

东亚飞蝗光谱特征波长筛选与龄期识别方法研究

李林 赵明明 王竹 彭帆 朱德海

(中国农业大学信息与电气工程学院, 北京 100083)

摘要: 利用高光谱图像采集系统在 400 ~ 1 000 nm 波段范围内采集东亚飞蝗成虫、5 龄、4 龄和 3 龄的前胸背甲光谱信息;每个龄期提取 15 像素 × 15 像素目标区域平均反射率信息作为样本信息;提出了一种基于 K 均值聚类 and 主成分分析(K-PCA)相结合特征波段提取方法,对比分析 K-PCA 和 SPA(投影连续变换)2 种特征波长提取方法,采用 Fisher 判别分析方法分别对 K-PCA 和 SPA 筛选的特征波长建立东亚飞蝗龄期识别判别模型,实验结果表明 K-PCA 筛选出的特征波长数少且正确识别率为 98.25%。K-PCA 筛选的特征波长为 468 nm、555 nm、635 nm、710 nm、729 nm、750 nm、786 nm 和 899 nm。本文提取的东亚飞蝗特征波长为东亚飞蝗的龄期识别奠定基础,进而对蝗灾的监测与预防提供了技术支持。

关键词: 东亚飞蝗; 高光谱图像; 特征波长; K 均值聚类; 主成分分析

中图分类号: S2 **文献标识码:** A **文章编号:** 1000-1298(2016)03-0249-05

Selection of Spectrum Feature Wavelength and Recognition of Different Ages of Manilensis

Li Lin Zhao Mingming Wang Zhu Peng Fan Zhu Dehai

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Manilensis is one of the major pests in China. A method for recognizing different ages of manilensis was presented based on K-means clustering and principal component analysis (PCA) with selected feature wavelength. The hyperspectral images in the range of 400 ~ 1 000 nm of manilensis back at different ages among adult, 5-age, 4-age and 3-age were collected and the average spectral information of target region on manilensis back with the size of 15 pixel × 15 pixel was extracted. A wavelength selecting method with combined PCA algorithm and K-means clustering (K-PCA) was proposed. The model for identifying manilensis ages was built by using Fisher algorithm and then compared with K-PCA algorithm and successive projections algorithm (SPA). The experiment results showed that the K-PCA algorithm needed fewer wavelengths but with the higher accuracy of 98.25%. The final feature wavelengths of K-PCA algorithm were 468 nm, 555 nm, 635 nm, 710 nm, 729 nm, 750 nm, 786 nm and 899 nm. The proposed method provides a certain technology support for manilensis monitoring and prevention.

Key words: manilensis; hyperspectral image; characteristic wavelength; K-means clustering; principal component analysis

引言

蝗虫灾害是严重影响农牧业生产安全和农村社

会稳定的重大生物灾害,与水灾、旱灾并称为三大自然灾害^[1]。目前我国蝗虫常年发生面积为 0.187 亿 hm²,分布在 20 个省的 1 100 多个县。其

收稿日期: 2015-10-08 修回日期: 2015-11-18

基金项目: 国家自然科学基金项目(31471762)

作者简介: 李林(1963—),女,教授,博士生导师,主要从事软件工程和软件自动化研究,E-mail: lilincau@126.com

通信作者: 朱德海(1962—),男,教授,博士生导师,主要从事 3S 技术在农业、国土资源领域应用研究,E-mail: zhudehai@cau.edu.cn

中飞蝗(东亚飞蝗、亚洲飞蝗)常年发生面积有200万hm²。飞蝗在成虫期具有爆发性、迁飞性和毁灭性的特点^[2],3龄期飞蝗的预防工作可有效降低飞蝗的爆发性^[3]。因此需要一种快速、高效的飞蝗龄期识别监测方法,来辅助预防飞蝗的爆发。

目前,光谱分析技术广泛应用在农业病虫害监测领域,光谱可反映生物内部元素的结构和含量特征,不易受外部因素影响。基于光谱分析建立的判别模型,可以快速有效地识别农业病害和虫害。高光谱分析技术可获得害虫不均匀表面所有化学信息^[4],得到化学成分的空间分布情况,弥补了近红外光谱分析技术的不足。目前,国内外在基于光谱技术的害虫识别的研究和特征波长提取方法研究方面均有一定的研究成果^[5-14]。

综上所述,基于光谱分析技术实现东亚飞蝗龄期识别的关键在于蝗虫光谱特征波长的提取。本文以东亚飞蝗为实验对象,采用高光谱分析技术(400~1000nm),提出一种基于K均值聚类 and 主成分分析(PCA)相结合的特征波长提取方法,对3龄、4龄、5龄和成虫4个龄期的东亚飞蝗进行识别。

1 实验

1.1 高光谱采集装置

高光谱采集装置如图1所示,由高光谱相机、光源和翻拍架3部分组成。其中,高光谱相机光谱范围400~1000nm,光谱分辨率为(4±0.5)nm,全幅像素1392(空间维)×1040(光谱维),具有自动对焦、自动扫描和自动曝光的功能;光源为2盏200W的钨灯,波长范围400~3000nm,光源支架长度为710mm,角度(水平方向)0°~180°均可调;翻拍架规格为600mm×500mm×1030mm。以图像采集软件ScanViewer采集图像样品图像,实验中采用Matlab 2014、ENVI 4.7和Excel 2013软件对数据进行处理。

1.2 高光谱数据采集

本实验的蝗虫样本是由中国农业大学农学与生物技术学院提供的东亚飞蝗成虫、5龄、4龄和3龄各30只。饲养温度28~32℃,相对湿度60%,12h光照时间与12h黑暗时间交替进行,每天用新鲜的小麦、玉米叶等喂食,长至所需龄期。采集图像参数设置:物距23cm,自动调焦和自动曝光,光源角度30°(垂直方向)。东亚飞蝗样品共30组,每组成虫、5龄、4龄和3龄各1只,利用粘纸将蝗虫固定在采样台上,如图2所示,使用高光谱相机采集蝗虫俯视图方向图像,采集前进行黑白校正。采集的每组数据中,共256个波长(400~1000nm),波长间隔为2.35nm,每个波长采集一幅图像。

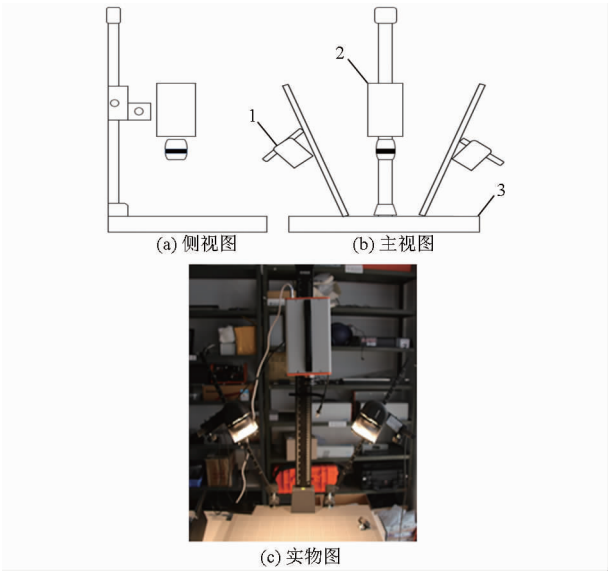


图1 高光谱采集装置

Fig. 1 Hyperspectral imaging system

1. 钨灯 2. 光谱仪 3. 翻拍架



图2 不同龄期东亚飞蝗

Fig. 2 Picture of manilensis at different ages

1.3 数据预处理

样本数据以东亚飞蝗前胸背甲为目标区域,东亚飞蝗表皮属于不均匀物质,从目标区域内以15像素×15像素的正方形内取反射率平均值,从30组样本数据中,每个龄期提取110组前胸背甲反射率的平均值,波段范围400~1000nm,波长间隔2.8nm。用高斯平滑处理方法和矢量归一化对光谱数据进行处理,剔除异常数据得到最终的样本数据,如图3所示。

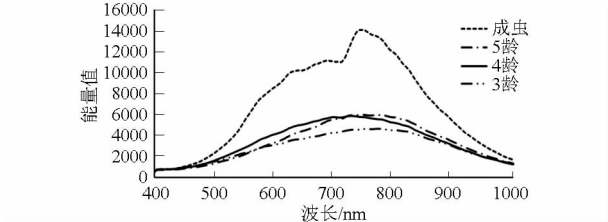


图3 东亚飞蝗不同龄期光谱能量图

Fig. 3 Spectral reflectance curve at different manilensis ages

2 结果分析

2.1 特征波长提取

2.1.1 连续投影算法

连续投影算法(SPA)^[15]是近几年提出的一种

特征波长提取算法,能够从光谱信息中充分寻找含有最低限度冗余信息的变量组,使得变量之间满足最小共线性原则。该算法能降低模型复杂度,从原始光谱数据中提取最能代表样品光谱信息的特征光谱。采用 SPA 对样本数据进行特征波长筛选时,对筛选的特征波段光谱区间设定最大有效波长数^[16] ($m_{\max}$),本实验设定为 10、20 和 30,SPA 算法对光谱筛选结果如表 1 所示,筛选光谱如图 4 所示。

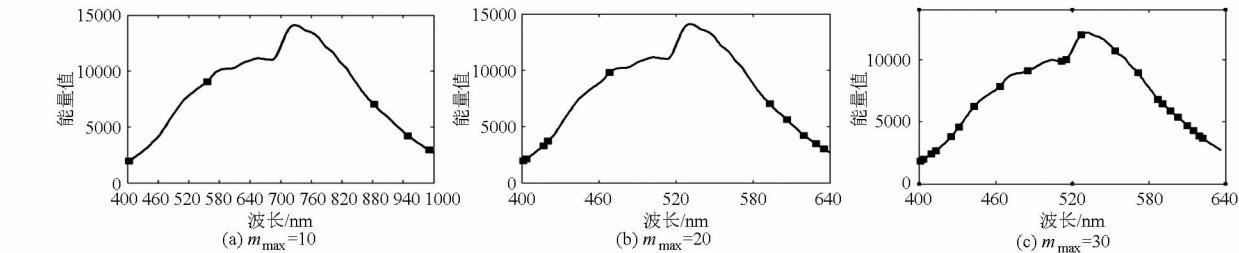


图 4 不同 $m_{\max}$ 值下 SPA 特征波长筛选结果

Fig.4 Selected wavelengths using SPA with different $m_{\max}$ values

2.1.2 PCA 与 K 均值聚类结合方法 (K - PCA)

主成分分析 (PCA) 是一种常用的经典特征波长提取方法,主要是将样本中含有的大量信息通过线性组合的方式用较少的相互独立变量来表示。根据权重系数进行特征波长的选取,其原理为

$$P_k = \sum_{i=1}^n a_i p_i \tag{1}$$

式中 P_k ——第 k 主成分图像
 a_i ——第 i 个权重系数
 p_i ——第 i 个波长图像
 n ——原始图像波长数,本文取 256

经主成分分析后,样本数据划分成 3 个主成分,第 1 主成分方差 88.9% 大于 85%,说明第 1 主成分具有一定的代表性。第 1 主成分的权重系数如图 5 所示,在波段 400 ~ 1 000 nm 之间,权重曲线变化不明显,没有明显的极值点,在波段 500 ~ 1 000 nm 之间权重系数的平均值为 0.987 88,方差为 0.000 041 3,进而证实离散度小,不能根据权重系数筛选特征波段。说明实验样本数据波段间相关性较大,导致波段权重系数基本一致。针对此问题,本文采用 K 均值聚类算法^[17],将样本数据分成 K 类,再针对每一个类别采用主成分分析法进行深度特征波段提取。

K 均值聚类属于一种典型的划分聚类方法,具体过程步骤:①选取 K 个对象作为初始的聚类中心。②根据聚类中心值,将每个对象赋给最相似的簇。③重新计算每个簇中对象的平均值,用此平均值作为新的聚类中心。④重复执行步骤②和步骤

表 1 SPA 算法特征波长筛选结果

Tab.1 Results of SPA with different $m_{\max}$ values

编号	最大有效 波长数	波长 个数	波长/nm
1	10	5	403、556、883、949、991
2	20	10	403、409、442、451、571、883、916、949、973、988
3	30	22	403、409、424、433、463、478、508、559、613、679、688、718、769、829、868、877、892、907、925、937、949、955

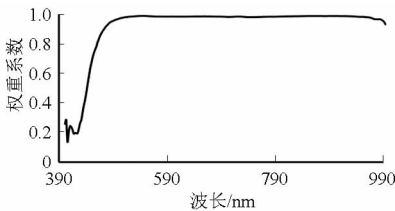


图 5 东亚飞蝗前胸背甲第 1 主成分权重图

Fig.5 First principal component weight chart of manilensis back

③,直到聚类中心不再变化为止^[18]。采用 K 均值聚类方法时,本实验设定 K 为 4,将 256 个波长划分成 4 类,对每一类进行主成分分析,得到成分权重系数曲线,如图 6 所示,得出 468、555、635、710、729、750、786、899 nm 共 8 个特征波长。

2.2 结果对比

根据 K - PCA 和 SPA 提取到的特征波长,对原始预处理过的数据进行随机分组,分成训练集 100 组和测试集 50 组,建立 Fisher 判别分析模型,判别分析结果如表 2 所示:判别函数的正相关系数越大,表明该函数越具有区别力,由表可知 K - PCA 方法建立的判别分析模型明显优于由 SPA 方法建立判别分析模型;SPA 随着 $m_{\max}$ 的增大,正确识别率增加,但筛选的波长数也随之增加,最后 SPA 确定的最优 $m_{\max}$ 为 30,训练集和预测集的正确识别率分别为 99.6% 和 97.25%,波长数为 22。K - PCA 提取的波长数为 8 个,准确识别率分别为 100% 和 98.25%;K - PCA 筛选的特征波长少于最优 $m_{\max}$ 下的 SPA,正确识别率高于最优 $m_{\max}$ 下的 SPA。说明 K - PCA

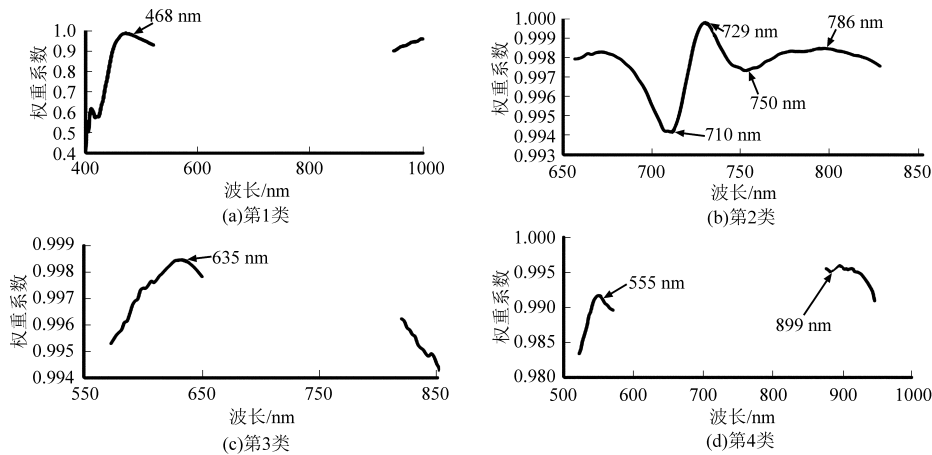


图 6 K-PCA 筛选特征波长结果
Fig. 6 Selected wavelengths using K-PCA

表 2 判别分析结果
Tab.2 Identification results of model based on Fisher algorithm

方法	正相关系数	波长数	龄期	正确识别率/%	
				训练集	测试集
K-PCA	99.80	8	3	100	98.00
			4	100	98.00
			5	100	97.00
			成虫	100	100
			平均	100	98.25
SPA-10	87.20	5	3	100	95.00
			4	100	97.00
			5	100	72.50
			成虫	96.70	97.50
			平均	99.20	90.50
SPA-20	88.40	10	3	100	92.00
			4	100	97.00
			5	100	92.50
			成虫	98.30	97.00
			平均	99.60	94.63
SPA-30	90.00	22	3	100	95.00
			4	100	98.00
			5	100	98.00
			成虫	98.30	98.00
			平均	99.60	97.25

注:SAP-10、SAP-20、SPA-30 分别表示 $m_{\max}$ 值为 10、20、30。

筛选的特征波长及建立的判别模型效果较好,优于 SPA。最终由 K-PCA 建立的 Fisher 判别函数为

$$F_1 = 0.247x_{468} + 1.295x_{555} - 1.603x_{635} - 1.559x_{710} + 4.275x_{729} + 1.854x_{750} - 4.179x_{786} + 0.117x_{899} \quad (2)$$

$$F_2 = 0.776x_{468} + 1.593x_{555} - 1.596x_{635} + 0.656x_{710} - 2.79x_{729} + 2.256x_{750} + 0.335x_{786} - 0.943x_{899} \quad (3)$$

$$F_3 = -0.27x_{468} + 0.359x_{555} + 0.954x_{635} + 2.903x_{710} - 4.791x_{729} - 5.093x_{750} + 7.447x_{786} - 0.461x_{899} \quad (4)$$

3 结束语

在实验室条件下利用高光谱成像技术对不同龄期的东亚飞蝗进行特征波长筛选,提出一种 K 均值聚类和主成分分析相结合的特征波长提取方法(K-PCA),解决了 PCA 筛选波长时出现权重系数离散度低、不易筛选波长的问题,并与连续投影方法(SPA)进行实验对比,结果证明:K-PCA 筛选的特征波长少,准确率高,识别效果优于 SPA,最后筛选的东亚飞蝗特 8 个特征波长分别是 468、555、635、710、729、750、786、899 nm,对不同龄期东亚飞蝗的正确识别率为 100%。

参 考 文 献

1 LI L, ZHU D, YE S, et al. Design and implementation of geographic information systems, remote sensing, and global positioning system-based information platform for locust control[J]. Journal of Applied Remote Sensing, 2014, 8(6):3597-3599.

2 张杰. 我国蝗虫绿色防控新技术推广模式研究与应用[D]. 北京:中国农业大学, 2014.

3 YE S J, ZHU D H, YAO X C, et al. Development of a highly flexible mobile gis-based system for collecting arable land quality data[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2014, 7(11):4432-4441.

4 卢辉, 韩建国, 张录达. 光谱分析技术在蝗虫监测中的应用[J]. 光谱学与光谱分析, 2008, 28(12):2808-2811. LU Hui, HAN Jianguo, ZHANG Luda. Applications of spectral analysis technique to monitoring grasshoppers[J]. Spectroscopy and Spectral Analysis, 2008, 28(12):2808-2811. (in Chinese)

5 刘桂霞, 王新谱, 李秀敏. 近红外光谱检测技术在害虫检测中的应用[J]. 光谱学与光谱分析, 2009, 29(7):1856-1859. LIU Guixia, WANG Xinpu, LI Xiumin. Application of NIR spectroscopy technology in the field of insect pests detection[J]. Spectroscopy & Spectral Analysis, 2009, 29(7):1856-1859. (in Chinese)

6 JOEL P M, THRONE J E, DOWELL F E, et al. Chronological age-grading of three species of stored-product beetles by using

- near-infrared spectroscopy. [J]. Journal of Economic Entomology, 2004, 97(3): 1159 – 1167.
- 7 PEREZ-MENDOZA J, DOWELL F E, BROCE A B, et al. Chronological age-grading of house flies by using near-infrared spectroscopy[J]. Journal of Medical Entomology, 2009, 39(3):499 – 508.
- 8 MAGHIRANG E B, DOWELL F E, BAKER J E, et al. Detecting single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy[J]. Transactions of the ASAE, 2002, 46(4): 1277 – 1288.
- 9 许章华, 李聪慧, 刘健, 等. 马尾松毛虫害等级的 Fisher 判别分析[J]. 农业机械学报, 2014, 45(6): 275 – 283.
- XU Zhanghua, LI Conghui, LIU Jian, et al. Fisher discriminant analysis of dendrolimus punctatus walker pest levels [J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(6): 275 – 283. (in Chinese)
- 10 黄健熙, 卓文, 杨春喜, 等. 基于 Landsat8 卫星数据的蝗虫遥感监测方法研究[J]. 农业机械学报, 2015, 46(5): 258 – 264.
- HUANG Jianxi, ZHUO Wen, YANG Chunxi, et al. The locust remote sensing monitoring methods research based on Landsat8 satellite data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(5): 258 – 264. (in Chinese)
- 11 徐瑞清, 周强, 王书茂. 蝗虫灾害的机械化捕集治理技术发展[J]. 农业机械学报, 2005, 36(11): 165 – 167.
- XU Ruiqing, ZHOU Qiang, WANG Shumao. Technical development on mechanized control of locust[J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(11): 165 – 167. (in Chinese)
- 12 毛文华, 郑永军, 苑严伟, 等. 基于色度和形态特征的蝗虫信息提取技术[J]. 农业机械学报, 2008, 39(9): 104 – 107.
- MAO Wenhua, ZHENG Yongjun, YUAN Yanwei, et al. Locust information extraction using hue and shape feature [J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(9): 104 – 107. (in Chinese)
- 13 宋革联, 余俊霖, 刘飞, 等. 基于高光谱成像技术的菜青虫生命状态检测研究[J]. 光谱学与光谱学分析, 2014, 34(8): 2225 – 2228.
- SONG Gelian, YU Junlin, LIU Fei, et al. Study on the live state of pieris rapaes using near infrared hyperspectral imaging technology[J]. Spectroscopy & Spectral Analysis, 2014, 34(8): 2225 – 2228. (in Chinese)
- 14 ZHU Fengle, PENG Jiyu, GAO Junfeng, et al. Determination and visualization of fat contents in salmon fillets based on visible and near-infrared hyperspectral imagery[J]. Transactions of the CSAE, 2014, 30(23): 314 – 323.
- 15 成忠, 张立庆, 刘赫扬, 等. 连续投影算法及其在小麦近红外光谱波长选择中的应用[J]. 光谱学与光谱分析, 2010, 30(4): 949 – 952.
- CHENG Zhong, ZHANG Liqing, LIU Heyang. Successive projections algorithm and its application to selecting the wheat near-infrared spectral variables[J]. Spectroscopy & Spectral Analysis, 2010, 30(4): 949 – 952. (in Chinese)
- 16 魏新华, 吴姝, 范晓冬, 等. 基于高光谱成像分析的冬枣微观损伤识别[J]. 农业机械学报, 2015, 46(3): 242 – 246.
- WEI Xinhua, WU Shu, FAN Xiaodong, et al. Identification of slight bruises on winter jujube based on hyperspectral imaging technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(3): 242 – 246. (in Chinese)
- 17 苏红军, 盛业华, Du Qian, 等. 高光谱影像的改进 K-均值监督式聚类分析方法[J]. 武汉大学学报: 信息科学版, 2012, 37(6): 640 – 643.
- SU Hongjun, SHENG Yehua, DU Qian, et al. Supervised K-means clustering analysis for hyperspectral imagery[J]. Geomatics and Information Science of Wuhan University, 2012, 37(6): 640 – 643. (in Chinese)
- 18 周世兵, 徐振源, 唐旭清. 新的 K-均值算法最佳聚类数确定方法[J]. 计算机工程与应用, 2010, 46(16): 27 – 31.
- ZHOU Shibing, XU Zhenyuan, TANG Xuqing. Method for determining optimal number of clusters in K-means clustering algorithm[J]. Computer Engineering & Applications, 2010, 46(16): 27 – 31. (in Chinese)