融合光谱、纹理及形态特征的水稻种子品种高光谱图 像单粒鉴别

邓小琴 朱启兵 黄 敏

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 水稻种子品种的单粒鉴别对于防止制种时的混杂、掺假现象,保证种子纯度具有重要意义。利用高光谱图像技术研究了水稻种子品种的单粒快速鉴别方法。采集了10类水稻种子在400~1000 nm 范围内的高光谱反射图像并提取其光谱、纹理和形态特征;结合偏最小二乘判别分析模型比较了不同特征及其组合下的分类精度,并利用多次递进 无信息变量消除算法结合偏最小二乘投影分析方法筛选最优波段。结果显示,在仅利用23个最优波段情况下,融合 均值、熵、能量和形态特征所建立的鉴别模型获得了令人满意的识别精度,其训练集、测试集精度分别为99.22%、 96%。结果表明,高光谱特征融合可以在少量波段情况下有效地提高水稻种子品种单粒鉴别的精度,基本满足国家 标准对种子纯度的检测要求。

关键词 图像处理;高光谱图像;特征融合;水稻种子;品种鉴别;偏最小二乘判别分析
 中图分类号 TP391.41
 文献标识码 A
 doi: 10.3788/LOP52.021001

Variety Discrimination for Single Rice Seed by Integrating Spectral, Texture and Morphological Features Based on Hyperspectral Image

Deng Xiaoqin Zhu Qibing Huang Min

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

Abstract Variety discrimination for single rice seed is important to prevent the mixing and adulteration during seed production and to ensure the seed purity. A fast discrimination method for single rice seed by using the hyperspectral imaging technology is investigated. Hyperspectral images of rice seeds from ten varieties are collected over the wavelength region of 400~1000 nm, and the spectral, texture and morphological features of rice seeds are extracted. The discrimination accuracy of different features and their combinations is compared by using the partial least squares discriminant analysis, and the multiple progressive uninformative variable elimination algorithm combined with the partial least squares projection analysis algorithm is used for optimal waveband selection. The results show that the satisfactory discrimination accuracy, which is 99.22% and 96% for the training set and test set respectively, is achieved when mean, entropy and power features for the 23 optimal wavebands and morphological features are integrated. It suggests that multiple hyperspectral feature integration can effectively improve discrimination accuracy for single rice seed at the case of a small amount of wavebands, which basically meets the requirements of the national standards on the seed purity identification.

Key words image processing; hyperspectral image; feature integration; rice seed; variety discrimination; partial least squares discriminant analysis

OCIS codes 100.4145; 100.2960; 110.4234

1 引 言

民以食为天,食以稻为先。水稻是我国最主要的粮食作物之一。全国水稻种植面积约占粮食作物面积

收稿日期: 2014-07-14; 收到修改稿日期: 2014-09-04; 网络出版日期: 2015-01-06

基金项目:国家自然科学基金(61271384,61275155)、江苏省自然科学基金(BK2011148)

作者简介:邓小琴(1988—),女,硕士研究生,主要从事高光谱图像无损检测方面的研究。

E-mail: deng_xiao_qin@163.com

导师简介:朱启兵(1973—),男,博士,副教授,主要从事农副产品无损检测方面的研究。E-mail: zhuqib@163.com

(1)

的30%,产量接近粮食总产量的50%。随着杂交技术的广泛应用,在水稻制种过程中的种子间混杂现象越来 越突出。此外,在种子销售环节中,受经济利益的驱使,在优质稻种中掺杂劣质稻种甚至以劣质稻种冒充优 质稻种的事件时有发生,给农民造成了严重的损失。如何对水稻种子品种实现快速、无损鉴别已成为水稻 育种、销售等环节的重要课题。

自1960年梅曼研制的第一台红宝石激光器问世以来,光学进入了快速发展阶段,应用非常广泛^[1-4]。自20世纪90年代以来,具有实时、无损检测特点的机器视觉技术、近红外光谱分析技术、高光谱图像技术等在种子品种检测中得到了广泛的研究^[5-9]。种子品种分类鉴别依赖于对特征信息的充分获取,由于高光谱图像技术集成了光谱技术和图像技术的优点,可较为充分地获取种子品种外在形态特征、内部结构特征和化学成分等多种特征信息,在种子分类鉴别领域表现出了良好的应用潜力。刘飞等^[10]利用874~1734 nm 波长范围内的近红外高光谱图像研究了4个品种共121个西瓜种子品种的识别问题,建立的识别模型精度对于建模集和测试集均为100%。但是上述研究无法实现品种的单粒识别,导致对于种子制种过程的混杂现象和销售过程中的掺假问题缺乏识别能力。黄敏等^[11-12]利用高光谱反射图像技术,提取了玉米种子在563.6~911.4 nm 共55个波段范围内的12个形状特征信息,对9个品种共432粒玉米种子,获得了测试集内93.98%的识别精度。朱启兵等^[11-12]分别利用光谱特征、熵特征等方法研究了玉米种子品种的单粒识别问题,也取得了很高的分类精度。但这些研究都是基于单一特征的识别模型,没有充分利用高光谱图像图谱合一的优势,存在着有用信息丢失的可能性。

本文利用高光谱图像技术研究了水稻种子的单粒识别问题。在比较光谱特征、纹理特征和形态特征对 分类精度影响的基础上,采用光谱、纹理、形状特征等多特征选择融合的识别方法;并针对高光谱图像技术 波段数目众多带来的数据传输、存储困难,在线应用实时性差等问题,引入了多次递进无信息变量消除算法 (MP-UVE)结合偏最小二乘投影分析算法进行最优波段选择,旨在波段数目较少的条件下,实现水稻种子品 种的单粒准确识别。

2 材料与方法

2.1 样本来源

实验样本来自于种子公司提供的苏北12号(SB)、皖稻87号(WD)、屉优267(TY)、培根(PG)、国亩20 (GM)、国丰112(GF)、九州(JZ)、两优3905(LY)、II优371(IIY)、丰两优香一号(FLY)共10类水稻品种,每类各 120粒,共1200个样本。所选样本的颜色、形状相似。每类样本中按照4选1的方法随机抽取90个样本作为 建模的训练集,30个作为预测集,最终获得900个训练集样本,300个预测集样本。

2.2 高光谱图像采集与校正

采用文献[12]中的高光谱图像采集系统进行高光谱图像采集。为避免图像失真,确定的最佳系统参数为: 曝光时间 250 ms,线扫描步长 80 μm,扫描宽度 60 mm,光谱压缩倍数设置为 10,即实际波段间隔为 6.4 nm。 图像采集时将水稻种子样本放置在黑色载物板上,整个采集过程在密闭黑箱中进行以减弱外部光源的干扰。 每隔3幅图像采集一次标准白板和全黑标定图像对原始图像进行校正,以降低光源变化及系统噪声的影响, 校正公式为

 $R_{\rm c} = \frac{R_{\rm s} - R_{\rm b}}{R_{\rm m} - R_{\rm b}},$



图1 培根品种在718 nm 波长下校正后的高光谱图像

Fig.1 Hyperspectral image of PG after calibration at the wavelength of 718 nm

激光与光电子学进展

式中 R_{a} 为校正图像, R_{s} 为原始图像, R_{u} 和 R_{b} 分别为白板和黑板图像。后续图像处理与分析均在 R_{a} 上进行。图 1为培根品种在 718 nm 波长下校正后的高光谱图像。

2.3 高光谱图像分割

为了实现水稻种子的单粒识别,需要将种子图像从背景中分割出来。基于区域主动轮廓模型的图像分割算法具有对光照和噪声不敏感的优势,对于弱边界、多目标等复杂图像都具有良好的分割效果^[13-14],本文采用区域主动轮廓算法实现种子图像的分割,图2为利用区域主动轮廓模型提取的培根品种(PG)在718 nm 波长下的轮廓图像。和图1对比可知,区域主动轮廓模型较好地提取了水稻种子的轮廓图像,为后续的特征准确提取提供了保证。将718 nm 波长下提取的种子轮廓图像投射到400~1000 nm 范围内的其他波段下,即可实现不同波段下的种子轮廓分割。



图2 培根品种在718 nm 波长下的高光谱图像分割

Fig.2 Hyperspectral image segmentation of PG at the wavelength of 718 nm

2.4 高光谱图像特征提取

为了充分挖掘高光谱图像中的目标信息,分别从光谱、纹理、形态三个方面进行特征提取。主要提取每 粒水稻种子光谱的均值(*M*)和标准差(SD)作为光谱特征描述符,用于反映水稻种子的化学信息及其分布状况。纹理是图像的一种局部结构化特征,反映了目标图像灰度的性质及其空间拓扑关系。与颜色和形态特征相比,纹理特征中关于被测物体的化学和物理属性信息更为丰富^[15-16]。主要提取了不同波段下图像的角 二阶矩(ASM)、熵(*E*)^[17]和能量(*P*)三种纹理特征,其公式为

$$E = -\sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j) \log_2 p(i,j),$$
(2)

$$M_{\rm ASM} = \sum_{i=1}^{m} \sum_{j=1}^{n} p^2(i,j), \tag{3}$$

$$P = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{x(i,j) - X_{\text{con}}}{\sqrt{(i-a)^2 + (j-b)^2}},$$
(4)

式中 $p(i,j) = x(i,j)/f_{sum}(i,j)$, $f_{sum}(i,j) = \sum_{i=1}^{m} \sum_{j=1}^{n} x(i,j)$, X_{con} 为质心的光强值, $a \downarrow b$ 分别为质心的横向和纵向像素值。

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、离散度和紧凑度等是 反映不同品种形态差异的基本物理特征^[18]。由于在图像分割时,采用了种子轮廓图像的投射方法,各波段下 的形态特征是相同的,因此仅利用单一波段下的上述12个形态特征作为水稻种子基本的物理特征描述符。 2.5 高光谱图像波段选择

高光谱图像波段数目众多,在提供大量待测目标特征信息的同时,也给数据传输、存储以及处理与分析带来了挑战,因此通常需要通过波段选择来提高它在实际应用中的实时性。无信息变量消除算法(UVE)^[19] 是一种基于回归系数稳定性分析的变量选择方法,能够有效剔除无用的信息变量,防止预测模型过拟合,改善模型的预测性能。

多次递进无信息变量消除是在UVE的基础上进行改进的算法,即将前一次经UVE选择所得的变量重新进行UVE筛选,达到消除低信息变量,选出高信息变量,简化模型,提高模型预测精度的目的。经过多次递进后,所选出的变量数目在一定范围内基本保持不变,可以有效地减小UVE算法不稳定性的影响。但有时通过多次递进后所得的变量数依然较多,可以考虑将筛选出的变量作为偏最小二乘(PLS)²⁰¹投影分析算法的输入变量,进一步筛选出高信息变量,减少变量数。本文采用MP-UVE-PLS算法对高光谱图像数据进行

波段选择。

2.6 偏最小二乘判别分析(PLSDA)分类模型

PLSDA 是一种基于判别分析的 PLS 回归算法^[21],将光谱数据域分类变量进行线性回归,在构造因素时 考虑到了辅助矩阵以代码形式提供的类成员信息。将 PLS 中的浓度变量换作类别变量,PLSDA 用来计算特 征矩阵 *X* 与类别矩阵 *Y* 的相关关系,获得最大协方差 $f_{cov}(X,Y)$,特征矩阵 $X \in R^{n \times n}$,类别矩阵 $Y \in R^{n \times p}$,其中 n 为样本数,m 为变量数,p 为类别数。PLSDA 模型分别提供了训练阶段和预测阶段正确分类的百分比^[22], 特别适合于解释变量多且存在着多重共线性、干扰噪声大的情况。实验中的主要工作在 Matlab2009 环境下 完成。

3 结果与分析

3.1 水稻种子的高光谱特征

图3给出了10类水稻品种在400~1000 nm 波长范围内的均值、标准差、能量、熵、角二阶矩共5个特征的 曲线图。图中的每条曲线代表一个类别的特征平均值。由于形态特征只在单一波段下提取,故图3中没有 给出。从图中可以看出这5种特征曲线中均存在部分种类间特征曲线的相互交叉重合,从视觉上很难分析 各类特征对水稻种子品种的鉴别能力,因此需要通过建模分析来进行进一步判断。



图3 水稻种子的5种特征曲线

Fig.3 Five feature curves of rice seeds

3.2 应用PLSDA建立水稻种子品种鉴别模型

分别将400~1000 nm 波长范围内共94个波段下能量、均值、熵、标准差、角二阶矩这5个特征作为输入 技能建立 PLSDA模型,鉴别结果如图4(a)所示。分析可知,不同的特征对水稻种子品种的鉴别能力不同。 能量特征和均值特征均获得了超过95%的测试精度,熵特征和角二阶矩特征的测试精度均高于85%,而标准 差的测试精度最低,为78%。原因可能在于实验的水稻种子都有坚硬不规则的外壳,导致标准差特征在同类 别样品中的差异性变化较大,从而使分类性能下降。由(2)式和(3)式可以看出,熵特征和角二阶矩特征都 忽略了测试样本的平均光谱信息,而平均光谱信息是种子化学成分的直接体现,也是种子类别区分的重要 特性,因此也导致了分类精度有一定的下降^[23]。

特征的选择往往对分类的结果产生很大的影响,合理地选择特征可以提高分类的准确性。分类的类别 越多,各类别之间的特征越接近,分类的难度就越大。利用图像的单一特征无法全面反映样本间的相似性 与差异性,识别时需要综合考虑样本的多种特征^[24]。将全波段下的能量、均值、熵、标准差、角二阶矩这5个

激光与光电子学进展

特征以不同的组合进行融合,并作为输入量建立 PLSDA模型,鉴别结果取最优的4组,如图4(b)所示。可以 看出,将这5个特征全都融合在一起时模型的训练精度最高,为99.33%。但是将能量、均值、熵这3个特征融 合在一起时模型的测试精度最高,为98.33%,其训练精度也达到了98.78%。融合3个特征较融合5个特征时 数据量减少了40%,可以简化模型,提高数据处理效率。而且将能量、均值、熵这3个特征融合在一起对水稻 种子品种的识别能力高于单一特征。



图4 全波段下水稻种子品种的平均分类精度



3.3 波段选择

过多的波段数目,将带来实时应用的困难,同时考虑到能量、均值和熵特征融合具有最高的分类精度,因此对能量、均值、熵3种特征分别采用多次递进UVE算法结合PLS波段投影的方法进行波段选择,每种特征递进次数设定为5次,PLS投影系数的截断阈值为投影系数最大值的20%。在选出每种特征的最优波段后,再选择3种特征的共有波段作为最终选择的最优波段。共选择出23个最优波段,其对应的中心波长分别为479、512、518、550、557、576、621、627、640、660、673、679、698、718、724、750、782、801、879、892、898、911、962 nm,从中可以看出这些波长主要分布在蓝绿光(476~550 nm)、红光(620~760 nm)和近红外(760~1000 nm)部分。实验中采用的水稻颜色为黄橙色,其互补色为蓝绿色;在红光和近红外波段则主要反映水稻中的O—H、C—H等化学键的吸收,这些波段是不同种子化学成分、物理状态差异性体现的主要波段。所选择的波段数目约占总波段数的24%,达到了减少变量数目,简化模型的目的。

3.4 水稻种子的光谱特征、纹理特征与形态特征融合

将这23个最优波段下的能量、均值和熵特征作为模型的输入量建立PLSDA模型,所得训练精度为 90.67%,测试精度为85.33%,如图5所示。这一结果相比于全波段条件下的分类精度有了极大的下降,可能 的原因在于利用MP-UVE-PLS算法后波段数目下降过多,导致部分分类信息的丢失。为了弥补分类信息 丢失的问题,将单一波段下的12个形态特征增加到23个最优波段下的能量、均值和熵特征中,并作为输入 量建立PLSDA模型。从图5中可以看出,增加12个形态特征后,无论是训练精度还是测试精度都有了极大 的提高,所得训练精度为99.22%,测试精度为96%。可以在大幅度降低波段数目、提高计算速度的同时,满足 国家标准对种子纯度不得低于95%的质量要求。





Fig.5 Mean accuracy of variety discrimination for rice seeds after feature integration

4 结 论

由于水稻种子的单一特征无法全面反映种子间的相似性与差异性,因此采用多特征融合结合波段选择的方法,利用较少波段下的融合特征信息实现了单粒水稻种子的准确识别。采集了10个水稻种子品种的高光谱图像,提取了能量、均值、熵、标准差和角二阶矩等17个特征,并通过MP-UVE-PLS算法选取最优波段。研究结果表明,融合光谱、纹理和形态特征所建立的PLSDA模型对水稻种子品种识别精度较高。但限于样本种类、数量及样本本身的纯度无法达到100%,模型的有效性和稳健性还有待验证。同时为了适应现实应用的需要,即以尽可能少的波段获得较高的品种识别精度,波段选择算法还需要做进一步的研究。

参考文献

- 1 Liu Yang, Zhang Guoan. Study on modulation scheme of visible light communications and its performance [J]. Laser & Optoelectronics Progress, 2014, 51(9): 090601.
 - 刘 洋,章国安.可见光通信调制方式及其性能研究[J].激光与光电子学进展,2014,51(9):090601.
- 2 Cao Pengfei, Li Hongning, Luo Yanlin, et al.. Selection of feature bands for Phaseolus vulgaris leaves based on multispectral imaging [J]. Laser & Optoelectronics Progress, 2014, 51(1): 011101.
- 曹鹏飞,李宏宁,罗艳琳,等.基于多光谱成像选取四季豆叶片的特征波段[J].激光与光电子学进展,2014,51(1):011101.
- 3 Lü B, Wu J. Quantifying spectroscopic concentration ratio by polarization subtraction technique [J]. Chinese Optics Letters, 2013, 11(s2): s23001.
- 4 Quan J, Bai B, Jin S, *et al.*. Indoor positioning modeling by visible light communication and imaging [J]. Chinese Optics Letters, 2014, 12(5): 052201.
- 5 Dana W, Ivo W. Computer image analysis of seed shape and seed color for flax cultivar description [J]. Computers and Electronics in Agriculture, 2008, 61(2): 126–135.
- 6 Khatchatourian O, Padilha F R R. Soybean varieties recognition through the digital image processing using artificial neural network [J]. Engenharia Agricola, 2008, 28(4): 759-769.
- 7 Cheng Fang, Ying Yibin. Application of machine vision to quality evaluation of agricultural seed resources [J]. Transactions of the Chinese Society of Agricultural Engineering, 2001, 17(6): 175–179.

成 芳,应义斌.机器视觉技术在农作物种子质量检验中的应用研究进展[J].农业工程学报,2001,17(6):175-179.

8 Su Qian, Wu Wenjin, Wang Hongwu, *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, 邬文锦, 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.

- 9 Liu Tianling, Su Qiya, Sun Qun, *et al.*. Recognition of corn seeds based on pattern recognition and near infrared spectroscopy technology [J]. Spectroscopy and Spectral Analysis, 2012, 32(6): 1550-1553. 刘天玲, 苏琪雅, 孙 群, 等. 基于 NIR 分析和模式识别技术的玉米种子识别系统 [J]. 光谱学与光谱分析, 2012, 32(6): 1550-1553.
- 10 Zhang Chu, Liu Fei, Kong Wenwen, *et al.*. Fast identification of watermelon seed variety using near infrared hyperspectral imaging technology [J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(20): 270-277. 张 初,刘 飞,孔汶汶,等. 利用近红外高光谱图像技术快速鉴别西瓜种子品种[J]. 农业工程学报, 2013, 29(20): 270-277.
- 11 Huang Min, Zhu Xiao, Zhu Qibing, *et al.*. Morphological characteristics of maize seed extraction and identification based on the hyperspectral image [J]. Acta Photonica Sinica, 2012, 41(7): 868–873.

黄 敏,朱 晓,朱启兵,等.基于高光谱图像的玉米种子特征提取与识别[J].光子学报,2012,41(7):868-873.

- 12 Zhu Qibing, Feng Zhaoli, Huang Min, *et al.*. Maize seed classification based on image entropy using hyperspectral imaging technology [J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(23): 271-276. 朱启兵, 冯朝丽, 黄 敏, 等. 基于图像熵信息的玉米种子纯度高光谱图像识别[J]. 农业工程学报, 2012, 28(23): 271-276.
- 13 Li Xuanping, Wang Xue. Active contour model-based medical image segmentation method collaborative with fuzzy cmeans [J]. Chinese Journal of Scientific Instrument, 2013, 34(4): 860–865.

李宣平, 王 雪. 模糊聚类协作区域主动轮廓模型医学图像分割[J]. 仪器仪表学报, 2013, 34(4): 860-865.

14 Huang Min, Zhu Xiao, Zhu Qibing, et al.. Hyperspectral image classification of maize seeds based on active contour model [J]. Journal of Data Acquisition and Processing, 2013, 28(3): 289–293. 黄 敏,朱 晓,朱启兵,等.基于主动轮廓模型的玉米种子高光谱图像分类[J].数据采集与处理,2013,28(3):289-293.

15 An Bin, Chen Shuhai, Chen Hua, *et al.*. Application of texture feature in mutispectral image classification [J]. Laser & Infrared, 2002, 32(3): 188–190.

安 斌,陈书海,陈 华,等.纹理特征在多光谱图像分类中的应用[J].激光与红外,2002,32(3):188-190.

16 Wang Shuzhi, Zhang Jianhua, Feng Quan. Defect detection of mulskmelon based on texture features and color features [J]. Transactions of the Chinese Society of Agricultural Machinery, 2011, 42(3): 175–179.

王书志,张建华,冯 全.基于纹理和颜色特征的甜瓜缺陷识别[J].农业机械学报,2011,42(3):175-179.

- 17 Feng Xin, Zeng Zhoumo, Feng Hao, *et al.*. A method for evaluating the disturbance in distributed vibration sensor based on wavelet information entropy [J]. Acta Optica Sinica, 2013, 33(11): 1106005.
 冯 欣,曾周末,封 皓,等. 基于小波信息熵的分布式振动传感器系统的扰动评价方法 [J]. 光学学报, 2013, 33(11): 1106005.
- 18 Jin Feijian. Research of Feature Extraction and Recognition Method of Weed Image Based on Machine Vision [D]. Zhenjiang: Jiangsu University, 2007. 49-56.

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. 49-56.

- 19 Wu D, He Y, Nie P, *et al.*. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice [J]. Analytica Chimica, 2010, 659(1): 229-237.
- 20 Li S, Lin L, Milliken R, *et al.*. Hybridization of partial least squares and neural network models for quantifying lunar surface minerals [J]. Icarus, 2012, 221(1): 208-225.
- 21 He Bin. Research on Discriminant Analysis of Milk Adulteration by Infrared Spectroscopy [D]. Tianjin: Tianjin University, 2010. 38-47.

何 彬.基于红外光谱技术的牛奶掺杂判别方法的研究[D].天津:天津大学,2010.38-47.

- 22 Westerhuis J A, Van Velzen E J J, Hoefsloot H C J, *et al.*. Multivariate paired date analysis: Multi level PLSDA versus OPLSDA [J]. Metabolomics, 2010, 6(1): 119–128.
- 23 Zheng C, Sun D-W, Zheng L. Recent developments and applications of image features for food quality evaluation and inspection a review [J]. Trends in Food Science and Technology, 2006, 17(12): 642-655.
- 24 Kong Xiangbing, Shu Ning, Tao Jianbin, *et al.*. A new spectral similarity measure based on multiple features integration [J]. Spectroscopy and Spectral Analysis, 2011, 31(8): 2166–2170.

孔祥兵, 舒 宁, 陶建斌, 等. 一种基于多特征融合的新型光谱相似性测度[J]. 光谱学与光谱分析, 2011, 31(8): 2166-2170.

栏目编辑: 吴秀娟