

# 基于高光谱图像技术的菜用大豆厚度检测

黄 敏<sup>1,2</sup>, 万相梅<sup>2</sup>, 朱启兵<sup>2</sup>, 张 懋<sup>\*1</sup>

(1. 食品科学与技术国家重点实验室 江南大学, 江苏 无锡 214122; 2. 江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

**摘要:** 菜用大豆厚度是划分菜用大豆等级的重要衡量指标之一。采用高光谱图像技术对菜用大豆的厚度进行预测。实验中选取 200 个菜用大豆作为测试样本, 获取其高光谱反射图像, 同时用数字式游标卡尺测量厚度值。选取 400~1 000 nm 范围的光谱信息, 采用多元散射校正、标准归一化和导数计算对光谱数据预处理, 结合偏最小二乘和多元线性回归两种分析方法建立厚度校正模型和预测模型。研究发现基于多元散射校正的偏最小二乘方法的模型精度较优, 校正模型和预测模型的相关系数分别为 0.956 和 0.933, 均方根误差分别为 0.59 mm 和 0.70 mm。研究结果表明可以利用高光谱图像技术预测菜用大豆厚度。

**关键词:** 菜用大豆; 高光谱图像; 厚度检测; 建模

**中图分类号:** O 657.3 **文献标志码:** A **文章编号:** 1673—1689(2012)011—1142—06

## Thickness Measurement of Green Soybean Using Hyperspectral Imaging Technology

HUANG Min<sup>1,2</sup>, WAN Xiang-mei<sup>2</sup>, ZHU Qi-bing<sup>2</sup>, ZHANG Min<sup>\*1</sup>

(1. Key Laboratory of Food Science and Safety, Ministry of Education, Jiangnan University, Wuxi 214122, China; 2. Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi 214122, China)

**Abstract:** Thickness is an important index for the grade of green soybean. In this manuscript, hyperspectral imaging technology was used to determine the thickness of two hundreds green soybean. The hyperspectral reflectance images of the samples were acquired using hyperspectral imaging system and the instrumental thickness values were measured by digital caliper. The range of 400~1000 nm spectral information was preprocessed using multiplicative scatter correction algorithm (MSC), standard normal variate algorithm (SNV) and derivation calculation algorithm, respectively. Then, partial least squares (PLS) and multiple linear regressions (MLR) were used to develop calibration and prediction model. Results show that optimal models are obtained by MSC

收稿日期: 2012-04-01

基金项目: 国家自然科学基金项目(61271384, 61275155); 江苏省自然科学基金项目(BK2011148); 中国博士后基金项目(2011M500851)。

作者简介: 黄 敏(1974—), 女, 辽宁铁岭人, 工学博士, 副教授, 硕士研究生导师, 主要从事近红外成像和高光谱图像技术分析研究。

E-mail: huangmzqb@163.com

\* 通信作者: 张 懋(1962—), 男, 浙江平湖人, 工学博士, 教授, 博士研究生导师, 主要从事农副产品加工与贮藏研究。

E-mail: min@jiangnan.edu.cn

coupled with PLS method, which the correlation coefficients of calibration set and prediction set were 0.956 and 0.933 and the root mean square error is 0.59 mm and 0.70 mm, respectively. The above results demonstrated that hyperspectral imaging technology is suitable for detection of green soybean thickness.

**Keywords:** vegetable soybean, hyperspectral images, thickness measurement, modeling

菜用大豆,又叫毛豆,是以幼嫩荚果和子粒作为蔬菜食用的一种专用型大豆品种。由于其兼具营养与保健双重功能<sup>[1]</sup>,富含蛋白质、脂肪、矿物质和维生素等成分<sup>[2]</sup>,深受国内外消费者的青睐。目前,中国是最大的菜用大豆生产国与出口国<sup>[3]</sup>,具有良好的出口创汇前景。为延长鲜食菜用大豆的储藏期和货架期,通常采用密封冷藏,冷藏保存之前需对菜用大豆品质按等级进行分检处理。菜用大豆品质主要指外观品质、食用品质和营养品质,其中外观品质是最重要、最受到注重的商品品质之一。亚洲蔬菜研发中心明确规定了菜用大豆外观品质的标准并形成了外观分级体系,其中虫食程度一个极其重要的分级指标(关于菜用大豆虫食程度的研究在其他文章中另外探讨),而厚度又是一个极其重要的外观选择标准<sup>[4]</sup>。菜用大豆厚度与其饱满度有关,传统的菜用大豆厚度的分拣方法是用肉眼进行人工挑选,这种方法存在费时、费力、效率低等缺点,故根据其厚度对菜用大豆的等级进行更深一步的划分具有重要的现实意义。

高光谱图像技术作为一种新型的无损检测技术,结合了计算机视觉和光谱检测两种技术的优点<sup>[5]</sup>,能够很好的记录农产品外在特征和内部品质的丰富信息。已有研究表明高光谱图像技术已广泛的应用于农畜产品以及水果的无损检测中,如检测玉米含水率和含油率<sup>[6]</sup>,检测酸樱桃内部的果核<sup>[7]</sup>,评估腌制黄瓜的质量品质<sup>[8]</sup>,苹果粉质化检测<sup>[9]</sup>等,且取得了较好研究成果。光照射到水果和蔬菜等农产品是一个及其复杂的现象。当入射光作用于物体表面时,除有少数的入射光从物体表面被反射掉,发生镜面反射等现象外,剩余的光能量进入其细胞组织并在其内部发生光的吸收和漫反射现象<sup>[10]</sup>。光的吸收主要和物体的化学成分(如糖分,酸性)有关,而漫反射主要受到结构和物理特性(如密度,粒子尺寸和细胞结构)的影响。由于生长期不同,饱满(较厚)与不饱满(较薄)的菜用大豆的内部细胞组

织结构和物理特征也会有所不同,故光在其内部的衰减也会有所差别。

作者设想在不增加设备成本的前提下,对检测完有虫菜用大豆后对无虫菜用大豆样本厚度进行检测,从而实现更深一步地分拣,探求菜用大豆厚度与高光谱图像数据之间的定量关系,为实现高光谱图像对菜用大豆厚度在线检测提供理论基础。

## 1 材料与方法

### 1.1 试验材料

菜用大豆:2011年4月采购于浙江省慈溪市海通食品有限公司的毛豆种植基地。选取200个菜用大豆作为实验样本,实验前把菜用大豆的表皮杂物处理干净,依次进行编号、标记。

### 1.2 高光谱图像采集

实验中使用高光谱图像采集系统如图1所示。整个系统由图像采集单元、光源、样本输送平台3部分组成。其中,图像采集单元包括图像光谱仪(1003A-10140 HyperspecTM VNIR C-Series, Headwall Photonics Inc., USA)、CCD摄像头(pixelfly qe IC\*285AL, Cooke, USA)、镜头(10004A-21226 Lens, F/1.4 FL23mm, Standard Barrel, C-Mount, USA);光源为150W光纤卤素灯(150W EKE, 3250K(#20094145), Techniquip, USA)。整个系统装在一个暗箱中,避免在图像采集过程中受到外界光的影响。高光谱图像光谱仪的光谱范围为400~1000 nm,光谱分辨率为1.29 nm/pixel,空间分辨率是0.15 mm/pixel,波段间隔为0.64 nm,采集图像时对10个波段进行binning,即实际图像的波段间隔为6.4 nm。图像采集时的曝光时间为200 ms。实验中,先采集系统的全黑标定图像,再采集菜用大豆样本高光谱图像。每次将2个菜用大豆样本整齐的摆放在(20 cm×20 cm)黑色背景板上进行图像采集,最后获取一幅大小为1392×350×94的菜用大豆样本高光谱图像。同时,为了对采集的样本图像进

行校正,每采集 5 次样本图像,需采集一次校正白板(全白标定)的图像。

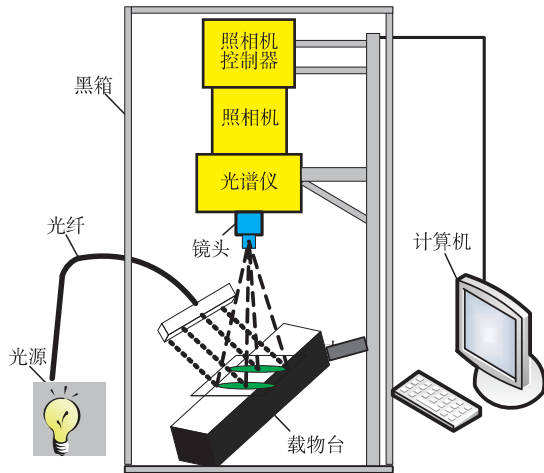


图 1 高光谱图像系统

Fig.1 Hyperspectral imaging system

### 1.3 样本标准仪器厚度值测量

利用数位式游标卡尺 (Digital Caliper, CD-6~CSX, 500-196-20, Mitutoyo Corporation, Japan, 精度为 0.01 mm) 测量菜用大豆样本的厚度,测量时选取菜用大豆中每个豆粒的赤道位置附近作为游标卡尺的测量位置,采用 5 次测量的平均值作为菜用大豆厚度的标准仪器测量值。

## 2 数据处理与分析

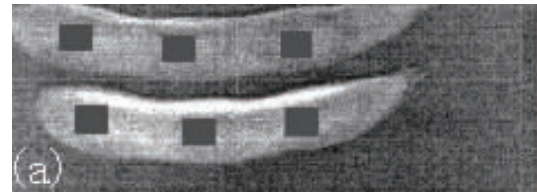
### 2.1 图像采集与数据处理平台

整个实验过程中,高光谱图像的采集是基于 Hyperspec™-M 图像采集软件平台;数据的处理基于 ENVI V. 4.3 (Research System, Inc., USA)、Excel 和 MATLAB (Math works, matlabR2009b, Inc., USA) 等软件平台。

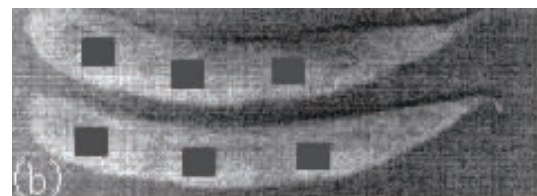
### 2.2 样本原始图像处理

首先采用 ENVI 软件处理采集到的高光谱图像,获得菜用大豆样本图像;然后用菜用大豆样本图像减去全黑标定图像,再除以全白标定图像减去全黑标定图像之差,从而获得最终相对高光谱图像<sup>[1]</sup>;再采用 3×3 中值滤波器对图像进行处理。因采集到的原始的高光谱图像过大,为便于后续处理,需从处理后的图像中提取感兴趣区域。图 2 为 506 nm 处的不同厚度值的菜用大豆图像,为了消除误差,在每个菜用大豆相对图像中每个豆粒所在位置中

心处取一个 60 pixel × 50 pixel 同样大小的感兴趣区域,将每个区域所有的光谱数值在各个波段上取平均值,即把平均反射光谱数据作为后期数据处理。



(a) 8.95 mm(up)和9.58 mm(down)



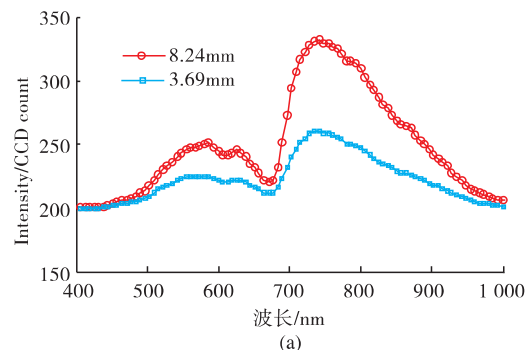
(b) 6.35 mm(up)和6.55 mm(down)

图 2 在 506 nm 处的不同厚度的菜用大豆图像

Fig.2 Green soybean images of different thickness at 506 nm

### 2.3 光谱曲线

图 3(a)所示为不同厚度值的菜用大豆的原始高光谱反射图像,从图中可以看出厚度值大的菜用大豆的曲线在厚度值小的菜用大豆的上方,即厚度值大的菜用大豆的反射光强大。光谱在 400~700 nm 有所波动,并且由于叶绿素的吸收大约在 675 nm 处有一波谷;在 700~1000 nm 波动较大,菜用大豆反射率的这一光谱特性主要由菜用大豆内部复杂结构决定的。图 3(b)为白板的光谱曲线,可以发现其曲线呈现正态分布的趋势。图 3(c)所示为菜用大豆样本的相对反射光谱曲线,可以发现经过白板校正后的不同厚度的菜用大豆光谱曲线也会有所不同,尤其在近红外范围内差别较大。



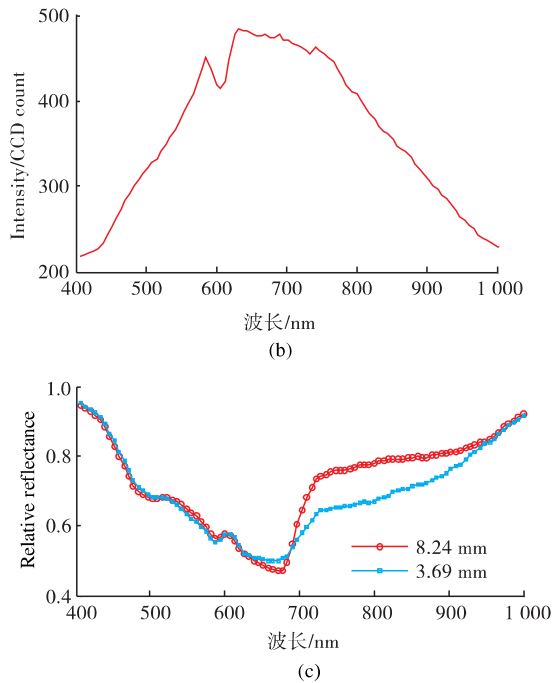


图3 不同厚度的菜用大豆高光谱原始曲线 (a), 白板曲线 (b)以及相对反射光谱曲线(c)

Fig.3 Curves of two green soybeans reflectance

## 2.4 样本光谱预处理与模型评价

由于实验操作等问题的存在会产生奇异样本,而奇异样本可能会对模型精度产生影响,需对200个样本进行去奇异样本处理,共去除2个奇异样本。将去奇异样本后的198个样本光谱曲线进行多元散射校正技术<sup>[12]</sup>(Multiplicative Scatter Correction, MSC)、标准归一化(Standard Normal Variate, SNV),并在此基础上分别获得一阶微分光谱。采用偏小二乘法(Partial Least Squares, PLS)、多元线性回归法(Multiple Linear Regression, MLR)分别建立菜用大豆厚度与菜用大豆反射光谱的定量分析模型。模型的校正指标<sup>[13]</sup>使用校正相关系数(Correlation Coefficient for Calibration,  $R_C$ ),校正均方差(Root

Mean Square Error for Calibration, RMSEC),交叉验证相关系数(Correlation Coefficient for Cross Validation,  $R_{CV}$ )和交叉验证均方差(Root Mean Square Error for Cross Validation, RMSECV)表示,预测指标采用预测相关系数(Correlation Coefficient for Prediction,  $R_P$ )和预测均方差(Root Mean Square Error for Prediction, RMSEP)表示。其中,预测相关系数越高,预测均方差越小,模型的性能越优。

## 3 实验结果与分析

### 3.1 样本集划分

对预处理后的198个样本进行样本集划分。Kennard-Stone算法<sup>[14]</sup>是化学计量学中常用的样本集划分算法,其基本原理是用变量空间之间的相对欧氏距离找出样本集中的有代表性的各个边缘样本直至找到设定的样本数,能使校正集样品分布更均匀,样品集划分更合理。利用该算法选取149个样本作为校正集,剩余49个样本作为预测集。样本厚度的最大值、最小值、平均值和标准差见表1。

表1 菜用大豆厚度测量值统计

Tab.1 Statistics of thickness measurements for green soybeans

| 项目  | 最大值/mm | 最小值/mm | 均值/mm | 标准差/mm |
|-----|--------|--------|-------|--------|
| 全部  | 10.77  | 1.72   | 6.89  | 2.03   |
| 校正集 | 10.77  | 1.72   | 7.07  | 2.01   |
| 预测集 | 9.85   | 2.53   | 6.38  | 1.91   |

### 3.2 PLS建模

PLS是对光谱矩阵和测量矩阵进行主成分分解,并以主因子进行回归,其主成分数影响到PLS模型的性能,故选取一个合适的主成分数是PLS建模过程中的一个关键。采用留一法交叉验证得到的最小均方根误差所对应的主成分数为最佳主成分数。表2是采用PLS进行建模的结果。

表2 PLS建模结果

Tab.2 Model results using PLS

| 方法       | 主成分数 | RC    | RMSEC/mm | RCV   | RMSECV/mm | RP    | RMSEP/mm |
|----------|------|-------|----------|-------|-----------|-------|----------|
| 未处理      | 9    | 0.958 | 0.60     | 0.931 | 0.74      | 0.929 | 0.73     |
| MSC      | 8    | 0.956 | 0.59     | 0.932 | 0.74      | 0.933 | 0.70     |
| MSC+一阶微分 | 5    | 0.957 | 0.59     | 0.930 | 0.75      | 0.930 | 0.72     |
| SNV      | 7    | 0.952 | 0.63     | 0.934 | 0.73      | 0.934 | 0.73     |
| SNV+一阶微分 | 5    | 0.952 | 0.62     | 0.924 | 0.78      | 0.929 | 0.71     |



### 3.3 MLR 建模

MLR 是由两个或者多个自变量的最优组合共同来预测或估计因变量。本文通过回归计算出单波段的显著性水平,然后按照显著性水平从大到小的顺序排列,然后通过相同的方法计算出双波段的显

著性水平,最后选取前 40 个波段进行回归。同样,采用留一法交叉验证得到的最小均方根误差所对应的特征波长数为最优特征波长数。表 3 为采用 MLR 的建模结果。

表 3 MLR 建模结果

Tab.3 Model results using MLR

| 方法 | 特征波长数 RC | RMSEC/mm | RCV   | RMSECV/mm | RP    | RMSEP/mm | 未处理      |
|----|----------|----------|-------|-----------|-------|----------|----------|
| 32 | 0.969    | 0.50     | 0.950 | 0.64      | 0.925 | 0.75     | MSC      |
| 31 | 0.962    | 0.54     | 0.953 | 0.62      | 0.930 | 0.71     | MSC+一阶微分 |
| 24 | 0.962    | 0.56     | 0.944 | 0.67      | 0.924 | 0.75     | SNV      |
| 30 | 0.972    | 0.48     | 0.955 | 0.61      | 0.924 | 0.75     | SNV+一阶微分 |
| 28 | 0.969    | 0.50     | 0.953 | 0.62      | 0.928 | 0.75     |          |

### 3.4 讨论

从表 2 可以看出:对于 PLS 模型而言,(1)MSC 和 SNV 这两种背景校正法对菜用大豆厚度预测结果有所改善, $R_p$  由 0.929 分别提高到 0.933 和 0.934, RMSEP 也由 0.73 mm 变为 0.70 mm 和 0.73 mm,并且主成分由 9 分别降低到 8 和 7,模型得到简化;(2)MSC 和 SNV 分别与背景扣除(一阶导数)法结合后对菜用大豆厚度预测模型的影响不尽相同,但是对模型效果的改善影响不大。

从表 3 可以看出:对于 MLR 模型,(1)MSC 法对模型厚度预测结果的改善有影响,RP 由 0.925 提高到 0.930, RMSEP 也由 0.75 mm 降低到 0.71 mm,而且特征波长数由 32 降低到 31;(2)SNV 法以及 MSC 和 SNV 分别与一阶导数法结合后对菜用大豆厚度预测模型效果均没有得到明显改善。

结合表 2 和表 3 的结果可以发现:(1)经 MSC 处理后的菜用大豆厚度的 PLS 模型的效果较优,原因是因为 MLR 使用光谱矩阵建立模型,并未考虑矩阵中的信息是否与真实模型相关,而 PLS 同时对光谱矩阵和测量矩阵进行主成分分解,并以主因子进行回归,进一步提高了模型的可靠性。(2)MSC 和 SNV 与一阶导数结合后对模型精度没有得到明显的改善。因为一阶导数法用来进行峰的特征点检测,菜用大豆的光谱曲线没有明显的波动,同时噪声对一阶导数的影响较大,由于对菜用大豆的原始图像进行了中值滤波以及对感兴趣区域的光强进行了平均,有效的去除了噪声的影响,故一阶导数对光谱的影响不大, MSC 与 SNV 结合一阶导数的

效果也就会不明显。(3)实验样本随机购买于菜市场,菜用大豆表面存在一些影响因素,如菜用大豆荚表面灰尘厚度影响、菜用大豆新鲜度保持的影响等等。MSC 可以消除由于样品表面颗粒分布不均匀及颗粒大小不同产生的散射对其光谱的影响,故模型精度较优。图 4 和图 5 为没有经过任何预处理以及经过 MSC 处理后的菜用大豆厚度的 PLS 模型的预测值与实测值相关图。

如果今后增加用于建模和预测的样品数,增加感兴趣区域的大小以及尝试不同的光检测方法(透射等)以及提取检测菜用大豆厚度的特征波段等等,将能得到更加稳固、适应性更强的定量分析模型,提高菜用大豆厚度的预测精度,这样更有利于实现在线检测。

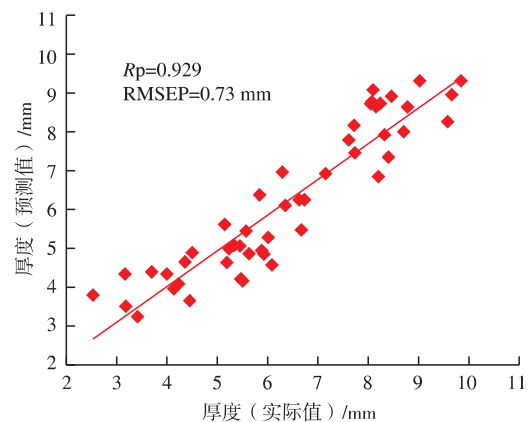


图 4 没有经过处理的菜用大豆厚度的 PLS 模型预测结果  
Fig.4 Thickness prediction results of PLS based on no any treatment for green soybeans

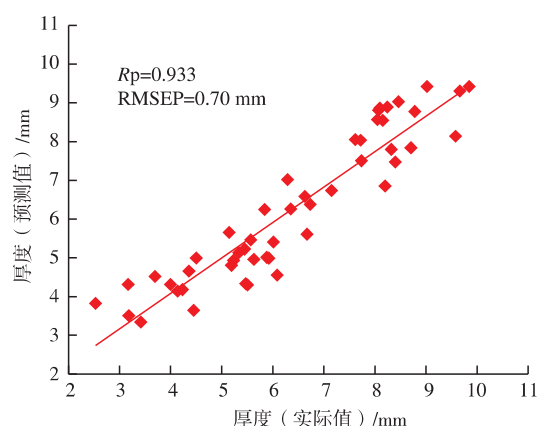


图5 MSC处理后的菜用大豆厚度的PLS模型预测结果

Fig.5 Thickness prediction results of PLS based on MSC for green soybeans

## 4 结语

以菜用大豆为研究对象,采用高光谱图像检测技术对菜用大豆厚度进行预测。通过对200个菜用大豆样品分别进行图像处理,去奇异点处理以及采用MSC和SNV等数据预处理方法,再结合PLS和MLR两种分析方法在波段400~1000 nm建立菜用大豆厚度的定量分析模型,发现基于MSC处理后的PLS模型的预测结果优于光谱没有经过任何处理后的结果, $R_p$ 由0.929提高到0.933,RMSEP也由0.73 mm降低到0.70 mm,并且主成分由9分别降低到8,模型得到简化。研究结果表明,基于高光谱图像技术无损检测菜用大豆厚度是可行的。

## 参考文献:

- [1] 张圣平. 菜用大豆荚果可溶性糖变化规律的研究[D]. 泰安: 山东农业大学, 2002.
- [2] 杨倩, 张慧, 李瑞杰. 加工条件对豆干质构的影响[J]. 食品与生物技术学报, 2011, 30(5): 683–686.  
YANG Qian, ZHANG Min, LI Rui-jie. Effect of processing conditions on texture of the dried beancard [J]. Journal of Food Science and Biotechnology[J]. 2011, 30(5): 683–686. (in Chinese)
- [3] 范秀凤. 不同生态条件下菜用大豆农艺性状与产量、品质的关系[D]. 上海: 上海交通大学, 2005.
- [4] 李之国. 大豆菜用性状评价、品种筛选鉴定及 AFLP 分析[D]. 保定: 河北农业大学, 2006.
- [5] 陆婉珍. 现代近红外光谱分析技术(第二版)[M]. 北京: 中国石化出版社, 2007.
- [6] Cogdill R P, Hurburgh C R, Rippke G R. Single-kernel maize analysis by near-infrared imaging[J]. Transactions of ASAE, 2004, 47: 311–320.
- [7] Lu R F, Qin J W. Detection of pits in tart cherries by hyperspectral transmission imaging[J]. Transactions of the ASAE, 2005, 48(5): 1963–1970.
- [8] Lu R F, Ariana D P. Quality evaluation of pickling cucumbers using hyperspectral reflectance and transmittance imaging[J]. Sensing and Instrumentation for Food Quality and Safety, 2008, 2(3): 144–151.
- [9] HUANG Min, LU R F. Apple mealiness detection using hyperspectral scattering technique[J]. 2010, 58(3): 168–175.
- [10] Qin, J W. Measurement of the optical properties of horticultural and food products by hyperspectral imaging[D]. USA: Michigan State University, 2007.
- [11] Maleki M R, Mouazen A M, Ramon H, et al. Multiplicative Scatter Correction during On-line Measurement with Near Infrared Spectroscopy[J]. Biosystems Engineering, 2007, 96(3): 427–433.
- [12] 郭恩有, 刘木华, 赵杰文, 等. 脐橙糖度的高光谱图像无损检测技术[J]. 农业机械学报, 2008, 39(5): 91–93.  
GUO En-you, LIU Mu-hua, ZHAO Jie-wen, et al. Nondestructive detection of sugar content on navel orange with hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(5): 91–93. (in Chinese)
- [13] 王爽, 黄敏, 朱启兵. 基于无信息变量和偏最小二乘投影分析的高光谱散射图像最优波段选[J]. 光子学报, 2011, 40(3): 428–432.  
WANG Shuang, HUANG Min, ZHU Qi-bing. Optimal wavelength selection of hyperspectral scattering images based on UVE-PLS projection analysis[J]. Acta Photonica Sinic, 2011, 40(3): 428–432. (in Chinese)
- [14] Macho S, Iusa R, Callao M P, et al. Monitoring ethylene content in heterophasic copolymers by near-infrared spectroscopy standardization of the calibration model[J]. Analytica Chimica Acta, 2001, 445(2): 213–220.