

# 基于动态集成的黄瓜叶部病害识别方法

王志彬<sup>1,2</sup> 王开义<sup>1,2</sup> 王书锋<sup>1,2</sup> 王晓锋<sup>1,3</sup> 潘守慧<sup>1,2</sup>

(1. 北京农业信息技术研究中心, 北京 100097; 2. 农业部农业信息技术重点实验室, 北京 100097;

3. 北京工业大学信息学部, 北京 100124)

**摘要:** 对作物病害类型的准确识别是病害防治的前提。为提高病害识别的准确度,以黄瓜叶部病害识别为例,提出一种基于动态集成的作物叶部病害种类的识别方法。首先利用图像分块策略提取病害图像的75维颜色统计特征,然后采用不一致度量方法对构建的10个BP神经网络单分类器进行差异性度量,并按照差异性大小进行排序,最后根据分类器的可信度,动态选择差异性大的分类器子集对病害图像进行集成识别。在由512幅白粉病、霜霉病、灰霉病和正常叶片4类黄瓜叶片组织图像构成的测试集上,所提方法的识别错误率为3.32%,分别比BP神经网络、SVM、Bagging、AdaBoost算法降低了1.37个百分点、1.56个百分点、1.76个百分点、0.78个百分点。试验结果表明:所提方法能够实现黄瓜叶部病害种类的准确识别,可为其它作物病害的识别提供借鉴。

**关键词:** 黄瓜; 叶部病害; 图像识别; 集成学习; 差异性度量; 动态选择

**中图分类号:** S431.9; TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2017)09-0046-07

## Recognition Method of Cucumber Leaf Diseases with Dynamic Ensemble Learning

WANG Zhibin<sup>1,2</sup> WANG Kaiyi<sup>1,2</sup> WANG Shufeng<sup>1,2</sup> WANG Xiaofeng<sup>1,3</sup> PAN Shouhui<sup>1,2</sup>

(1. Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China

2. Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100097, China

3. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

**Abstract:** Crop disease is one of the most important influencing factors for agricultural high yield and high quality. Accurate classification of diseases is a key and basic step for early disease monitoring, diagnostics and prevention. The optimal individual classifier design is currently the common limitation in most crop disease recognition methods based images. To improve the accuracy and stability of disease identification, a disease recognition method of cucumber leaf images via dynamic ensemble learning was proposed. The approach consisted of three major stages. Firstly, totally 75-dimension color features of leaf image were extracted with image block processing. Secondly, a disagreement approach was used to measure the diversity among 10 classifiers of neural networks with an ensemble technique, where the classifiers were ordered according to the diversity. Finally, with the confidence of classifiers, a classifier subset was dynamically selected and integrated to identify the images of crop leaf diseases. To verify the effectiveness of the proposed method, classification experiments were performed on images of four kinds of cucumber leaf tissues, including 512 samples composed of powdery milder, downy mildew, gray mold and normal leaf. The experimental results showed that the recognition error rate of the proposed method was 3.32%, compared with those of BP neural network, SVM, Bagging and AdaBoost methods, it was reduced by 1.37 percentage point, 1.56 percentage point, 1.76 percentage point and 0.78 percentage point, respectively. The proposed method identified the diseases accurately from cucumber leaf images. Moreover, the method was feasible and effective, and it can also be utilized and modified for the classification of other crop diseases.

**Key words:** cucumber; leaf diseases; image recognition; ensemble learning; diversity measure; dynamic selection

收稿日期: 2017-03-27 修回日期: 2017-07-13

基金项目: 国家自然科学基金项目(61403035、71301011)和北京市自然科学基金项目(9152009)

作者简介: 王志彬(1983—),男,助理研究员,博士,主要从事作物病虫害图像识别技术研究,E-mail: wangzb@nrcita.org.cn

通信作者: 王开义(1974—),男,研究员,博士,主要从事农业信息化技术研究,E-mail: wangky@nrcita.org.cn

## 引言

近年来,我国蔬菜产业得到稳步发展,播种面积呈逐年递增。据联合国粮农组织统计,我国蔬菜播种面积和产量分别占世界的43%和49%,均居世界第一<sup>[1]</sup>。在蔬菜生产过程中,病害是制约蔬菜高产、优质、高效益的重要因素,一般可造成蔬菜产量损失的10%~30%,病害严重时损失可达60%~70%甚至绝产<sup>[2-3]</sup>。因此,快速准确地识别作物病害,采取有效的防治措施,对确保蔬菜产量和质量具有十分重要的社会意义和经济价值。

传统的病害识别方法主要依靠植保专家的个人经验和病理学分析,存在实时性差、工作效率低、识别结果主观性强等问题,且往往会耽误病害防治,易造成农药的误用、滥用,缺乏科学合理的病害防治方法<sup>[4-5]</sup>。随着农业信息化进程的推进,基于机器视觉的作物病害识别方法能够显著克服上述问题,成为该领域当前研究的热点<sup>[6-9]</sup>。该方法采用机器视觉和图像分析技术在获取病害叶片图像的同时可将病害种类、危害程度和防治方法同步显示给用户,实现作物病害的无损检测、快速诊断。目前,常用于作物病害图像识别的方法主要包括Fisher法、D-S证据论法、分类规则法、神经网络法和支持向量机(Support vector machine, SVM)法等。

文献[10-20]对作物病害诊断时,大多采用单一分类器进行识别,这需要选取最优的单一特征或特征组合,同时设计出最优的单分类器才能取得最佳的识别效果。在实际应用中,存在最优特征集和最优分类器难以获取的问题,且识别算法极易受到参数、训练集等因素的影响,算法缺乏足够的稳健性。为此,研究者提出基于集成学习的病害识别方法,利用多个单分类器集成的方式来实现作物病害的识别,以获得更好的识别效果<sup>[21-23]</sup>。但是,该方法存在集成规模难以确定的问题,且多采用静态的集成方式,未能充分利用各个单分类器之间的互补性去除冗余实现最优集成识别。

本文以黄瓜白粉病、霜霉病、灰霉病3种病害为研究对象,提出一种基于动态集成的作物叶部病害种类的识别方法,以期为实现田间作物病害的准确识别提供新的技术和方法。

## 1 材料与方法

### 1.1 试验材料

试验选取黄瓜的3种常见病害:白粉病、霜霉病、灰霉病,和黄瓜的健康叶片图像进行识别。根据病情发展,可将病害分为发病初期、发病中期和发

病晚期3个阶段。由于发病初期病症往往难以察觉而发病晚期则存在叶片枯死的问题,因此,本文采用叶部病症明显的发病中期图像进行试验。

采用的黄瓜叶片图像由2个途径获取。一是在大田开放环境下,采用人工手持数码相机(SONY DSC-W35)拍摄方法收集了北京市小汤山国家精准农业研究示范基地的黄瓜病害叶片图像。在拍摄图像时,相机设置为微距模式,自动调节焦距和光圈,自动白平衡,关闭闪光灯。选择表面洁净的黄瓜叶片并在光照适中的晴天进行图像采集,避免光照直射和避开刮风、降雨的天气。二是从公共网站(中国农业网, <http://www.agronet.com.cn/>)<sup>[24]</sup>上收集黄瓜叶片图像。在植保专家的指导下,主要收集黄瓜白粉病、霜霉病、灰霉病3种病害中期的叶片图像和黄瓜健康叶片图像,并对病害的准确性和唯一性进行判别。部分黄瓜叶片图像如图1所示。



图1 黄瓜叶片图像

Fig. 1 Images of cucumber leaf

采用上述2种方式进行样本图像的收集,较容易获得不同地点、不同季节、不同生长时期的黄瓜叶片图像,以有效地验证本文算法。

### 1.2 特征提取方法

叶部病害识别时,经常采用的特征有颜色、形状、纹理等<sup>[10,25]</sup>。各种特征中,以颜色信息最为直观和有效。作物染病后,健康叶片和病斑的颜色往往存在一定的差异。目前,已有多种识别算法据此实现病害种类的准确识别,而且对于部分颜色类似的病斑也有较好的识别效果<sup>[26-28]</sup>。根据黄瓜白粉病、霜霉病、灰霉病和健康叶片图像的特点<sup>[3]</sup>,本文在RGB颜色空间上,利用分块图像的红色(R)、绿色(G)、蓝色(B)3个颜色分量的统计信息,构建病害识别算法的有效特征集,进而实现病害特征的量化、数字化表示。特征提取方法如下<sup>[29]</sup>:

首先,沿图像水平方向和垂直方向进行均匀划分,分成 $M \times N$ 个大小相等的子图像。经过图像划分后,图像 $I$ 可以用集合表示为

$$I = \{I_{11}, I_{12}, \dots, I_{MN}\} \quad (1)$$

其中,元素 $I_{ij}$ ( $i = 1, 2, \dots, M; j = 1, 2, \dots, N$ )是图像 $I$

经过划分后得到的每一个子图像。

然后,对任意子图像  $I_{ij}$ ,利用该图像在  $R$ 、 $G$ 、 $B$  颜色通道上的颜色值和的平均值作为其颜色特征。假设  $I_{ij}$  高为  $m$  像素,宽为  $n$  像素,则其颜色特征计算公式为

$$\begin{cases} \bar{r} = \frac{\sum_{l=1}^m \sum_{z=1}^n r_{lz}}{mn} \\ \bar{g} = \frac{\sum_{l=1}^m \sum_{z=1}^n g_{lz}}{mn} \\ \bar{b} = \frac{\sum_{l=1}^m \sum_{z=1}^n b_{lz}}{mn} \end{cases} \quad (2)$$

式中  $r_{lz}$ —— $I_{ij}$  中像素点  $(l, z)$  在  $R$  颜色通道上的颜色值

$g_{lz}$ —— $I_{ij}$  中像素点  $(l, z)$  在  $G$  颜色通道上的颜色值

$b_{lz}$ —— $I_{ij}$  中像素点  $(l, z)$  在  $B$  颜色通道上的颜色值

$\bar{r}$ —— $I_{ij}$  中像素点在  $R$  颜色通道上颜色值和的平均值

$\bar{g}$ —— $I_{ij}$  中像素点在  $G$  颜色通道上颜色值和的平均值

$\bar{b}$ —— $I_{ij}$  中像素点在  $B$  颜色通道上颜色值和的平均值

子图像  $I_{ij}$  就可以表示为一个三元向量

$$\mathbf{v}_{ij} = (\bar{r}, \bar{g}, \bar{b}) \quad (3)$$

最后,经过子图像划分和特征提取,图像  $I$  可以表示为

$$\mathbf{A}_I = \begin{bmatrix} \mathbf{v}_{11} & \mathbf{v}_{12} & \mathbf{v}_{13} & \cdots & \mathbf{v}_{1N} \\ \mathbf{v}_{21} & \mathbf{v}_{22} & \mathbf{v}_{23} & \cdots & \mathbf{v}_{2N} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{v}_{M-1,1} & \mathbf{v}_{M-1,2} & \mathbf{v}_{M-1,3} & \cdots & \mathbf{v}_{M-1,N} \\ \mathbf{v}_{M,1} & \mathbf{v}_{M,2} & \mathbf{v}_{M,3} & \cdots & \mathbf{v}_{M,N} \end{bmatrix} \quad (4)$$

式中  $\mathbf{A}_I$ ——图像  $I$  的特征矩阵

根据特征矩阵  $\mathbf{A}_I$ , 图像  $I$  可以表示为特征向量

$$\mathbf{V}_I = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_M) \quad (5)$$

其中,元素  $\mathbf{v}_i$  对应式(4)中矩阵的第  $i$  行。 $\mathbf{V}_I$  即为所提取的作物叶部病害特征向量。

上述特征提取方法能够兼顾图像的整体和局部特性,从而可以更准确地表示叶片不同组织之间的特征差异。

### 1.3 病害识别算法

在对作物叶部病害识别时,采用集成学习技术能够获得更好的识别效果<sup>[30-32]</sup>。但在实际应用中,

作物叶部病害识别是一个实时性过程。在该过程中,不但要保证识别的准确性,而且要保证识别结果的可靠性。若在集成系统中,采用全部的分类器对叶部病害组织进行识别,必然会影响系统的识别性能,缺乏识别的针对性,且对识别结果的可靠性缺少评估。为此,将动态选择集成<sup>[33]</sup>应用到病害识别中,设计了一种基于分类器动态选择集成的作物叶部病害识别方法,该方法主要包括2部分:分类器排序;分类器的动态选择集成。

#### 1.3.1 分类器排序

分类器的差异性在保证集成系统具有较高识别率和泛化能力的关键,是分类器动态选择的重要依据。现有的差异性度量方法主要分为2大类:成对差异性度量方法和非成对差异性度量方法<sup>[34]</sup>。其中,成对差异性度量方法中的不一致度量为常采用的度量方法。

不一致度量表示2个分类器  $c_q$ 、 $c_k$  在识别  $N$  个样本时,识别不一致时的样本实例占整个测试实例的比例,计算公式为

$$F = \frac{N^{01} + N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (6)$$

式中  $N^{01}$ ——分类器  $c_q$  识别错误、分类器  $c_k$  识别正确时样本所占比重

$N^{10}$ ——分类器  $c_q$  识别正确、分类器  $c_k$  识别错误时样本所占比重

$N^{11}$ ——分类器  $c_q$ 、 $c_k$  均识别正确时样本所占比重

$N^{00}$ ——分类器  $c_q$ 、 $c_k$  均识别错误时样本所占比重

$F$  的取值范围为  $[0, 1]$ ,  $F$  越大,分类器间的差异性越大。推广到  $L$  个分类器,则不一致度量计算公式为

$$\bar{F} = \frac{2}{L(L-1)} \sum_{q=1}^{L-1} \sum_{k=q+1}^L F_{qk} \quad (7)$$

式中  $F_{qk}$ ——分类器  $c_q$ 、 $c_k$  之间不一致度量值

根据上述差异性度量公式,本文将分类器的选择与差异性的度量进行结合,实现分类器子集按照差异性进行排序。分类器排序算法流程图如图2所示。

其排序过程为:首先选择识别率最高的分类器  $c$ ,放入临时变量  $D$  中并排在第1位,其次从剩余的候选分类器集合  $T$  中选择分类器  $e$  排在第2位,选择的标准是它与前面分类器(临时变量  $D$  中)所构成的  $\bar{F}$  最大,然后重复选择  $e$ ,直到所有候选分类器都被选入  $D$ ,各分类器入选的顺序即为排序结果  $P$ 。

按照该排序算法可以保证在对病害图像识别

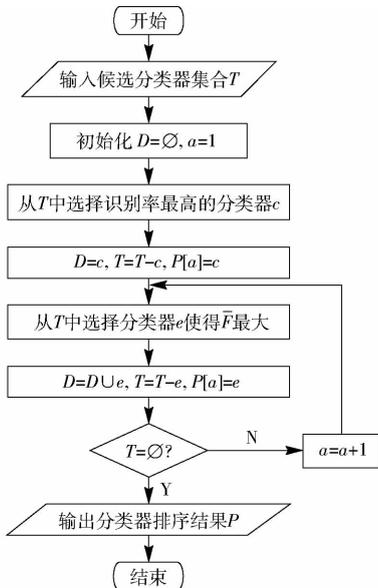


图 2 分类器排序算法流程图

Fig. 2 Flow chart of classifier sorting

时,当需要  $n$  个分类器时,序列  $P$  中的前  $n$  个分类器的组合就是互补性最强的分类器子集;当只需要 1 个分类器时,选择的是所有分类器中识别率最高的分类器。

### 1.3.2 动态选择集成

所采用的动态选择集成方法能够针对不同的待识别样本,挑选出不同数目的分类器进行集成识别,从而有效地提高系统的识别性能。当一个或少数几个分类器就能满足识别可信度要求时,则无需选择更多的分类器;否则依次添加分类器,直至所有分类器都入选;若所有分类器都已入选,仍不能满足识别可信度要求时,则将每次集成时的识别结果进行投票,得票最多的类别即为识别结果,算法流程图如图 3 所示。

在图 3 中,  $y_d$  是第  $d$  个参与集成的分类器的输出结果,  $Y = (o_1, o_2, \dots, o_\omega)^T$ ,  $\omega$  为类别数,其中每一维分量  $o$  表达了该分量所对应类别在集成后的可信度,  $S_{\max}$  为各分量中的最大值,所对应的类别即为识别结果,  $\theta$  为系统识别结果的可信度,其取值范围为  $[0, 1]$ ,  $Q$  为候选分类器的个数。

按照上述算法,对作物叶部病害识别过程为:首先根据识别精度的需要设定初始可信度阈值  $\theta_0$ ,然后根据分类器排序结果  $P$  选取第 1 个分类器对叶部病害图像样本  $x$  进行识别,当识别结果满足可信度的要求时,则输出识别结果  $R$ ,无需集成其它分类器;否则依次选入  $w$  ( $w \geq 2$ ) 个分类器,并对入选的分类器进行集成,保存集成识别结果  $R_w$ ,当满足输出条件  $S_{\max} \geq w\theta$  时,输出识别结果  $R$ ;若所有分类器都已选入仍不满足输出条件,则对每次集成时的识别结果  $R_w$  进行投票,得票最多的类别即为该叶部病

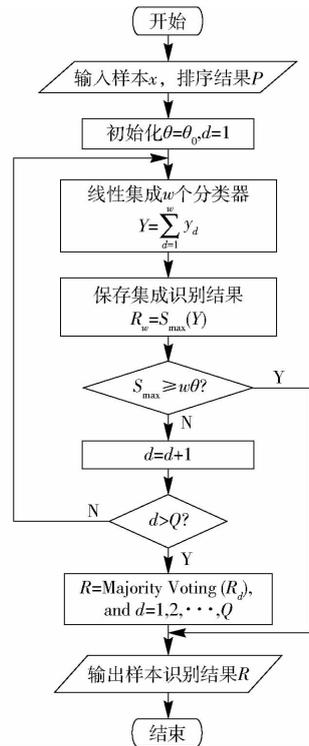


图 3 分类器动态选择集成算法流程图

Fig. 3 Flow chart of dynamic selective ensemble of classifiers

害图像样本的最终识别结果。

## 2 结果与分析

在 Matlab 2011 环境下,通过编程实现本文所提的病害识别算法,并以实际黄瓜叶片图像为测试对象进行试验验证。试验中所采用的计算机配置为: Intel Pentium 4 CPU 3.0 GHz 中央处理器, 2 GB 内存, Windows XP 操作系统。

### 2.1 样本集

试验用到的图像全部来源于所建立的黄瓜叶片图像库。在植保专家的指导下,从不同黄瓜叶片上截取部分典型组织图像,构成样本集。该样本集包括 3 部分:训练集、测试集和验证集,共 2 443 幅图像。其中,训练集包含 1 721 幅图像,测试集包含 512 幅图像,验证集包含 210 幅图像。样本集具体构成如表 1 所示。试验中,将黄瓜叶部组织图像分为 4 类:白粉病、霜霉病、灰霉病和正常叶片,采用 75 维的颜色特征。

表 1 黄瓜叶部病害样本集

Tab. 1 Sample sets of cucumber leaf diseases

| 病害种类 | 训练集   | 测试集 | 验证集 |
|------|-------|-----|-----|
| 白粉病  | 426   | 122 | 50  |
| 灰霉病  | 410   | 110 | 50  |
| 霜霉病  | 415   | 119 | 50  |
| 正常叶片 | 470   | 161 | 60  |
| 合计   | 1 721 | 512 | 210 |

## 2.2 单分类构建

本文采用3层BP神经网络作为单分类器对黄瓜叶片组织图像进行识别。其中,BP神经网络采用75-5-4三层结构,学习速率为0.5,惯性系数设为0.8,目标误差为0.001,迭代次数设置为500,输出层的神经元采用S型变换函数将输出结果转换为

$[0,1]$ 上的可信度。

为构建更多的单分类器,通过重采样技术对训练集有放回地抽取,为每个单分类器都构建一个跟训练集同样大小但各不相同的训练集,从而训练出不同的单分类器。本文随机产生10个BP单分类器,其识别错误率如表2所示。

表2 各单分类器在测试集上的识别错误率

Tab.2 Recognition error rate of individual classifiers for test set

| 分类器     | C1    | C2    | C3    | C4   | C5   | C6   | C7   | C8    | C9   | C10  |
|---------|-------|-------|-------|------|------|------|------|-------|------|------|
| 识别错误率/% | 12.11 | 25.59 | 12.89 | 4.69 | 4.69 | 8.20 | 9.77 | 15.43 | 5.27 | 6.45 |

由表2可知,利用重采样技术获得的10个BP神经网络单分类器各不相同,性能最高的BP神经网络在测试集上的识别错误率为4.69%,识别错误率最高为25.59%,平均识别错误率为10.51%。

## 2.3 动态集成识别结果

在单分类器设计完成后,采用验证样本集对上述单分类器进行差异性度量和排序,其结果如表3所示。

表3 分类器排序结果

Tab.3 Ranking results of classifier sorting

| 分类器  | C1    | C2    | C3    | C4    | C5 | C6    | C7    | C8    | C9    | C10   |
|------|-------|-------|-------|-------|----|-------|-------|-------|-------|-------|
| 排序结果 | 3     | 2     | 7     | 8     | 1  | 9     | 6     | 5     | 4     | 10    |
| 差异度  | 0.216 | 0.262 | 0.151 | 0.142 | 0  | 0.134 | 0.158 | 0.171 | 0.188 | 0.125 |

按照表3的排序结果,在测试样本集上,采用动态选择集成算法对黄瓜叶部病害图像进行识别,并与常用的病害识别算法SVM、经典集成学习算法Bagging、AdaBoost进行对比试验。试验结果如表4所示。

表4 各算法识别性能对比

Tab.4 Recognition performances comparison of different methods

| 分类器     | C5    | SVM   | Bagging | AdaBoost | 本文方法  |
|---------|-------|-------|---------|----------|-------|
| 识别错误率/% | 4.69  | 4.88  | 5.08    | 4.10     | 3.32  |
| 分类器数/个  | 1     | 1     | 10      | 10       | 1.041 |
| 算法耗时/s  | 0.066 | 0.106 | 0.308   | 0.318    | 0.544 |

在表4中,SVM采用运算复杂度小、稳定性高的径向基函数(Radial basis function, RBF)作为核函数,并通过试验确定惩罚因子 $C=30$ ,核函数 $g=5$ ;Bagging算法和AdaBoost算法则均采用BP神经网络结构作为单分类器,分别产生10个单分类器并对其进行线性集成。本文采用的动态选择集成算法的可信度阈值设置为 $\theta_0=0.6$ 。

由表4可以看出,本文提出的病害识别算法在测试集上的识别错误率为3.32%,分别比BP神经网络、SVM、Bagging、AdaBoost算法降低了1.37个百分点、1.56个百分点、1.76个百分点、0.78个百分点。所提算法具有更好的识别效果,其原因为:在对病害叶片图像进行识别时,本文采用动态集成的方

式,即针对每一病害叶片图像,均动态选择出性能最优的分类器或组合进行识别(参与集成的分类器个数不同),这样的识别更有针对性,获得了更好的识别效果。但是,该算法存在耗时较长的问题。Bagging算法与BP神经网络识别性能最优的C5相比,Bagging算法对产生的10个单分类器进行线性集成,但识别结果未能达到最优,则进一步说明产生的多分类器之间存在冗余,直接对所有分类器进行集成,会融合一些无效的决策而影响最终的识别结果。AdaBoost与Bagging算法相比,AdaBoost算法识别性能更优,错误率降低了0.98个百分点,其原因在于:AdaBoost对产生的分类器进行加权集成,即差异性大的分类器其集成权值更高,从而获得更高的识别性能。此外,本文所提算法在初始可信度为0.6时,对每一样本进行识别时,平均集成分类器个数为1.041个,这说明在此可信度下,大部分样本的识别精度均已满足可信度的要求,只有少数难以识别的样本需进行动态选择集成,因此所提算法结构更灵活、识别更有针对性。

可信度是影响算法识别性能的重要参数,为研究其与识别错误率、算法耗时之间的关系,本文进行了3种可信度情况下的算法性能测试,试验结果如表5所示。

由表5可以看出,算法的初始可信度越大,参与集成的分类器个数越多,算法的耗时越大,算法的识

表5 初始可信度对算法性能的影响

Tab.5 Influences of initial confidence on algorithm performance

| 可信度  | 分类器数/个 | 算法耗时/s | 识别错误率/% |
|------|--------|--------|---------|
| 0.60 | 1.041  | 0.544  | 3.32    |
| 0.95 | 1.932  | 0.611  | 5.08    |
| 1    | 10     | 0.926  | 5.27    |

别错误率越高,但识别结果越可靠。改变初始可信度会影响算法识别性能的原因在于:改变初始可信度阈值,会影响算法识别结果的可靠性。当对算法识别的可靠性要求越高时,单一分类器的识别结果若不能满足该识别精度要求时,必然要求参与集成的分类器个数增加,从而增加算法的执行时间,且识别可靠性要求高,易导致算法识别精度的降低;反之亦然。所以在实际应用中,应根据应用场景对识别结果可靠性设置合适的初始可信度,以兼顾识别精度和效率。

此外,颜色是区分作物病害种类的一个重要特征。传统人工诊断方法由于病害诊断标准含糊不清,加之人眼分辨颜色的能力有限,有时会妨碍对病害的判断。但是,通过颜色特征提取病害图像后,能够对病害特征进行数字化、精细化、量化的描述,从而有利于实现病害种类的准确识别。本文在RGB颜色空间上,利用分块图像的R、G、B 3个颜色分量的统计信息,构建了75维有效特征向量集,取得较好的识别效果。随着多特征融合技术的发展,利用多特征实现作物病害的准确识别已得到越来越多研究者的关注<sup>[12,19,35]</sup>。但是,在病害识别时,选择

何种特征以及采用何种融合方法更能表达作物病害信息,仍需要进一步的研究。此外,在算法实际应用时,田间光照强度的不均特别是光照直射,会影响算法的识别性能<sup>[22,25,36]</sup>,如何克服光照影响,提高算法的鲁棒性也将是本文算法改进的一个重要方向。

### 3 结论

(1)在512幅由白粉病、霜霉病、灰霉病和正常叶片4类黄瓜叶片组织图像构成的测试集上,所提出的识别算法的识别错误率为3.32%,分别比BP神经网络、SVM、Bagging、AdaBoost算法降低了1.37个百分点、1.56个百分点、1.76个百分点、0.78个百分点。

(2)在病害识别时,采用Bagging算法比10个BP单分类器均值的识别错误率低5.43个百分点,而AdaBoost算法比Bagging算法识别错误率降低了0.98个百分点,集成多个分类器通常能够获得更好的识别效果,且分类器之间差异性越大,集成识别效果越明显。

(3)所采用的特征提取方法能够很好的表达黄瓜叶部不同病害组织的颜色特征,且计算简单、实用性强。

(4)算法在可信度为0.6、0.95、1时,算法平均参与集成分类器个数依次为:1.041、1.932、10个;算法的平均耗时为0.544、0.611、0.926s;算法的识别错误率为:3.32%、5.08%、5.27%,算法的可信度越大,参与集成的分类器个数越多,算法的耗时越大,算法的识别错误率越高,但识别结果越可靠。

### 参 考 文 献

- 李宝聚. 蔬菜主要病害2013年发生概况及2014年发生趋势[J]. 中国蔬菜, 2014(2): 5-8.
- 腾葳, 柳琪, 郭栋梁. 我国蔬菜病虫害的基本概况及对蔬菜安全质量的影响分析[J]. 食品研究与开发, 2003, 24(5): 3-5.
- 王恒亮, 倪云霞, 李好海, 等. 蔬菜病虫害诊治原色图鉴[M]. 北京: 中国农业科学技术出版社, 2013: 110-177.
- SANKARAN S, MISHRA A, EHSANI R, et al. A review of advanced techniques for detecting plant diseases[J]. Computers and Electronics in Agriculture, 2010, 72(1): 1-13.
- 刘君, 王振中, 李宝聚, 等. 基于图像处理的作物病害自动识别系统的研究[J]. 计算机工程与应用, 2012, 48(13): 154-158.
- 赖军臣, 李少昆, 明博, 等. 作物病害机器视觉诊断研究进展[J]. 中国农业科学, 2009, 42(4): 1215-1221. LAI Junchen, LI Shaokun, MING Bo, et al. Advances in research on computer-vision diagnosis of crop diseases[J]. Scientia Agricultura Sinica, 2009, 42(4): 1215-1221. (in Chinese)
- DIAO Z, ZHAO C, WU G, et al. Review of application of mathematical morphology in crop disease recognition[C]// Computer and Computing Technologies in Agriculture, 2008, 2: 981-990.
- BARBEDO J G A. Digital image processing techniques for detecting, quantifying and classifying plant diseases[J]. SpringerPlus, 2013, 2:660.
- BARBEDO J G A. A review on the main challenges in automatic plant disease identification based on visible range images[J]. Biosystems Engineering, 2016, 144: 52-60.
- 贾建楠, 吉海彦. 基于病斑形状和神经网络的黄瓜病害识别[J]. 农业工程学报, 2013, 29(增刊): 115-121. JIA Jiannan, JI Haiyan. Recognition for cucumber disease based on leaf spot shape and neural network[J]. Transactions of the CSAE, 2013, 29(Supp.): 115-121. (in Chinese)
- PHADIKAR S, SIL J, DAS A K. Rice diseases classification using feature selection and rule generation techniques[J]. Computers and Electronics in Agriculture, 2013, 90: 76-85.
- 王献锋, 张善文, 王震, 等. 基于叶片图像和环境信息的黄瓜病害识别方法[J]. 农业工程学报, 2014, 30(14): 148-153. WANG Xianfeng, ZHANG Shanwen, WANG Zhen, et al. Recognition of cucumber diseases based on leaf image and environmental information[J]. Transactions of the CSAE, 2014, 30(14): 148-153. (in Chinese)
- 田凯, 张连宽, 熊美东, 等. 基于叶片病斑特征的茄子褐纹病识别方法[J]. 农业工程学报, 2016, 32(增刊): 184-189.

- TIAN Kai, ZHANG Liankuan, XIONG Meidong, et al. Recognition of phomopsis vexans in solanum melongena based on leaf disease spot features[J]. Transactions of the CSAE, 2016, 32(Supp.): 184-189. (in Chinese)
- 14 BARBEDO J G A, KOENIGKAN L V, SANTOS T T. Identifying multiple plant diseases using digital image processing[J]. Biosystems Engineering, 2016, 147: 104-116.
- 15 HU M, BU X, SUN X, et al. Rape plant disease recognition method of multi-feature fusion based on D-S evidence theory[J]. Mathematical and Computational Applications, 2017, 22(18): 1-7.
- 16 霍迎秋, 唐晶磊, 尹秀珍, 等. 基于压缩感知理论的苹果病害识别方法[J/OL]. 农业机械学报, 2013, 44(10): 227-232. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20131036&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20131036&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2013.10.036.
- HUO Yingqiu, TANG Jinglei, YIN Xiuzhen, et al. Apple disease recognition based on compressive sensing[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(10): 227-232. (in Chinese)
- 17 叶海建, 郎睿, 刘成启, 等. 基于视觉显著性图的黄瓜霜霉病识别方法[J/OL]. 农业机械学报, 2016, 47(5): 270-274. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20160536&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160536&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2016.05.036.
- YE Haijian, LANG Rui, LIU Chengqi, et al. Recognition of cucumber downy mildew disease based on visual saliency map[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(5): 270-274. (in Chinese)
- 18 CAMARGO A, SMITH J S. Image pattern classification for the identification of disease causing agents in plants[J]. Computers and Electronics in Agriculture, 2009, 66(2): 121-125.
- 19 马浚诚, 温皓杰, 李鑫星, 等. 基于图像处理的温室黄瓜霜霉病诊断系统[J/OL]. 农业机械学报, 2017, 48(2): 195-202. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20170226&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170226&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2017.02.026.
- MA Juncheng, WEN Haojie, LI Xinxing, et al. Downy mildew diagnosis system for greenhouse cucumbers based on image processing[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(2): 195-202. (in Chinese)
- 20 张建华, 孔繁洵, 李哲敏, 等. 基于最优二叉树支持向量机的蜜柚叶部病害识别[J]. 农业工程学报, 2014, 30(19): 222-231.
- ZHANG Jianhua, KONG Fantao, LI Zheming, et al. Recognition of honey pomelo leaf diseases based on optimal binary tree support vector machine[J]. Transactions of the CSAE, 2014, 30(19): 222-231. (in Chinese)
- 21 TIAN Y, ZHAO C, LU S, et al. Multiple classifier combination for recognition of wheat leaf diseases[J]. Intelligent Automation and Soft Computing, 2011, 17(5): 519-529.
- 22 许良凤, 徐小兵, 胡敏, 等. 基于多分类器融合的玉米叶部病害识别[J]. 农业工程学报, 2015, 31(14): 194-201.
- XU Liangfeng, XU Xiaobing, HU Min, et al. Corn leaf disease identification based on multiple classifiers fusion[J]. Transactions of the CSAE, 2015, 31(14): 194-201. (in Chinese)
- 23 马晓丹, 关海鹰, 祁广云, 等. 基于改进级联神经网络的大豆叶部病害诊断模型[J/OL]. 农业机械学报, 2017, 48(1): 163-168. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20170121&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170121&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2017.01.021.
- MA Xiaodan, GUAN Haiou, QI Guangyun, et al. Diagnosis model of soybean leaf diseases based on improved cascade neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(1): 163-168. (in Chinese)
- 24 MA J, LI X, WEN H, et al. A key frame extraction method for processing greenhouse vegetables production monitoring video[J]. Computers and Electronics in Agriculture, 2015, 111: 92-102.
- 25 耿长兴, 张俊雄, 曹峥勇, 等. 基于色度和纹理的黄瓜霜霉病识别与特征提取[J]. 农业机械学报, 2011, 42(3): 170-174.
- GENG Changxing, ZHANG Junxiong, CAO Zhengyong, et al. Recognition and features extraction of cucumber downy mildew based on color and texture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3): 170-174. (in Chinese)
- 26 岑喆鑫, 李宝聚, 石延霞, 等. 基于彩色图像颜色统计特征的黄瓜炭疽病和褐斑病的识别研究[J]. 园艺学报, 2007, 34(6): 1425-1430.
- CEN Zhexin, LI Baoju, SHI Yanxia, et al. Discrimination of cucumber anthracnose and cucumber brown speck base on color image statistical characteristics[J]. Acta Horticulturae Sinica, 2007, 34(6): 1425-1430. (in Chinese)
- 27 李旺, 唐少先, 陈荣. 基于颜色特征和支持向量机的黄瓜叶部病害识别[J]. 农机化研究, 2014, 36(1): 73-75.
- 28 田有文, 张长水, 李成华. 基于支持向量机和色度矩的植物病害识别研究[J]. 农业机械学报, 2004, 35(3): 95-98.
- TIAN Youwen, ZHANG Changshui, LI Chenghua. Study on plant disease recognition using support vector machine and chromaticity moments[J]. Transactions of the Chinese Society for Agricultural Machinery, 2004, 35(3): 95-98. (in Chinese)
- 29 SHEN H, HAO H, WEI L, et al. An image based classification method for cataract[C]//2008 International Symposium on Computer Science and Computational Technology, 2008: 583-586.
- 30 张春霞, 张讲社. 选择性集成学习算法综述[J]. 计算机学报, 2011, 34(8): 1399-1410.
- ZHANG Chunxia, ZHANG Jianshe. A survey of selective ensemble learning algorithms[J]. Chinese Journal of Computers, 2011, 34(8): 1399-1410. (in Chinese)
- 31 YIN X, HUANG K, HAO H, et al. A novel classifier ensemble method with sparsity and diversity[J]. Neurocomputing, 2014, 134:214-221.
- 32 KRAWCZYK B, MINKU L L, GAMA J, et al. Ensemble learning for data stream analysis: a survey[J]. Information Fusion, 2017, 37: 132-156.
- 33 YAN Y, YIN X, WANG Z, et al. Sorting-based dynamic classifier ensemble selection[C]//12th International Conference on Document Analysis and Recognition, 2013: 673-677.
- 34 杨春, 殷绪成, 郝红卫, 等. 基于差异性的分类器集成: 有效性分析及优化集成[J]. 自动化学报, 2014, 40(4): 660-674.
- YANG Chun, YIN Xucheng, HAO Hongwei, et al. Classifier ensemble with diversity: effectiveness analysis and ensemble optimization[J]. Acta Automatica Sinica, 2014, 40(4): 660-674. (in Chinese)
- 35 ZHANG S, WU X, YOU Z, et al. Leaf image based cucumber disease recognition using sparse representation classification[J]. Computers and Electronics in Agriculture, 2017, 134: 135-141.
- 36 OUYANG C, LI D, WANG J, et al. The research of the strawberry disease identification based on image processing and pattern recognition[C]//6th International Conference on Computer and Computing Technologies in Agriculture, 2012: 69-77.