

doi: 10.3788/gzxb20124107.0868

基于高光谱图像的玉米种子特征提取与识别

黄敏, 朱晓, 朱启兵, 冯朝丽

(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:玉米种子的形态特征是玉米品种识别的重要因素之一. 采用高光谱成像系统获取 9 个品种共 432 粒玉米种子的高光谱反射图像, 对图像进行校正和预处理, 提取每个样本在 563.6~911.4 nm 共 55 个波段范围内的形状特征. 分别利用单波段、多波段和全波段下的玉米种子形状特征结合偏最小二乘判别法进行模型分类. 结果显示, 全波段范围内训练集和测试集的平均正确识别率达到 98.31% 和 93.98%, 均优于多波段和单波段的正确识别率. 研究表明, 该方法能充分利用高光谱图像中可见光和近红外区域的有效特征信息, 较准确地鉴别玉米品种, 为玉米品种的自动识别领域提供了一种新方法.

关键词:高光谱图像; 玉米种子; 形态特征; 品种识别; 偏最小二乘判别分析

中图分类号: TP391.4; S513

文献标识码: A

文章编号: 1004-4213(2012)07-0868-6

0 引言

玉米种子的纯度直接影响到种子检验、农业生产以及作物育种等各方面, 研究玉米种子品种的无损识别已成为国内外的研究热点之一^[1-2]. 随着计算机图像处理和光谱技术的发展, 基于机器视觉检测和近红外光谱分析技术的玉米种子识别方法在玉米品种的鉴定上得到了较为广泛的应用, 也取得了较好的效果^[3-6]. 但是, 传统的机器视觉技术获得的是种子可见光的形态学特征信息, 近红外光谱分析技术获得的是种子的光谱特征信息, 两种方法获取的种子特征信息较少, 随着种子品种数目增多, 种子特征的重叠现象就越严重, 导致特征的可区分性变差, 从而影响识别效果^[7].

高光谱图像技术是集成光谱技术与图像技术的无损检测新方法, 相比于机器视觉技术和近红外光谱分析技术, 高光谱图像技术提供了包含被测对象形态学特征、内部结构特征和化学成分特征的所有信息. 近年来, 高光谱图像在农产品内外部品质预测、损伤识别等领域得到了较为成功的运用^[8-11].

本文采用高光谱图像对 9 个品种的玉米种子进行特征提取和识别, 比较了不同波段下形态特征差异性, 并结合偏微分最小二乘判别法分别得到在单波段、多波段和全波段条件下的分类准确度, 为高光谱图像在玉米种子自动识别的应用提供理论依据.

1 实验材料与方法

1.1 实验材料

实验以玉米种子作为研究对象, 测试样本从不同的种子公司购得, 选取超甜 2008、高品乐、黄妃、黄金糯玉米、黄糯一号、绿色超人、香甜一号、早鲜金甜香糯、金糯玉米等 9 个不同品种的玉米, 主色调均为黄色, 每个品种选取样本 48 粒, 样本总数为 432.

1.2 高光谱成像系统

实验中使用的白光光谱图像采集系统如图 1. 整个系统由图像采集单元、光源、样本输送平台三部分组成. 其中, 图像采集单元包括图像光谱仪(1003A-

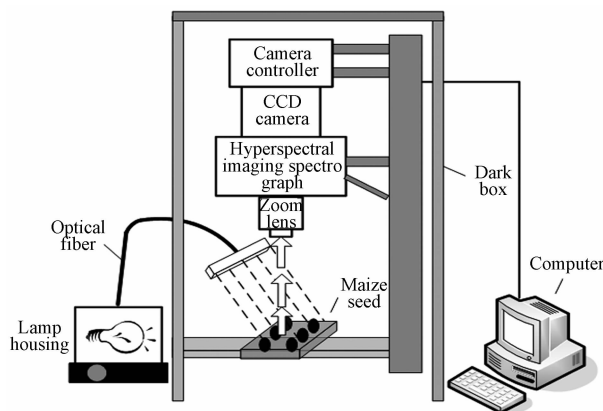


图 1 高光谱成像系统
Fig. 1 Hyperspectral imaging system

基金项目:国家自然科学基金(No. 60805014)、江苏省自然科学基金(No. BK2011148)、中国博士后基金(No. 2011M500851)和中央高校基本科研业务费专项基金(No. JUSRP21132)资助

第一作者:黄敏(1974-), 女, 副教授, 博士, 主要研究方向为农产品无损检测. Email: huangmzqb@163.com

收稿日期: 2011-12-31; **修回日期:** 2012-04-17

10140 Hyperspc™ 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 FL23 mm, Standard Barrel, C-Mount, USA);光源为 150 W 可调功率光纤卤素灯 (150 W 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日光的影响. 高光谱图像光谱仪的狭缝宽度为 25 μm,光谱范围为 400~1 000 nm,光谱分辨率为 1.29 nm,光谱采样的波段间隔为 0.64 nm,采集图像时 binning10 即实际图像的波段间隔为 6.4 nm,空间分辨率 0.15 mm.

1.3 高光谱图像采集

高光谱图像数据采集之前,根据光源的强度设定光谱摄像头的曝光时间来保证得到的图像清晰,并调整好输送装置的速度以免图像尺寸和空间分辨率失真. 经过多次调试,最终确定相关参量为:物距 25 cm,曝光时间 350 ms,光源功率设定为 60 W. 实验中,每次选取同一个品种中的 16 粒种子,将其反面朝上(无胚面),按照固定次序放置在黑色背景板 (20×20 cm²)上,然后一起放到载物台上.

图像采集时,线阵探测器在光学焦面的垂直方向作横向扫描,随着载物台的移动,线阵探测器扫描出整个平面的图像从而完成样本图像的采集. 所有的图像由 Hyperspectral Scanning and Image Rendering Software, Rev A. 2. 1. 3 (Headwall Photonics, Inc., USA)图像采集软件来完成.

2 数据分析

2.1 高光谱图像校正

不同波段下,光源的强度分布不均匀以及摄像头中存在暗电流的影响,会导致光源分布较弱的波段所含的噪音过大,因此,必须要对高光谱图像进行校正. 实验中,为了降低光源因为温度变化造成对图像的干扰,本实验每采集 15 幅样本图像,进行一次全白标定图像 R_i 和全黑标定图像 R_d 的采集. 最后

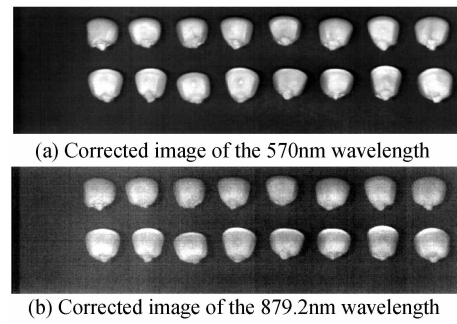


图 2 朝鲜金甜香糯在可见光和近红外波段的两幅校正后图像 Fig. 2 Two corrected images in visible and NIR wavelengths

根据式(1)得到校正后的高光谱图像. 图 2 为朝鲜金甜香糯在波段为 570 nm 和 879.2 nm 的校正后图像.

$$R = (I - R_d) / (R_i - R_d) \tag{1}$$

式中, I 为原始的高光谱图像; R 为校正后的高光谱图像.

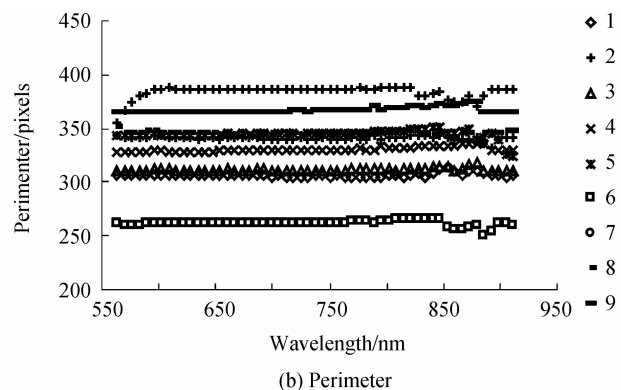
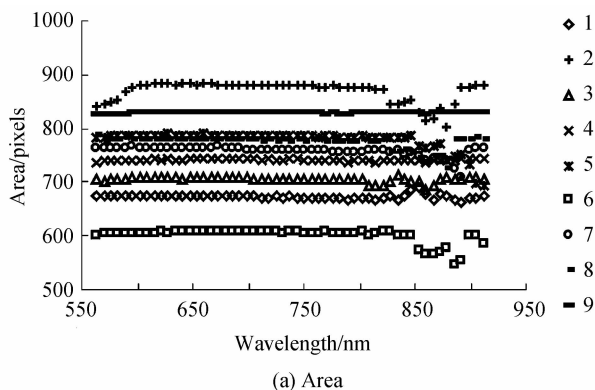
2.2 高光谱图像预处理和特征提取

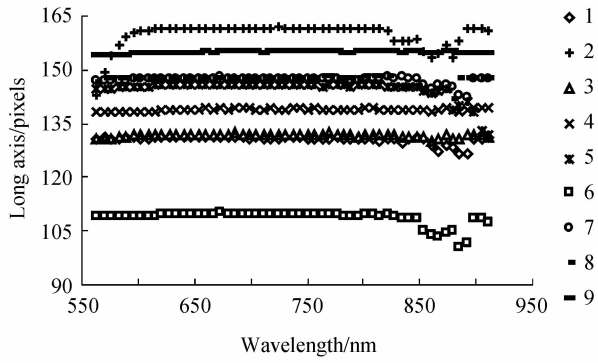
对高光谱图像进行特征提取时,每个波段下种子图像显示的灰度值都是不同的,因此,需要分别对各波段的图像进行特征提取. 在图像特征提取之前,对样本图像进行必要的图像预处理,本实验进行了图像增强、阈值分割、开闭合运算、边缘检测、空洞填充等预处理^[12].

本文提取了包括面积、周长、长轴、短轴、离心率、当量直径、长宽比、圆形度、矩形度、内切圆半径、离散度、紧凑度共 12 个形状特征信息^[13-14]. 由于受光源功率、binning 及量子效率等综合因素的影响,采集后的光谱图像在小于 563.6 nm 和大于 911.4 nm 的光谱数据噪音较大,实验仅采用 563.6~911.4 nm 共 55 个波段(波段编号为 26~80)的形状特征来进行数据处理和分析.

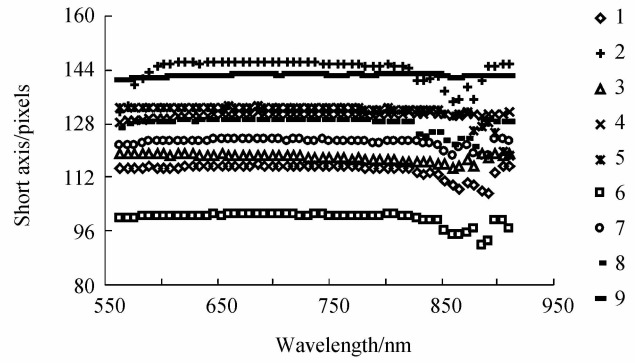
2.3 高光谱图像特征参量分析

图 3 给出了 9 个品种玉米在 55 个波段下形状特征的统计特性. 图中横轴为波段编号,纵轴为形状特征的类内统计平均值,1~9 分别代表超甜 2008、高品乐、黄妃、黄金糯玉米、黄糯一号、绿色超人、香甜一号、朝鲜金甜香糯、金糯玉米等 9 个品种的玉米.

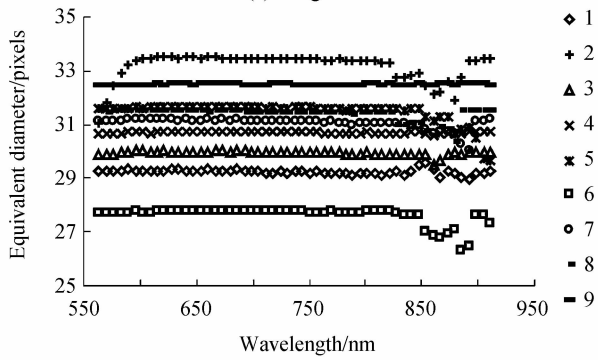




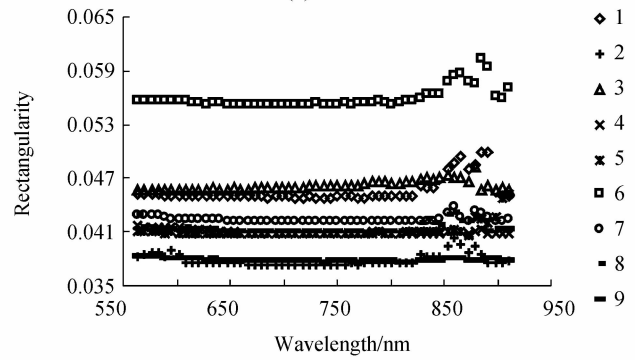
(c) Long axis



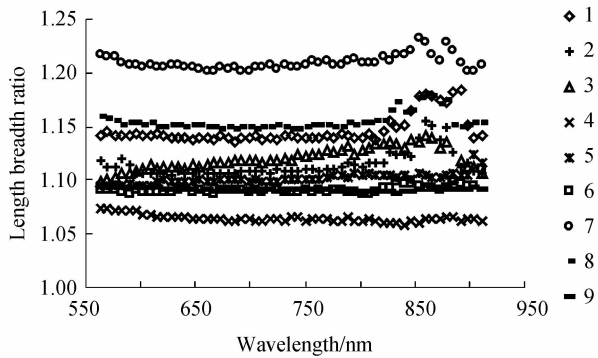
(d) Short axis



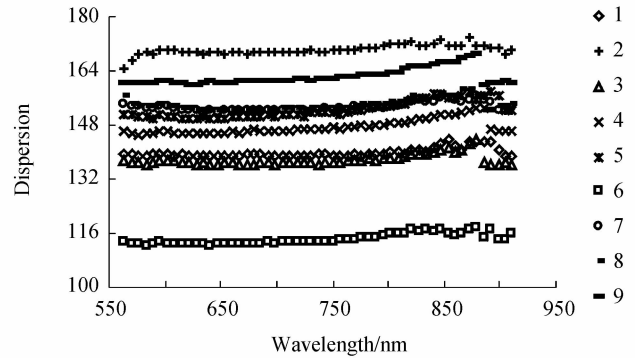
(e) Equivalent diameter



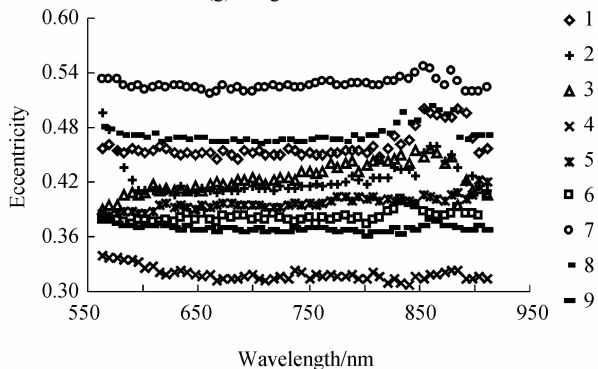
(f) Rectangularity



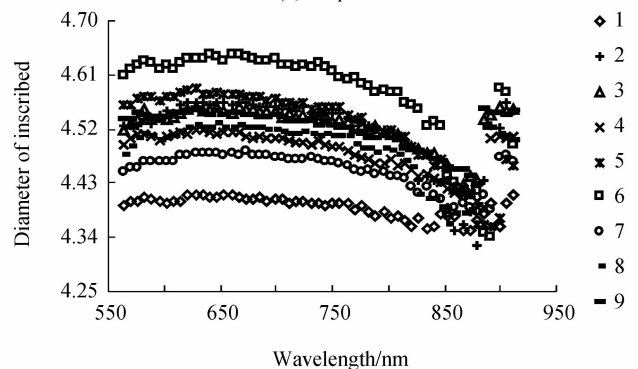
(g) Length breadth ratio



(h) Dispersion



(i) Eccentricity



(j) Diameter of inscribed circle

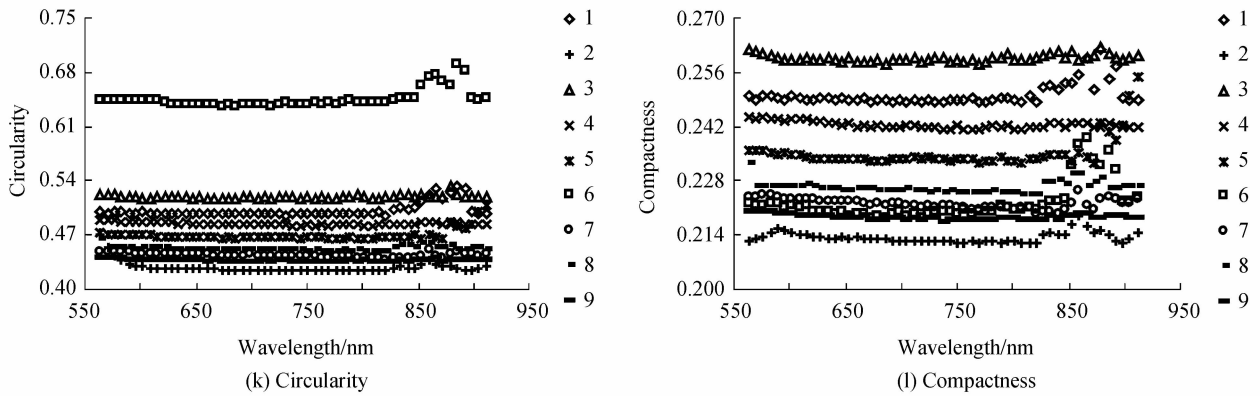


图 3 每个特征的类内统计均值在全波段中的散点分布图
 Fig. 3 The scatterplot distribution for statistical mean value of each feature in full wavelengths

由图 3 分析可知,面积、周长、长轴、短轴等特征参量的类内统计均值特征在第 26~61 波段(563.6~789 nm)时,同一品种的特征参量在各个波段下的差异性较小,几乎成一条直线.主要由于在第 26~61 波段下,图像的光谱区域主要处于可见光范围,与传统的机器视觉技术一样,获取的种子形态学信息特征受到其它因素影响较小,使得提取的形态特征参量在不同波段下表现出恒定性.

特征参量在第 62~80 波段(795.4~911.4 nm)下的类内统计均值呈不规则变化,波动较明显.主要原因在于:在第 62~80 波段下,图像的光谱区域处于近红外光谱范围内,此时 CCD 的量子效率减小;同时种子内部的化学信息成分对光的反射和吸收作用较强;从而导致种子图像的信噪比减小,最终导致提取特征的波动性.

3 结果与讨论

3.1 单波段 PLSDA 建模分类

为了避免由于不同特征之间参量差异性过大而对建模结果产生影响等因素,本文对玉米种子的各个特征参量进行归一化处理.对所有样本进行 4 选 1 选取每个品种 36 个样本作为分类器的训练样本,剩下的 12 个样本作为分类器的测试样本.利用每个波段下玉米种子的 12 个形状特征结合偏最小二乘判别法进行建模分类.建模过程中,采用留一交叉验证法产生的交叉验证错误分类率(Classification Error Rate for Cross-Validation, CLASSERRCV)的最小值确定最佳主成分数,同时,为了减少样本对分类结果产生的影响,选取 4 次不同的测试样本进行分类,并对 4 次结果作平均值.图 4 给出了每个波段下 9 类玉米种子的平均正确识别率.

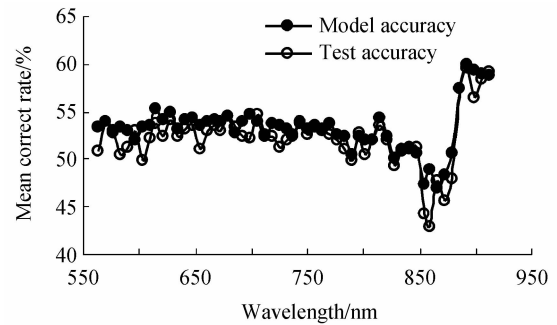


图 4 单波段下 9 类玉米种子的平均正确识别率
 Fig. 4 Mean value of correct recognition rate for nine kinds of maize seeds in each wavelength

由图 4 分析可知,利用单波段的特征进行 PLSDA 建模分类,9 个品种的玉米平均正确识别率大多集中在 40%~60%之间.训练集的最大平均正确识别率为 59.95%,测试集的最大平均正确识别率为 59.49%,整体识别率较低主要因为单波段下的品种特征数目较少,特征所能反映的品种信息较单一,从而不利于玉米品种的识别.

3.2 多波段、全波段条件下 PLSDA 建模分类

选取与单波段建模相同数目的训练样本和测试样本,分别利用第 26~61 波段(563.6~789 nm)、第 62~80 波段(795.4~911.4 nm)和全波段(563.6~911.4 nm)的特征结合偏最小二乘判别法进行建模分类,最佳主成分数确定方法和测试样本的选取与单波段时相同,三个分类模型的平均正确识别准确度如表 1.

由表 1 分析可知:1)第 26~61 波段(563.6~789 nm)的特征建模分类结果显示,9 个品种的平均训练准确度和测试准确度分别达到 71.22%和 65.43%,效果优于单波段的特征建模分类.结果说明,利用多波段特征数目多的优势,可以提高玉米品种的分类正确率.但是由于第 26~61 波段范围内,

表 1 三个分类模型的平均正确识别准确度

Table 1 Mean value of correct recognition accuracy of three classification models

Maize seed varieties	Number of training sample	Number of testing sample	No. 26~No. 61 wavelengths		No. 62~No. 80 wavelengths		Full wavelengths	
			Model	Test	Model	Test	Model	Test
			accuracy	accuracy	accuracy	accuracy	accuracy	accuracy
Chao Tian 2008	36	12	78.50	83.33	97.92	83.33	99.31	91.67
Gao Pin Le	36	12	97.92	95.84	95.83	89.59	100.00	95.84
Huang Fei	36	12	73.61	52.08	93.75	83.33	96.53	83.33
Huang Jin Nuo	36	12	72.92	66.67	100.00	95.83	100.00	95.84
Huang Nuo Yi Hao	36	12	70.14	64.58	100.00	97.92	100.00	100.00
Jin Nuo	36	12	98.61	97.92	100.00	95.84	100.00	97.92
Lv Se Chao Ren	36	12	50.00	43.75	100.00	100.00	100.00	97.92
Xiang Tian Yi Hao	36	12	33.33	24.31	89.58	83.33	93.06	93.75
Zao Xian Jin	36	12	65.97	60.42	90.97	81.25	95.83	89.59
Mean Correct Rate	36	12	71.22	65.43	96.45	90.05	98.31	93.98

不同波段下的同类种子的特征差异性较小(参见图3),导致该波段范围内的大量形态特征难以提供充足的有效分类特征,使得分类识别准确率提高有限。

2)第62~80波段(795.4~911.4 nm)的特征建模分类结果显示,9个品种的平均训练准确度和测试准确度分别达到96.45%和90.05%,效果均优于第26~61波段的特征建模分类。结果说明:相比于第26~61波段范围内的形态特征,在近红外光谱区域内,种子形态特征受到化学信息影响而产生的特征差异性增大,提供了更为充分的有效的分类特征信息,从而提高分类识别率。

3)全波段(563.6~911.4 nm)的特征建模分类结果显示,单个品种的训练准确度均在93%以上,测试准确度均在83%以上,9个品种的平均训练准确度和测试准确度分别达到98.31%和93.98%,效果均优于其它三类特征建模分类。结果说明:在高光谱区域范围内,充分结合可见光和近红外区域的有效特征信息,比单一区域内特征信息分类的效果显著,分类结果更具普遍性。

4 结论

高光谱图像中种子在不同波段的形状特征各有差异,可见光和近红外区域所呈现的形状特征差异更加明显。利用这个特点,本文将不同波段种子的特征信息应用到品种分类中以达到更好的分类效果。实验表明,利用不同波段范围的特征信息,分类识别能力不同,识别效果全波段和多波段均优于单波段,并且在充分利用高光谱图像中可见光和近红外区域的有效特征信息的基础上,全波段的平均正确识别率更高。该方法具有较高的识别准确度,为高光谱图像在玉米品种自动识别上的应用提供了重要的参考价值。

参考文献

- [1] SU Qian, WU Wen-jin, WANG Hong-wu, *et al.* Fast discrimination of varieties of corn based on near infrared spectra and biomimetic pattern recognition[J]. *Spectroscopy and Spectral Analysis*, 2009, **29**(9): 2413-2416.
苏谦, 邹文锦, 王红武, 等. 基于近红外光谱和仿生模式识别玉米品种快速鉴别方法[J]. *光谱学与光谱分析*, 2009, **29**(9): 2413-2416.
- [2] ZHANG Yuan, ZHANG Lu-da, BAI Qi-lin, *et al.* Nondestructive identification of different oil content maize kernels by near-infrared spectra [J]. *Spectroscopy and Spectral Analysis*, 2009, **29**(3): 686-689.
张愿, 张录达, 白琪林, 等. 近红外光谱法快速无损识别普通、高油、超高油玉米籽粒[J]. *光谱学与光谱分析*, 2009, **29**(3): 686-689.
- [3] QUAN Long-zhe, ZHU Rong-xin, LEI Pu, *et al.* Recognition method of Maize cultivars based on K-L transform and LS-SVM [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, **41**(4): 168-172.
权龙哲, 祝荣欣, 雷涛, 等. 基于K-L变换与LS-SVM的玉米品种识别方法[J]. *农业机械学报*, 2010, **41**(4): 168-172.
- [4] CHENG Hong, SHI Zhi-xing, YAO Wei, *et al.* Corn breed recognition based on support vector machine[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, **40**(3): 180-183.
程洪, 史智兴, 么炜, 等. 基于支持向量机的玉米品种识别[J]. *农业机械学报*, 2009, **40**(3): 180-183.
- [5] WANG Yu-liang, LIU Xian-xi, SU Qing-tang, *et al.* Maize seeds varieties identification based on multi-object feature extraction and optimized neural network[J]. *Transactions of the Chinese Society for Agricultural Engineering*, 2010, **26**(6): 199-204.
王玉亮, 刘贤喜, 苏庆堂, 等. 多对象特征提取和优化神经网络的玉米种子品种识别[J]. *农业工程学报*, 2010, **26**(6): 199-204.
- [6] YAN Xiao-mei, LIU Shuang-xi, ZHANG Chun-qing, *et al.* Purity identification of maize seed based on color characteristics [J]. *Transactions of the Chinese Society for Agricultural Engineering*, 2010, **26**(Supp. 1): 46-50.
闫小梅, 刘双喜, 张春庆, 等. 基于颜色特征的玉米种子纯度识别[J]. *农业工程学报*, 2010, **26**(Supp. 1): 46-50.
- [7] YANG Jin-zhong, HAO Jian-ping, DU Tian-qing, *et al.* Discrimination of numerous maize cultivars based on seed image process[J]. *Acta Agronomica Sinica*, 2008, **34**(6):

- 1069-1073.
杨锦忠,郝建平,杜天庆,等. 基于种子图像处理的大数目玉米品种形态识别[J]. 作物学报, 2008, **34**(6): 1069-1073.
- [8] LU R. Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images[J]. *Sensing and Instrumentation for Food Quality and Safety*, 2007, **1**(1): 19-27.
- [9] HUANG Min, ZHU Qi-bing. Feature extraction of hyperspectral scattering image for apple mealiness based on singular value decomposition[J]. *Spectroscopy and Spectral Analysis*, 2011, **31**(3): 767-770.
黄敏,朱启兵. 基于SVD的苹果粉质化高光谱散射图像特征提取[J]. 光谱学与光谱分析, 2011, **31**(3): 767-770.
- [10] HONG Tian-sheng, QIAO Jun, NING Wang, et al. Nondestructive inspection of chinese pear quality based on hyperspectral imaging technique[J]. *Transactions of the Chinese Society for Agricultural Engineering*, 2007, **23**(2): 151-155.
洪添胜,乔军, Ning Wang, 等. 基于高光谱图像技术的雪花梨品质无损检测[J]. 农业工程学报, 2007, **23**(2): 151-155.
- [11] LU R, PENG Y. Hyperspectral scattering for assessing peach fruit firmness[J]. *Biosystems Engineering*, 2006, **93**(2): 161-171.
- [12] GONZALEZ R C, WOODS R E, EDDINS S L. Digital image processing using matlab[M]. RUAN Qiu-qi, transl. Beijing: Publishing House of Electronics Industry, 2005: 252-320.
冈萨雷斯 R C, 伍兹 R E, 艾丁斯 S L. 数字图像处理 (MATLAB 版)[M]. 阮秋琦, 译. 北京: 电子工业出版社, 2005: 252-320.
- [13] WANG Yu-liang. Maize seed identification and detection based on the machine vision [D]. Tai'an: Shandong Agriculture University, 2008.
王玉亮. 基于机器视觉的玉米种子品种识别与检测研究[D]. 泰安: 山东农业大学, 2008.
- [14] JIN Fei-jian. Feature of weeds extraction and identification based on the machine vision [D]. Zhenjiang: Jiangsu University, 2007.
金飞剑. 基于机器视觉的杂草图像特征提取及识别研究[D]. 镇江: 江苏大学, 2007.

Morphological Characteristics of Maize Seed Extraction and Identification Based on the Hyperspectral Image

HUANG Min, ZHU Xiao, ZHU Qi-bing, FENG Zhao-li

(Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education),
Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract: Morphological characteristic of maize seed is an important factor in identifying maize varieties. Hyperspectral images of 432 maize seeds including nine varieties were acquired using hyperspectral imaging system. The images were corrected and pre-processed, and then shape features of each sample were extracted in the range of 563.6 ~ 911.4 nm including 55 wavelengths. The classification models were developed using the shape features of maize seeds from single-wavelength, multi-wavelengths and full wavelengths coupled with partial least squares discriminant analysis (PLSDA), respectively. Simulation results indicate that the average correct identification rate of training set and testing set with full wavelengths is 98.31% and 93.98%, which are better than single-wavelength and multi-wavelengths. Therefore, that is the accurate mean for identifying maize varieties using the feature information of visible and near-infrared region from hyperspectral images.

Key words: Hyperspectral images; Maize seed; Morphological characteristics; Species identification; Partial least squares discriminant analysis