,将未熟期样本赋值为 1,半熟期样本赋值为 2,成熟期样本赋值为 3,过熟期样本赋值为 4。图 1 为四种成熟期的样本图。

1.2 高光谱数据采集系统

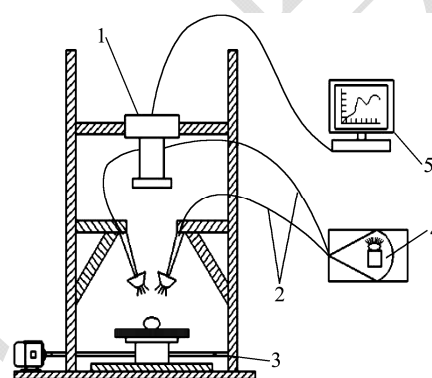


图2 高光谱采集系统

Fig.2 Hyperspectral imaging system

注: 1.摄像头; 2.光纤; 3.输送平台; 4.光源调节器; 5.计算机。

图 2 为高光谱成像系统结构示意图。仪器中有两个信号检测模块,分别为 MSC 621 VIS II 和 MCS 611 NIR 1.7 HR,采集的数据分别为 420~1000 nm 和 980~1700 nm,积分时间分别为 1.2 s 和 3.2 s。本实验采用 420~1000 nm 的 MSC 621 VIS II 模块,其中采集信息物距(对象至镜头边缘)为 30 cm,曝光时间为 3000 μ s,平台运行速度为 20 mm/s,总的运行距离为 400 mm。

1.3 光谱校正

在采集样本数据时由于光程不一致,对光谱数据的采集造成影响,并且由于采集中难免有随机噪声的产生,也会影响数据的正确采集。为了减少这种影响,能够在最大程度上挖掘样本光谱信息中的有效信息,提取到正确的光谱数据而进行光谱校正。常用的平滑校正方法有移动拟合法(Moving Average Smothing)、高斯平滑滤波法(Gaussian Filter Smothing)及卷积平滑滤波法(Savitzky Golay Smothing),另一种是多元散射校正(Multiplicative Scatter Correction)。本文运用平滑处理(Smoothing)与标准正态变量校正(SNV)相结合对光谱数据进行预处理。

2 结果与分析

2.1 光谱曲线和预处理

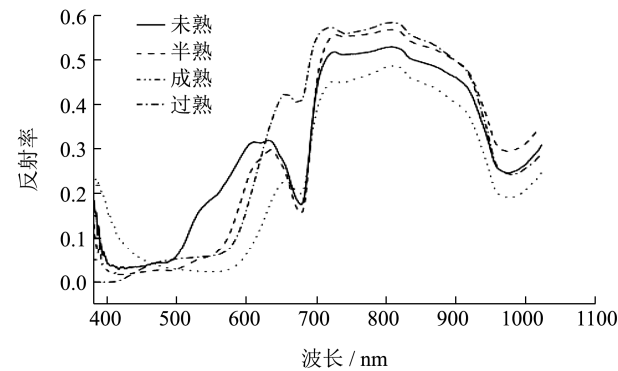


图3 原始光谱图

Fig.3 Original spectra

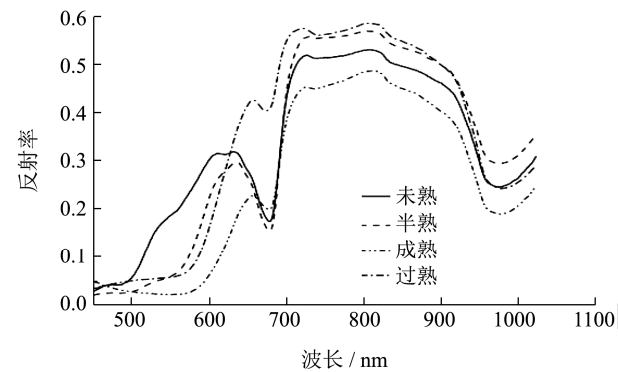


图4 处理光谱图

Fig.4 Spectrum processing

如图 3 所示为四种成熟度李果实的原始光谱曲线,由图可知其光谱曲线在 420~450 nm 存在大量噪声,本研究采用平滑处理(Smoothing)与标准正态变量校正(SNV)相结合的方法对光谱数据进行预处理,如图 4 所示为去除噪声后光谱曲线。本文将以此作为研究样本曲线,由图 4 可知,四种不同成熟度的李果实其光谱曲线还是存在一定的差距,波长范围在

700~900 nm 之间其区别比较明显,可能是由于李果实不同成熟度时内部碳水化合物、糖、氨基酸等成分所引起的。

2.2 基于全波段光谱的分类判别模型

将四类样本光谱数据经过预处理后,建立基于全波段光谱的李果实成熟度 PLS 模型,并对预测集样本的成熟度进行判别。由于模型预测结果为非整数型,并且考虑不同成熟度李果实设定的值,我们确定以 0.5 作为最大偏离值,即预测值和实际值的差值在 0.5 以内,我们认为预测结果准确,超出这个值,我们认为预测错误。所以,本文对李果实未熟样本的预测值应在 0.5~1.5 之间,半熟样本的预测值应在 1.5~2.5 之间,成熟样本的预测值应在 2.5~3.5 之间,过熟样本的预测值应在 3.5~4.5 之间。预测集判别结果如图 5 所示,表 1 为全波长 PLS 预测结果统计情况。

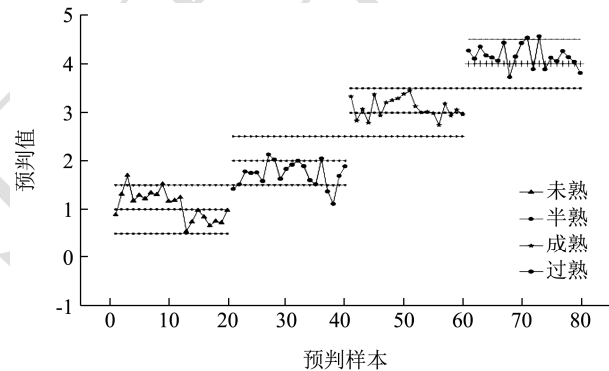


图5 全波段 PLS 模型预测集判别结果

Fig.5 Full band PLS prediction results

由图 5 和表 1 看出,基于全波段光谱对李果实成熟度进行校正时,未熟、半熟误判个数分别为 12、8 个,而成熟、过熟为 1、0 个。基于全波段光谱对李果实成熟度进行预测时,未熟和过熟误判数均为 1,半熟误判数为 3,而成熟误判数为 0。最后得出,FS-PLS 模型对李果实四种成熟度总体判别率为 91.88%。

表 1 全波长 PLS 预测结果统计表

Table 1 Discriminating results of PLS mode

全波长 PLS 模型预测结果		总数/个	误判数/个	识别率/%	综合识别率/%
未熟	校正集	60	12	80.0	87.50
	预测集	20	1	95.0	
半熟	校正集	60	8	86.7	85.85
	预测集	20	3	85.0	
成熟	校正集	60	1	98.3	99.15
	预测集	20	0	100	
过熟	校正集	60	0	100	97.50
	预测集	20	1	95.0	
总体		320	26		91.88

2.3 基于主成分的分类判别模型

2.3.1 主成分分析

主成分分析是研究如何通过少数的几个主成分来解释多变量的方差-协方差结构的分析方法,它是指把一组变量通过线性的变换,转换为一组不相关的变量。在这种变换中,保持变量的总方差不变,同时具有最大方差,我们称为第一主成分,具有次大方差,为第二主成分,以此类推,它主要侧重在信息贡献影响力的综合评价。

2.3.2 提取主成分

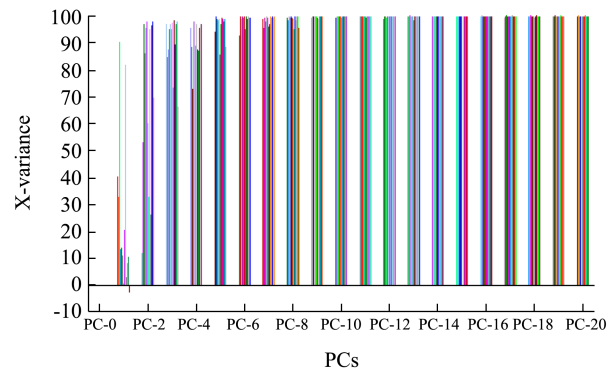


图 6 主成分累积贡献率

Fig.6 Cumulative contribution of the principal components

表 2 前 8 个主成分的累积贡献率

Table 2 Cumulative contribution of the previous 8 principal components

主成分	累积贡献率/%	主成分	累积贡献率/%
PC_01	43.0265	PC_05	97.8284
PC_02	77.6848	PC_06	98.9633
PC_03	88.9190	PC_07	99.2786
PC_04	94.1775	PC_08	99.5076

通过提取全波段光谱数据的主成分值,得到前 8

个主成分的累计贡献率如图 6,表 2 为数字化每一主成分的累计贡献率。由图 6 和表 2 可知,前 5 个主成分的累计贡献率已经达到 97.8284%。因此选取前 5 个主成分值用来作为建模的输入参数。

2.3.3 基于主成分的判别模型

以经过平滑滤波和多元散射校正的光谱数据主成分值作为输入参数,其类别值作为输出参数,建立 PCA-PLS 模型并对样本的成熟度进行判别。预测集判别结果如图 7 所示,表 3 为 PCA-PLS 预测结果统计情况。

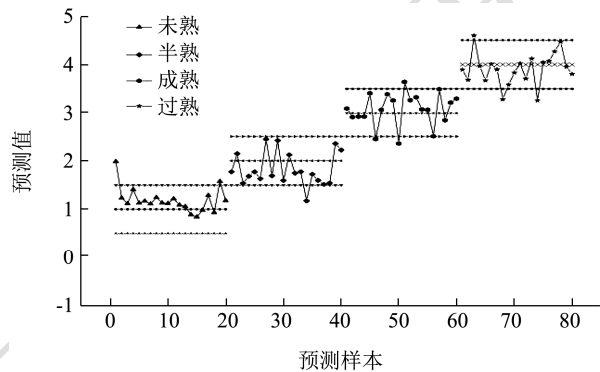


图 7 主成分 PLS 模型预测集判别结果

Fig.7 Classifying results of PLS model based on the principal component

由图 7 和表 3 可知,PCA-PLS 模型对李果实成熟度进行预测时,总体误判个数为 9。对未熟样本的误判数为 2,半熟为 1,而成熟和过熟误判数各为 3;基于主成分建立 PLS 模型对李果实成熟度进行校正时,未熟和半熟样本的误判数为 14 个和 15 个,而成熟和过熟误判数各为 11,其识别率分别为 83.35%、85.00%、83.35%和 83.35%,整体平均判别准确率为 81.25%。从中可以看出基于主成分对样本成熟度的预测都存在一定的误判。

表 3 主成分 PLS 预测结果统计表

Table 3 Predicting results of the PLS model based on the principal components

主成分 PLS 模型预测结果		总数/个	误判数/个	识别率/%	综合识别率/%
未熟	校正集	60	14	76.7	83.35
	预测集	20	2	90.0	
半熟	校正集	60	15	75.0	85.00
	预测集	20	1	95.0	
成熟	校正集	60	11	81.7	83.35
	预测集	20	3	85.0	
过熟	校正集	60	11	81.7	83.35
	预测集	20	3	85.0	
总体		320	60	81.25	

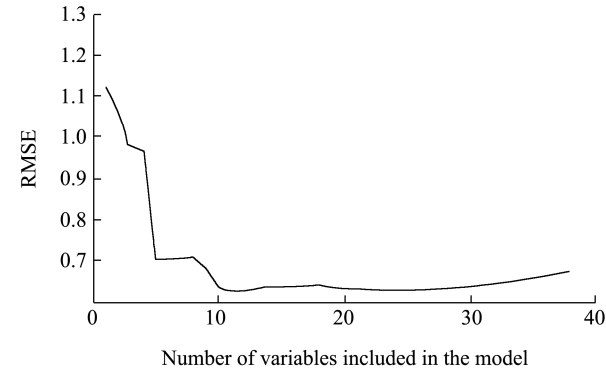


图8 连续投影算法的误差图

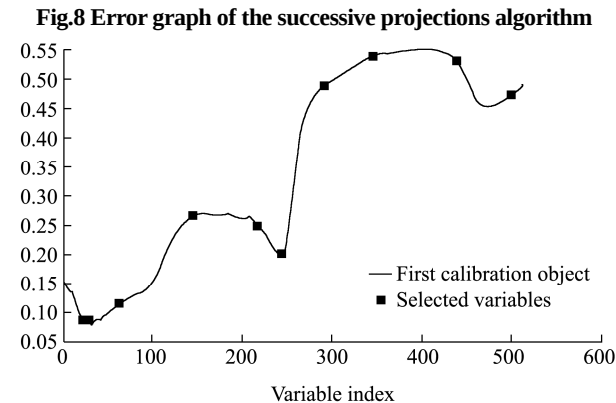


图9 连续投影算法提取的特征波长

Fig.9 Characteristic wavelengths extracted by the successive projections algorithm

2.4 基于特征波长的分类判别模型

2.4.1 特征波长提取

特征波长的提取是高光谱数据提取的一个中心环节,它直接影响到建模后结果的准确性。因为相邻波段之间难免具有一定的相关性,实验中通过高光谱成像技术获取的光谱数据中就或多或少存在一些冗余信息变量组,即多余的信息。这些冗余信息变量组不仅会降低运算速度还会影响建模效果,利用波长优选获取的特征波长建立模型可以大幅提高运算速度,提升

模型的预测精度。

2.4.2 特征波长提取

采用连续投影算法(SPA)对四类样本(未熟、半熟、成熟、过熟)的512个波段数据提取特征波长。由图8可知,当特征波长个数为10时, RMSE为0.63547,且为最小值,本文中将图9所示的10个波段作为最优特征波段,分别对应的波长为381.08、382.25、386.95、408.17、442.73、493.56、596.25、812.94、962.75及1007.86 nm。

2.4.3 特征波长建模

以经过预处理后的平滑的特征波长光谱数据及其类别值为输入参数,建立 SPA-PLS 模型并对样本的成熟度进行判别。预测集判别结果如图10所示,表4为 SPA-PLS 预测结果统计情况。

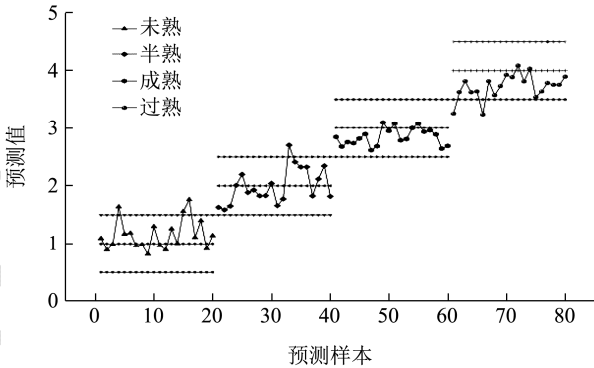


图10 特征波长 PLS 预测集判别结果

Fig.10 Predicting results of PLS model based on the characteristic wavelengths

由图10和表4可以看出, SPA-PLS 模型对李果实成熟度进行校正时,未熟和半熟误判个数分别为9个和6个,而成熟和过熟为2个和5个。SPA-PLS 模型对李果实成熟度进行预测时,未熟和过熟误判数分别为3个和1个,半熟误判数为0,而成熟误判数为2。综合以上, SPA-PLS 模型对李果实各种成熟度的总体判别正确率为91.25%。

表4 特征波长 PLS 模型预测结果统计表

Table 4 Predicting results of PLS model based on the characteristic wavelengths

特征波长 PLS 模型预测结果		总数/个	误判数/个	识别率/%	综合识别率/%
未熟	校正集	60	9	81.7	83.35
	预测集	20	3	85.0	
半熟	校正集	60	6	90.0	92.50
	预测集	20	1	95.0	
成熟	校正集	60	2	96.7	98.35
	预测集	20	0	100	
过熟	校正集	60	5	91.7	90.85
	预测集	20	2	90.0	
总体		320	28	91.25	

3 结论

基于高光谱成像技术对不同成熟度的李果实(未熟、半熟、成熟、过熟)样本的光谱信息(420~1000 nm)进行采集,对 Smoothing+SNV 方法预处理后的光谱,分别建立 FS-PLS、PCA-PLS、SPA-PLS 预测判别模型。结果表明,FS-PLS 建立的模型判别准确率最高,综合准确率达到 91.88%;但考虑实验计算量及建模复杂程度来说,SPA-PLS 建立的模型判别准确率最优,综合准确率达到 91.25%。该研究为不同成熟度的李果实判别检测提供了新的理论基础。

参考文献

- [1] 迟茜,王转卫,杨婷婷,等.基于近红外高光谱成像的猕猴桃早期隐形损伤识别[J].农业机械学报,2015,46(3):234-241
CHI Qian, WANG Zhuan-wei, YANG Ting-ting, et al. Recognition of early hidden bruises on kiwifruits based on near-infrared hyperspectral imaging technology [J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(3): 235-241
- [2] 张雪茹,冯美臣,李广信,等.低温胁迫下冬小麦叶绿素含量高光谱估测[J].山西农业科学,2017,45(4):526-529
ZHANG Xue-ru, FENG Mei-chen, LI Guang-xin, et al. Hyperspectral estimation on the chlorophyll content of winter wheat under low temperature stress [J]. Journal of Shanxi Agricultural Sciences, 2017, 45(4): 526-529
- [3] 章海亮,高俊峰,何勇,等.基于高光谱成像技术的柑橘缺陷无损检测[J].农业机械学报,2013,44(9):177-181
ZHANG Hai-liang, GAO Jun-feng, HE Yong, et al. Nondestructive detection of citrus defection using hyper-pectra imaging technology [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(9): 177-181
- [4] 薛建新,张淑娟,张晶晶.基于高光谱成像技术的沙金杏成熟度判别[J].农业工程学报,2015,31(11):300-307
XUE Jian-xin, ZHANG Shu-juan, ZHANG Jing-jing. Ripeness classification of Shajin apricot using hyperspectral imaging technique [J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(11): 300-307
- [5] 余克强,赵艳茹,李晓丽,等.基于高光谱技术的鲜枣裂纹识别研究[J].光谱学与光谱分析,2014,34(21):532-537
YU Ke-qiang, ZHAO Yan-ru, LI Xiao-li, et al. Study on identification the crack feature of fresh jujube using hyperspectral imaging [J]. Spectroscopy and Spectral Analysis, 2014, 34(21): 532-537
- [6] 强锋,王芹志,何建国,等.基于 VIS/NIR 高光谱技术的灵武长枣 SSC 含量的无损检测研究[J].现代食品科技,2017,33(4):283-288
QIANG Feng, WANG Qin-zhi, HE Jian-guo, et al. Application of Vis/NIR hyperspectral imaging technology in non-destructive measurement of soluble solid content in lingwu jujube [J]. Modern Food Science and Technology, 2017, 33(4): 283-288
- [7] 李丽丽,王斌,张学豪,等.基于高光谱图像信息的李果实成熟度判别[J].现代食品科技,2017,33(12):213-218
LI Li-li, WANG Bin, ZHANG Xue-hao, et al. Discrimination on maturity of plums based on hyperspectral imaging information [J]. Modern Food Science and Technology, 2017, 33(12): 213-218
- [8] 罗阳.基于 NIR 高光谱成像技术的长枣虫害及可溶性固形物无损检测研究[D].宁夏:宁夏大学,2013
LUO Yang. Non-destructive Detection for insect-infested and SSC of Changzao based on NIR hyperspectral imaging technique [D]. Ningxia: Ning Xia University, 2013
- [9] 黄双萍,齐龙,马旭,等.基于高光谱成像的水稻穗瘟病害程度分级方法[J].农业工程学报,2015,31(1):212-219
HUANG Shuang-ping, QI Long, MA Xu, et al. Grading method of rice panicle blast severity based on hyperspectral image [J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(1): 212-219
- [10] 孙静涛,马本学,董娟,等.高光谱技术结合特征波长筛选和支持向量机的哈密瓜成熟度判别研究[J].光谱学与光谱分析,2017,37(7):2184-2191
SUN Jing-tao, MA Ben-xue, DONG Juan, et al. Study on maturity discrimination of hami melon with hyperspectral imaging technology combined with characteristic wavelengths selection methods and SVM [J]. Spectroscopy and Spectral Analysis, 2017, 37(7): 2184-2191
- [11] P Rajkumar, G Elmasry, G S V Raghavan, et al. Studies on banana fruit quality and using hyperspectral imaging [J]. Journal of Food Engineering, 2012, 108(1): 194-V200
- [12] Nayeli Velez Rivera, Juan Gomez-Sanchis, Jorge Chanona-Perez. Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning [J]. Biosystems Engineering, 2014, 122: 91-98
- [13] Lalit Mohan Kandpal, Santosh Lohumi, Moon S. Kim, et al. Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds [J]. Sensors and Actuators, 2016, 229(28): 534-544

现代食品科技