

基于可见光光谱分析的黄瓜白粉病识别研究

王翔宇^{1,2}, 朱晨光¹, 傅泽田¹, 张领先¹, 李鑫星^{1*}

1. 中国农业大学信息与电气工程学院, 食品质量与安全北京实验室, 北京 100083
2. 长治学院电子信息与物理系, 山西 长治 046011

摘要 白粉病是黄瓜常见病害之一, 传播速度极快, 严重时可造成黄瓜大量减产, 对其进行快速准确识别, 对黄瓜白粉病诊断和防治具有重要意义, 应用可见光谱技术, 结合主成分分析和支持向量机算法, 实现对黄瓜白粉病的快速识别。配制白粉病菌孢子悬浮液, 并人工接种于科研温室内的黄瓜叶片上, 以诱发黄瓜白粉病, 待白粉病有一定面积暴发后, 利用海洋光学 USB2000+型便携式光谱仪对黄瓜叶片光谱信息进行采集, 利用五点取样法采集样本, 在5个检查点, 每点选取2株黄瓜进行调查, 每株选取4枚感病叶片, 每枚叶片随机选取5个感病区域进行光谱采集, 共计采集200个感病叶片光谱样本, 同样采集200个健康叶片样本作为对照。通过 Ocean Optics Spectra-Suite 软件采集漫反射标准白板信息和光谱仪暗电流实现光谱仪校正, 调节积分时间、扫描次数以及平滑度等参数来实现光谱曲线平滑处理, 以有效抑制光谱噪声, 对光谱特征进行分类识别, 去掉首尾噪声较大的波段, 保留光谱的可见光波段进行研究, 最终选取450~780 nm 波段范围作为研究对象。利用主成分分析对所研究波段范围内的高维光谱数据(947维)进行降维处理, 根据主成分的累计贡献率, 选取前5个主成分作为分类模型的输入, 以白粉病和健康叶片的判别结果作为输出, 利用支持向量机算法, 通过对样本的分类学习训练构建黄瓜白粉病和健康叶片的分类识别模型, 随机选取120个样本作为训练集用于分类模型构建, 其余80个样本作为测试集用于模型检验, 并通过选取不同的核函数来获得最优模型。利用混淆矩阵对分类识别模型的准确率进行评价, 当选取径向基核函数时, 分类识别模型对黄瓜健康叶片和白粉病叶片的识别准确率最高, 分别为100%和96.25%, 总准确率为98.125%, 具有较高的准确率。结果表明, 利用可见光光谱信息并结合主成分分析和支持向量机算法, 可以实现对黄瓜白粉病的快速准确识别, 为黄瓜病害诊断提供了方法和参考依据。

关键词 可见光谱; 病害识别; 主成分分析; 支持向量机

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)06-1864-06

引言

黄瓜白粉病是保护地黄瓜的常见病害之一, 各地均有分布, 通常在黄瓜生长的中、后期发病严重, 造成黄瓜减产甚至提前拉秧, 在一般年份, 可造成减产10%左右, 在流行年份造成减产20%~40%^[1]。对于黄瓜白粉病识别, 传统检测方法过度依赖于植保专家的经验知识或实验室检测, 前者具有较强的主观性^[2], 而后者是破坏性取样, 且时间成本较高, 降低了病害识别的实际意义。因此, 实现黄瓜白粉病快速准确识别, 对于黄瓜病害防控具有重要意义, 可以有效减少黄瓜产量损失。

目前, 图像处理技术是作物病害识别的主要方法。马浚诚等^[3]对黄瓜霜霉病病斑进行图像分割并转换到HSV颜色空间, 通过提取病斑颜色、纹理和形状等25个特征, 结合支持向量机算法构建了黄瓜霜霉病识别模型, 实现了对黄瓜霜霉病的准确识别与诊断。贾建楠等^[4]利用最大类间方差法对黄瓜叶片角斑病和霜霉病病斑进行分割并提取了面积、周长、矩形度等10个形状特征, 通过神经网络训练构建了病害识别模型, 识别准确率达100%。Zhang等^[5]通过提取黄瓜病斑的颜色、形状等特征信息, 结合K均值聚类和稀疏表示法对黄瓜主要病害进行了识别, 总体识别率达85.7%。图像处理技术发展较为成熟, 且处理方法多, 选择性大, 但其所需特征变量多, 运算量大且运算复杂。近年来, 随着光谱技

收稿日期: 2018-05-07, 修订日期: 2018-10-21

基金项目: 国家自然科学基金面上项目(31271618)资助

作者简介: 王翔宇, 1989年生, 中国农业大学信息与电气工程学院博士研究生 e-mail: wangxiangyu@cau.edu.cn

* 通讯联系人 e-mail: lxxcau@cau.edu.cn

术在农作物病害检测与诊断领域中的发展与应用,作物病害识别也逐步向快速、无损方向发展。李金萍等和王惟萍等利用傅里叶变换红外光谱技术结合 PCR 技术,对大白菜根肿病进行了快速检测与定量测评。柴阿丽等利用傅里叶变换红外光谱技术,选取 $1\ 735$, $1\ 545$ 和 $1\ 240\ \text{cm}^{-1}$ 三个敏感谱带进行研究,实现了对黄瓜褐斑病的检测。赵艳茹等利用共聚焦拉曼光谱技术,结合小波变换和偏最小二乘分析,构建了油菜菌核病的判别模型,并达到了 100% 的识别准确率。孙旭东等利用高光谱成像技术,结合偏最小二乘法,构建了柑桔缺素黄化叶片和黄龙病叶片的判别模型,并具有较高的分类准确率。

目前,红外光谱、高光谱和拉曼光谱技术均在作物病害识别研究中取得了一定进展,而可见光谱在该领域的应用较少。可见光谱只需对所提取的特征波段进行分析,即可构建病害识别模型,所需变量少且运算简单易于实现,因此,本研究主要针对黄瓜白粉病的可见光谱进行分析,所采集光谱数据由天津市农业科学院信息研究所和天津市植物保护研究所等合作单位提供,通过主成分分析对光谱特征波段进行降维处理,应用支持向量机算法对采集到的光谱样本进行分类训练,以构建黄瓜白粉病与健康叶片的分类识别模型,并利用混淆矩阵完成模型准确率评估。该模型可以对黄瓜白粉病进行快速准确识别,为黄瓜病害诊断提供了方法和参考依据。

1 实验部分

1.1 实际方案设计

首先对温室黄瓜叶片进行病原菌接种。黄瓜白粉病由黄瓜白粉病菌(*Sphaerotheca fuliginea*) 侵染所致^[6], 采摘带有白粉病菌的叶片用无菌水冲洗, 配制成孢子悬浮液, 将其人工接种于科研温室中的黄瓜叶片上, 以诱发黄瓜白粉病。

待白粉病有一定面积爆发后, 利用便携式光谱仪及采样设备进入温室对白粉病叶片进行实地采样。通过五点取样法来获取白粉病样本, 并利用相同方法对黄瓜健康叶片进行采样以形成对照组。为避免黄瓜叶片失水萎蔫, 影响采样和实验效果, 每枚叶片的采样应在 $5\ \text{min}$ 内完成。

对样本数据进行积分、平滑等预处理后, 结合主成分分析对光谱样本的主成分进行选择, 利用支持向量机算法对所选主成分构成的新样本进行学习训练, 以得到黄瓜白粉病和健康叶片的分类识别模型。

1.2 仪器

采用的光谱仪为美国海洋光学(Ocean Optics)公司生产的 USB2000+型光纤光谱仪, 光谱范围为 $200\sim 1\ 100\ \text{nm}$, 分辨率为 $0.3\ \text{nm}$, 光谱探测器像素值为 $2\ 048$ 。采样设备为内置光源的 ISP-REF 型反射用积分球, 光源为内置式卤钨灯光源, 采样孔为圆形, 直径为 $10.32\ \text{mm}$ (面积 $83.65\ \text{mm}^2$)。通过采集漫反射标准白板信息和光谱仪暗电流实现光谱仪校正。光谱仪与采集设备如图 1 所示。

1.3 样本光谱采集与预处理

利用五点取样法进行样本采集, 对于感病叶片, 在 5 个

检查点, 每个检查点选取 2 株黄瓜进行调查, 每株黄瓜选取 4 枚感病严重的叶片, 每枚叶片上随机选取 5 个感病区域(即病斑区域)进行光谱采集, 共计采集 200 个感病叶片光谱样本; 对于健康叶片, 采集方法同感病叶片, 共采集 200 个健康叶片光谱样本。其中, 对于黄瓜白粉病, 根据植保专家经验知识, 将病斑面积占整个叶片 40% 以上的叶片定义为感病严重的叶片^[7]。

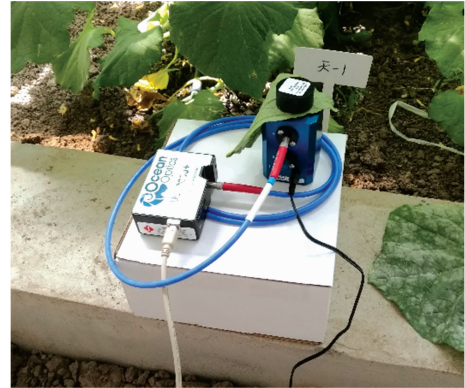


图 1 光谱仪与采样设备

Fig. 1 Spectrograph acquisition equipment

通过海洋光学提供的 Ocean Optics SpectraSuite 软件对采集的光谱进行预处理。通过调整参数, 得到合适的积分时间为 $6\ \text{ms}$, 设定扫描次数为 100 次, 平滑度为 7, 来实现对光谱曲线的平滑处理。

1.4 光谱特征分类识别

利用光谱仪及采样设备采集到的黄瓜健康叶片和白粉病叶片感病区域的反射光谱曲线如图 2 所示。

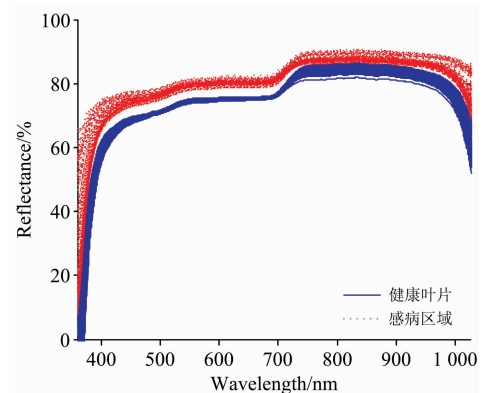


图 2 反射光谱曲线

Fig. 2 Reflection spectrum curve

由图 2 可知, 黄瓜白粉病叶片感病区域与健康叶片的光谱曲线变化趋势相同, 在 $470\sim 520$, $530\sim 580$ 和 $700\sim 780\ \text{nm}$ 区间出现波峰, $780\sim 900\ \text{nm}$ 区间内的感病叶片与健康叶片光谱曲线变化平缓且出现重叠部分。由于光谱波峰特征不明显, 且光谱首尾噪声较大, 因此去除首尾噪声较大的波段, 保留光谱的可见光波段进行研究, 则所选择的研究波段范围为 $450\sim 780\ \text{nm}$ 。

1.5 数据分析

主成分分析法(principal component analysis, PCA)是一种有效的数据降维方法,可通过线性变换将可能存在相关性的变量进行重新组合,保留方差大、含信息多的分量而形成新的不相关的综合变量,从而降低数据的维数^[8-10]。本研究所包含的特征波段较多,利用主成分分析法可以对特征波段进行有效降维,减少了输入变量的数量,从而有效提高了算法运行速度。

支持向量机(support vector machine, SVM)是一种基于统计的学习方法,通过非线性映射将低维空间样本映射到高维特征空间,使线性不可分问题转化为线性可分问题^[11]。支持向量机算法是对结构风险最小化归纳原则的近似,能较好地解决小样本、非线性、高维数和局部极小点等实际问题^[12-15]。本研究采用支持向量机算法构建分类模型,以实现黄瓜健康叶片和白粉病叶片的识别。

2 结果与讨论

2.1 主成分选择

在 450~780 nm 波段范围内,由于每条光谱曲线包含了 947 个数据点,即每个样本的数据维数为 947,数据维数很高。为有效降维,同时避免信息冗余,采用主成分分析法来完成光谱数据主成分提取,实现对光谱信息的高效压缩,同时消除原始数据的相关性和信息冗余^[16-17]。

利用 Matlab R2014a 对光谱数据进行主成分分析处理,得到各主成分的贡献率与累计贡献率如表 1 所示。

表 1 主成分贡献率与累计贡献率
Table 1 Contribution and cumulative contribution rate of PCs

主成分	贡献率/%	累计贡献率/%
PC1	95.875 8	95.875 8
PC2	3.156 2	99.032 0
PC3	0.900 7	99.932 7
PC4	0.056 4	99.989 1
PC5	0.003 9	99.993 0

通过表 1,可知前 5 个主成分的累计贡献率达 99.99% 以上,已经足够用来解释光谱数据中的绝大部分信息。因此选取 PC1—PC5 这 5 个主成分作为新的变量代替原来的变量。

2.2 分类模型构建

利用支持向量机算法完成黄瓜白粉病及健康叶片的分类模型构建。

对于支持向量机分类的准确性,核函数的选取是关键,常用的核函数包括线性核函数、多项式核函数、径向基核函数和 Sigmoid 核函数^[18],其数学表达式如表 2 所示。

表 2 常用核函数数学表达式

核函数名称	数学表达式
线性核函数	$K(x, y) = xy$
多项式核函数	$K(x, y) = (xy + 1)^q$
径向基核函数	$K(x, y) = \exp\left\{-\frac{ x-y ^2}{\sigma^2}\right\}$
Sigmoid 核函数	$K(x, y) = \tanh(v(xy) + c)$

以 PC1—PC5 这 5 个主成分作为输入变量,给定健康叶片的类别标签为 -1,白粉病叶片的类别标签为 1。将采集到的 200 个健康叶片样本和 200 个白粉病叶片样本进行分类训练,其中随机选择 120 个健康叶片和白粉病叶片样本作为训练集用于分类模型构建,其余 80 个样本作为测试集用于模型检验。

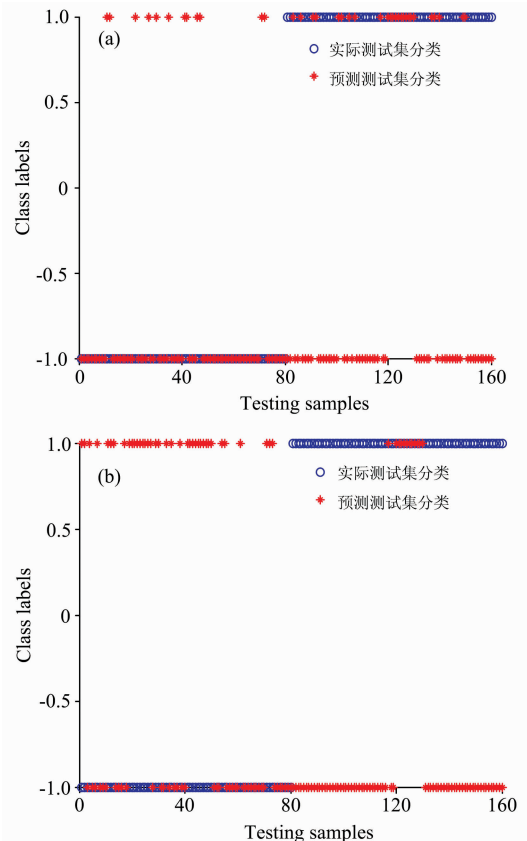
通过 Matlab R2014a 结合 libSVM 工具箱,实现对黄瓜白粉病叶片和健康叶片的分类识别。

2.3 识别结果及评价

当选取不同核函数时,分类识别模型的识别结果如图 3 所示。

图 3 中(a)—(d)依次为选取线性核函数、多项式核函数、径向基核函数和 Sigmoid 核函数时,分类识别模型测试集的分类结果。

测试结果通过二元分类问题混淆矩阵来进行评价,混淆矩阵模型如图 4 所示。



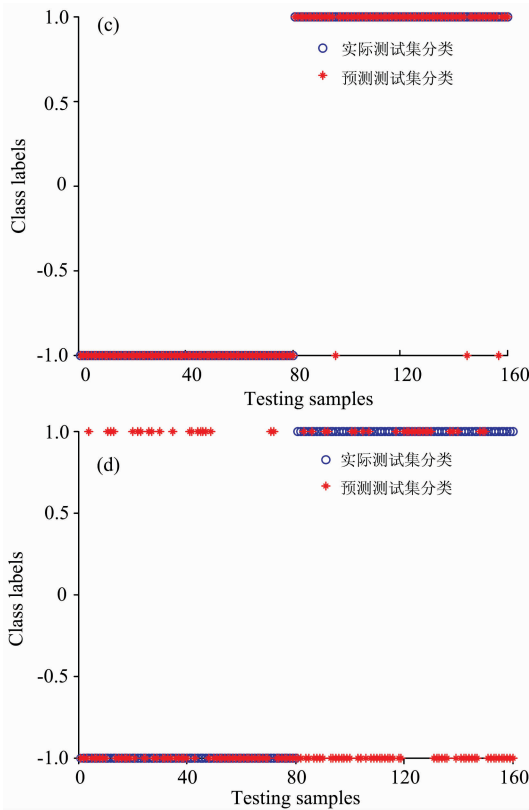


图 3 测试集的分类结果

(a): 线性核函数; (b): 多项式核函数;
(c): 径向基核函数; (d): Sigmoid 核函数

Fig. 3 Classification result of testing samples

(a): Linear kernel function; (b): Polynomial kernel function;
(c): Radial basis kernel function; (d): Sigmoid kernel function

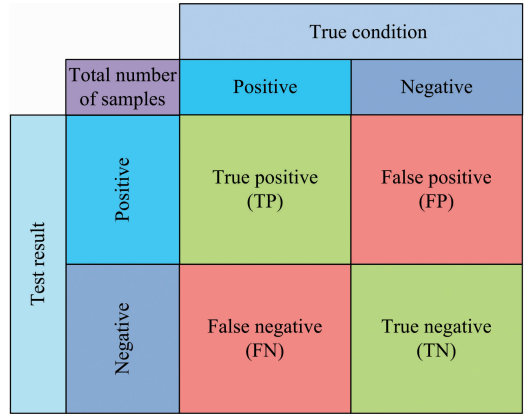


图 4 混淆矩阵模型

Fig. 4 Confusion matrix model

在混淆矩阵中, 真实情况 (True condition) 分为阳性 (Positive) 和阴性 (Negative) 两类, 同理, 分类模型的预测结果 (Test result) 也分为阳性和阴性两类。实际是阳性, 预测成阳性的样本数, 称作“真阳性”(true positive, TP); 实际是阴性, 预测成阳性的样本数, 称作“假阳性”(false positive, FP); 实际是阳性, 预测成阴性的样本数, 称作“假阴性”(false negative, FN); 实际是阴性, 预测成阴性的样本数, 称作“真阴性”(true negative, TN)。由 TP, FP, FN 和 TN 所组成的矩阵即为混淆矩阵。将 N 记作总样本数 (total number of samples), 则混淆矩阵的准确率 (accuracy, ACC) 如式 (1)

$$ACC = (TP + TN) / N \quad (1)$$

利用混淆矩阵对分类识别模型的准确率进行评价, 其分类准确率如表 3 所示。

表 3 不同核函数的分类准确率

Table 3 Classification accuracy with different kernel functions

类别	线性核函数		多项式核函数		径向基核函数		Sigmoid 核函数	
	健康	白粉病	健康	白粉病	健康	白粉病	健康	白粉病
健康	68	55	42	68	80	3	60	54
白粉病	12	25	38	12	0	77	20	26
健康样本识别准确率/%	85		52.5		100		75	
白粉病样本识别准确率/%	31.25		15		96.25		32.5	
总准确率/%	58.125		33.75		98.125		53.75	

由表 3, 当选取径向基核函数时, 基于支持向量机的分类识别模型对健康叶片和白粉病叶片的分类准确率最高, 分别为 100% 和 96.25%, 总体识别准确率达到 98.125%。因此, 以径向基函数为核函数的支持向量机分类识别模型具有较好的分类准确率, 能够实现黄瓜白粉病叶片和健康叶片的识别。

3 结论

以黄瓜白粉病叶片为研究对象, 选取 450~780 nm 可见

光波段光谱数据进行分析, 结合主成分分析与支持向量机构建黄瓜白粉病叶片和健康叶片的识别模型, 实现了对黄瓜白粉病的快速准确识别。

首先采用主成分分析法对黄瓜霜霉病叶片和健康叶片的光谱特征波段进行降维处理, 根据主成分累计贡献率, 以前 5 个主成分作为分类识别模型的输入变量。实验共采集到 200 个健康叶片样本和 200 个白粉病叶片样本, 利用支持向量机, 随机选择 120 个样本作为训练集进行分类模型构建, 其余 80 个样本作为测试集用于模型检验。利用混淆矩阵对所构建的分类模型进行评价, 结果表明, 当选取径向基核函

数时, 分类识别模型对黄瓜健康叶片和白粉病的识别准确率分别为 100% 和 96.25%。识别总准确率为 98.125%。结果表明, 分类模型具有较高的准确率, 该方法可以实现对黄瓜白粉病的快速准确识别, 为黄瓜病害诊断提供了方法和参考依据。

本研究说明可见光谱技术对黄瓜白粉病识别具有较好效果, 对于识别模型的构建, 本研究仅仅选取了可见光谱信息进行学习训练, 在今后的研究中, 可以加入图像信息等, 对训练数据进行充实完善, 同时优化分类识别算法, 使分类识别模型具有更快的响应速度, 并且具有更好的泛化能力。

References

- [1] ZHANG Peng, ZHU Yu-qiang, WANG Li-li, et al(张 鹏, 朱育强, 王丽莉, 等). Chinese Agricultural Science Bulletin(中国农学通报), 2017, 33(21): 134.
- [2] GUAN Hui, ZHANG Chang-li, ZHANG Chun-yuan(关 辉, 张长利, 张春媛). Journal of Agricultural Mechanization Research(农机化研究), 2010, (3): 94.
- [3] MA Jun-cheng, WEN Hao-jie, LI Xin-xing, et al(马浚诚, 温皓杰, 李鑫星, 等). Transactions of The Chinese Society of Agricultural Machinery(农业机械学报), 2017, 48(2): 195.
- [4] JIA Jian-nan, Ji Hai-yan(贾建楠, 吉海彦). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(S1): 115.
- [5] Zhang S, Wu X, You Z, et al. Computers and Electronics in Agriculture, 2017, 134: 135.
- [6] SUN Xu-dong, LIU Yan-de, XIAO Huai-chun, et al(孙旭东, 刘燕德, 肖怀春, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(2): 551.
- [7] WANG You-ping, ZHU Jin-ying, GAO Ping-yin, et al(王友平, 朱金英, 高平银, 等). Journal of Changjiang Vegetables, 2009, (1): 37.
- [8] GAO Shi-gang, LUO Jin-yan, ZENG Rong, et al(高士刚, 罗金燕, 曾 蓉, 等). Journal of Plant Protection, 2017, 44(5): 779.
- [9] ZHANG Yin, ZHOU Meng-ran(张 银, 周孟然). Infrared Technology(红外技术), 2007, 29(6): 345.
- [10] LI Hong-lian, GONG Dong-jun, CAI Duan-bo, et al(李红莲, 贡东军, 蔡端波, 等). Laser Journal, 2015, 36(10): 65.
- [11] ZHANG Jian-hua, KONG Fan-tao, LI Zhe-min, et al(张建华, 孔繁涛, 李哲敏, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2014, 30(19): 222.
- [12] WANG Jin-jing, ZHAO De-an, JI Wei(王津京, 赵德安, 姬 伟). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2009, 40(1): 148.
- [13] Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.
- [14] Cortes C, Vapnik V. Machine Learning 1995, 20(3): 273.
- [15] Cherkassky V, Mulier F. Learning from Data: Concepts, Theory and Methods. New York: John Wiley & Sons, 1997.
- [16] ZHANG Xue-gong(张学工). Acta Automatica Sinica(自动化学报), 2000, 26(1): 36.
- [17] ZHU Yi-ning, YANG Ping, YANG Xin-yan, et al(朱毅宁, 杨 平, 杨新艳, 等). Chinese Journal of Analytical Chemistry(分析化学), 2017, 45(3): 336.
- [18] YIN Fei, FENG Da-zheng(尹 飞, 冯大政). Computer Technology and Development(计算机技术与发展), 2008, 18(10): 31.

Research on Cucumber Powdery Mildew Recognition Based on Visual Spectra

WANG Xiang-yu^{1,2}, ZHU Chen-guang¹, FU Ze-tian¹, ZHANG Ling-xian¹, LI Xin-xing^{1*}

1. Beijing Laboratory of Food Quality and Safety, College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Department of Electronic Information and Physics, Changzhi University, Changzhi 046011, China

Abstract Powdery mildew is one of the common diseases of cucumber, which has a rapid propagation speed and can cause a large reduction of cucumber. Quick and accurate recognition of cucumber powdery mildew has great significance for the diagnosis and control of cucumber diseases. Utilize visible spectrum technology combined with principal component analysis and support vector machine algorithm can realize the quick recognition of cucumber powdery mildew. *Sphaerotheca fuliginea* was used to make spore suspension and inoculated it into the cucumber leaves in a scientific research solar greenhouse to induce powdery mildew. When the powdery mildew occurred, the spectral information of cucumber leaves was collected by the Ocean Optics USB2000+ portable spectrometer. Five point sampling method was used to collect samples, two cucumber plants were inspected at each point and four leaves were checked on each plant, and five areas were chosen randomly on each leaf to use to spectral information acquisition. Then 200 samples of cucumber powdery mildew leaves were got, and 200 healthy leaf samples were collected as contrast by the same method. the standard white plate and dark current was Utilized to calibrate the spectrometer. The integral time and the scanning times were adjusted and the smoothness parameters of Ocean Optics Spectra-Suite software was used to smooth spectral curves and suppress noise. Through classification and recognition of spectral features, the spectral bands with big noise was removed and the 450~780 nm visible light band was chosen as the research range. The principal component analysis (PCA) was used to reduce the dimension of high-dimensional spectral data (947 dimension). According to the cumulative contribution rate of principal components, the former 5 principal components were chosen as input variables and the discriminant results as the output to build the classification model. We utilized support vector machine (SVM) algorithm and randomly took 120 samples as the training set to build the classification model, and the rest 80 samples as testing set for model checking, and the optimal model was obtained by selecting different kernel functions. The confusion matrix was used to evaluate the accuracy of the classification model, when the radial basis kernel function was selected, the recognition accuracy of the classification model for cucumber healthy leaves and powdery mildew leaves were respectively 100% and 96.25%, and the total accuracy was 98.125%. The results showed that the visible light spectrum analysis combined with PCA and SVM algorithm could be used to identify cucumber powdery mildew quickly and accurately, which provides a method and reference for the diagnosis of cucumber diseases.

Keywords Visible spectrum; Disease recognition; Principal component analysis; Support vector machine

(Received May 7, 2018; accepted Oct. 21, 2018)

* Corresponding author