

基于分段主成分分析和高光谱技术的大豆品种识别

刘瑶^{1,2}, 谭克竹¹, 陈月华¹, 王志朋¹, 谢红², 王立国²

(1. 东北农业大学 电气与信息学院 黑龙江 哈尔滨 150030; 2. 哈尔滨工程大学 信息与通信工程学院 黑龙江 哈尔滨 150001)

摘要: 为了实现大豆品种的快速且无损鉴别, 对大豆高光谱图像中的光谱信息进行研究分析。利用高光谱图像采集系统采集波长范围为 400 ~ 1 000 nm 的 6 类共 660 粒大豆样本的高光谱图像, 从每粒大豆样本的中心区域上提取感兴趣区域并以此区域的平均光谱信息代表此粒大豆的光谱信息。对光谱曲线进行多元散射校正 (multiple scattering correction, MSC) 后, 根据相关系数矩阵图, 将整个高光谱波段分解为 3 个子分段, 分别在每个子分段上做主成分分析 (principal component analysis, PCA), 提取 1 ~ 20 个主成分作为光谱特征, 利用极限学习机 (extreme learning machine, ELM) 和随机森林 (random forests, RF) 模型进行大豆品种识别。结果表明: 在第二分段 (510.6 ~ 685.4 nm) 进行 PCA 变换, 识别效果优于全波段 PCA 变换。因此, 应用分段 PCA 变换和高光谱技术对大豆品种进行无损识别是可行的。

关键词: 高光谱; 分段主成分分析; 大豆; 品种识别

中图分类号: TP391.41; TP274.5 文献标识码: A DOI: 10.11861/j.issn.1000-9841.2016.04.0672

Variety Recognition of Soybeans Using Segmented Principal Component Analysis and Hyperspectral Technology

LIU Yao^{1,2}, TAN Ke-zhu¹, CHEN Yue-hua¹, WANG Zhi-peng¹, XIE Hong², WANG Li-guo²

(1. College of Electric and Information, Northeast Agricultural University, Harbin 150030, China; 2. College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: In order to realize rapid nondestructive recognition for soybeans varieties, spectral information of hyperspectral image for soybeans is analyzed. Hyperspectral images for 660 soybeans including 6 varieties from 400 to 1 000 nm are acquired by hyperspectral image system. The interested region for each sample is extracted and average spectral information is obtained. Spectrum curve is conducted multiple scattering correction (MSC). According to the correlation coefficient matrix, the entire hyperspectral bands are segmented into three highly relevant sub-segments. Principal component analysis (PCA) is respectively used in each sub-segments. The first twenty principal components are extracted as spectral features of soybean samples. Identification model is developed using Extreme Learning Machine (ELM) and Random Forests (RF) models. Experimental results indicate that the identification results using PCA in the second sub-segment (510.6 ~ 685.4 nm) is outperform than using PCA in the entire bands. Nondestructive recognition and classification for soybean varieties using segmented PCA and hyperspectral technology is effective.

Keywords: Hyperspectral; Segmented principal component analysis; Soybean; Variety identification

大豆是黑龙江省主要农作物之一, 在国民经济发展中占有重要地位^[1]。1985 - 2011 年, 由黑龙江省品种审定委员会审定通过的大豆品种达到 285 种, 这些新品种大豆的推广和应用对黑龙江省, 乃至全国大豆生产的发展起到了重要的推动作用。随着大豆品种数量的增加, 在大豆的种植、运输和储藏过程中品种易混杂, 种子质量降低, 给国家和农民造成较大的利益损失。快速、准确地鉴别大豆品种, 对保护农民利益, 鉴定种子质量以及保障粮食安全具有重要意义, 并且可以为大豆育种技术的发展提供技术支持。

大豆品种识别相关研究目前还很少, 主要集中

在应用近红外技术。洪庆红等^[2]利用傅里叶变换红外光谱法测定了大豆品种的子叶和外表皮的红光谱, 研究发现不同大豆品种的傅里叶变换红外光谱在 1 800 ~ 1 200 cm^{-1} 范围内有较大的差异, 可以用于不同的大豆品种的鉴别, 但文中并未给出具体的鉴别方法, 属于探索性的研究。朱大洲等^[3]采用近红外光谱仪采集了单粒大豆的漫反射光谱, 并结合软独立建模分类方法建立了大豆的定性分析模型, 实现对两个品种大豆的鉴别, 虽然鉴别的准确率高, 但品种较少。杨东风等^[4]对 16 份大豆品种的近红外透射光谱进行分析, 用主成分分析和离散多带小波变换提取光谱特征, 作为 BP 神经网络的

收稿日期: 2015-12-08

基金项目: 国家自然科学基金资助项目 (60802059) 黑龙江省自然科学基金重点项目 (ZD201303)。

第一作者简介: 刘瑶 (1982-), 女, 博士, 讲师, 主要从事高光谱图像处理技术研究。E-mail: liuyao0904@163.com。

通讯作者: 谢红 (1962-), 女, 教授, 博导, 主要从事信号与信息处理技术研究。E-mail: xiehong@hrbeu.edu.cn。

输入变量,建立大豆品种识别模型,训练集和测试集识别准确率分别能达到 98.125% 和 95.93%,识别准确率较高,但是没有考虑高光谱波段之间的相关性。柴玉华等^[5]利用主成分分析后的高光谱图像的纹理特征对大豆进行分级,预测准确率达到 92%,主要利用的是高光谱的图像特征,没有涉及到光谱特征。

高光谱图像技术是一种结合图像技术与光谱技术的无损检测新方法,具有多波段、图谱合一、光谱分辨率高等特点。目前,高光谱图像技术应用在农业检测领域已成为国内外的一个研究热点^[6-11],已广泛应用在农产品的品种分类、损伤识别、内外品质预测、安全检测等领域。本研究选取由东北农业大学大豆研究所选育的 6 种大豆样本进行高光谱图像品种识别,运用相关矩阵图划分光谱波段,在每个分段上做主成分分析提取光谱数据特征,利用极限学习机和随机森林建模,实现大豆品种分类特征信息的充分提取,以期达到提高品种分类精度的目的。

1 材料与方法

1.1 材料与设备

供试大豆材料为东农 42、东农 43、东农 47、东农 51、东农 53 和东农 61。每个品种中选取 110 粒正常完好的籽粒,共计 660 个样本。本试验所用 6 种大豆样品由东北农业大学国家大豆研究中心提供。

试验中所用仪器是美国 Headwall 公司的高光谱图像采集系统,该系统包括图像光谱仪(Hyperspec VNIR-A)、CCD 像头、镜头;光源为 150 W 可调功率光纤卤素灯;样本输送平台由步进电机控制^[12]。高光谱图像光谱仪的狭缝是 25 μm ,有效光谱的范围 400 ~ 1 000 nm,光谱分辨率 2 ~ 3 nm,光谱采样间隔为 0.74 nm。

1.2 高光谱图像采集与校正

在对所有样本进行高光谱图像采集试验时,将大豆样本留有一定间隔地摆放在全黑背景的移动平台上进行高光谱图像采集。

在高光谱成像系统中,光源强度分布不均匀,存在一些噪声,如暗电流,严重影响采集的高光谱图像质量,因此需要对每幅样本图像进行校正,校正公式为:

$$R_c = \frac{R_s - R_b}{R_w - R_b}$$

式中 R_c 为校正图像; R_s 为原始图像; R_w 为标准白板图像; R_b 为标准黑板图像。

后续实验中的图像处理与分析均在 R_c 上进行。图 1 为大豆样本东农 51 在 706.15 nm 波段下的校正后图像。

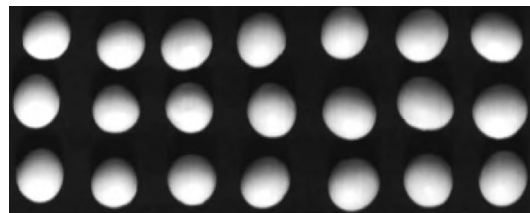


图 1 706.15 nm 波段下东农 51 的校正后图像

Fig. 1 Image of Dongnong 51 after correction in 706.15 nm

1.3 高光谱数据的提取与预处理方法

在高光谱数据提取时,首先需要在样本中心区域内手动选取大小为 100 像素的椭圆形区域,作为感兴趣区域(region of interest, ROI),ROI 选取的好坏直接影响后期所建模型的性能。本试验使用 ENVI 软件的感兴趣提取工具进行提取,然后计算 ROI 内所有像素点的光谱反射值的平均值,得到每个样本的平均高光谱数据。

高光谱数据中除了包含样品的自身信息外,还包含其他无用的信息和噪声,因此在建模之前对原始光谱数据进行预处理用以消除这些因素的影响是有必要的。光谱预处理的方法主要有基线补偿、去趋势、多元散射校正、变量标准化、一阶求导处理和二阶求导处理等,本试验中采用多元散射校正和矢量归一化相结合的预处理方法^[13]。

1.4 分段主成分分析

高光谱数据的特点是波段多、数据量大、冗余性强,若建模中直接使用原始高光谱数据可能会发生 Hughes 现象,所建立的模型性能差、效率低。因此需要对这些高光谱数据进行处理,用维数较少的并且不相关的数据来取代原始的高光谱数据信息,使用变换后的数据进行建模,从而达到解决问题的目的,而这些高维数据变换通常采用主成分分析的方法^[14]。

假定有 n 个样本,每个样本共有 P 个变量,构成一个 $n \times P$ 阶的矩阵:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1P} \\ x_{21} & x_{22} & \cdots & x_{2P} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nP} \end{bmatrix}$$

原始数据 X 的协方差矩阵为:

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T$$

其中 μ 是原始数据中的全部样本的均值矢量。

协方差矩阵 C 的特征值 λ 和特征向量 U 可以通过求解特征方程来获取。

对特征值进行降序排列, 即 $\lambda_1 > \lambda_2 > \dots > \lambda_m$, 其相应的特征向量为 $U_1 > U_2 > \dots > U_m$ 。可以得到 PCA 变换矩阵 $W = [U_1, U_2, \dots, U_m]$ 。

最终的 PCA 变换可以表示为 $Y = W^T X$, 其中 Y 是变换后的数据。

在主成分分析中, $V_m = \sum_{i=1}^m \lambda_i / \sum_{i=1}^P \lambda_i$, 为主成分 Y_1, Y_2, \dots, Y_m 的累计贡献率, 它是特征提取算法中用来选择特征的主要依据。基于主成分分析的降维算法是光谱图像数据降维中最常用的方法之一。该方法用累计贡献率 V_m 大于某一值(如 99%)的前 m 个主成分代替原来的图像数据, 实现降维。

当数据之间有很强的相关性时, 使用主成分变换方法效果较好, 但当数据相关性较弱的时候, 主成分的变换不是很有效^[15-16]。因此, 应尽可能把基于主成分变换应用于相关性较强的数据上。

计算第 i 个光谱波段和第 j 个光谱波段的相关系数:

$$r_{ij} = \frac{(y_i - \bar{y}_i)^T (y_j - \bar{y}_j)}{|y_i - \bar{y}_i| \cdot |y_j - \bar{y}_j|}$$

式中 \bar{y}_i 为 y_i 的均值向量。

在计算任意 2 个波段光谱相关系数后, 得到高光谱数据的所有 N 个波段的相关系数矩阵为:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1N} \\ r_{21} & r_{22} & \dots & r_{2N} \\ \vdots & \vdots & & \vdots \\ r_{N1} & r_{N2} & \dots & r_{NN} \end{bmatrix}$$

根据相关系数矩阵, 将整个高光谱图像空间分解为多个高度相关的子空间, 然后分别对每个子空间进行主成分变换, 并选择适当数量的主成分因子作为后续建模的输入向量^[17]。由于分段主成分变换是分段进行, 可避免局部重要的波段被忽略。

1.5 ELM 和 RF 分类器

人工神经网络作为一种典型的分类建模方法, 已经得到了深入而广泛的研究。其中单隐层前馈神经网络由于其强大的学习能力而在很多领域得到了广泛的应用。然而, 反向传播算法(back-propagation, BP)作为单隐层前馈网络的常用学习算法有着明显的缺陷, 即基于梯度下降的迭代方法收敛速度慢、容易陷入局部极小等。为了克服这些问题,

极限学习机理论由 Huang 等^[18]提出。不同于传统的神经网络, ELM 中所有的输入层与隐含层间的连接权值及隐含层神经元的阈值均为随机产生, 不需要烦琐的迭代过程; 其输出权值则通过求解矩阵的广义逆得到。因此, 相较于传统的学习方法, 在保证学习性能的基础上, ELM 具有参数选择容易, 学习速度快等优点。

随机森林是一种基于分类树的集成学习算法^[19]。20 世纪 80 年代 Breiman 等发明了分类和回归树(classification and regression tree, CART)的算法, 通过反复二分数据进行分类或回归, 计算量大大降低。2001 年 Breiman 和 Culter 借鉴贝尔实验室的 Ho 所提出的随机决策森林的方法, 把分类树组合成随机森林。随机森林是一种统计学习理论, 它是利用 bootstrap 重抽样方法从原始样本中抽取多个样本, 对每个 bootstrap 样本进行决策树建模, 然后组合多棵决策树的预测, 通过投票得出最终预测结果。随机森林在训练过程中的每次 bootstrap 抽样, 将有约三分之一的数据未被抽中, 这部分数据被称为袋外(out-of-bag, OOB)数据。随机森林利用这部分数据进行内部的误差估计, 产生 OOB 误差。Breiman 通过实验证明, OOB 误差是无偏估计, 近似于交叉验证得到的误差。随机森林分类器的优点在于具有高的预测精度、泛化误差可控、收敛速度快以及调节参数少, 可有效避免“过拟合”现象发生, 适用于各种数据集的运算, 尤其适用于超高维特征向量空间。

2 结果与分析

2.1 大豆的原始光谱曲线和预处理后光谱曲线

在波长 400 ~ 1 000 nm 范围内, 应用高光谱图像采集系统可得到 203 个波段下 660 个大豆样本的高光谱数据(图 2)。每种大豆中所含的脂肪、蛋白质和大豆异黄酮等成分的含量都不相同, 这些都能反映在高光谱曲线上。由图 2 可以看出, 虽然曲线密集, 但仍然有一定的差异性, 为使这种差异性表现较为明显, 分别求出每一个品种的大豆样本的平均光谱值, 图 3 为不同品种大豆样本平均光谱曲线图, 可以看出 6 种大豆样本具有一定的可分性, 这为后面试验中的建模分类提供了一定的前提基础。

2.2 光谱数据波段的划分

由各个波段的相关系数矩阵可得到高光谱图像的相关矩阵图, 以 3 种大豆样本东农 42、东农 51 和东农 61 为例, 图 4 给出了它们的相关矩阵图, 灰度值范围为 0 ~ 1, 亮度越高表示波段之间的相关性越强, 从图中可以看出, 高光谱图像相邻波段之间

的相关性较强,而相隔较远的波段之间相关性较弱。尽管每个品种的大豆相关矩阵图略有差异,但在对角线上都有明显的分块结构。按照对角线上的分块结构,以东农 51 大豆为例将 203 个波段分成 3 个子段,划分结果如图 5 所示。表 1 给出了每一个分段的具体信息。图 5 中同时也给出了各个分块的平均相关系数,位于对角线上的 3 个子块的平均相关系数都在 0.92 以上,大于其他非对角线上的各个子块。每个分段内各个波段之间的相关性是很强的,因此在每个分段内进行主成分变换可以得到较好的效果。

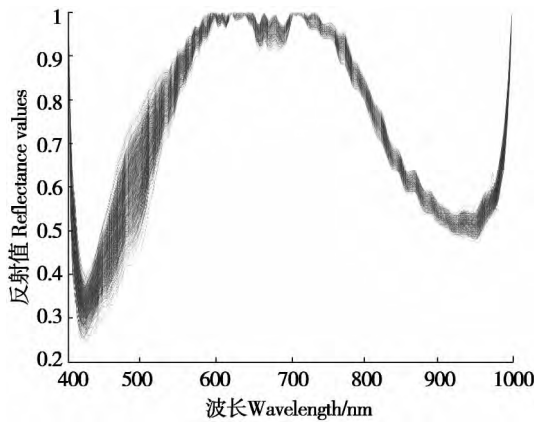


图 2 所有大豆样本的原始光谱曲线
Fig. 2 Original spectrum curves of all the soybean samples

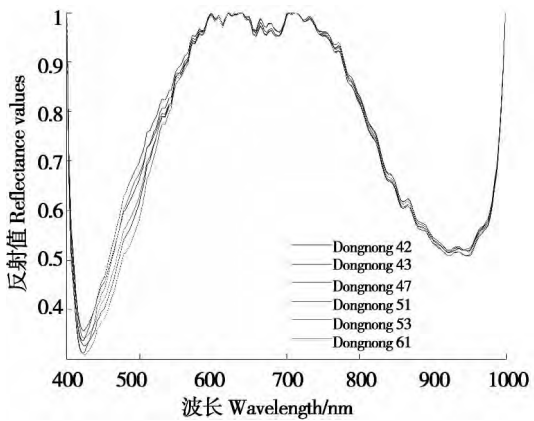
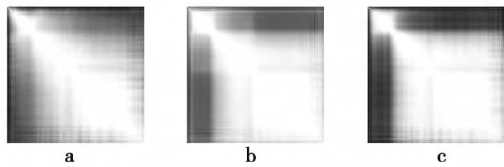


图 3 6 种大豆的平均光谱曲线
Fig. 3 Average spectrum curves of six varieties soybean



a: 东农 42; b: 东农 51; c: 东农 61。
a: Dongnong 42; b: Dongnong 51; b: Dongnong 61.
图 4 不同品种大豆的相关矩阵图
Fig. 4 Correlation matrices of different varieties soybean

表 1 分段信息			
Table 1	Information for segmenting		
	第一分段 Subgroup1	第二分段 Subgroup2	第三分段 Subgroup3
波段范围 Waveband range	Band 1 ~ 37	Band 38 ~ 97	Band 98 ~ 203
光谱范围 Spectral range/nm	400.9 ~ 507.6	510.6 ~ 685.4	688.4 ~ 999.5
波段数量 Waveband number	37	60	106

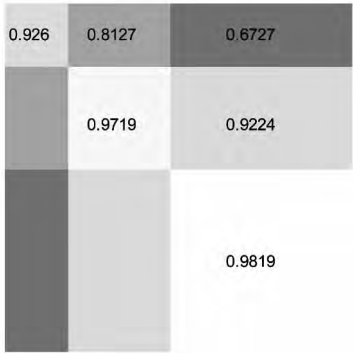


图 5 子块的平均相关系数
Fig. 5 Average correlation coefficient of sub-segments

2.3 高光谱数据的主成分分析

为了比较不同波段的光谱特征对大豆品种分类的影响,分别选择 400.9 ~ 507.6 nm, 510.6 ~ 685.4 nm, 688.4 ~ 999.5 nm 和全波段光谱特征作为大豆品种的分类特征,在每个波段范围内对光谱数据做主成分变换,可得到各主成分累计贡献率(表 2)。累计贡献率反映了前 n 个主成分对原数据信息的表征能力^[20]。除第三分段外,其余各个分段前 10 个主成分的累计贡献率都达到 98% 以上,说明这 10 个变量可以很好地表征原始数据信息。

2.4 基于不同主成分因子数建立的分类模型

在使用 PCA 变换进行波段降维时,一个非常重要的问题是 PCA 变换具体变换到多少维的空间,也就是有多少个主成分因子数,如果变换后的维数过低,会影响后期的建模效果,如果维数过大,则会影响后期建模的效率。通过改变主成分因子数的大小(从 1 ~ 20),观察主成分因子数对样本识别率的影响。

本文分别建立了 ELM 和 RF 分类模型。ELM 建模分析中先分别选取“Sigmoidal”“Sine”和“Hardlim”函数作为 ELM 模型隐含层激励函数。经多次反复试验,当隐含层激励函数选为“Sine”函数时,隐含层神经元个数设定为 200 时,模型品种识别性能较稳定,且具有较高的识别准确率。使用 RF 分类

器建模时,随机森林中决策树的个数设为 500。

表 2 各主成分累计贡献率

Table 2 Cumulative percent of principal components (PCs) (%)

主成分 因子 PCs	第一分段 Subgroup1	第二分段 Subgroup2	第三分段 Subgroup3	全波段 Entire Bands
PC1	84.9240	72.7123	80.4561	77.0251
PC2	94.6086	83.2752	88.7219	85.3767
PC3	97.3836	90.2631	92.5020	90.8478
PC4	98.1252	94.9163	94.0248	93.3439
PC5	98.5566	96.6666	94.8454	95.2715
PC6	98.7886	97.7123	95.5049	96.4445
PC7	98.9878	98.3624	96.0129	97.0450
PC8	99.1475	98.7108	96.4027	97.4305
PC9	99.2789	98.9999	96.7298	97.6910
PC10	99.3776	99.1512	97.0059	98.0198
PC11	99.4535	99.2275	97.2412	98.1157
PC12	99.5172	99.2896	97.4325	98.2685
PC13	99.5658	99.3407	97.6074	98.3923
PC14	99.6131	99.3890	97.7715	98.5032
PC15	99.6549	99.4307	97.9221	98.5981
PC16	99.6936	99.4690	98.0530	98.6737
PC17	99.7291	99.5012	98.1797	98.7412
PC18	99.7611	99.5290	98.2989	98.8067
PC19	99.7888	99.5527	98.4023	98.8681
PC20	99.8129	99.5755	98.4890	98.9249

把全部 660 个大豆样本分为训练集和测试集,从每个品种大豆样本中各随机选取 90 个样本共计 540 个样本用作建模,每个品种中选取 20 个样本共 120 个样本用作测试。ELM 和 RF 模型对于各个分段以及全波段的测试集样本的识别准确率如图 6 和 7 所示。从图中可以看出,无论是使用 ELM 分类模型,还是 RF 分类模型,当主成分因子数多于 7 个时,第二分段的识别准确率大于全波段的识别准确率,第三分段的识别效果略差,第一分段的效果最差。当选择相同数量的主成分因子数时,ELM 分类模型的效果优于 RF 分类模型。随着主成分因子数的增加,从整体上看大豆品种识别的准确率曲线是上升的,但并不是单调上升,而是有所波动,例如在第二个分段的准确率曲线中,用 14 个主成分因子进行 ELM 建模的准确率是 94.166 7%,而用 15 个主成分因子进行建模的准确率反而降低为 92.5%,这个现象在其他分段以及全波段曲线上也都有所体

现。这说明,在建模中使用更多的主成分因子不一定能得到更好的分类性能。我们要根据分类准确率选择合适数量的主成分因子数。对于 ELM 分类模型,当主成分因子数为 16 个时,对测试集 120 个样本采用第二分段做主成分分析,品种识别率达到 95%,而采用全波段做主成分分析,品种识别率只有 90%,第一分段和第三分段识别率只有 73.333 3% 和 76.666 7%,识别率较低。对第二分段中的 60 个波段进行主成分分析能够达到比对全部 203 个波段做主成分分析更好的效果,但是主成分分析时的输入变量减少了 70%,从而减少了计算量,节约了计算时间,有助于实现大豆品种的在线分类识别。

从图 6 中可以看出,当主成分因子数为 10 个时,采用第二分段做主成分分析的 ELM 模型的识别准确率已经能达到 91% 以上,我们以这种情况为例,详细的说明训练集和测试集中各种样本的分类识别情况(表 3)。从试验结果可以看出,对第二分段进行主成分分析,在测试集中东农 42、东农 51 和东农 47 各有一个样本被误分类;东农 61 和东农 43 有两个样本被误分类;东农 53 有 3 个样本被误分类。

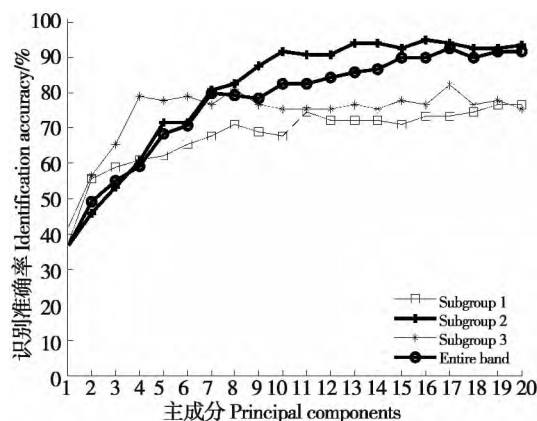


图 6 ELM 模型识别准确率

Fig. 6 Identification accuracy based on ELM

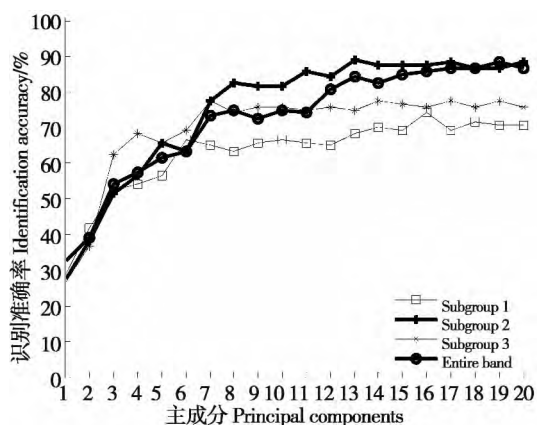


图 7 RF 模型识别准确率

Fig. 7 Identification accuracy based on RF

3 结论与讨论

本文提出了一种基于高光谱技术和分段 PCA 变换的大豆品种识别方法。该方法运用波段间的相关性矩阵把整个波段分为高度相关的 3 个子段, 在每个子段上分别进行 PCA 变换, 再应用极限学习机和随机森林建立品种识别模型, 本算法有效的减少了输入变量数, 可以简化模型, 降低计算复杂度, 同时避免了过拟合问题。对第二分段(510. 6 ~ 685. 4 nm) 的高光谱信息进行 PCA 变换, 在训练集和测试集中识别效果都较理想, 优于在全部波段上进行 PCA 变换。故可用第二分段中 60 个波段的 PCA 代替全部 203 个波段 PCA 变换进行大豆品种的识别。

本研究首次实现了应用高光谱技术进行大豆品种快速无损识别, 为提高大豆种子的质量和纯度, 加工业中功能型大豆的选择, 健康大豆和真菌感染大豆的分类提供理论基础, 具有一定的理论和实际应用价值。以本课题作为基础, 建立农产品高光谱无损检测系统, 并应用于生产实践中, 大大地提高目标对象的检测准确率, 满足农畜产品综合品质的在线、快速、无损检测。

在今后的研究中将考虑研究更多的大豆品种, 建立更为稳健以及适用范围更广的品种判别模型。同时也要探索一些其他分类方法, 如支持向量机^[21]和仿生模式识别^[22]等, 寻找更简洁、稳定、通用的分类方法。

表 3 不同分段和全波段下训练集和测试集样本识别结果

Table 3 Identification results for training set and testing set using three different segments and the entire bands								
样本 Samples	训练集 Training set				测试集 Validation set			
	第一分段 Subgroup1	第二分段 Subgroup2	第三分段 Subgroup3	全波段 Entire bands	第一分段 Subgroup1	第二分段 Subgroup2	第三分段 Subgroup3	全波段 Entire bands
东农 42 Dongnong 42	90	90	90	90	14	19	15	17
东农 51 Dongnong 51	90	90	90	90	13	19	17	17
东农 61 Dongnong 61	90	90	90	90	15	18	16	16
东农 43 Dongnong 43	90	90	89	90	13	18	14	16
东农 47 Dongnong 47	90	90	90	90	12	19	14	17
东农 53 Dongnong 53	89	90	89	90	13	17	15	16
平均识别率 Identification accuracy /%	99. 81	100	99. 63	100	66. 67	91. 67	75. 83	82. 50

参考文献

[1] 李文霞,李柏云,薛红,等. 黑龙江省不同生态区大豆品种育种性状的主成分分析[J]. 大豆科学, 2013, 32(6) : 731-734. (Li W X ,Li B Y ,Xue H ,et al. Principal components analysis of breeding traits in various ecological regions in Heilongjiang province[J]. Soybean Science, 2013, 32(6) : 731-734.)

[2] 洪庆红,李丹婷,郝朝运. 应用 FTIR 直接测定法鉴定大豆的品种[J]. 光谱学与光谱分析, 2005, 25(8) : 1246-1249. (Hong Q H ,Li D T ,Hao C Y. Identification of soybean varieties by direct determination of FTIR spectrum[J]. Spectroscopy and Spectral Analysis, 2005, 25(8) : 1246-1249.)

[3] 朱大洲,王坤,周光华,等. 单粒大豆的近红外光谱特征及品种鉴别研究[J]. 光谱学与光谱分析, 2010, 30 (12) : 3217-3221. (Zhu D Z ,Wang K ,Zhou G H ,et al. The NIR spectra based variety discrimination for single soybean seed [J]. Spectroscopy and Spectral Analysis, 2010, 30(12) : 3217-3221.)

[4] 杨冬风,朱洪德. 基于近红外透射光谱分析和 BP 神经网络的大豆品种识别[J]. 大豆科学, 2013, 32(2) : 249-253. (Yang D F ,Zhu H D. Recognition of soybean varieties based on near infrared transmittance spectroscopy and BP neural network[J]. Soybean Science, 2013, 32(2) : 249-253.)

[5] 柴玉华,侯升飞,彭长禄. 基于高光谱图像技术的大豆分级识别方法研究[J]. 东北农业大学学报, 2014, 45(4) : 107-112. (Chai Y H ,Hou S F ,Peng C L. Identification of different soybean grades based on hyperspectral imagery [J]. Journal of Northeast Agricultural University, 2014, 45(4) : 107-112.)

[6] Zhang X L ,Liu F ,He Y ,et al. Application of hyperspectral imaging and chemometric calibrations for variety discrimination of

- maize seeds[J]. *Sensors* 2012 ,12(12) : 17234-17246.
- [7] 鲍一丹,陈纳,何勇,等. 近红外高光谱成像技术快速鉴别国产咖啡豆品种[J]. *光学精密工程*, 2015 ,23(2) : 349-355. (Bao Y D ,Chen N ,He Y ,et al. Rapid identification of coffee bean variety by near infrared hyperspectral imaging technology [J]. *Optical Precision Engineering* ,2015 ,23(2) : 349-355.)
- [8] 杨仁杰,刘蓉,杨延荣,等. 用二维相关近红外谱和多维主成分分析判别掺杂牛奶[J]. *光学精密工程*, 2014 ,22(9) : 2352-2358. (Yang R J ,Liu R ,Yang Y R ,et al. Classification of adulterated milk by two-dimensional correlation near-infrared spectroscopy and multi-way principal component analysis [J]. *Optical Precision Engineering* 2014 22(9) : 2352-2358.)
- [9] 王斌,薛建新,张淑娟. 基于高光谱成像技术的腐烂、病害梨枣检测[J]. *农业机械学报* 2013 44(S1) : 205-209. (Wang B ,Xue J X ,Zhang S J. Detection of decay and disease pear jujube based on hyperspectral imaging technology [J]. *Transactions of the Chinese Society for Agricultural Machinery* 2013 44(S1) : 205-209.)
- [10] Siedliska A ,Baranowski P ,Mazurek W. Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data[J]. *Computers and Electronics in Agriculture* ,2014 ,106: 66-74.
- [11] Nakariyakul S ,Casasent D P. Classification of internally damaged almond nuts using hyperspectral imagery[J]. *Journal of Food Engineering* 2011 ,103(1) : 62-67.
- [12] Sakla W ,Chan A ,Ji J ,et al. An SVDD-based algorithm for target detection in hyperspectral imagery [J]. *Geoscience and Remote Sensing Letters* ,IEEE 2011 8(2) : 384-388.
- [13] 严衍禄,陈斌,朱大洲. 近红外光谱分析的原理、技术与应用[M]. 北京: 中国轻工业出版社, 2013. (Yan Y L ,Chen B ,Zhu D Z. Near infrared spectroscopy-principles ,technologies and applications[M]. Beijing: China Light Industry Press 2013.)
- [14] Shahin M A ,Symons S J. Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis[J]. *Computers and Electronics in Agriculture* 2011 75(1) : 107-112.
- [15] Jia X ,Richards J A. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification[J]. *Geoscience and Remote Sensing* ,1999 37(1) : 538-542.
- [16] Du Q ,Zhu W ,Yang H ,et al. Segmented principal component analysis for parallel compression of hyperspectral imagery [J]. *Geoscience and Remote Sensing Letters* , IEEE ,2009 ,6 (4) : 713-717.
- [17] Borges J S ,Marcal A R S ,Dias J M B. Evaluation of feature extraction and reduction methods for hyperspectral images[C]//Proceedings of the 26th EARSeL Symposium-New Developments and Challenges in Remote Sensing 2007: 255-264.
- [18] Huang G B ,Zhu Q Y ,Siew C K. Extreme learning machine: Theory and applications [J]. *Neurocomputing* ,2006 ,70(1) : 489-501.
- [19] Breiman L. Random forests [J]. *Machine Learning* ,2001 ,45(1) : 5-32.
- [20] Vidal M ,Amigo J M. Pre-processing of hyperspectral images. Essential steps before image analysis[J]. *Chemometrics and Intelligent Laboratory Systems* ,2012 ,117: 138-148.
- [21] 朱启兵,冯朝丽,黄敏,等. 基于高光谱图像技术和SVDD的玉米种子识别[J]. *光谱学与光谱分析*, 2013 (2) : 517-521. (Zhu Q B ,Feng Z L ,Huang M ,et al. Maize seed identification using hyperspectral imaging and SVDD algorithm [J]. *Spectroscopy and Spectral Analysis* 2013 (2) : 517-521.)
- [22] 郭文川,王铭海,谷静思,等. 近红外光谱结合极限学习机识别贮藏期的损伤猕猴桃[J]. *光学精密工程* ,2013 21(10) : 2720-2727. (Guo W C ,Wang M H ,Gu J S ,et al. Identification of bruised kiwifruits during storage by near infrared spectroscopy and extreme learning machine [J]. *Optical Precision Engineering* ,2013 21(10) : 2720-2727.)

欢迎订阅 2017 年《大豆科学》

全国中文核心期刊

中国科技核心期刊

中国农业核心期刊

RCCSE 中国学术核心期刊

中国科学引文数据库来源期刊

CAB International(英) 收录期刊

CA 化学文摘(美) 收录期刊

JST 日本科学技术振兴机构数据库(日) 收录期刊

《大豆科学》主要面向从事大豆科学研究的科技工作者、大专院校师生、各级农业技术推广部门的技术人员及科技种田的农民。主要刊登有关大豆遗传育种、品种资源、生理生态、耕作栽培、植物保护、营养肥料、生物技术、食品加工、药用功能及工业用途等方面的学术论文、研究报告、研究简报、国内外研究述评、学术活动简讯和新品种介绍等。

《大豆科学》为双月刊,逢单月 20 日出刊,刊号: ISSN 1000-9841 CN 23-1227/S,广告经营许可证号: 2301030000004,邮发代号: 14-95。国内每期定价: 20.00 元,全年 120.00 元,国外每期定价: 20.00 美元(含邮资),全年 120.00 美元,国外代号: Q5587。全国各地邮局均可订阅,也可向编辑部直接订购。

地址: 哈尔滨市南岗区学府路 368 号《大豆科学》编辑部(邮编: 150086) 电话: 0451-86668735

网址: www.haasep.cn E-mail: ddkxbjb@126.com