高光谱图像感兴趣区域对苹果糖度模型的影响

郭志明[1,2]，黄文倩[2]，彭彦昆[1]，王秀[2]，李江波[2]
（1.中国农业大学工学院，北京 100083）（2.国家农业信息化工程技术研究中心，北京 100097）

摘要：高光谱图像技术作为一种强有力的新兴技术，已应用于食品农产品品质与安全检测研究，然而高光谱图像中感兴趣区域形状大小的选择直接影响着检测的精度和稳定性。首先采集苹果 330~1100 nm 的高光谱图像，分别提取不同大小的圆形感兴趣区域和方形感兴趣区域的平均光谱，经光谱预处理以消除噪声及无关信息的影响，然后采用偏最小二乘法分别建立苹果的糖度定量分析模型，并以独立样本的预测集进行验证，分析感兴趣区域形状大小对高光谱图像建模精度的影响。结果表明，提取直径为 150 像素的圆形感兴趣区域建立的苹果糖度模型精度最高，预测能力最强，校正集相关系数 R_c 为 0.9305，校正均方根误差 RMSEC 为 0.4331，预测集相关系数 R_p 为 0.9232，预测均方根误差 RMSEP 为 0.4568。研究表明，针对研究对象选择合适形状和大小的感兴趣区域，对提高模型精度、发挥高光谱图像的技术优势具有重要意义。

关键词：高光谱成像技术；感兴趣区域；偏最小二乘法；苹果；糖度
文章篇号：1673-9078(2014)8-59-63

Impact of Region of Interest Selection for Hyperspectral Imaging and Modeling of Sugar Content in Apple

GUO Zhi-ming[1,2], HUANG Wen-qian[2], PENG Yan-kun[1], WANG Xiu[2], LI Jiang-bo[2]
(1.College of Engineering, China Agricultural University, Beijing 100083, China)
(2.National Engineering Research for Information Technologies in Agriculture, Beijing 100097, China)

Abstract: Hyperspectral imaging is an effective imaging technique that has been applied to quality and safety inspection of food and agricultural products. However, the selected shape and size of the region of interest (ROI) in hyperspectral imaging directly affects the accuracy and stability of the measurement. In this study, a hyperspectral imaging system was developed for wavelengths spanning from 330 to 1100 nm to acquire hyperspectral images of apple samples. Mean reflectance spectra of round and square ROIs with different sizes were extracted. After the spectral pretreatment to eliminate the impacts of noise and irrelevant information, models for quantitative analysis of the sugar content in apple were developed using partial least squares method. The models were externally verified with the prediction set consisting of independent samples, in order to analyze the impact of the shape and size of ROIs on the accuracy of hyperspectral imaging system modeling. The results showed that the model of sugar content in apple, constructed using a round ROI with 150 pixels diameter, yielded the highest accuracy and predictive capability. The correlation coefficient of calibration set (R_c) was 0.9305, the root mean square error of calibration (RMSEC) was 0.4331, the correlation coefficient of prediction set (R_p) was 0.9232, and the root mean square error of prediction (RMSEP) was 0.4568. These results demonstrate that selecting an ROI with an appropriate shape and size is important to improve the accuracy of modeling and harness the full potential of the hyperspectral imaging technique.

Key words: hyperspectral imaging technique; region of interest; partial least squares; apple; sugar content

光谱分析技术已广泛应用于食品农产品品质与安全无损检测领域，如近红外光谱可以较好的表征食品农产品内部品质信息[1]，但通过积分球或光纤传输获取的近红外光谱一般仅能采集测试对象单点的光谱信息，在测试空间广度上受到很大限制，且样本表面颜色、组织均匀程度和品质分布特性等方面的差异均能够影响近红外光谱检测的全面性和可靠性[2~3]。高光谱成像技术由于能同时获得待测物的图像和光谱信息，图像技术能反映待测对象的外在特征，光谱技术能检测待测对象的物理结构和化学成分，因此能同时对待测对象的外观特性和内部成分进行检测。高光谱成像

收稿日期：2014-03-14
基金项目：国家自然科学基金资助项目（31301236）；北京市农林科学院科技创新基金（CXJJ201314）
作者简介：郭志明，男，博士生，主要从事高光谱图像技术的农产品质量安全检测研究
通讯作者：彭彦昆，教授，博士生导师，主要从事农产品无损检测技术与装备研究

技术将传统的机器视觉检测技术和光谱分析技术有机的结合在一起，具有高分辨率和图谱合一的优点，已被引入到食品农产品品质与安全无损检测领域[4~7]。

据 FAO 统计数据，2012 年我国苹果种植面积达到 220 万公顷，苹果总产量 3700 万吨，种植面积和产量均为世界第一，在农业经济发展中占有重要地位。国内外研究人员先后利用高光谱成像技术检测苹果的内外品质指标。国外，Xing 等[8]利用高光谱图像的第一主成分得分图识别苹果的轻微损伤和果梗花萼。ElMasry 等[9~10]采用高光谱图像技术选择有效波段，结合模式识别方法可以准确检测苹果擦伤，还可以识别苹果冷害。Peng 等[11]利用高光谱散射图像的洛伦兹拟合参数建立了苹果硬度和糖度含量的预测模型。Mendoza 等[12]综合利用高光谱散射数据的光谱和图像特征信息建立苹果的硬度和糖度的偏最小二乘分析模型。国内，赵杰文等[13]利用高光谱成像技术结合支持向量机建立了苹果硬度的预测模型。单佳佳等[14]利用高光谱空间散射曲线的洛伦兹拟合参数对苹果的硬度和可溶性固溶物含量进行同时检测。Wang 等[15]选择高光谱散射图像的特征波长后结合 PLS 和 ANN 建立苹果硬度预测模型。但目前还没有开展高光谱成像的感兴趣区域选择方法的系统化研究，前期研究一般以经验选择正方形的区域，像素大小选择因人而异。高光谱图像中感兴趣区域的选择是高光谱图像源头数据的选择问题，将直接影响着定性或定量分析模型的精度和稳定性。研究中分别提取不同大小的圆形感兴趣区域和方形感兴趣区域的平均光谱，采用稳健的偏最小二乘法分别建立苹果的糖度定量分析模型，分析感兴趣区域形状大小对高光谱图像建模精度的影响。

1 材料与方法

1.1 试验材料

试验用苹果为富士品种，样本于 2013 年 10 月中旬在苹果试验示范基地手工采摘（分别位于北京昌平和山东泰安）。选择形状大小均匀、无缺陷、损伤或污染物的苹果共 173 个，果径 75~80 mm，均属优等品（依据 GB/T 10651-2008）。随机挑选 115 个（70 个山东、45 个北京）作为校正集，其余的 58 个（35 个山东、23 个北京）作为预测集，将样本编号后置于 4 ℃冰柜中贮藏。试验前，将苹果从冰柜取出置于实验室中 12 h，以使苹果样本整体温度与环境温度一致，同时避免苹果表面有水分影响高光谱图像的采集，试验过程保持实验室温湿度基本不变（20 ℃，相对湿度 55%）。

1.2 高光谱图像采集系统

苹果样本的高光谱图像数据通过基于光谱仪的高光谱图像采集系统获得，如图 1 所示，该系统主要包括高光谱相机、光源组、移动平台、计算机和专用高光谱图像采集软件。其中高光谱相机由成像光谱仪（ImSpector V10E, Spectral Imaging Ltd., Oulu, Finland）、EMCCD 相机（Luca-R, Andor Technology, Belfast, UK）和镜头（OLE23 f-2.4/23 mm, Spectral Imaging Ltd., Oulu, Finland）构成。有效光谱范围 330~1100 nm，入射狭缝宽度 30 μm，光谱分辨率 2.8 nm。面阵 EMCCD 相机分辨率为 1004 (H)×1000 (V)，半导体制冷（-20 ℃）。光源选用 150 W 卤钨灯，配置 Y 型光纤线光源（3900-ER, Illumination Technologies, Inc., USA）。高光谱图像数据采集通过 HyperSpectral Image（Isuzu Optics Corp., Taiwan）软件完成。整个系统置于定制的黑箱内，可有效避免高光谱图像采集时环境杂散光的干扰。

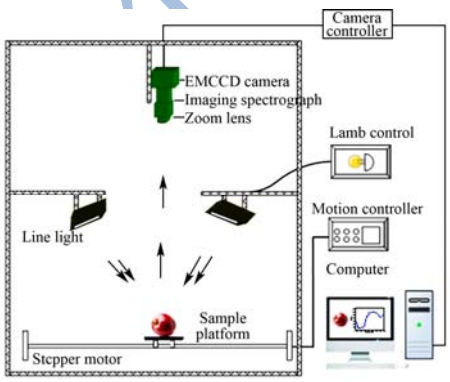


图 1 高光谱成像系统结构示意图

Fig.1 Schematic representation of the hyperspectral imaging system

1.3 高光谱图像获取与校正

苹果样本置于移动平台的水果托盘上，使果梗和花萼处于同一水平面上，推扫式数据采集，调整相机的曝光时间、平台移动速度、光源强度和样本与镜头距离等参数间的匹配，保证采集的高光谱图像数据的质量。在苹果样本高光谱图像数据采集过程中，参数设置为曝光时间 55 ms，平台移动速度 0.70 mm/s，镜头到样本距离 400 mm，相机制冷温度恒定在-20 ℃。每个苹果样本的图像分辨率为 700 pixel×1004pixel，采集得到 1000 个波长下的图像，最终得到一个大小 700 pixel×1004 pixel×1000 的高光谱图像数据立方体。

高光谱图像采集过程中，因光源在各波段下强度分布的差异性和相机暗电流噪音的影响，会夹杂部分噪音信息。这些噪音信息会影响高光谱图像的质量，

进而影响高光谱图像定性或定量分析模型的精度和稳定性。因此有必要对高光谱图像进行校正，以消除噪声的影响。研究采用黑白校正方法，采用公式(1)进行校正，其中 D 为关闭光源未打开镜头盖时采集的全黑参考高光谱图像，W 为采集的聚四氟乙烯标准白板得到的全白参考高光谱图像，S 为苹果样本原始高光谱图像，R 为校正后的高光谱图像，下标 λ，n 表示在波长 λ 下第 n 个像素点。高光谱图像采集系统所用相机为 14 位精度，校正公式最后乘以 16384 使每个校正后的像素点保留原来的相对强度值。

$$R_{\lambda,n}=\frac{S_{\lambda,n}-D_{\lambda,n}}{W_{\lambda,n}-D_{\lambda,n}}\times 16384 \quad (1)$$

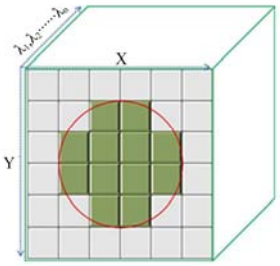


图 2 高光谱图像圆形感兴趣区域选择图解

Fig.2 Diagrammatic illustration of pixel selection by circular region of interest from hyperspectral image

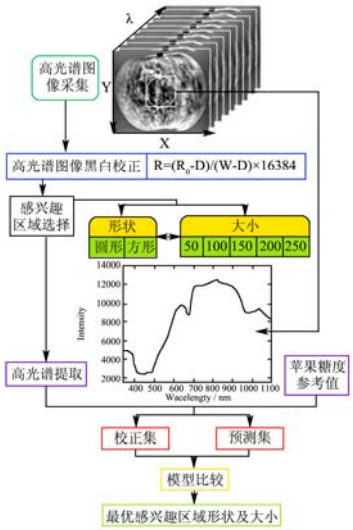


图 3 高光谱图像感兴趣区域选择流程图

Fig.3 Flowchart of the main steps involved in the whole procedure of hyperspectral image analysis

1.4 感兴趣区域选择

高光谱图像技术进行样本内部品质分析时，首先是特征光谱的提取环节，感性兴趣区域的形状和大小都会影响下一步光谱的预处理、特征选择、建模分析和预测精度。前期研究均采用方形感兴趣区域，正方形内的各像素点都包含在内。圆形感兴趣区域的选择如图 2 所示，单个像素点的面积超过一半在圆圈内则入选感兴趣区域。研究中感兴趣区域的选择均以苹果平放上部的空间顶点为中心，提取不同大小的圆形感兴趣区域和方形感兴趣区域的平均光谱，感兴趣区域横向大小设置为 50、100、150、200 和 250，对应圆形感兴趣区域 1960、7852、17652、31380 和 49040 个像素，对应方形感兴趣区域 2500、10000、22500、40000 和 62500 个像素。研究的思路如图 3 所示，对提取的不同感兴趣区域分别采用偏最小二乘法分别建立苹果的糖度定量分析模型，以分析感兴趣区域形状和大小对模型建立结果的影响。

1.5 糖度值测定

苹果样本经高光谱图像采集后，参照 GB/T 10651-2008，分别测定可溶性固形物含量值（水果生产者和消费者一般称为糖度）。采用 Arias 500 型阿贝折光仪（Reichert Inc., USA），自动温度补偿为 20 ℃。从苹果样本中挤滤出汁液 1~2 滴，滴在棱镜平面的中央，迅速关合辅助棱镜，明暗分界线与十字准线点同线的读数即苹果在 20 ℃下可溶性固形物的百分率。作为高光谱图像苹果样本糖度预测的标准值。表 1 列出了被测苹果可溶性固形物实测值的变化范围、平均值和标准差等统计量。

表 1 苹果样本校正集和预测集糖度检测统计结果

Table 1 Statistics of sugar content (Brix°) for the calibration and prediction sets

Subsets	Samples	Ranges /Brix°	Mean /Brix°	Standard deviation /Brix°
Calibration set	115	11.56~16.46	13.732	1.186
Prediction set	58	11.62~16.28	13.784	1.195

2 结果与讨论

2.1 感兴趣区域光谱特征分析与预处理

苹果高光谱图像有效波段范围的不同大小圆形感兴趣区域的反射光谱如图 4 所示，图中以波长为横坐标，校正后的响应强度为纵坐标，方形感兴趣区域的反射光谱图与此类似。不同大小感兴趣区域的反射光谱整体趋势一致，但随着感兴趣区域增大，反射光谱的响应强度降低，特别是在图 4 中放大部分更能明显的看出，这主要是苹果类球形表面光照强度分布不均引起的。

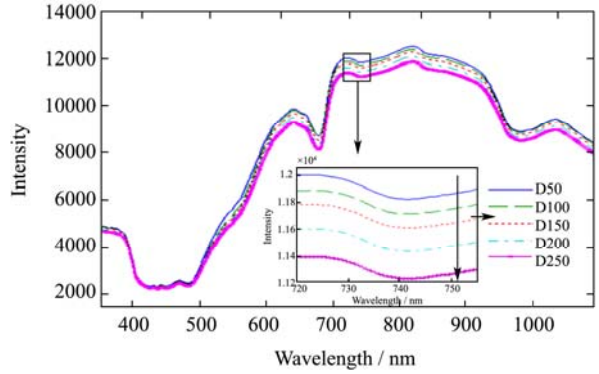


图 4 不同大小圆形感兴趣区域的平均光谱比较

Fig.4 Mean relative reflectance spectra of same apple sample with different size of circular regions of interest

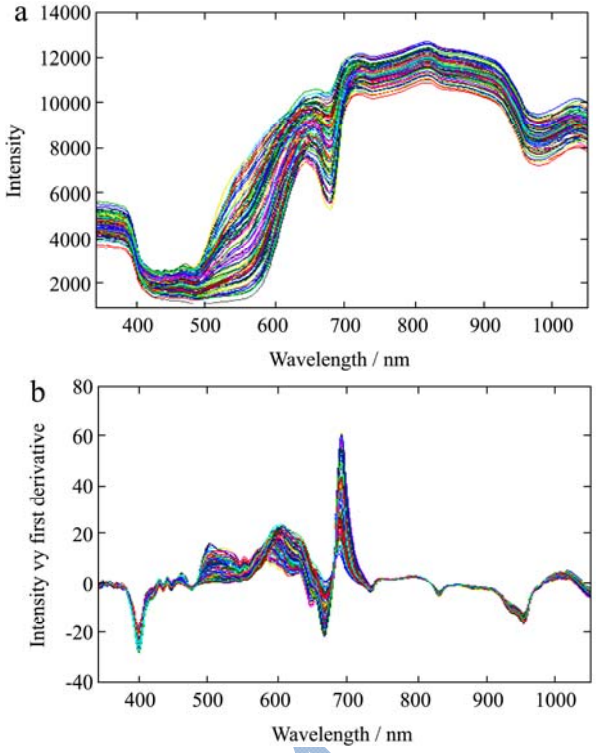


图 5 苹果高光谱图像感兴趣区域原始光谱和一阶导数光谱

Fig.5 Raw relative reflectance spectra (A) and first derivative processing spectra (B) from hyperspectral image of ROI in apple samples

校正集样本感兴趣区域的原始光谱如图 5a 所示，光谱在 330~400 nm 和 1050~1100 nm 两个末端光谱噪声较大，光谱信噪比低；在 400~500 nm 区间内曲线平滑，吸光度变化不大；500 nm 后反射强度快速升高，波峰为 650 nm、720 nm、824 nm 和 1039 nm，波谷为 679 nm 和 980 nm。经对比分析多元散射校正、标准正交变量变换、S~G 一阶微分、S~G 二阶微分和极值标准化等光谱预处理方法[16]，选用 S~G 卷积平滑后的一阶微分预处理以消除噪声及无关信息的影响。从 S~G 一阶微分光谱预处理的光谱图 5b 可以看出，485~640 nm 范围内光谱波动较大，主要可能受颜色和水果大小的影响。研究选取 640~1050 nm 区间范围的光谱数据用于建立苹果的糖度定量分析模型。

2.2 偏最小二乘分析模型

表 2 不同大小和形状的苹果糖度偏最小二乘预测结果比较

Table 2 PLS models for prediction of sugar content in apple with different shapes and sizes

Shape	Size	PLS factor	Calibration set		Prediction set	
			R_c	RMSEC/%	R_p	RMSEP/%
Roundness	50	6	0.9142	0.4778	0.8972	0.5366
	100	6	0.9244	0.4528	0.9184	0.4695
	150	6	0.9305	0.4331	0.9232	0.4568
	200	7	0.9177	0.4684	0.9097	0.5034
	250	6	0.8955	0.5256	0.8802	0.6007
Square	50	6	0.9108	0.4856	0.902	0.5224
	100	7	0.9263	0.4503	0.9193	0.4682
	150	7	0.9184	0.4698	0.9082	0.5084
	200	7	0.9021	0.51	0.8826	0.6003
	250	6	0.8745	0.6017	0.8587	0.6454

苹果的糖度定量分析模型采用稳健的偏最小二乘法建立，以含量的实测值与预测值的相关系数 R、校正均方根误差（root mean square error of calibration, RMSEC）和预测均方根误差（root mean squared error of prediction, RMSEP）对模型进行综合评价。在偏最小二乘分析模型建立过程中，以均方根误差最小来确定主因子数。如果选择的主因子数太小，将会丢失原始光谱较多的有用信息，拟合不充分；如果选取主因子数太大，会将噪音过多的吸收进来，出现过拟合现象，所建模型的预测误差会显著增大。

对苹果不同形状和大小的感兴趣区域的光谱数据分别采用偏最小二乘法建立模型，结果如表 2 所示。圆形感兴趣区域横向最大选择 150 个像素点时得到的模型结果最好，校正集相关系数 R_c 为 0.9305，RMSEC 为 0.4331，预测集相关系数 R_p 为 0.9232，RMSEP 为 0.4568，相对分析误差 RPD 为 2.60，表明定标效果良好。图 6 为该模型校正集样本和预测集样本的预测值与实测值之间的散点图，可以看出预测值与实测值有很好的相关性。从表 2 可以看出，在采集相同的高光谱图像的条件下，圆形感兴趣区域选择太小和太大都不利于建立高精度的定量分析模型。方形感兴趣区域选择 100×100 pixel 时得到的模型结果最好，校正集相关系数 R_c 为 0.9263，RMSEC 为 0.4503，预测集相关系数 R_p 为 0.9193，RMSEP 为 0.4628，相对分析误差 RPD 为 2.54，结果稍差于圆形 150 像素点的模型。圆

形和方形感兴趣区域这两种提取方式获得的高光谱图像苹果糖度的预测均方根误差均小于 0.5 Brix°，表明高光谱图像技术对苹果糖度进行无损检测是行之有效的。

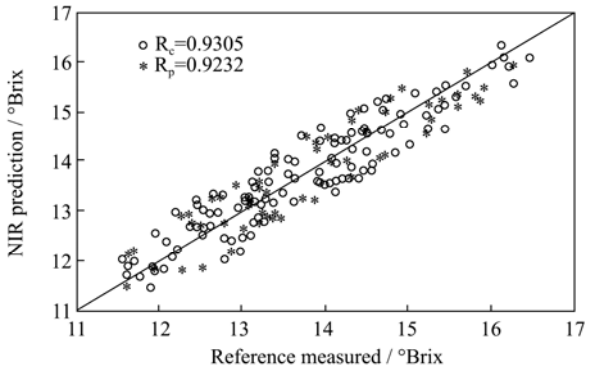


图 6 苹果糖度分析模型校正集和预测集的散点图

Fig.6 Apple sugar content prediction by partial least square for the calibration (o) and prediction sets (*)

2.3 感兴趣区域选择对模型的影响分析

分析表 2 可以看出，感兴趣区域的大小影响模型的精度和预测能力，选择感兴趣区域的平均光谱有利于消除各像素点的随机噪声，降低光谱选择的随机性。当感兴趣区域选择偏小时，不能包含足够的光谱信息及对应的内部品质信息；感兴趣区域选择偏大时，又会融入不相关信息，经平均处理会降低光谱信息与内部品质的相关关系。另外，圆形感兴趣区域的最优建模和预测结果优于方形感兴趣区域的结果。苹果是类球形体，特别是果形不规则时引起表面光照分布不均匀，会引起平均光谱的波动。另一个原因是苹果轴向糖度呈不均匀分布，经感官评定和折光仪糖度测定验证发现，苹果花萼部糖度测定值高而感官评定低，因果汁相对含量少；赤道部糖度测定值次之但多汁，口感好；果梗部糖度测定值最低。通过选择合适大小的圆形感兴趣区域，利用高光谱图像可以建立精度高稳定性好的苹果糖度定量分析模型。

3 结论

对高光谱图像分别提取不同大小的圆形感兴趣区域和方形感兴趣区域的平均光谱，经 S~G 平滑的一阶微分预处理以消除噪声及无关信息的影响，然后采用稳健的偏最小二乘法分别建立苹果的糖度定量分析模型，并用独立样本的预测集进行验证，分析感兴趣区域形状大小对高光谱图像建模精度的影响。圆形感兴趣区域的最优建模和预测结果优于方形感兴趣区域的结果，其中提取直径为 150 像素的圆形感兴趣区域建立的苹果糖度模型精度和预测能力最强，校正集相关系数 R_c 为 0.9305，RMSEC 为 0.4331，预测集相关系数 R_p 为 0.9232，RMSEP 为 0.4568。研究表明，根据研究对象选择合适形状和大小的感兴趣区域，对提高模型精度、发挥高光谱图像的技术优势具有重要意义。感兴趣区域的形状和大小的选择研究可为高光谱成像的品质分析与安全检测提供理论依据和方法参考。

参考文献

[1] 王若兰,王春华,黄亚伟.小麦脂肪酸值的近红外光谱快速测定研究[J].现代食品科技,2013,29(2):393-396
WANG Ruo-lan, WANG Chun-hua, HUANG Ya-wei. Rapid determination of fatty acid content in wheat by Near-Infrared Spectroscopy [J]. Modern Food Science & Technology, 2013, 29(2): 393-396

[2] Qin J, Chao K, Kim M S, et al. Hyperspectral and multispectral imaging for evaluating food safety and quality [J]. Journal of Food Engineering, 2013, 118(2): 157-171

[3] Wu D, Sun D. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review part ii: applications [J]. Innovative Food Science & Emerging Technologies, 2013, 19: 15-28

[4] Kim M S, Chen Y R, Mehl P M. Hyperspectral reflectance and fluorescence imaging system for food quality and safety [J]. Transactions of the American Society of Agricultural Engineers, 2001, 44(3): 721-729

[5] 刘木华,赵杰文,郑建鸿,等.农畜产品品质无损检测中高光谱图像技术的应用进展[J].农业机械学报,2005,36(9):139-143
LIU Mu-hua, ZHAO Jie-wen, ZHENG Jian-hong, et al. Review of hyperspectral imaging in quality and safety inspections of agricultural and poultry products [J]. Transactions of the Chinese Society of Agricultural Machinery [J]. 2005, 36(9): 139-143

[6] 彭彦昆,张雷蕾.农畜产品品质安全高光谱无损检测技术进展和趋势[J].农业机械学报,2013,44(4):137-145
PENG Yan-kun, ZHANG Lei-lei. Advancement and trend of hyperspectral imaging technique for nondestructive detection of agro-product quality and safety [J]. Transactions of the Chinese Society of Agricultural Machinery, 2013, 44(4): 137-145

[7] Gowen A A, O'Donnell C P, Cullen P J, et al. Hyperspectral imaging an emerging process analytical tool for food quality and safety control [J]. Trends in Food Science & Technology, 2007, 18(12): 590-598

[8] Xing J, Baerdemaeker J. Bruise detection on ‘Jonagold’

apples using hyperspectral imaging [J]. Postharvest Biology and Technology, 2005, 37(2): 152-162

[9] ElMasry G, Wang N, Vigneault C, et al. Early detection of apple bruises on different background colors using hyperspectral imaging [J]. LWT-Food Science and Technology, 2008, 41(2): 337-345

[10] ElMasry G, Wang N, Vigneault C. Detecting chilling injury in red delicious apple using hyperspectral imaging and neural networks [J]. Postharvest Biology and Technology, 2009, 52(1): 1-8

[11] Peng Y, Lu R. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content [J]. Postharvest Biology and Technology, 2008, 48(1): 52-62

[12] Mendoza F, Lu R, Ariana D, et al. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content [J]. Postharvest Biology and Technology, 2011, 62(2): 149-160

[13] ZHAO Jiewen, CHEN Quansheng, Saritporn V, et al. Determination of apple firmness using hyperspectral imaging technique and multivariate calibrations [J]. Transaction of the CSAE, 2009, 25(11): 226-231

[14] 单佳佳,吴建虎,陈菁菁,等.基于高光谱成像的苹果多品质参数同时检测[J].光谱学与光谱分析,2010,30(10): 2729-2733

SHAN Jia-jia, WU Jian-hu, CHEN Jing-jing, et al. Rapid nondestructive detection of apple quality attributes using hyperspectral scattering images [J]. Spectroscopy and Spectral Analysis, 2010, 30(10): 2729-2733

[15] Wang S, Huang M, Zhu Q. Model fusion for prediction of apple firmness using hyperspectral scattering image [J]. Computers and Electronics in Agriculture, 2012, 80: 1-7

[16] 庞艳苹,刘坤,闫军颖,等.近红外光谱法快速鉴别成安草莓[J].现代食品科技,2013,29(5):1160-1162

PANG Yan-ping, LIU Kun, YAN Jun-ying, et al. Rapid identification of cheng'an strawberry with near infrared spectroscopy [J]. Modern food science & technology, 2013, 29(5): 1160-1162

现代食品科技