

文章编号: 1000-7032(2019)06-0808-07

基于近红外的柚子品种判别和糖度检测通用模型

李 雄, 刘燕德*, 欧阳爱国, 孙旭东, 胡 军, 姜小刚, 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: 针对厚皮果透光性差、不同柚子品种糖度在线检测要单独建模等问题, 本文以柚子为研究对象, 采集有效光谱 325 条, 对比分析不同柚子品种在 710 nm 和 800 nm 附近的两个吸收峰光谱响应特性。550 ~ 970 nm 全波段范围内的光谱采用 SPA 无信息消除后建立柚子偏最小二乘判别模型误判率为 1.25%; 偏最小二乘法在 550 ~ 970 nm 全波段范围和去差异化后 750 ~ 930 nm 波段范围的预测相关系数分别为 0.58 和 0.86, 预测均方根误差 RMSEP 分别为 0.84 和 0.55。实验结果表明, 连续投影法结合偏最小二乘判别模型可以实现不同柚子品种的定性判别, 变异系数法对光谱去差异化后结合最小二乘模型对厚皮果柚子可溶性固形物的定量检测效果最佳, 该研究为不同品种的厚皮果在线分选技术提供了参考和理论依据。

关键词: 近红外光谱; 在线检测; 模型; 可溶性固形物; 品种判别

中图分类号: O657.33

文献标识码: A

DOI: 10.3788/fgxb20194006.0808

A General Model for Judging and Brix Detection of Grapefruit Variety Based on Near Infrared

LI Xiong, LIU Yan-de*, OUYANG Ai-guo, SUN Xu-dong, HU Jun, JIANG Xiao-gang, OUYANG Yu-ping

(School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China)

* Corresponding Author, E-mail: jxliuyd@163.com

Abstract: In view of the poor light transmission of thick skin fruit and the independent modeling of the sugar content of different grapefruit varieties, the grapefruit was the research object, and 325 effective spectra were collected to compare and analyze the spectral responses of two absorption peaks of different grapefruit varieties near 710 nm and 800 nm. In the 550 – 970 nm full-band spectrum, the false positive rate of the grapefruit partial least squares discriminant model is 1.25% after SPA no information elimination; the partial least squares method is in the 550 – 970 nm full-band range and the de-differential 750 – 930 nm band. The prediction correlation coefficients of the range are 0.58 and 0.86, respectively, and the predicted root mean square error RMSEP is 0.84 and 0.55, respectively. Experiments show that the continuous projection method combined with the partial least squares discriminant model can realize the qualitative discrimination of different grapefruit varieties. The coefficient of variation method is the best to quantify the soluble solids of the thick fruit grapefruit with the least squares model. This study provides a reference and theoretical basis for the online sorting technology of different varieties of thick fruit.

收稿日期: 2018-07-25; 修订日期: 2018-10-18

基金项目: 国家自然科学基金(61640417); 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)资助项目

Supported by National Natural Science Foundation of China(61640417); Center of The Technology and Equipment of The Intelligent Management for The Southern Mountain Orchard Collaborative Innovation(2014-60); Jiangxi Advantage Science and Technology Innovation Team Construction Project(20153BCB24002)

Key words: near-infrared spectroscopy; on-line detection; universal model; soluble solids; variety discrimination

1 引言

柚子不仅味道甜美,而且营养价值高,富含各种维生素,主要产于我国广东、福建、江西、广西、湖南等南方地区^[1-2]。产地品种及可溶性固形物是评价口感和价格差异的主要指标^[3],因不同产地的柚子外观相似,如价格2元/斤的井冈蜜柚与40元/个的梅州金柚外观极为相似,缺乏经验的消费者无法区分价格低廉、昂贵的柚子,传统的人工分选也难以判断柚子品种及等级差异,从而影响柚子产业的发展^[4-8]。

目前有不少关于近红外光谱技术对薄皮水果快速分选的研究,如苹果^[9-10]、柑橘^[11]、梨^[12]等都曾作为典型的薄皮水果用来分类研究。柚子作为一种具有代表性的厚皮果,由于透光性差,可见/近红外光谱技术检测内部品质及种类判别还未见相关研究报道,因此攻克厚皮果透光性的技术难题,研究出一种能够快速、无损、批量检测不同产地柚子品种及可溶性固形物的通用模型具有十分重要的意义。

本研究以不同产地的厚皮果柚子(江西井冈蜜柚、江西上饶马家柚和广东梅州金柚)为研究对象,建立不同品种柚子定性判别模型,并在柚子的有效波段内通过变异系数法进行了去差异化处理和无信息变量消除,完成3种柚子的可溶性固形物在线检测通用模型的建立,为厚皮果实际加工生产提供技术支持和参考依据。

2 材料与方法

2.1 试验材料

试验所用柚子分别来自2017年10月在广东省梅州市某果园采摘的梅州金柚145个、2017年11月在江西省吉安市某果园采摘的井冈蜜柚100个和2017年12月在江西省上饶市广丰区某果园采摘的马家柚79个。样品实验前编号处理,在室内温度25℃存放一天后采集光谱,并在相同位置采集柚子的可溶性固形物真值,取平均真值和光谱后共得到325个真值与光谱,分为建模集245个,预测集80个。图1(a)、(b)、(c)分别是江西井冈蜜柚、江西上饶马家柚、广东梅州金柚,井冈蜜柚和梅州金柚外观极为相似,而上饶马家柚个

体大、皮厚,且同一个柚子的果皮厚薄程度不一,最厚的果皮可达6cm。

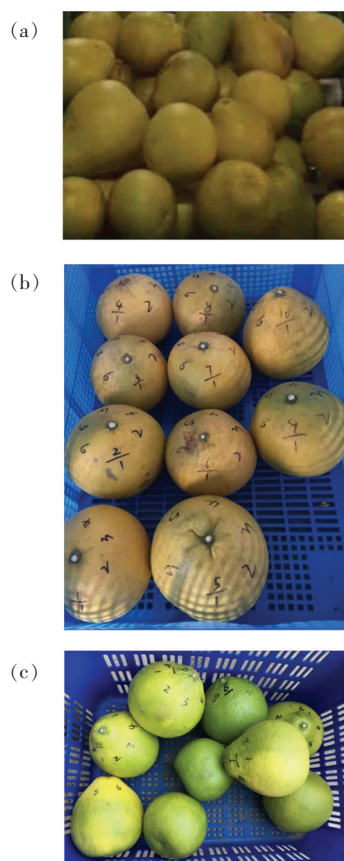


图1 不同产地柚子样品实物图

Fig. 1 Physical map of grapefruit samples from different origins

2.2 光谱采集

如图2所示,在线检测系统包含光谱仪、计算机、数模转换器、PLC、传感器、探头、果杯、柚子样品、传输链、光纤、卤素灯。光谱采集采用漫透射方式,将样品放置在传送带上,卤素灯照射的光透过柚子,传送带下方的光纤探头接收光谱信息。传送带每秒传输5个柚子,积分时间100ms,卤钨灯光照强度为100W,光谱采集前装置预热30min,待光源稳定后用白色聚四氟乙烯球作为参比,采集的参比球能量谱标准差小于1%开始采集光谱。由于柚子具有果柄,故将柚子果柄与传送带运动方向一致、缝合面垂直于水平面放置。传感器安装在传送带下方2mm处,果杯上转动一次触发传感器发出高电位信号,光谱仪采集一条光谱,保存在计算机软件中。

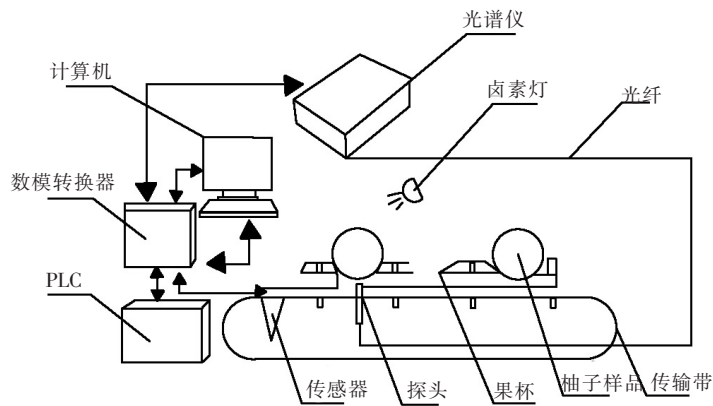


图 2 可见近红外漫透射在线检测装置

Fig. 2 Visible/near-infrared diffuse transmission online detection device

2.3 可溶性固形物含量的测定

柚子样品的可溶性固形物的测量采用折射式数字糖度计 (PR-101a) 测量, 测量时, 将标记处 (去皮后) 的 2~3 mm 的果肉切下, 挤出适量的柚子汁测量可溶性固形物, 每次测量后用纯净水标定糖度 0% 后继续测量下一个点。每个柚子标记 4 个面, 取平均值作为该柚子的可溶性固形物值。

2.4 数据分析

将采集的柚子光谱数据导入 MATLAB 和 Unscrambler 软件完成建模, 模型建立后导入分选软件, 实现柚子的定性定量在线检测。模型的优劣由相关系数、校正均方根误差和预测均方根误差进行评价, 一个合适的模型表现为较高的相关系数和较小的均方根误差。

3 结果与讨论

3.1 柚子光谱特征分析

图 3 为井冈蜜柚、上饶马家柚和梅州金柚 3 种柚子的代表性近红外光谱, 从中可以看出 710 nm 和 800 nm 附近有两个明显的吸收峰。柚子内部含有各种复杂的化合物, 而化合物中又由 C—H、N—H、O—H 等官能团组成。前者吸收峰有可能是由 C—H 键的伸缩震动四级倍频导致的, N—H 键伸缩震动和变形震动一级倍频致使后者出现吸收峰。而 310 nm 处的直线可能是由于光谱采集时机器的杂散光或者噪声导致。

3 种柚子皮的厚度不同, 造成 3 种样品的吸光度在 600~900 nm 之间有所差异。上饶马家柚称为棉花柚, 皮厚, 透光性弱, 故导致吸光度最低, 在 710 nm 和 800 nm 处吸收峰明显低于井冈蜜柚

和梅州金柚; 井冈蜜柚皮薄果肉紧实, 透光性较好, 特别是在 710 nm 左右; 梅州金柚在 800 nm 左右吸光度强于井冈蜜柚可能是由于梅州地区光照时间充足, 光合作用最强, 化合物中的 N—H 数量偏多, 伸缩震动和变形震动强度高于井冈蜜柚, 所以 800 nm 处的吸收峰最强^[13]。

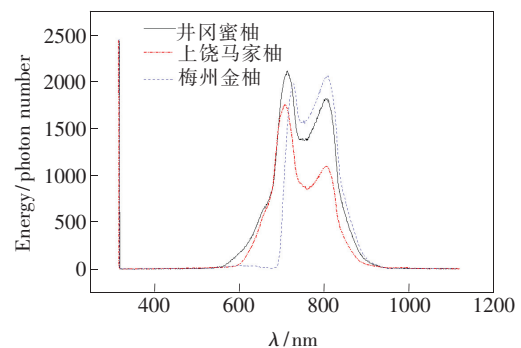


图 3 柚子近红外漫透射原始光谱

Fig. 3 Near-infrared diffuse transmission spectra of grapefruit

3.2 品种判别

3.2.1 连续投影算法

由于噪声和杂散光导致光谱中含有一些无用信息, 在建模之前, 通过光谱预处理方法消除无关信息是非常重要的过程。连续投影法是一种有效降低模型复杂度的新型特征波长筛选算法, 目的是在重叠的光谱信息中提取有效信息、降低光谱变量之间的共性, 以此来简化模型、提高建模速度。连续投影法 (SPA) 在 MATLAB2012 中运行, 选择的变量数在 10~50 之间, 结果如图 4 所示, 得到 716, 827, 709, 737, 784, 769, 695, 663, 856, 812 nm 共 11 个波长为特征波长, 并组成波长组合, 每次运行结果后得到的特征波长组合相同。

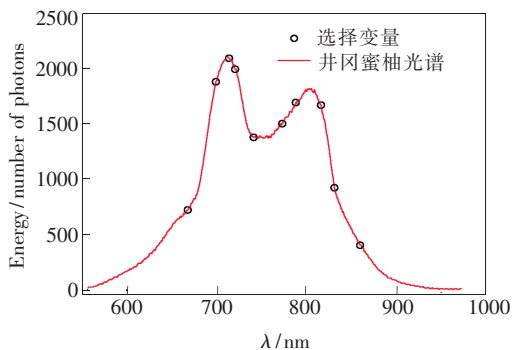


图 4 SPA 运行结果

Fig. 4 SPA operation results

表 1 全波段和 SPA 无信息消除后 PLS-DA 测试结果

Tab. 1 PLS-DA test results after full spectrum and SPA no information elimination

特征变量选择	主成分分子数	建模相关系数	预测相关系数	建模均方根误差	预测均方根误差	误判率/%
SPA	5	0.96	0.96	0.48	0.44	1.25
全波段	7	0.94	0.94	0.58	0.58	2.5

从表 1 中可以发现, SPA 筛选变量后作为模型输入值时, 仅有一个值被误判, 误判率为 1.25%, 而全波段作为输入时误判率为 2.5%。全波段的误判率高于 SPA 筛选变量后的输入值可能是因为全波段数据中含有无用信息或冗余信息变量导致分类效果不佳。

特征波长建立的偏最小二乘判别模型 PLS-DA 均方根误差与主成分数关系如图 5 所示, 从图中可以看出在主成分为 5 时预测集均方根误差 RMSEP 最小, 误判率最低。图 6 是建模集与预测集柚子分类变量与预测值相关图, 无信息变量筛选后定性判别模型建模相关系数 R_c 为 0.96, 预测相关系数 R_p 为 0.96, 建模均方根误差为 0.44,

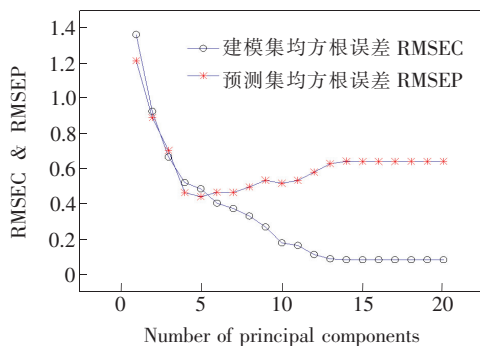


图 5 主成分因子数图

Fig. 5 Number of main component factors

3.2.2 PLS 定性模型建立与验证

PLS-DA 是在 PLS 基础上衍生出的一种定性分析方法, 通常将分类值代替真值, 对预测结果进行分类, 如公式(1)所示:

$$y = \sum_{i=1}^n k_i x_i + b, \quad (1)$$

其中 y 是分类量, 根据柚子样品的种类试验中设为 [1 4 7], 其中 1 为井冈山柚, 4 为马家柚, 7 为梅州金柚, 两两分类值中间的平均值为分类的阈值。柚子近红外全波段范围和 SPA 消除无信息变量后的结果结合分类值分别建立的柚子 PLS-DA 判别模型结果如表 1 所示。

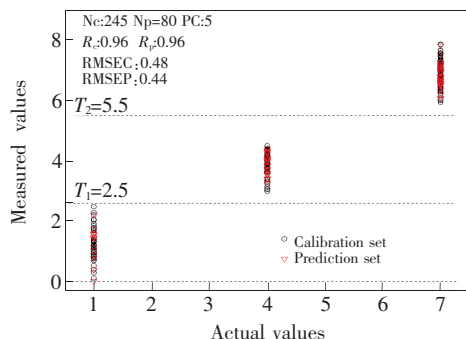


图 6 建模集与预测集柚子分析变量与预测值相关图

Fig. 6 Correlation between model set and prediction set of grapefruit analysis variables and forecast values

预测均方根误差为 0.48。第一类和第二类样品之间的阈值为 2.5, 第二和第三类样品之间的阈值为 5.5, 若预测值大于 0 小于 2.5 时判断为井冈山柚, 若预测值大于 2.5 小于 5.5 判断为上饶马家柚, 若预测值大于 5 判断为梅州金柚, 预测的柚子样品中只有 1 个第一类样品(井冈山柚)被误判。

3.3 可溶性固形物定量模型建立与验证

3.3.1 光谱差异化

差异性分析时采用主成分分析法将柚子的近红外光谱矩阵转换成主成分得分矩阵, 最大主成分数为 20。图 7 是前 2 个主成分的分散点图, 并

冈蜜柚、上饶马家柚、梅州金柚大致能够分为 2 类,前 2 个主成分累计贡献率 99%。不同品种的柚子光谱存在一定的差异性,这 3 类柚子样品存在的聚类现象,需要进一步处理以提高定量模型的精确性。

图 8 是 3 种柚子光谱经过求导、MSC、平滑后

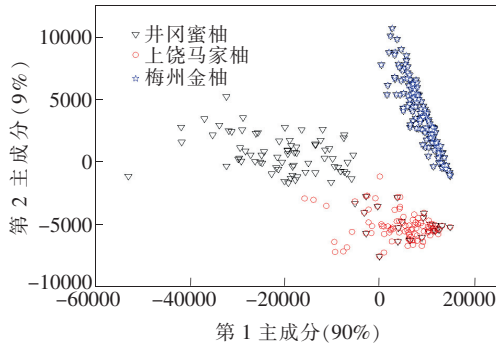


图 7 柚子差异性分析散点图

Fig. 7 Scatter plot of grapefruit ariation plots difference analysis

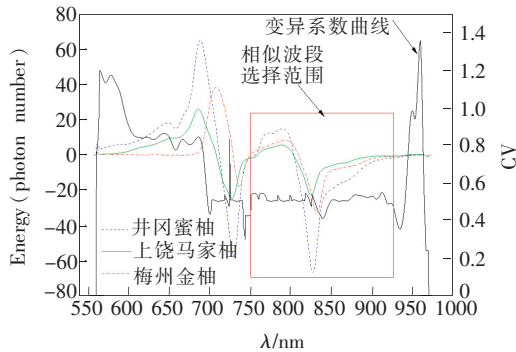


图 8 柚子预处理后的平均光谱与变异系数曲线图

Fig. 8 Mean spectral and coefficient of variation plots of vgrapefruit after pretreatment

的平均光谱与变异系数曲线的对比图。变异系数 (CV) 是不同光谱变化程度的参考量,系数越小且越稳定则表明差异性越小。黑线表示变异系数曲线,红蓝绿分别表示 3 种柚子预处理后的平均光谱。在 550 ~ 750 nm 波长范围内,变异系数值波动较大,则这 3 种样品光谱的离散程度较大,差异化较大时影响光谱的稳定性和准确性。在变异系数小于 2 的范围内,750 ~ 930 nm 波段范围内变异系数整体较小,波动也较为平整,不存在明显的波峰和波谷,不会影响建模时模型的稳定性和准确性,因此图中红色线框内选择的波段范围是合适的。

3.3.2 模型建立与预测

最小二乘支持向量机通过调节惩罚因子和核函数建立光谱数据和化学组分间的定量分析通用模型。所有样品光谱去差异化后建立最小二乘支持向量机定量模型,在 750 ~ 930 nm 波段范围共 180 个波长点建模,讨论最小二乘支持向量机建模 Lin-kernel、RBF-kernel 等函数对模型的影响。

柚子糖度定量检测模型效果如表 2 所示,最小二乘支持向量机采用线性核函数 Lin-kernel 全波段和变异系数法选波段后的相关系数分别为 0.63 和 0.68,径向基核函数 RBF-kerne 全波段和选波段后的相关系数分别为 0.62 和 0.65。选择变异系数变化稳定的波段 750 ~ 930 nm,偏最小二乘通过变异系数法去差异化之后预测相关系数有显著的提高,预测样本数为 80,预测均方根误差为 0.55,相关系数为 0.86。偏最小二乘结合变异系数选波段法适合不同柚子品种糖度的定量检测。

表 2 柚子糖度定量模型效果比较

Tab. 2 Comparison of quantitative effects of grapefruit sugar quantitative model

模型	核函数	输入量	波段范围/nm	预测样本数	RMSEP	相关系数
Ls-svm	Lin-kernel	全波段	550 ~ 970	80	4.06	0.63
		选波段	750 ~ 930	80	1.32	0.68
	RBF-kernel	全波段	550 ~ 970	80	4.01	0.62
		选波段	750 ~ 930	80	1.31	0.65
PLS	*	全波段	550 ~ 970	80	0.84	0.58
	*	选波段	750 ~ 930	80	0.55	0.86

图 9 和图 10 分别为偏最小二乘模型的测量值和真实值散点图和 PLS 定量模型的主成分因

子数和均方根误差关系图。在 750 ~ 930 nm 波段范围共 180 个波长点划分好的建模集和预测集,

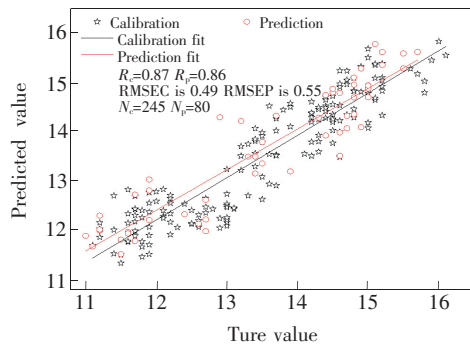


图9 PLS模型的测量值和真实值散点图

Fig. 9 Scatter plots of measured values and true values for the PLS model

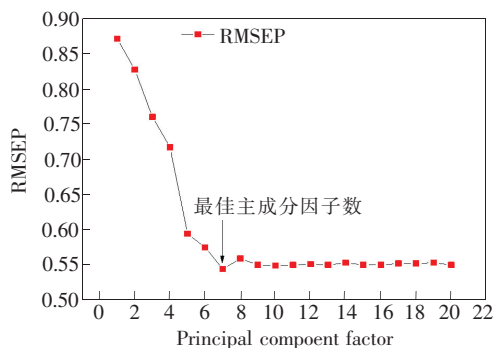


图10 主成分因子数和均方根误差关系图

Fig. 10 Relationship between principal component factor and root mean square error

建立柚子的可溶性固形物偏最小二乘定量模型,相关系数为 0.87。建模集均方根误差为 0.49。建模过程选择的主成分因子数(PC)过低会因为覆盖掉柚子光谱中的有效信息从而导致预测精度过低,选择 PC 为 4 或 5 时出现“欠拟合”现象。当 PC 选择过高时,如 9 或 10 时又将出现过拟合现象,PC 选择 6 或 8 时预测精度尚未达到最优,所以本次定量建模的最佳主成分因子数为 7。

4 结 论

本文采用近红外光谱漫透射技术结合偏最小二乘和最小二乘支持向量机模型,实现井冈蜜柚、上饶马家柚、梅州金柚等不同品种柚子的定性判别和可溶性固形物的在线无损检测。其中连续投影算法筛选无信息变量后偏最小二乘判别模型误判率为 1.25%,能够准确地将不同品种的柚子推入分级入口。经过求导、MSC、平滑处理后获得柚子的可见/近红外光谱,用变异系数法对光谱去差异化,比较了 PLS 建模与 LS-SVM 建模结果,最终采用相似波段建立的 PLS 定量模型为最佳建模方式,预测相关系数达到 0.86,预测均方根误差为 0.55。本研究简化了同种水果不同品种生产加工过程,减少了模型的建立和维护成本,对柚子的生产加工在线分级具有参考意义。

参 考 文 献:

- [1] 刘勇,周群,刘德春,等. 中国柚类生态分布多样性研究[J]. 江西农业大学学报, 2006,28(3):332-335.
LIU Y,ZHOU Q,LIU D C,*et al.*. A study on the ecological distribution of pummelo in China by means of AFLP [J]. *Acta Agric. Univ. Jiangxiensis*, 2006,28(3):332-335. (in Chinese)
- [2] 杨宁. 柚子全果综合利用及生物活性研究进展[J]. 广州化工, 2015,43(5):9-11.
YANG N. Research progress of comprehensive utilization and biological activity of grapefruit fruit [J]. *Guangzhou Chem. Ind.*, 2015,43(5):9-11. (in Chinese)
- [3] KRINTIRAS G A,DIAZ J G,VAN DER GOOT A J,*et al.*. On the use of the Couette Cell technology for large scale production of textured soy-based meat replacers [J]. *J. Food Eng.*, 2016,169:205-213.
- [4] 宋彩英,覃志豪. 吉安市井冈蜜柚产业发展研究[J]. 中国农业资源与区划, 2014,35(5):113-116.
SONG C Y,QIN Z H. Study on the development of Jinggang candied pomelo industry of Ji'an city [J]. *Chin. J. Agric. Resour. Reg. Plan.*, 2014,35(5):113-116. (in Chinese)
- [5] 郭樱花,李艳梅. 梅州金柚品牌建设的经验与启示[J]. 现代农业科技, 2013(6):293-294.
GUOY H,LI Y M. Experience and enlightenment of brand construction of Meizhou golden pomelo [J]. *Mod. Agric. Sci. Technol.*, 2013(6):293-294. (in Chinese)
- [6] 耿一曼. X射线无损检测技术在柚子品质检测中的应用研究[D]. 福州:福建农林大学, 2012.
GENG Y M. *Application Study of X-ray NDT in Grapefruit Quality Testing* [D]. Fuzhou:Fujian Agriculture and Forestry University, 2012. (in Chinese)

- [7] 颜静,唐成,梁亚雄,等. 柚子原产地溯源鉴定技术 [J]. 食品科学, 2014,35(4):248-252.
YAN J,TANG C,LIANG Y X,*et al.*. Tracing the geographical origin of grapefruit [J]. *Food Sci.*, 2014,35(4):248-252. (in Chinese)
- [8] 陈兆杰,黄慧俐,韦婕,等. 色谱-质谱联用技术在我国 3 种柚子组分鉴定及指纹图谱建立中的应用 [J]. 色谱, 2016,34(6):558-566.
CHEN Z J,HUANG H L,WEI J,*et al.*. Application of chromatography-mass spectrometry technology in the component identification and fingerprint establishment of the three kinds of pomelo in China [J]. *Chin. J. Chromatogr.*, 2016, 34(6):558-566. (in Chinese)
- [9] 刘燕德,马奎荣,孙旭东,等. 梨和苹果糖度在线检测通用数学模型研究 [J]. 光谱学与光谱分析, 2017,37(7): 2177-2183.
LIU Y D,MA K R,SUN X D,*et al.*. The fruits soluble solids content detection online using universal mathematical model [J]. *Spectrosc. Spect. Anal.*, 2017,37(7):2177-2183. (in Chinese)
- [10] 陆辉山,陈鹏强,闫宏伟,等. 基于近红外光谱漫透射技术的苹果可溶性固形物含量在线检测 [J]. 食品与机械, 2015,31(3):40-42.
LU H S,CHEN P Q,YAN H W,*et al.*. On-line detection research of apple's soluble solids content of diffuse transmission technology based on near infrared spectroscopy [J]. *Food Mach.*, 2015,31(3):40-42. (in Chinese)
- [11] 刘燕德,陈兴苗,欧阳爱国. 可见/近红外光谱无损检测赣南脐橙可溶性固形物 [J]. 光学学报, 2008,28(3): 478-481.
LIU Y D,CHEN X M,OUYANG A G. Non-destructive measurement of soluble solid content in Gannan navel oranges by visible/near-infrared spectroscopy [J]. *Acta Opt. Sinica*, 2008,28(3):478-481. (in Chinese)
- [12] 刘燕德,朱丹宁,吴明明,等. 玉露香梨可溶性固形物近红外漫透射光谱在线检测 [J]. 食品与机械, 2016, 32(10):115-119.
LIU Y D,ZHU D N,WU M M,*et al.*. Online detection of soluble solids contents for "Yuluxiang" pear by visible-near infrared diffuse transmission spectroscopy [J]. *Food Mach.*, 2016,32(10):115-119. (in Chinese)
- [13] 王加华,韩东海. 便携式近红外水果糖度分析模型简化研究 [J]. 食品安全质量检测技术, 2009,1(1):32-38.
WANG J H,HAN D H. Simplification of NIR model for determination of SSC in fruits by portable Vis/NIR instrument [J]. *Food Saf. Qual. Detect. Technol.*, 2009,1(1):32-38. (in Chinese)



李雄(1992-),男,江西抚州人,硕士,助教,2018年于华东交通大学获得硕士学位,主要从事光电检测技术的研究。

E-mail: 1598657020@qq.com



刘燕德(1967-),女,江西泰和人,博士,教授,2006年于浙江大学获得博士学位,主要从事智能农机装备、光电检测技术的研究。

E-mail: jxliuyd@163.com