

高光谱成像的机采籽棉杂质分类检测

常金强, 张若宇*, 庞宇杰, 张梦芸, 扎 亚

石河子大学机械电气工程学院/农业农村部西北农业装备重点实验室, 新疆 石河子 832003

摘 要 机采籽棉杂质分类检测为调整棉花清理机械加工参数和工序提供参考依据, 对提升皮棉品质具有重要意义。但由于籽棉棉层分布不均匀, 使得图像检测难度增大, 使用传统的检测方法无法有效检测各类杂质。采用高光谱成像方法对机采籽棉中的棉叶、棉枝、地膜和铃壳(内外)五种杂质进行分类判别检测。首先采集 120 个机采籽棉样本的高光谱图像, 选取感兴趣区域获取平均光谱曲线。发现由于物质构成的差异, 不同杂质体现出不同的吸收和反射特性, 不同种类物质之间的光谱差异大于同类物质。对提取的平均光谱曲线进行主成分分析(PCA), 结果显示棉花、残膜和铃壳外与其他三类相比, 有较好的聚集性和可分性, 但是棉叶、铃壳内和棉枝三类相互叠加在一起, 空间分布存在严重交叉重叠。以提取的平均光谱曲线为训练样本, 选择线性判别分析(LDA)、支持向量机(SVM)和神经网络(ANN)三种分类判别算法, 对算法参数进行寻优, 并建立机采籽棉杂质分类判别模型。其中, 经过 LDA 模型降维后的样本空间较 PCA 表现出了更好的聚集性和可分性, 采用正则化防止过拟合, 得到训练集准确率为 86.4%, 测试集准确率为 86.2%; SVM 模型的参数寻优结果为 $C=10^5$, $g=0.1$, 其训练集准确率为 83.42%, 测试集准确率为 83.40%; ANN 模型参数寻优得到隐含层数和神经元个数分别为 2 和 17, 训练集准确率为 82.9%, 测试集准确率为 81.8%。对三种模型分类效果和检测用时进行比较, LDA 模型结果最优。通过对高光谱图像进行像素等级分类判别, 结果显示棉花识别效果较好, 植物性杂质都被有效检测, 但是地膜和棉花存在误识别, 分类效果与杂质光谱的分类判别模型结果一致。因此, 采用高光谱成像技术可以快速、无损的检测和识别籽棉杂质, 为棉花加工装备提供反馈参数, 对棉花加工机械化和智能化有重要意义。

关键词 机采籽棉; 杂质检测; 高光谱成像; 分类判别

中图分类号: S24 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)11-3552-07

引 言

近年来棉花全程机械化生产比例增加, 机采籽棉需要在后续加工过程中进行多道清理工艺, 但是清理机械会对棉花纤维造成损失, 降低加工所得皮棉的品质, 影响最终产品价格和经济效益。因此对棉花杂质进行检测, 并将杂质进行分类判别, 为调整棉花清理机械加工参数和工序提供参考依据, 对提升皮棉品质具有重要实际生产价值和意义。

由于皮棉中异性纤维含量对价格影响较大, 国内的研究主要集中在异性纤维检测^[1-2]。张志峰等^[3]提出了一种基于改进的自适应迭代阈值法皮棉疵点快速检测方法; 张林等^[4]采用 LED 与线激光的双光源一次成像方法, 可以检测出各种颜色

的异性纤维; 张成梁等^[5-6]、王昊鹏等^[7]提取机采籽棉可见光图像中杂质的颜色、形状和纹理特征, 对各类植物杂质进行分类检测; 倪超等^[8]采用深度学习方法对短波近红外高光谱图像中的地膜进行检测。

国外的研究主要集中在植物性杂质的检测, Wang 等^[9]采用基于自动视觉检测系统的伪异性纤维检测方法, 提高了棉花中异性纤维的分类精度。Fortier 等^[10]建立棉花中植物杂质的近红外光谱库, 进行杂质光谱分类识别。Li 等^[11-15]基于高光谱成像技术, 采用反射、透射和荧光等成像方式, 应用降维、特征波段选择、分类判别算法等分析方法, 对皮棉中多种植物和异纤杂质进行检测。

上述研究对象主要是皮棉, 由于皮棉经过杂质清理和轧花去籽处理, 杂质含量小, 棉层均匀易于图像中杂质的检

收稿日期: 2020-08-28, 修订日期: 2021-01-11

基金项目: 国家重点研发计划项目(2018YFD0700403), 兵团重点领域科技攻关计划项目(2020AB006), 八师石河子市重点研发计划项目(2019GY13)和十二师科技计划项目(SR2020026)资助

作者简介: 常金强, 1992 年生, 石河子大学机械电气工程学院硕士研究生 e-mail: changjinq@163.com

* 通讯作者 e-mail: ry248@163.com

测；而机采籽棉中不仅含有较多杂质，且棉籽导致棉层不均匀，使得图像检测难度增大，使用传统的检测方法无法有效检测各类杂质。

基于高光谱成像检测技术，根据棉花和各类杂质的光谱特征，针对机采籽棉中存在的植物和残膜杂质建立分类判别模型；并充分利用光谱图像的空间信息，实现对机采籽棉各类杂质的像素等级分类判别，为棉花加工设备提供快速信息反馈。

1 实验部分

1.1 样本的制备

共取籽棉棉 10 kg，其中籽棉取自棉花加工企业，地膜取自采收后的棉花地。将籽棉和杂质手动混合均匀，每个样本(30±0.5) g，使用电子天平称重(量程 1 000 g，分度值 0.01 g)，共 120 个籽棉样本。样本中检测的杂质有棉叶，棉枝，铃壳(内和外)和地膜共 5 种杂质，如图 1 所示。



图 1 机采籽棉和主要杂质

Fig. 1 Machine-harvested seed cotton and main impurities

1.2 高光谱成像系统和图像采集

高光谱图像采集系统如图 2 所示，由成像光谱仪 (ImSpectral V10E-QE, Finland)、CCD 相机 (C8484-05G, Hamamatsu Photonics, Japan)、镜头、光源(150 W 卤素灯, China)、电动位移平台 (PSA200-11-X, Zolix) 和电动位移平台控制器 (CS300-1A, Zolix)、暗箱、PC 计算机等组成；在 PC 上用 Spectral 软件进行图像采集软件控制。高光谱成像系统光谱范围为 360~1 000 nm，光谱分辨率为 2.7 nm，采集的图像有 256 个波段。

为保证视野足够，调节镜头和样本的间距为 25.5 cm；为矫正速度不匹配带来的空间畸变，使用一张打印有一个圆圈的 A4 纸调试平台的速度，转速设定为 940 pulses · s⁻¹；曝光时间为 3.5 ms。

将样本置于内部大小为 15 cm×20 cm×3 cm 的样本盒中，分布均匀，将样本盒固定于移动平台上进行图像采集。样本盒覆盖有黑色背景纸，有利于后期掩膜去除背景以及后续处理。

为减少光源光强分布不均匀导致的图像信息噪声影响，使用的高光谱成像系统在采集图像之前需要进行黑白校正。扫描聚四氟乙烯白板获得白校正图像；镜头拧上镜头盖并关闭光源采集黑校正图像，该图像包含有相机暗电流噪声信息。图像采集后用软件 SpecView(V2.9.2.7)按式(1)进行校

正

$$I_a = \frac{I - I_b}{I_w - I_b} \quad (1)$$

其中： I 为原始图像， I_b 为黑校正图像， I_w 为白校正图像， I_a 为获取校正后的图像。

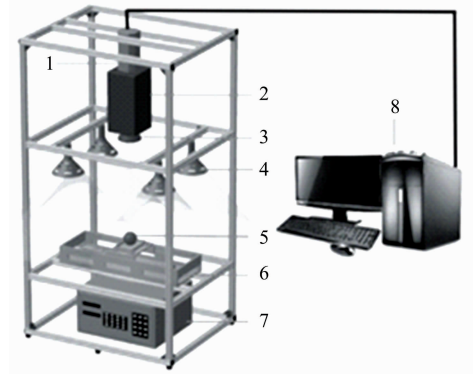


图 2 高光谱图像采集系统

1: CCD 相机; 2: 光谱仪; 3: 调焦镜头; 4: 卤素灯;
5: 样品; 6: 位移台; 7: 控制器; 8: 计算机

Fig. 2 Hyperspectral imaging system

1: CCD camera; 2: Spectrograph; 3: Lens; 4: Halogen lamps;
5: Sample; 6: Stage; 7: Controller; 8: Computer

1.3 机采籽棉数据分析和杂质多分类模型

使用 PCA(principal component analysis, PCA)对平均光谱数据进行分析，将成百个相互高度相关波段数据降维至少数个新的主成分变量上，用来代替原来数据的大部分信息，并通过绘制分布散点图体现原光谱数据的分类识别可行性。

采用 LDA, SVM 和 ANN 三种有监督的分类判别分析方法建立机采籽棉杂质多分类判别模型。模型训练的过程为：首先将提取的平均光谱数据按照 7:3 的比例，随机划分为训练集和测试集；然后根据不同模型的参数特点和数据特性，使用训练集采用 5 折交叉验证，确定最佳的模型参数，并使用测试集对模型结果进行评估。

2 结果与讨论

2.1 高光谱图像光谱曲线数据提取与分析

2.1.1 平均光谱曲线提取与变化规律

经过黑白校正后的图像，在可见至近红外波段上，共有 256 个波段。意味着在空间域上每个像素具有 256 个特征，这些特征组成该像素对应的光谱曲线。因高光谱图像中存在噪声，单一像素对应的光谱曲线可能在噪声的影响下，表现出较大的变化。因为光谱成像仪的特性，高光谱图像在首尾的波段图像噪声较大，有用信息较少，所以将这些波段剔除，即去除 395 nm 以前和 970 nm 以后的光谱图像波段，将 395~970 nm 区间共 226 个光谱波段的数据作为后续分析数据。

从每幅图像中提取 10 条平均光谱曲线，共 1 200 条光谱曲线，其中棉叶、残膜、铃壳外、铃壳内、棉枝和棉花分别为 457, 173, 88, 193, 63 和 226 条。绘制机采籽棉中具有代表

性的棉花和各类杂质的平均光谱曲线,如图 3 所示:各类物质在 430 nm 处附近反射率均为最小,吸收最强;棉花的反射率较其他物质在大部分波段范围高;残膜整体上和棉花变化趋势一致,但是数值比棉花低,验证了从图像上检测残膜的难度较大;铃壳内的反射率在 750 nm 前低于棉花和残膜,但是在 750 nm 后超过了棉花和残膜;棉叶、棉枝和铃壳外在趋势和数值上都比较相似,但是棉叶在 680 nm 处出现了吸收峰,此现象对应了叶绿素的吸收波段。从 630 nm 开始到近红外波段范围内,铃壳外的反射率比棉叶和棉枝都高。

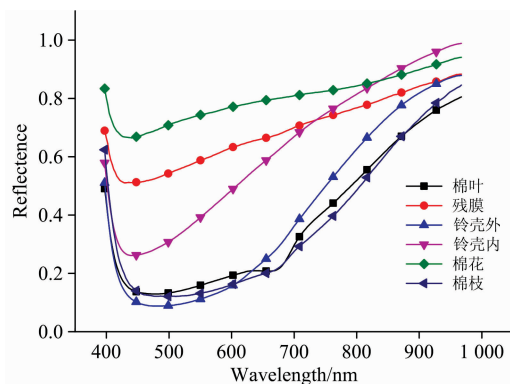


图 3 机采棉和各类杂质平均光谱曲线

Fig. 3 Mean spectra of machine-harvested cotton and impurities

综上所述,虽然棉花和各类杂质的光谱曲线趋势相同,但还是体现出不同的吸收和反射特性。不同种类物质(棉花、化学纤维和植物)之间的差异大于同类物质之间的光谱差异,同种物质之间的差异不能通过单个波段进行判别,所以需要进行数据分析和建模。

2.1.2 机采籽棉光谱曲线 PCA 分析

对提取的平均光谱曲线进行 PCA 变换,如图 4 所示,前 2 个主成分的累计贡献率达到了 97.2%,前 6 个主成分的累计贡献率达到了 99.9%,能够代表原始光谱数据的大部分信息。PCA 前两个主成分的散点图如图 5 所示,6 类物质光谱

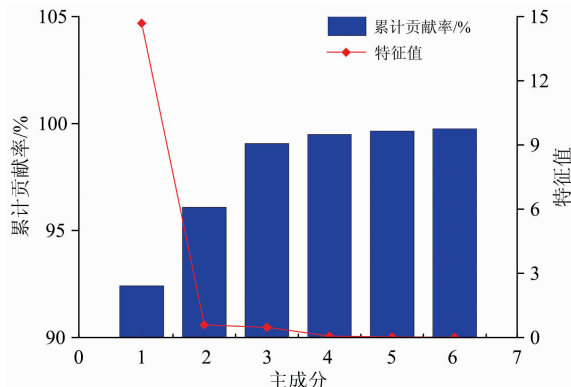


图 4 前 6 个主成分的特征值和累计贡献率

Fig. 4 Eigenvalues and cumulative contribution rates of the first 6 principal components

变换后的新变量分布于整个空间中。由图可知,棉花、残膜和铃壳外与其他三类相比,有较好的聚集性和可分性,但是由于棉叶、铃壳内和棉枝三类的物质组成(纤维素和木质素)相似性较高,光谱特征相似,导致相互叠加在一起,空间分布存在严重交叉,无法有效区分类别。由于 PCA 为无监督降维方法,无法有效利用分类信息,因此需要使用有监督的数据建模方法,对光谱分类数据进行学习拟合,实现对杂质类别的准确识别。

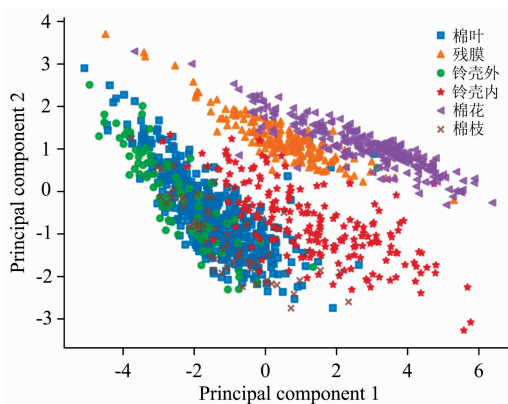


图 5 前 2 个主成分分类散点图

Fig. 5 Scatter clusters of the first 2 principal components

2.2 机采籽棉杂质光谱多分类模型

2.2.1 线性判别分析(LDA)模型

线性判别分析(linear discriminant analysis, LDA)是将原始数据投影到更低的维度上,减少特征之间的线性相关性导致的特征冗余问题。通过 LDA 进行降维,可以达到提升分类准确率的目的。

与 PCA 中的分布相比,图 6(a)中棉花、残膜和铃壳外有更好的聚集性和可分性,表明有监督的 LDA 模型降维方法变换后的数据具有更好的可分性;但是棉叶、铃壳内和棉枝这三类还是相互叠加在一起,空间分布存在严重交叉,无法有效区分类别。因此针对该三类重新进行了 LDA 降维,见图 6(b)中的棉叶、铃壳内和棉枝表现出了较高的可分性,验证了 LDA 模型在机采籽棉多分类上的可行性。

因 LDA 易出现过拟合,因此在 LDA 模型构件中采用正则化防止过拟合,建立分类模型,得到训练集准确率为 86.4%,测试集准确率为 86.2%,其差值较小,未出现过拟合现象。

2.2.2 支持向量机(SVM)模型

支持向量机(support vector machine, SVM)广泛应用于建立分类判别模型。在 SVM 分类模型构建中采用 RBF 径向基函数构建了分类模型,对 $\gamma(g)$ 和 $\text{cost}(C)$ 两个参数进行寻优,将 $\text{Lg}(g)$ 和 $-\text{Lg}(c)$ 参数区间设置为 $[0, 10]$ 。由图 7 可知,在 $C=10^5$ 、 $\gamma=0.1$ 时,交叉验证集的准确率最高达到 95.19%。根据最优参数模型得出训练集准确率为 83.42%,测试集准确率为 83.40%,两者差值较小,未出现过拟合现象。

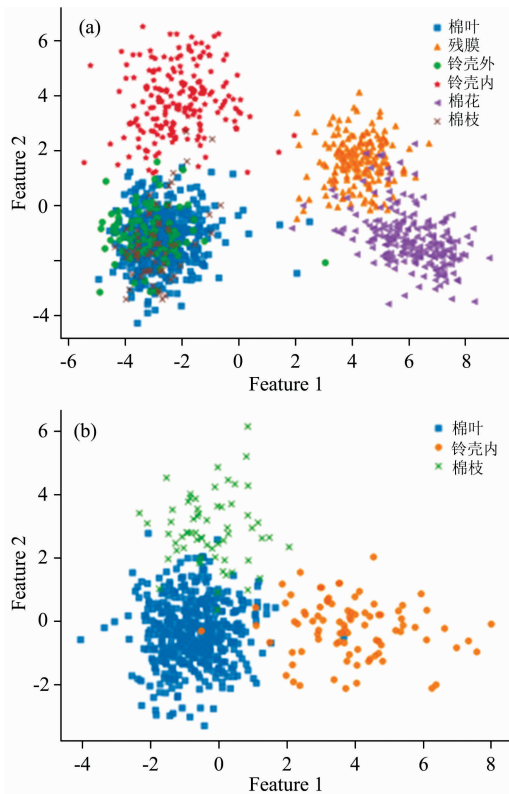


图 6 LDA 前两个特征类别散点图

(a): 全部 6 类物质; (b): 棉叶、铃壳内和棉枝

Fig. 6 Scattering clusters of the first 2 variables of LDA

(a): All 6 types of materials; (b): Leaf, bell shell inner, stem

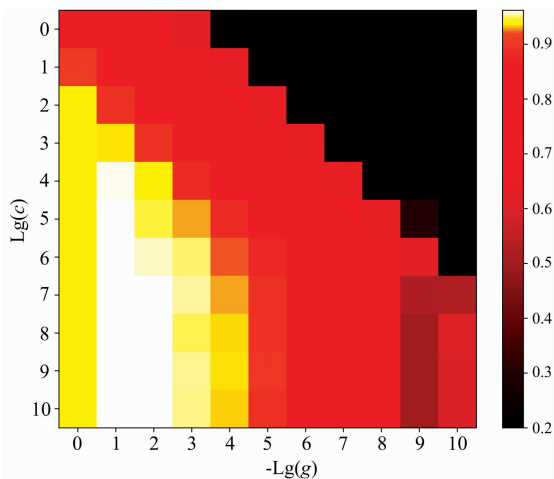


图 7 SVM 模型寻优结果

Fig. 7 Parameter optimization results of SVM model

2.2.3 人工神经网络(ANN)模型

人工神经网络 (artificial neural network, ANN) 是一种影响强、分类效果好的神经网络分类算法, 在解决非线性问题上具有较强能力。在 ANN 分类模型构建中, 设置隐含层数区间为 [1, 10], 隐含层神经元个数区间为 [1, 18], 激活函数选择 Relu 函数进行参数寻优。由图 8 可知, 在隐含层

数为 2, 隐含层神经元个数为 17, 交叉验证集的准确率达到最高为 73.92%。以寻优所得到的参数, 建立 ANN 分类模型并输出, 训练集准确率为 82.9%, 测试集准确率为 81.8%, 没有发生过拟合。

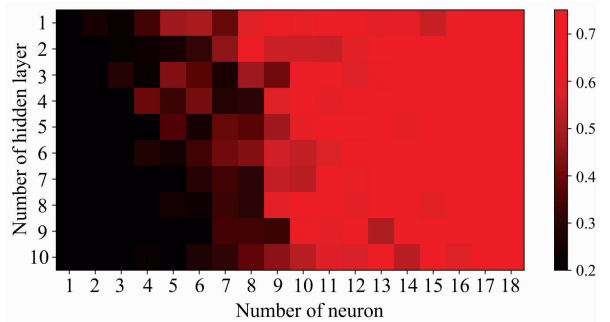


图 8 ANN 参数寻优结果

Fig. 8 Parameter optimization results of ANN model

2.3 模型效果比较

对上述的多分类模型准确率性能进行对比, 如表 1 所示, 结果显示 LDA 模型的准确率高于 SVM 模型和 ANN 模型, 训练集和预测集的准确率达到了 86.4% 和 86.2%。由于高光谱波段之间具有较高的相关性, 分类模型无法有效筛选信息, 会引起误差的产生。LDA 在分类前对光谱特征进行了降维, 减少了特征之间的相关性, 保留了大部分类间信息, 因此在多分类问题中, 相较于 SVM 和 ANN 具有更好的效果。

表 1 光谱曲线分类模型准确率和时间

算法	训练集/%	测试集/%	全部数据/%	检测用时/s
LDA	86.4	86.2	86.3	1.86
SVM	83.4	83.4	83.4	73.65
ANN	82.9	81.8	82.6	2.58

三个模型预测效果如图 9 所示。在 LDA 模型中, 地膜、铃壳(内和外)和棉花的准确率较高, 均高于 90%; 棉叶和棉枝的准确率较低, 分别为 59.84% 和 77.08%, 其中有 26.77% 的棉叶被识别为棉枝, 9.72% 的棉枝被识别为棉叶, 9.72% 的棉枝和 8.66% 的棉叶被识别为铃壳内; 与 LDA 模型相比较, SVM 模型和 ANN 模型的铃壳内准确率有所降低, 误差类别分布一致但较高。分析认为这些识别错误的原因主要是棉叶、棉枝和铃壳内的物质成分相似度高, 导致在波段范围内表现出光谱曲线相似的特点。

2.4 像素等级杂质分类判别

根据三种算法对 120 个高光谱图像进行检测分类, 并将运行时间进行平均, 得到每个模型检测高光谱图像所需运行时间。结果如表 1 所示, SVM, LDA 和 ANN 的运行时间分别为 73.65, 1.86 和 2.58 s, 综合比较, LDA 的分类准确率较高且运行时间少, 确定 LDA 分类模型为最优模型。

使用训练的 LDA 模型对高光谱图像进行像素等级分类, 分类效果如图 10 所示。可看出棉花识别效果较好; 部分棉叶

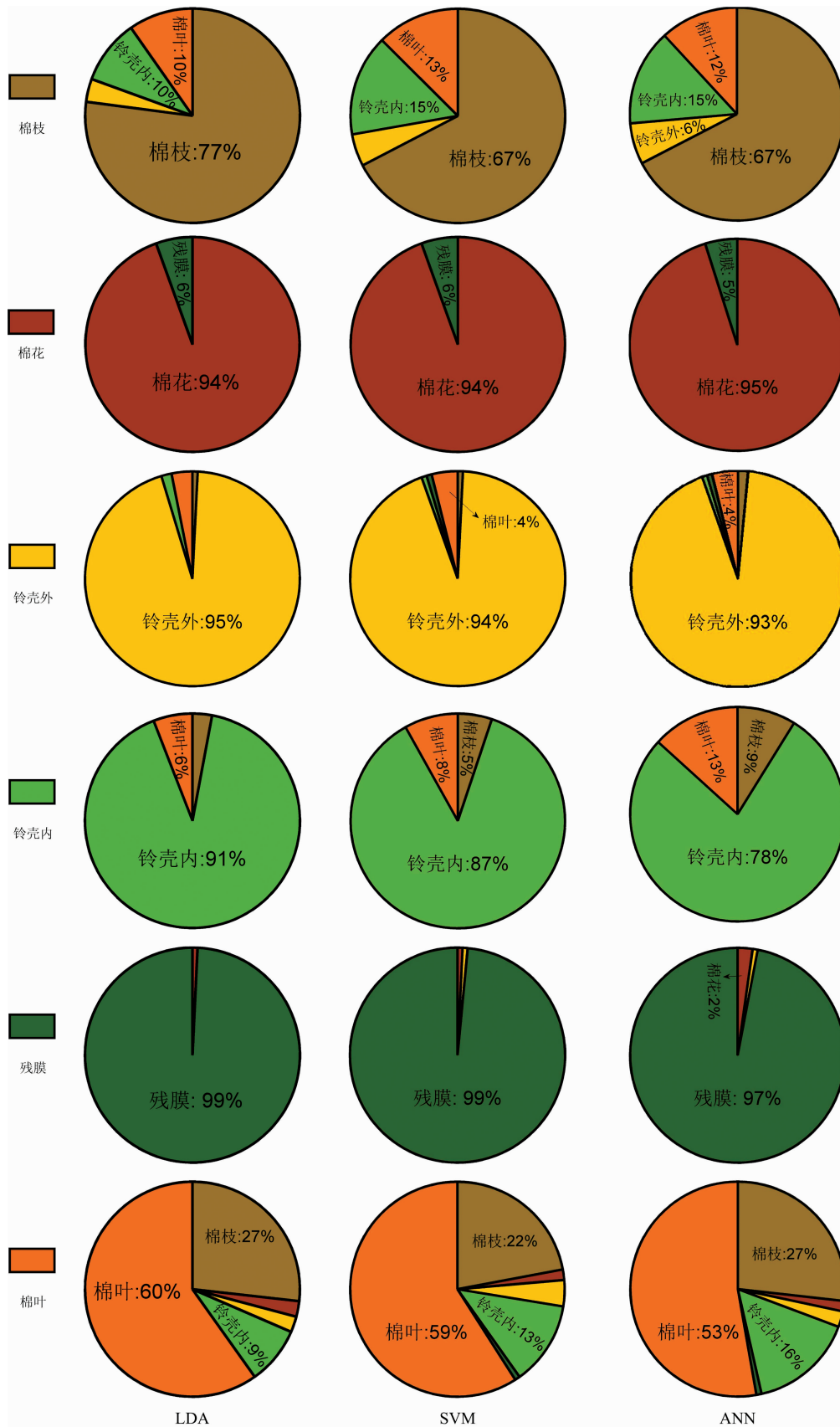


图 9 分类模型预测集效果
Fig. 9 Pie chart of prediction

和棉枝不能有效识别；地膜虽然被检测出来，但因地膜的光谱曲线在大部分波段上和棉花相似，亮度较棉花低，所以部

分棉花中表面不平导致的亮度较低的区域被识别为地膜。上述分类效果与杂质光谱的分类判别模型结果一致。

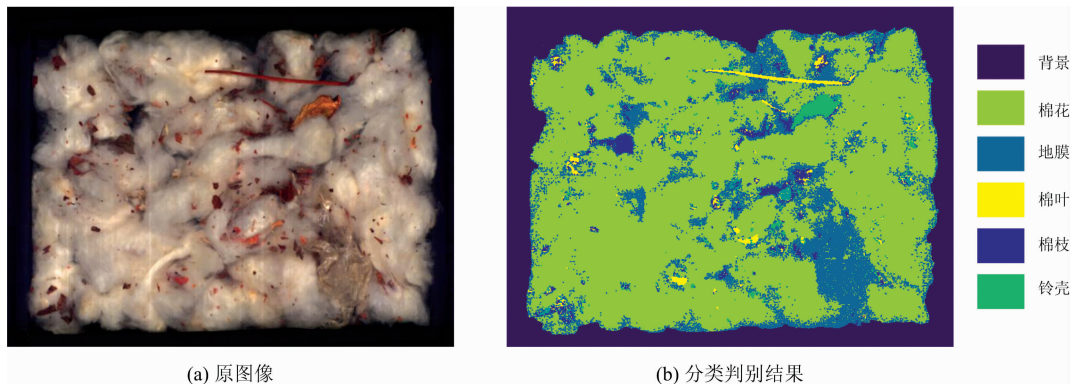


图 10 高光谱图像像素等级分类识别结果

(a): 原图像; (b): 分类判别结果

Fig. 10 Hyperspectral image classification results in pixel level

(a): Original image; (b): Classification result

3 结 论

(1) 通过参数优化，建立了三种机采籽棉杂质分类判别模型。其中 LDA 的分类准确率较高，训练集和测试集的准确率分别为 86.4% 和 86.2%。由于棉叶和棉枝的物质成分相似，光谱曲线相似，导致棉叶和棉枝杂质的分类准确率较低。

(2) 对于像素等级杂质检测，该方法能够识别大部分杂

质，检测效果明显。LDA 算法需要的时间约为 1.86 s，少于 ANN 的 2.58 s，且远少于 SVM 的 73.65 s，能够满足实际生产对于检测的需求，因此 LDA 为最佳模型。

(3) 在后续研究中可以基于该方法，增加样本数量，选择覆盖范围更大的波段和加入纹理特征，提升棉叶和棉枝的检测效果；并根据光谱图像数据分析提取特征波段，开发多光谱成像检测系统，实现更效率的机采籽棉杂质实时检测。

References

- [1] WANG Xin, LI Dao-liang, YANG Wen-zhu, et al(王 欣, 李道亮, 杨文柱, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2015, 46(8): 7.
- [2] Zhou W H, Xu S D, Liu C J, et al. Applied Spectroscopy Reviews, 2016, 51(4): 298.
- [3] ZHANG Zhi-feng, ZHAI Yu-sheng, GUO Ying-ying, et al(张志峰, 翟玉生, 郭莹莹, 等). Laser & Optoelectronics Progress(激光与光电子学进展), 2015, 52(3): 154.
- [4] ZHANG Lin, WEI Ping, WU Jian-bo, et al(张 林, 韦 平, 伍剑波, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(15): 289.
- [5] ZHANG Cheng-liang, LI Lei, DONG Quan-cheng, et al(张成梁, 李 蕾, 董全成, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(24): 189.
- [6] ZHANG Cheng-liang, LI Lei, DONG Quan-cheng, et al(张成梁, 李 蕾, 董全成, 等). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2016, 47(7): 28.
- [7] WANG Hao-peng, LI Hui(王昊鹏, 李 慧). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(3): 236.
- [8] Ni Chao, Li Zhenye, Zhang Xiong, et al. IEEE Access, 2020, 8: 93028.
- [9] Wang Xin, Yang Wenzhu, Li Zhenbo. Computers and Electrical Engineering, 2015, 46: 500.
- [10] Fortier C, Cintron M S, Rodgers J. AATCC Journal of Research, 2015, 2(6): 1.
- [11] Jiang Yu, Li Changying. PLOS ONE, 2015, 10(3): e0121969.
- [12] Jiang Y, Li C. Computers and Electronics in Agriculture, 2015, 119: 191.
- [13] Mustafic A, Li C. Textile Research Journal, 2015, 85(12): 1209.
- [14] Zhang R Y, Li C Y, Zhang M Y, et al. Computers and Electronics in Agriculture, 2016, 127: 260.
- [15] Zhang M Y, Li C Y, Yang F Z. Computers and Electronics in Agriculture, 2017, 139: 75.

Classification of Impurities in Machine-Harvested Seed Cotton Using Hyperspectral Imaging

CHANG Jin-qiang, ZHANG Ruo-yu*, PANG Yu-jie, ZHANG Meng-yun, ZHA Ya

College of Mechanical and Electrical Engineering, Shihezi University/Key Laboratory of Northwest Agricultural Equipment, Ministry of Agriculture, Shihezi 832003, China

Abstract The classification and detection of impurities in machine-harvested seed cotton provides a reference for adjusting cotton cleaning mechanical processing parameters and has important significance for improving lint quality. However, the uneven distribution of seed cotton makes image detection more difficult, and traditional detection methods cannot effectively detect various impurities. The hyperspectral imaging method was used to discriminate the five impurities (cotton leaf, cotton stem, plastic film, hull inner, and hull outer) in the machine-harvested seed cotton. The hyperspectral images of 120 machine-harvested seed cotton samples were collected, and the region of interest was selected to obtain the average spectral curve. Due to the difference in the composition of materials, various impurities showed different spectral absorption and reflection characteristics, and the spectral difference of the characteristics of different materials was greater than that of similar materials. Principal component analysis (PCA) of the extracted average spectral curve showed that cotton, plastic film and hull outer were better separable than the other three types. However, the spectral distributions of cotton leaf, hull inner, and cotton stem overlapped seriously. Based on the extracted average spectral curve as the training sample, three discrimination algorithms of linear discriminant analysis (LDA), support vector machine (SVM) and neural network (ANN) were used to optimize the algorithm parameters and finally established the impurity detection model. Among them, the sample space after dimensionality reduction of the LDA model shows better separability than PCA. Regularization was used to prevent overfitting in LDA, and the accuracy rate of the training set was 86.4%, and the accuracy of the test set was 86.2%. The parameter optimization result of the SVM model was $C=10^5$, $g=0.1$. The accuracy of the training set was 83.42%, and the accuracy of the test set was 83.40%. The parameter optimization result of the ANN model was that the number of hidden layers and neurons were 1 and 10, respectively. The accuracy rate of the training set was 82.9%, and the accuracy rate of the test set was 81.8%. Comparing the classification accuracy and detection time of the three models, the results of the LDA model were all the best. Through the pixel level discrimination of hyperspectral images, the results show that both cotton and botanical impurities were effectively detected. However, there were misidentifications between plastic film and cotton, which was consistent with the results of the impurity spectrum classification discrimination model. Therefore, hyperspectral imaging technology can detect and identify seed cotton impurities quickly and non-destructively and provide feedback parameters for cotton processing equipment, which is of great significance to the mechanization and intelligence of cotton processing.

Keywords Seed cotton; Impurities detection; Hyperspectral imaging; Classification

(Received Aug. 28, 2020; accepted Jan. 11, 2021)

* Corresponding author