

## 可见光谱图像联合区间的黄瓜白粉病分割与检测

白雪冰, 余建树, 傅泽田, 张领先, 李鑫星\*

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

**摘要** 黄瓜白粉病是黄瓜病害中爆发频率较高的一种, 传播速度极快, 常常导致产量降低, 无法获得预期的经济效益。特别是在病害爆发严重的年份, 黄瓜减产量高达 20%。提出了一种基于可见光谱图像联合区间的偏最小二乘回归判别模型(SI-PLSR), 用于黄瓜白粉病无损检测。采用 Canon EOS 800D 和 Ocean Optics USB2000+ 光纤光谱仪采集了 200 个黄瓜白粉病感病叶片的可见光谱图像和反射率曲线。首先, 采用基于小波降噪和 H 分割的分水岭分割算法从实时采集的黄瓜白粉病感病叶片可见光谱图像中提取目标叶片; 其次, 通过高斯拟合优化的 Otsu 算法分割目标叶片的可见光谱图像, 获取白粉病病斑; 然后, 对 350~1100 nm 全波段光谱反射率曲线建立偏最小二乘回归模型并计算交叉验证均方根误差  $RMSE_{CV}$ , 同时将全波段等分为 20 个子区间, 分别建立偏最小二乘回归模型, 选取  $RMSE_{CV}$  小于全波段反射率曲线建模  $RMSE_{CV}$  的子区间组成联合区间; 最后, 将光谱联合区间与白粉病病斑分割结果融合建立 SI-PLSR 模型。从实验结果可知, 感病目标叶片的提取成功率高达 94.00%, 200 幅感病叶片可见光谱图像中成功提取 188 幅, 其中 157 幅目标叶片的完整性参数高于 95%, 31 幅目标叶片完整性参数在 90%~95% 之间。188 幅目标叶片的病斑分割结果显示, 平均错分率为 5.81%, 其中平均 False negative 为 1.55%, 平均 False positive 为 4.26%。对 20 个子区间分别建立偏最小二乘回归模型发现, 第 5, 6, 7, 11, 12, 13 和 19 子区间的  $RMSE_{CV}$  值小于全波段光谱反射率曲线建模的  $RMSE_{CV}$  值, 说明这 7 个子区间的光谱信息对白粉病的判别有较大的贡献, 这与呈现波峰的 470~520, 530~580 和 700~780 nm 波段相对应, 因此选取这 7 个子区间的光谱反射率曲线建立联合区间。对联合区间建立 SI-PLSR 模型, 其主成分数为 7, 校正集和验证集的相关系数和标准误差分别是 0.975 2, 0.907 3 和 0.919 5, 1.091。与全波段 PLSR 模型相比, SI-PLSR 的相关系数更接近于 1, 且标准误差更小。结果表明, 所提出的 SI-PLSR 模型有效去除了可见光谱数据中冗余信息, 加强了模型的稳定性, 可以实现对黄瓜白粉病的快速无损准确识别, 为黄瓜病害诊断提供了方法和参考依据。

**关键词** 可见光谱; 联合区间; 偏最小二乘回归模型; 计算机视觉

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3592-07

### 引言

黄瓜作为温室蔬菜主要种植品种之一, 常年受到病害的危害导致产量大幅度降低, 农民无法获得预期的经济效益<sup>[1-2]</sup>。白粉病是黄瓜病害中爆发频率较高的一种, 在一般年份会造成减产 10%, 在病害爆发严重的年份会造成减产高达 20%<sup>[3]</sup>。传统的病害诊断方法主要包括理化检测和感官诊断。理化检测的诊断结果较为准确, 但是需要在实验室环境下进行, 诊断过程繁琐且具有破坏性<sup>[4-5]</sup>。感官诊断依靠植保专家的专业知识, 由于专家人数和精力的限制, 往往无法

兼顾大多数地区的病情诊断<sup>[6-7]</sup>。利用作物感病叶片上病斑与正常叶片在可见光范围的光谱和图像差异来进行病害信息的快速获取及病害的无损检测, 成为农业病害诊断的研究热点。

Bai<sup>[8]</sup>利用病害可见光谱图像的空间信息建立二维向量, 改进了 FCM 算法, 成功提取了黄瓜霜霉病病斑信息。Ma<sup>[9]</sup>提出了一种利用超红特征(ExR)、H 分量和  $b^*$  分量三种颜色特征融合的黄瓜病斑分割方法, 克服了光照条件不均匀对病斑分割的影响。Sui<sup>[10]</sup>采集了健康叶片和黄瓜霜霉病感病叶片的光谱信息, 对象建立了基于黄瓜霜霉病光谱特征指标的预测模型。Li<sup>[11]</sup>通过求解表征光谱的线性最优化方程, 实

收稿日期: 2018-09-02, 修订日期: 2019-02-04

基金项目: 国家自然科学基金面上项目(31271618)资助

作者简介: 白雪冰, 1991 年生, 中国农业大学信息与电气工程学院博士研究生 e-mail: 464161695@qq.com

\* 通讯联系人 e-mail: lxxcau@cau.edu.cn

现黄瓜病害分布和严重性的定量检测。

计算机视觉能够直观的观察病斑大小和形态,但是缺乏区分病斑的能力。光谱技术能定性的分析病害类型,但是缺乏对病斑形态的直观性描述。将二者结合可以兼顾定性分析和直观性的优势。因此,通过高斯拟合优化 Otsu, 提取病斑可见光谱图像;并将反射率曲线等分为 20 个子区间,选取  $RMSE_{CV}$  小于全波段建模  $RMSE_{CV}$  的子区间组成联合区间;建立基于可见光谱图像联合区间的偏最小二乘回归判别模型 (SI-PLSR), 用于黄瓜白粉病无损检测。

## 1 实验部分

### 1.1 试验数据采集

黄瓜白粉病植株是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单位协助培植。试验数据采集于天津市农科院植保所农业创新基地 5 号黄瓜白粉病试验温室,采集时间为 2017 年 6 月 9 日 8:00—17:00, 共计 200 个感病叶片。图像数据获取设备为 CanonEOS 800D, 采用自动对焦, 原始图片像素尺寸为  $3\ 984 \times 2\ 988$ , 将采集的感病叶片图像统一编辑为像素  $800 \times 600$ , 以 JPG 格式储存; 光谱数据采集设备为 Ocean Optics USB2000+ 光纤光谱仪, 光谱范围  $200 \sim 1\ 100\ \text{nm}$ , 分辨率为  $0.3\ \text{nm}$ , 光谱像素值为 2 048。参数设置为: 积分时间:  $6\ 000\ \mu\text{s}$ , 平均次数为 100 次, 平滑度为 7。

### 1.2 病斑分割

在温室实时采集的可见光谱图像一般包含较为复杂的背景以及环境因素带来的噪声, 对病斑的提取造成较大的影响。故先基于小波变换和 HSV 颜色模型中 H 分量对光照变化不敏感的特点, 采用分水岭算法提取黄瓜白粉病目标叶片的可见光谱图像。后采用基于高斯拟合优化的 Otsu 迭代算法, 获取最佳阈值, 实现病斑分割。以初次 Otsu 算法的阈值  $th_0$  为界将直方图分成两部分  $C_0$  和  $C_1$  并计算均值 ( $\mu_0$  和  $\mu_1$ ) 和方差 ( $\sigma_0^2$  和  $\sigma_1^2$ ), 将它们拟合成两个高斯分布  $N_0(\mu_0, \sigma_0^2)$  和  $N_1(\mu_1, \sigma_1^2)$ 。将  $N_0$  和  $N_1$  中方差较小的高斯分布记为  $N_{Begin}(\mu_{Begin}, \sigma_{Begin}^2)$ , 方差较大的记为  $N_{End}(\mu_{End}, \sigma_{End}^2)$ 。以  $\mu_{Begin}$  为起点, 向  $\mu_{End}$  方向遍历, 当概率密度值  $p(x_i | Begin) \leq p(x_i | End)$  时, 本次迭代得到新的最佳阈值为  $th_{i+1} = x_i$ 。直至  $|th_{i+1} - th_i| < \Delta$  迭代结束, 其中  $\Delta$  为可接受的误差。具体流程如图 1 所示。

### 1.3 光谱数据分析

可见光谱检测的优势在于较短时间获取大量数据信息, 但是其中也包含了噪声和背景等, 这些噪声信号无法通过预处理消除<sup>[12-13]</sup>。将全部信号都用于建模必然降低识别的效率, 同时噪声信号会影响模型的精度和鲁棒性。研究表明, 通过筛选剔除原始自变量中不相关的信号, 从而实现减少输入变量, 可以简化模型, 有效提高检测的效率和鲁棒性。应用逐步回归法, 筛选原始自变量, 建立了基于最佳变量联合区间与病斑图像融合的偏最小二乘回归模型 (简称 SI-PLSR)。

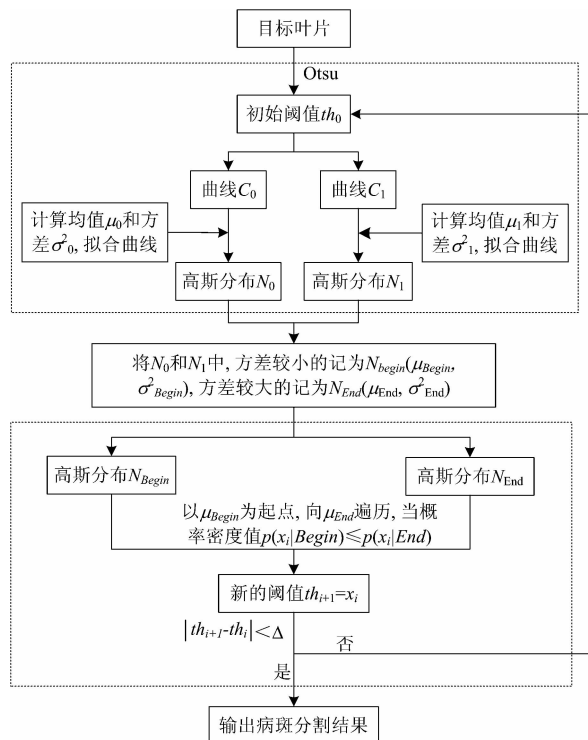


图 1 分割方法流程图

Fig. 1 Flow chart of the segmentation method

偏最小二乘回归模型 (简称 PLSR) 的主成分数选取不超过 15。对光谱全波段按主成分数从 1 到 15 分别建立回归模型, 并获取交叉验证均方根误差  $RMSE_{CV}$ , 如式 (1) 所示。将整个光谱波段等间隔划分为 20 个子区间, 对每个子区间分别建立 PLSR, 筛选出  $RMSE_{CV}$  值小于全波段  $RMSE_{CV}$  值的子区间, 组合成联合区间, 并建立 SI-PLSR。

$$RMSE_{CV} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - z_i)^2} \quad (1)$$

式中  $y_i$  为样品理化指标的标准测量值,  $z_i$  为模型预测值,  $N$  为样品总数。

### 1.4 模型分析

采用相关系数法和标准误差法分析联合区间回归模型的有效性和鲁棒性。相关系数  $R$  反映了感病叶片反射光谱测量值和模型预测值的相关度,  $|R|$  越近似于 1, 测量值和预测值越相关, 计算式 (2)

$$R = \frac{\sum_{i=1}^N (z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2)$$

式中  $y_i$  为样品理化指标的标准测量值,  $\bar{y}$  为标准测量平均值,  $z_i$  为模型预测值,  $\bar{z}$  为模型预测平均值,  $N$  为样品总数。

在相关系数  $R$  相近时, 标准差能够有效反映出模型的鲁棒性。选取校正标准误差  $RMSEC$  和验证标准差  $RMSEV$  对联合区间的回归模型进行评价,  $RMSEC$  和  $RMSEV$  的值越小, 说明模型越稳定<sup>[14]</sup>。  $RMSEC$  和  $RMSEV$  的计算方法如下

$$\text{RMSEC} = \sqrt{\frac{1}{m-k-1} \sum_{i=1}^m (y_i - z_i)^2} \quad (3)$$

$$\text{RMSEV} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - z_i)^2} \quad (4)$$

式中  $y_i$  为样品理化指标的标准测量值,  $z_i$  为模型预测值,  $m$  为校正集,  $n$  为验证集,  $k$  为主成分数,  $N$  为样品总数。

## 2 结果与讨论

### 2.1 目标叶片完整性分析

引用完整性参数 Solidity 作为目标叶片提取结果的评价指标<sup>[15]</sup>, 如式(5)所示。式中, Extracted Area 表示提取的面积, Convex Area 表示目标叶片外接最小多边形的面积。

$$\text{Solidity} = \frac{\text{Extracted Area} \times 100\%}{\text{Convex Area}} \quad (5)$$

根据叶片完整性参数, 将目标叶片提取结果分为三类: Perfect 为 Solidity > 95%, 表示提取结果与真实叶片高度一致; Good 为 90% ≤ Solidity ≤ 95%, 表示提取结果存在缺损,

但可用于病斑分割; Fail 为 Solidity < 90%, 表示提取结果与真实叶片存在较大的形状或面积差异, 提取失败。基于上述原则, 对 129 幅感病叶片图像的目标叶片提取结果进行评价, 如表 1 所示。

表 1 目标叶片提取结果  
Table 1 Target leaf extraction results

|      | Success |      | Fail |
|------|---------|------|------|
|      | Perfect | Good |      |
| 叶片数  | 157     | 31   | 12   |
| 比例/% | 78.5    | 15.5 | 6.00 |

根据表 1 可知, 在 200 幅感病叶片的可见光谱图像中, 成功提取目标叶片共计 188 幅, 成功率为 94.00%, 其中 157 幅提取结果与真实叶片高度一致, 31 幅提取结果存在缺损, 但是对病斑分割影响较小, 可用于后续试验。为更直观的观测实验结果, 随机选取了 6 幅感病叶片的可见光谱图像的目标叶片提取结果, 如图 2 所示。

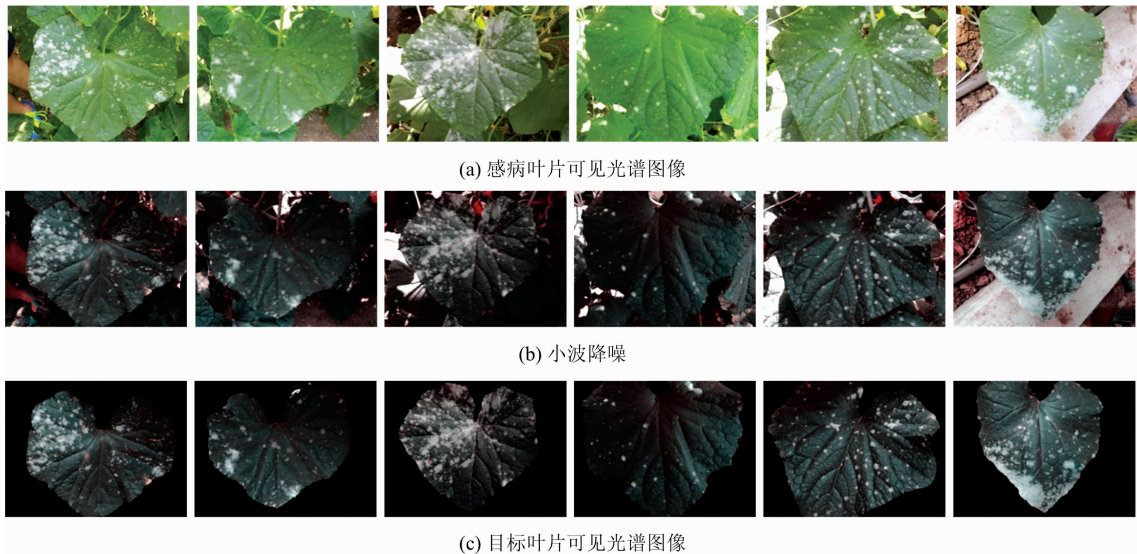


图 2 目标叶片提取

(a): 感病叶片可见光谱图像; (b): 小波降噪; (c): 目标叶片可见光谱图像

Fig. 2 Target leaf extraction

(a): Visible spectrum images of susceptible leaves; (b): Images denoised by wavelet transform;

(c): Visible spectrum images of target leaf

### 2.2 病斑分割结果评价

分别采用 Otsu, 3D Otsu, FCM 以及本算法分割目标感病叶片的可见光谱图像, 并引用 Mizushima<sup>[16]</sup> 提出的方法, 对比分析四种算法所得结果的 False negative 和 False positive, 评价各算法的准确性。False negative 表示病斑误分割为背景的比例; False positive 表示背景被误分割为病斑的比例。以上文随机选取的 6 幅感病叶片可见光谱图像为例进行定性和定量分析, 如图 3 和表 2 所示。

从图 3 和表 2 可知, Otsu 的错分率在 41.90% ~ 69.22% 之间, 实验结果中包含大量正常叶片部分和背景, 无法从中辨认病斑的位置和形状, 说明 Otsu 作为一种简单

的阈值分割算法, 只适合分割较为简单的图片, 并不适用于实时采集的感病叶片可见光谱图像。3D Otsu 处理前 5 幅图像的错分率在 17.91% ~ 20.00% 之间, 但是处理第 6 幅图像的错分率为 43.55%, 从实验结果中也能够观察到病斑的位置和形状, 但是该算法无法区分背景中与病斑类似的物体, 造成第 6 幅图像错分率较大的原因。3D Otsu 引入了像素点的邻域均值和邻域中值, 有效矫正了奇异像素点, 但是实验结果并不理想。FCM 处理前 5 幅图像的错分率在 16.11% ~ 18.07% 之间, 对第 6 幅图像的错分率为 42.91%, 因为 FCM 将病斑与病斑类似的物体聚为一类, 导致错分率较大。FCM 再处理上述 6 幅图像时迭代次数都达到最高值, 用时大于

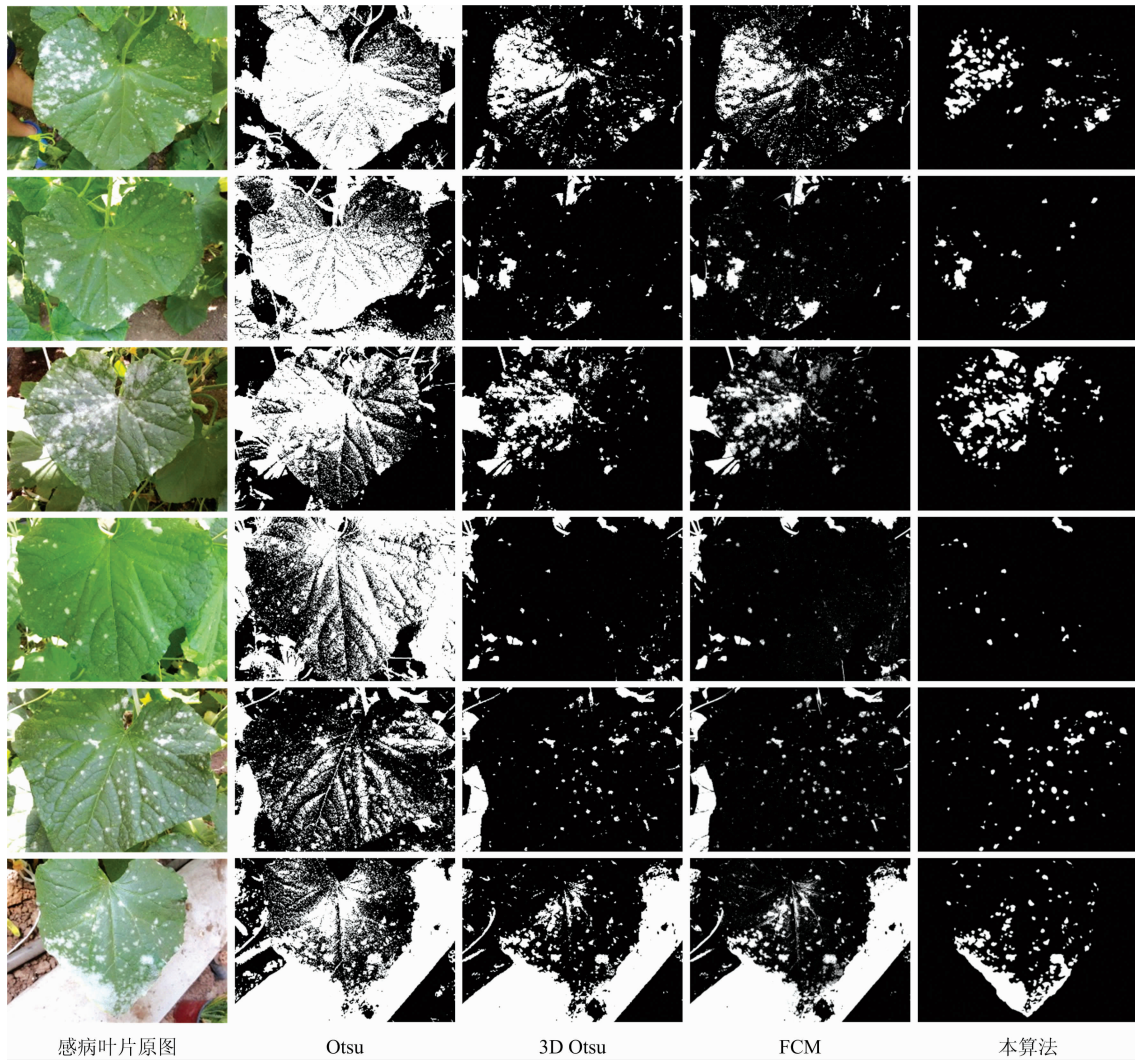


图 3 白粉病分割结果

Fig. 3 Segmentation results of powdery mildew

表 2 4 种算法的错分率 (%)

Table 2 Segmentation errors for four algorithms (%)

|         | 叶片 1           |                |       | 叶片 2           |                |       | 叶片 3           |                |       |
|---------|----------------|----------------|-------|----------------|----------------|-------|----------------|----------------|-------|
|         | False negative | False positive | Total | False negative | False positive | Total | False negative | False positive | Total |
| Otsu    | 0.00           | 69.22          | 69.22 | 0.00           | 67.14          | 67.14 | 0.12           | 51.54          | 51.66 |
| 3D Otsu | 0.81           | 19.19          | 20.00 | 1.10           | 18.35          | 19.45 | 1.33           | 17.72          | 19.05 |
| FCM     | 0.55           | 16.54          | 17.09 | 1.81           | 16.26          | 18.07 | 3.13           | 12.98          | 16.11 |
| 本算法     | 1.31           | 3.76           | 5.07  | 1.85           | 2.28           | 4.13  | 1.89           | 4.13           | 6.02  |
|         | 叶片 4           |                |       | 叶片 5           |                |       | 叶片 6           |                |       |
|         | False negative | False positive | Total | False negative | False positive | Total | False negative | False positive | Total |
| Otsu    | 0.09           | 62.11          | 62.20 | 1.54           | 40.36          | 41.90 | 0.10           | 57.91          | 57.01 |
| 3D Otsu | 1.15           | 16.76          | 17.91 | 0.97           | 17.68          | 18.65 | 0.71           | 42.84          | 43.55 |
| FCM     | 1.01           | 16.06          | 17.07 | 0.75           | 16.68          | 17.43 | 0.48           | 42.43          | 42.91 |
| 本算法     | 0.95           | 3.71           | 4.66  | 1.01           | 3.78           | 4.79  | 1.70           | 4.08           | 5.78  |

13 s, 运算效率较低。本算法的错分率在 4.13%~6.02% 之间, 准确率较高。从实验结果中可知, 本算法提取目标叶片有效避免了背景中与病斑类似的物体对算法的影响; 高斯拟

合优化了阈值, 得到良好的分割结果, 为下一步白粉病检测提供了可靠的数据。

在成功提取目标叶片的 188 幅感病叶片可见光谱图像的

实验中,平均 False negative 为 1.55%,平均 False positive 为 4.26%,平均错分率为 5.81%。平均运行时间为 5.56 s,其中高斯拟合优化的迭代平均次数为 20 次。实验结果说明本算法有较高的鲁棒性和效率。

### 2.3 主成分选择

利用光谱仪分别采集白粉病叶片和健康叶片的反射光谱曲线,如图 4 所示。

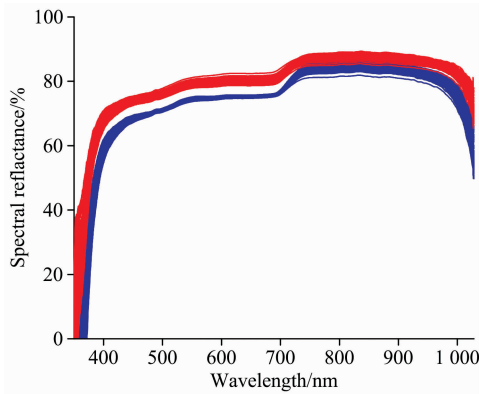


图 4 白粉病反射光谱

Fig. 4 Reflectance spectra curve of powdery mildew

从图 4 可知,白粉病叶片与健康叶片的光谱曲线变化趋势相同,在 470~520, 530~580 和 700~780 nm 区间出现波峰,780~900 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。对光谱全波段按主成分数由 1~15 分别建立 PLSR,并计算各主成分数对应的  $RMSE_{CV}$ ,如图 5 所示。

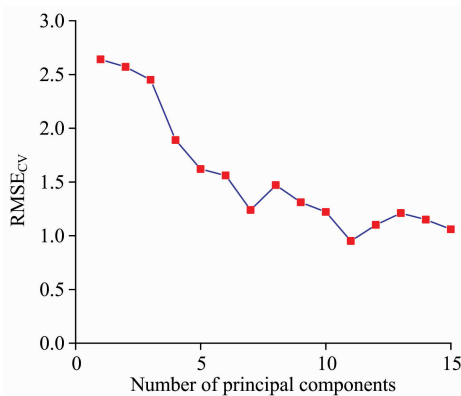


图 5 主成分变化图

Fig. 5 Change of principal component

从图 5 可知,当主成分数为 11 时,对应的  $RMSE_{CV}$  值最小,为 0.95,所以基于可见光谱波段的黄瓜白粉病 PLSR 的最佳主成分数为 11。将白粉病光谱波段等间隔划分为 20 个子区间,分别对每个子区间建立 PLSR,各子区间的分布状况和  $RMSE_{CV}$  如图 6 所示。

图 6 中,红色曲线为黄瓜白粉病叶片的平均反射率光谱曲线,蓝色曲线为正常叶片的平均反射率光谱曲线,虚线为整个可见光谱参与建模时最佳主成分数对应的  $RMSE_{CV}$  值。

从图中可知,第 5, 6, 7, 11, 12, 13 和 19 子区间的  $RMSE_{CV}$  值小于全波段建模的  $RMSE_{CV}$  值,说明这 7 个子区间的光谱信息对白粉病的判别有较大的贡献,这与呈现波峰的 470~520, 530~580 和 700~780 nm 波段相对应。其他区间的  $RMSE_{CV}$  值都大于全波段建模的  $RMSE_{CV}$  值,原因在于单个区间包含信息有限,无法反应白粉病光谱反射特征,造成判别模型不稳定。因此,选用第 5, 6, 7, 11, 12, 13 和 19 子区间作为联合区间,结合病斑图像,从重建 SI-PLSR。

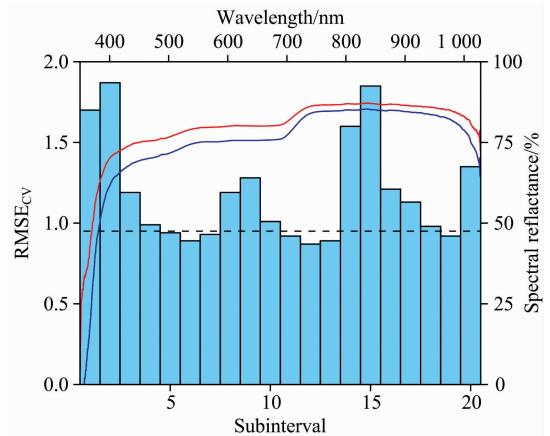


图 6 平均光谱曲线及子区间  $RMSE_{CV}$  值

Fig. 6 Mean spectral curve and RMSECV value of Subinterval

### 2.4 识别结果分析

将成功提取出病斑的 188 幅感病叶片按 3:2 的比例分为校正集和验证集,校正集包含 113 个叶片,验证集包含 75 个叶片。基于 350~1 100 nm 建立全波段 PLSR,并结合第 5, 6, 7, 11, 12, 13 和 19 子区间及病斑图像分割结果建立 SI-PLSR,模型结果如表 3 所示。

表 3 判别模型结果分析

Table 3 Analysis of prediction models

| 区间组合    | 主成分 | 校正集     |         | 验证集     |             |      |
|---------|-----|---------|---------|---------|-------------|------|
|         |     | $R_C$   | RMSEC   | $R_V$   | RMSEV 误判率/% |      |
| 全波段     | 11  | 0.957 8 | 0.953 5 | 0.822 3 | 1.507       | 9.33 |
| SI-PLSR | 7   | 0.975 2 | 0.907 3 | 0.919 5 | 1.091       | 2.67 |

从表 3 可知,SI-PLSR 的主成分数为 7,小于全波段 PLSR 的主成分数,说明 SI-PLSR 排除了冗余的光谱信息,精简了判别模型。SI-PLSR 对校正集和验证集的相关系数和标准误差分别是 0.975 2, 0.907 3 和 0.919 5, 1.091。与全波段 PLSR 相比,SI-PLSR 的相关系数更接近于 1,且标准误差更小,说明 SI-PLSR 输入了更多有效的光谱信息,同时提高了模型的鲁棒性。SI-PLSR 的误判率为 2.67%,对 75 个验证集判别中,仅误判 2 个,准确度明显高于全波段 PLSR。

## 3 结论

以黄瓜白粉病 350~1 100 nm 波段的可见光谱图像为依

据, 基于可见光谱图像联合区间建立 SI-PLSR 判别模型, 实现了黄瓜白粉病快速无损检测。结果显示:

(1) 基于小波降噪和 H 分量分割的分水岭算法提取感病目标叶片的成功率高达 94.00%。在对 188 幅目标叶片可见光谱图像的分割结果中显示, 平均错分率为 5.81%, 其中平均 False negative 为 1.55%, 平均 False positive 为 4.26%。说明本算法对实时采集的白粉病感病叶片图像分割具有较高的准确性。

(2) 通过对比分析子区间与全波段建模的  $RMSE_{CV}$  值,

第 5, 6, 7, 11, 12, 13 和 19 子区间的  $RMSE_{CV}$  值小于全波段建模的  $RMSE_{CV}$  值。因此, 与之相对应的 470~520, 530~580 和 700~780 nm 波段的光谱信息对白粉病的判别有较大的贡献。

(3) 选取对白粉病的判别有较大的贡献的子区间光谱信息组成联合区间建立的 SI-PLSR 白粉病判别模型主成分数减少为 7, 同时判别结果的相关系数和标准误差均优于全波段 PLSR, 说明本模型具有更高的效率和鲁棒性。

## References

- [ 1 ] MA Jun-cheng, WEN Hao-jie, ZHANG Ling-xian, et al(马浚诚, 温皓杰, 张领先, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2017, 48(2): 195.
- [ 2 ] Lebeda A. Journal of Phytopathology, 2010, 108(1): 71.
- [ 3 ] 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.
- [ 4 ] Vatchev T, Maneva S. Crop Protection, 2012, 42(4): 16.
- [ 5 ] Joe M M, Islam M R, Karthikeyan B, et al. Crop Protection, 2012, 42: 141.
- [ 6 ] HUANG Shuang-ping, QI Long, MA Xu, et al(黄双萍, 齐 龙, 马 旭, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(1): 212.
- [ 7 ] Kuska M T, Mahlein A K. European Journal of Plant Pathology, 2018, 152(4): 1.
- [ 8 ] Bai X, Li X, Fu Z, et al. Computers & Electronics in Agriculture, 2017, 136: 157.
- [ 9 ] Ma J, Du K, Zhang L, et al. Computers & Electronics in Agriculture, 2017, 142(142): 110.
- [10] Sui Y Y, Wang Q Y, Yu H Y. Spectroscopy & Spectral Analysis, 2016, 36(6): 1779.
- [11] Li H N, Feng J, Yang W P, et al. Spectrum-Based Method for Quantitatively Detecting Diseases on Cucumber Leaf. 4th International Congress on Image and Signal Processing. IEEE, 2011, 4: 1971.
- [12] West A G, Goldsmith G R, Matimati I, et al. Rapid Communications in Mass Spectrometry, 2011, 25(16): 2268.
- [13] LIU Yan-de, XIAO Huai-chun, SUN Xu-dong, et al(刘燕德, 肖怀春, 孙旭东, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(2): 528.
- [14] ZHAO Juan, PENG Yan-kun(赵 娟, 彭彦昆). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(7): 279.
- [15] TANG Xiao-dong, LIU Man-hua, ZHAO Hui(汤晓东, 刘满华, 赵 辉, 等). Journal of Electronic Measurement and Instrument(电子测量与仪器学报), 2010, 24(4): 385.
- [16] Mizushima A, Lu R. Computers & Electronics in Agriculture, 2013, 94(94): 29.

# Segmentation and Detection of Cucumber Powdery Mildew Based on Visible Spectrum and Image Processing

BAI Xue-bing, YU Jian-shu, FU Ze-tian, ZHANG Ling-xian, LI Xin-xing\*

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

**Abstract** Powdery mildew, as a kind of cucumber disease with high outbreak frequency, spreads very fast, often leads to yield reduction and can't achieve the expected economic benefits. Especially in serious years of disease outbreak, the reduction of cucumber in some areas was as high as 20%. This paper proposed a subinterval interval partial least squares regression (SI-PLSR) based on visible spectrum image for cucumber powdery mildew non-destructive detection. We used Canon EOS 800D and Ocean Optics USB2000+ optical fiber spectrometer to collect visible spectral images and reflectivity curves of 200 cucumber powdery mildew leaves. Firstly, we used wavelet transform and watershed algorithm to extract the target leaves from the real-time visible spectral images of cucumber powdery mildew leaves. Secondly, The Otsu algorithm optimized by Gauss fitting was used to segment the powdery mildew lesion. Thirdly, we established the PLSR in 350~1100 nm band and calculated the cross validation root-mean-square error ( $RMSE_{CV}$ ). At the other hand, 350~1100 nm was divided into 20 sub-intervals, and established the PLSR independently. The sub-intervals of  $RMSE_{CV}$  smaller than the full band were selected to form the joint interval. Finally, the SI-PLSR model was established based on powdery mildew lesions images and joint interval. Results show that 188 target leaves were extracted from 200 susceptible leaves visible spectral images successfully of which 157 were more than 95% and 31 were between 90% and 95%. The success rate was 94.00%. The average misclassification rate of powdery mildew was 5.81%. The average false negative was 1.55% and the average false positive was 4.26%. PLSR was established for 20 sub-intervals, and the results showed that the  $RMSE_{CV}$  values of the 5, 6, 7, 11, 12, 13 and 19 sub-intervals were lower than those of the full-band modeling, indicating that the spectral information of these seven sub-intervals contributed greatly to the identification of powdery mildew, which was relative to the wavebands of 470~520, 530~580 and 700~780 nm showing peaks. Therefore, these 7 sub intervals should be selected to establish the joint interval. The principal component number of SI-PLSR model was 7.  $R_c$ ,  $R_v$  and  $RMSEC$ ,  $RMSEV$  were 0.975 2, 0.907 3 and 0.919 5, 1.091. Compared with the full band PLSR model, the  $R_c$  and  $R_v$  of SI-PLSR was closer to 1, and the  $RMSEC$  and  $RMSEV$  were smaller. The above results showed that the SI-PLSR model proposed in this paper which effectively removed redundant information in visible spectral data and enhanced the stability of the model can be used to identify cucumber powdery mildew quickly and accurately, providing a method and reference for the diagnosis of cucumber diseases.

**Keywords** Visible spectrum; Subinterval interval; SI-PLSR; Computer vision

(Received Sep. 2, 2018; accepted Feb. 4, 2019)

\* Corresponding author