

# 基于机器视觉的茶小绿叶蝉识别方法研究

陈晶, 朱启兵, 黄敏, 郑阳

江南大学轻工过程先进控制教育部重点实验室, 江苏 无锡 214122

**摘要** 将机器视觉技术引入到了茶小绿叶蝉的自动识别领域, 以实现茶园中茶小绿叶蝉的准确及时预报。采集了自然场景下黄色诱虫板的图像, 利用超像素分割算法和多 DBSCAN 聚类图像融合的方法对采集的图像进行区域分割, 保证了目标区域的准确性和完整性。在此基础上, 提取了目标图像子区域的  $L$ 、 $a$ 、 $b$  均值和标准差特征, 构建了最小二乘支持向量机(LSSVM)自动识别模型。为解决训练样本中茶小绿叶蝉和其他害虫数量不均衡带来的分类超平面偏移问题, 采用改进 SMOTE 算法和 KS 算法来提高模型对茶小绿叶蝉小样本的识别精度。结果表明, 该算法的整体识别精度可达到 99.03%, 茶小绿叶蝉的查准率可达 91.76%, 为茶小绿叶蝉的实时检测提供了有效途径。

**关键词** 机器视觉; 茶小绿叶蝉; 自动识别; 超像素分割; SMOTE

中图分类号 S431.9

文献标识码 A

doi: 10.3788/LOP55.011502

## Recognition of *Empoasca Flavescens* Based on Machine Vision

Chen Jing, Zhu Qibing, Huang Min, Zheng Yang

Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi, Jiangsu 214122, China

**Abstract** The machine vision technique is introduced to recognize the *Empoasca flavescens* automatically on the yellow sticky traps in natural scenes in order to realize the accurate and timely forecast of *Empoasca flavescens* in tea garden. The superpixel segmentation algorithm and DBSCAN (density-based spatial clustering of applications with noise) cluster algorithm are employed to separate the interesting target regions from background, which ensures the accuracy and completeness of the target area. Then, six classification features, including mean value of  $L$ ,  $a$ , and  $b$  and their standard deviation, are extracted from the marked area in target image. Last, LSSVM (least squares support vector machine) is developed to identify *Empoasca flavescens* from other insects that are captured by sticky traps. As the imbalanced sample number between *Empoasca flavescens* and other insects results in the low classification accuracy, the improved SMOTE (synthetic minority over-sampling technique) algorithm and KS (Kennard-Stone) algorithm are used to improve recognition accuracy of *Empoasca flavescens*. The proposed algorithm achieves 99.03% of the overall recognition accuracy, and the identification accuracy of *Empoasca flavescens* reaches 91.76%. The proposed algorithm can provide an effective way for real-time detection of *Empoasca flavescens*.

**Key words** machine vision; *Empoasca flavescens*; automatic recognition; superpixel segmentation; synthetic minority over-sampling technique

**OCIS codes** 100.3008; 150.1135

## 1 引言

茶小绿叶蝉是我国茶区分布最广、危害最重的一

种茶树害虫, 其成虫和若虫吸取芽叶汁液, 会导致茶树芽叶生长迟缓、焦边、焦叶, 造成茶叶减产<sup>[1]</sup>。喷洒农药是目前茶园针对该虫害采取的主要防治手段<sup>[2]</sup>,

收稿日期: 2017-07-17; 收到修改稿日期: 2017-08-30

基金项目: 国家自然科学基金(61271384, 61275155)、中央高校自主科研项目(JUSRP51730A)、江苏省政策引导类计划(产学研合作)-前瞻性联合研究项目(BY2016022-32)

导师简介: 朱启兵(1973—), 男, 博士, 教授, 硕士生导师, 主要从事农业信息化方面的研究。

E-mail: zhuqib@163.com(通信联系人)

但盲目使用化学农药会造成环境污染和农药残留。因此,及时准确地做好茶小绿叶蝉预测预报和动态掌握茶园虫情是做到适时用药、掌握适合的用药浓度和用药次数的基础性工作<sup>[3]</sup>。传统的预报技术主要是在茶园中放置黄板,利用黄板诱集茶小绿叶蝉<sup>[4]</sup>,通过肉眼观察黄板上茶小绿叶蝉的数量来获取虫情信息。但诱虫板上昆虫的类型和数目众多,且茶小绿叶蝉体积小,给人工视觉观察带来极大困难。

机器视觉技术具有速度快、信息量大的特点,被广泛应用于农业信息化与病虫害检测<sup>[5-8]</sup>,但该技术茶园虫害自动识别领域的研究还未见报道。本文将机器视觉技术引入到茶小绿叶蝉的自动识别领域,采用简单线性迭代聚类(SLIC)超像素分割算法并结合具有噪声的基于密度的空间聚类(DBSCAN)<sup>[9]</sup>实现诱虫板图像的分割,针对复杂环境下茶小绿叶蝉易漏分割的问题,采用多聚类图像融合方法获取目标图像,提取分割后目标图像的 $L$ 、 $a$ 、 $b$ ( $L$ 表示照度,相当于亮度, $L$ 取值为 $0\sim 100$ ;  $a$ 表示从绿色至红色的范围, $a$ 取值为 $-128\sim +127$ ;  $b$ 表示从蓝色至黄色的范围, $b$ 取值为 $-128\sim +127$ )均值和标准差特征,最后利用最小二乘支持向量机(LSSVM)构建分类器对茶园中的小

绿叶蝉进行识别。

茶小绿叶蝉数量和其他害虫数量不均衡导致的分类超平面偏移是影响茶小绿叶蝉识别精度的最主要原因<sup>[10]</sup>。本文利用改进的合成少数类过采样技术(SMOTE)方法生成少数茶小绿叶蝉类样本,并利用 Kennard-Stone(KS)算法对多数类样本(其他昆虫)进行筛选,以减少多数类样本之间的冗余信息,缩减训练样本中正、负样本之间的数量差异。实验结果表明,改进后的方法明显提高了茶小绿叶蝉的识别率,为分析不同时期茶小绿叶蝉的识别数量和预测虫害的实际程度提供理论依据。

## 2 材料与方法

### 2.1 图像获取

本研究所用图片均是采用 Nikond Digital D5200 彩色照相机(分辨率为 $4000\text{ pixel}\times 3000\text{ pixel}$ )在江苏省无锡市茶叶研究所管理的茶园中于自然环境下拍摄的。将黄板放置在茶园的不同区域,在虫害高发期(2016年6月21—27日)每两天拍摄一次。为获取清晰的图像,消除运动模糊的干扰,拍摄时将黄板固定,以减少抖动;拍摄角度多样,拍摄距离范围为 $10\sim 50\text{ cm}$ 。部分实验图片如图1所示。

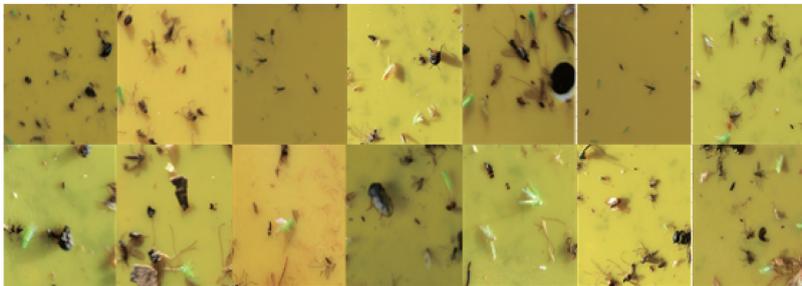


图1 部分实验图片

Fig. 1 Partial experimental pictures

### 2.2 基于 SLIC 超像素分割和多 DBSCAN 聚类图像融合的分割方法

超像素方法对颜色渐变区域能准确地检测到边缘,并且对处理随机不确定场景图像具有很强的稳健性<sup>[11-14]</sup>。SLIC算法相比其他超像素分割算法具有边缘贴合度高、数量可控的优点<sup>[15]</sup>。因此本研究使用SLIC算法对实验图片进行初分割,将背景和不同种类的害虫准确分离,再使用DBSCAN聚类算法<sup>[16]</sup>进行二次聚类。但聚类时阈值过小将导致聚类后的图像包含过多的背景信息,阈值过大则会使一些茶小绿叶蝉聚类成为背景而导致漏分割。因此本研究融合了阈值在两种

不同取值下的聚类结果,保证了分割的准确性和完整性,从而得到去除背景后的目标图像。算法的具体实现过程如下:

1) 初始化聚类中心。假设图像有 $N$ 个像素点,预先设定分割出 $K$ 个超像素。设间距 $S=N/K$ ,选择 $K$ 个初始种子(像素点)。为了将害虫之间相互分离,应尽可能地减少计算时间,这里设置 $K$ 为500。

2) 相似性度量。在颜色空间(CIELAB)提取 $N$ 个像素点,并在 $X$ - $Y$ 坐标系中提取五维特征向量 $c_i=(L_i, a_i, b_i, x_i, y_i)$ 。对五维特征向量构造相似性度量,定义如下:

$$d(i, k) = d_{lab} + \frac{m}{S} d_{xy}, \quad (1)$$

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2}, \quad (2)$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}, \quad (3)$$

式中  $m \in [1, 20]$  为平衡参数, 是平衡颜色值与空间信息在相似度量中的比重, 本研究中  $m$  的值设置为 10;  $d_{lab}$ 、 $d_{xy}$  为像素点之间的色差和空间距离;  $d(i, k)$  为第  $i$  个像素点与第  $k$  个聚类中心之间的相似度;  $l_i$ 、 $a_i$ 、 $b_i$  分别为  $Lab$  空间中第  $i$  个像素点的  $L$  分量、 $a$  分量、 $b$  分量;  $l_k$ 、 $a_k$ 、 $b_k$  分别为  $Lab$  空间中第  $k$  个聚类中心的  $L$  分量、 $a$  分量、 $b$  分量;  $x_i$ 、 $y_i$  分别为第  $i$  个像素点在  $X$ - $Y$  坐标系上的坐标值;  $x_k$ 、 $y_k$  分别为第  $k$  个聚类中心在  $X$ - $Y$  坐标系上的坐标值。

3) 更迭聚类中心。假定在  $X$ - $Y$  平面上聚类中心的关联像素点位于其  $2S \times 2S$  区域内, 则所有像素点被关联到最近的聚类中心后, 新聚类中心更改为同类别中所有像素点五维向量的平均值。重复此过程, 直到收敛, 停止迭代。将聚类中心标签赋给像素点, 形成  $K$  个超像素。

4) DBSCAN 聚类。如果两个超像素之间相连, 就根据(2)式计算两者之间的颜色距离  $d_{lab}$ 。当  $d_{lab}$  小于阈值时, 将两个超像素聚为一类并更新像素点的类别标签。通过试验可知: 当阈值设置为 8 时, 可以得到较完整的目标图像  $I_1$ ; 当阈值设置为 12 时, 可以去除掉绝大多数背景, 得到较为准确的目标图像  $I_2$ 。

5) 多聚类图像融合。对  $I_1$  中的所有连通域进行检测得到标记后的子区域, 遍历  $I_1$  中所有的子区域  $A_i$ , 若子区域与  $I_2$  没有相交的部分, 则将该子区域添加到  $I_2$  中, 即  $I_1 = [A_1, A_2, \dots, A_i, \dots, A_n]$  ( $n$  为  $I_1$  中连通域的个数), 如果  $A_i \cap I_2 = \emptyset$ , 则  $I_2 = A_i + I_2$ 。重复此过程, 就得到融合后的目标图像。该目标图像去除了背景信息, 仅包含含虫区域的信息。

### 2.3 特征提取

茶小绿叶蝉具有不同的大小和形状, 在黄板上的形态各异, 并会与一些昆虫彼此重叠, 因此难以通过形状特征来识别它们。通过观察发现, 茶小绿叶蝉身体的颜色主要呈现为绿色, 但颜色的深浅会随着生长阶段和光照的强弱而发生变化。 $Lab$  颜色空间可以准确描述颜色深浅以及亮度的变化, 体现色相间的差异。李冠林等<sup>[17]</sup>利用  $a$ 、 $b$  层二维数据空

间的颜色差异, 综合运用  $K$ \_means 聚类 and 图像形态学方法对葡萄病害彩色图像实现了较准确的分割。因此本研究将分别提取目标图像中各子区域所有像素点的  $L$ 、 $a$ 、 $b$  均值及其对应标准差 ( $\sigma_L$ 、 $\sigma_a$ 、 $\sigma_b$ ) 用于茶小绿叶蝉的识别。图 2 给出了茶小绿叶蝉区域和其他害虫区域的特征差异盒图。从图 2 中可以看出, 茶小绿叶蝉和其他害虫在这些特征上基本能够相互分离, 在  $L$ 、 $a$ 、 $b$  和  $\sigma_a$  这四个特征上表现得尤为明显。说明利用六个颜色特征可以较好地识别茶小绿叶蝉。

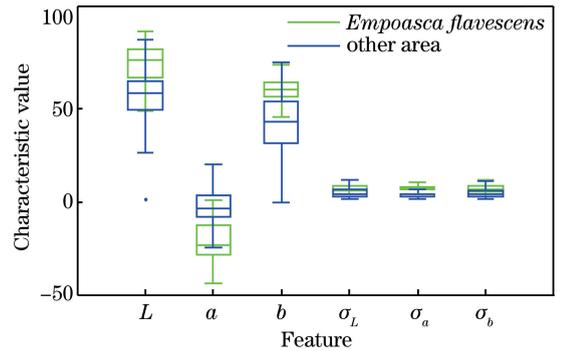


图 2 茶小绿叶蝉区域和其他害虫区域的特征值盒图

Fig. 2 Box-plot of color features for *Empoasca flavescens* and other insects

### 2.4 基于改进的 SMOTE 算法和 KS 算法相结合的茶小绿叶蝉识别

支持向量机是模式识别和机器学习的一种重要的分类方法<sup>[18]</sup>。LSSVM 将求解二次规划问题转化为求解线性方程组问题, 提高了求解问题的速度和收敛精确度<sup>[19]</sup>, 是支持向量机的一种改进。本研究利用 LSSVM 算法, 将目标图片中各子区域的六个特征值 ( $L$ 、 $a$ 、 $b$ 、 $\sigma_L$ 、 $\sigma_a$ 、 $\sigma_b$ ) 作为输入向量建立分类模型对茶小绿叶蝉进行识别, 核函数为径向基函数。该算法包括两个主要参数, 即  $c$  和  $\sigma$ , 前者是正则化参数, 后者表示核函数的径向宽度, 使用十折交叉验证法对这两个参数进行优化。

在目标图像中, 茶小绿叶蝉的样本个数较少, 而其他昆虫的样本个数较多, 使得分类超平面向少数类样本偏移, 并最终导致对茶小绿叶蝉的识别精度较低。为了解决分类问题中数据不均衡的问题, 基于样本均衡化思想的 SMOTE 被广泛运用, 以提高不均衡样本条件下分类器的分类精度<sup>[20-21]</sup>。

传统的 SMOTE 算法是利用全体少数类样本进行样本生成的, 并最终达到样本均衡。由于最终决定分类结果的是分类面附近的少量样本, 因此

SMOTE 的合成机制会产生部分冗余样本,存在不合理的地方。衣柏衡等<sup>[22]</sup>利用 SMOTE 算法对原始分类器的错分样本进行生成,克服了传统算法盲目生成少数类样本的缺陷。本研究采用该方法对错分样本进行人工生成。与此同时,为了减少多数类样本的冗余性,引入 KS 算法来选择具有代表性的多数类样本,以改善不同类别样本的均衡性<sup>[23]</sup>,具体算法设计如下:

1) 设原始训练集  $T = \{T_p, T_n\}$ , 其中  $T_p$  为正类样本集且  $T_p = \{(x_1, 1), (x_2, 1), \dots, (x_m, 1)\}$ ,  $T_n$  为负类样本集且  $T_n = \{(x_{m+1}, -1), (x_{m+2}, -1), \dots, (x_i, -1)\}$ 。

2) 用 LSSVM 对  $T$  进行分类,并用原始数据  $T$  对模型进行验证,将错分样本加入集合  $P_{\text{mis}}$  中。

3) 判断  $P_{\text{mis}}$  是否为空集。若是,结束算法;若不是,令集合  $S = T, P_{\text{MIS}} = P_{\text{mis}}$ ,进入下一步。

4) 用 SMOTE 算法通过表达式  $x_{\text{new}} = x_i + \text{rand}(0, 1) \times (x_i - x_j)$  对  $P_{\text{MIS}}$  中每个样本人工合成一次,将合成的新样本  $x_{\text{new}}$  添加到集合  $S$  中。其中,  $x_i$  为  $P_{\text{MIS}}$  中的样本,  $x_j$  为从  $x_i$  的  $k$  个近邻同类样本中随机选取的一个样本,  $\text{rand}(0, 1)$  表示取  $0 \sim 1$  之间的随机数。

5) 用 LSSVM 模型对  $S$  进行分类,并用原始数据  $T$  对模型进行验证,找出所有的错分样本,更新集合  $P_{\text{mis}}$ ,且  $P_{\text{MIS}} = \{P_{\text{MIS}}, P_{\text{mis}}\}$ 。

6) 重复 4) 和 5),直到集合  $P_{\text{mis}}$  不再产生新的错分样本或  $P_{\text{mis}}$  为空集。

7) 以样本间的欧氏距离为基础,利用 KS 算法将选择的样本与剩余样本之间的欧氏距离最大化,筛选出分布范围广泛且代表性强的样本。从  $T_n$  中选出具有代表性的样本  $T_{n1}$ ,与  $T_p, P_{\text{MIS}}$  构成新的训练集  $T_1$ ,即  $T_1 = \{T_{n1}, T_p, P_{\text{MIS}}\}$ 。

8) 使用 LSSVM 对新的训练集  $T_1$  进行建模。

总体算法流程如图 3 所示。

## 2.5 精确度测量标准

对于均衡的数据集,通常将整体分类误差作为模型性能的评估指标。但对于非均衡数据集,整体误差难以反映分类器在少数类样本中的表现。为了评价分类器对非均衡数据集的分类效果,引入少数类样本查全率 TPR、多数类样本查全率 TNR、少数类样本查准率 Precision、综合 G-mean、少数类样本 F-measure<sup>[22]</sup> 以及整体识别精度 Percent\_test 来多角度评价分类性能,这些指标的值越大,分类效果越优。

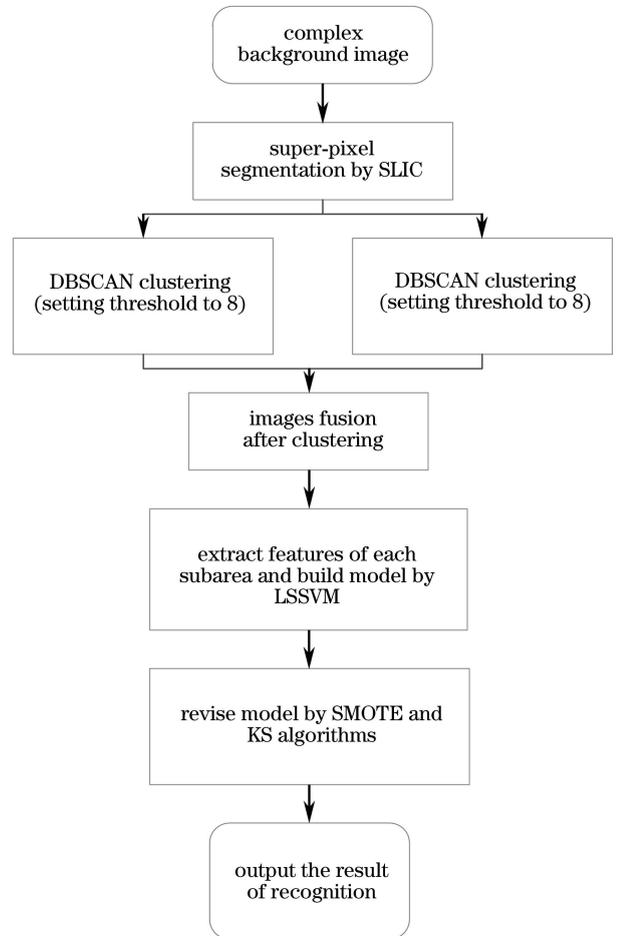


图 3 茶小绿叶蝉识别流程

Fig. 3 Process of *Empoasca flavescens* recognition

## 3 结果与讨论

### 3.1 所提分割算法与其他分割算法的比较

图 4 为传统 Ostu 算法以及 K\_means 聚类算法的分割比较结果。原始图像如图 4(a) 所示,图中红圈标记的区域为要识别的茶小绿叶蝉。图 4(b) 为利用 Ostu 算法求得自适应阈值对图像进行分割的结果。从图 4(b) 中可以看出,当光照比较强烈时,Ostu 算法不仅不能分割出茶小绿叶蝉区域,而且还会因为分割不准确而导致目标区域也包含了很多背景。图 4(c) 为基于 K\_means 聚类的分割结果,一类为背景,一类为目标区域。从图 4(c) 中可以看出,分割出的区域包含较少的背景信息,但还是存在茶小绿叶蝉被漏分割的情况。图 5 为结合超像素分割和 DBSCAN 聚类的分割结果。从图 5(a)、(b) 可以看出,在进行聚类时,阈值的不同取值会使聚类结果产生较大差异。图 5(a) 是阈值为 8 时的聚类结果,虽然能保证目标区域都能被分割出来,但包含了过多的背景信息,难以保证特征提取的有效

性。图 5(b)是阈值为 12 时的分割结果,可以看出分割出的目标区域较为准确,仅包含有少量的背景。但是在光照比较强烈时,会出现某些茶小绿叶蝉区域被漏分割的情况。当对两种不同阈值所得到的分割结果进行融合时,图 5(c)所示的多聚类融合图像既能保证目标区域分割的准确性,又能使易被漏分

割的茶小绿叶蝉区域被分割出来。

表 1 为该三种分割算法对 151 幅图像进行分割后得到的目标区域中茶小绿叶蝉的个数。从表 1 中可以看出,所提多聚类图像融合的方法能有效避免茶小绿叶蝉被漏分割,为后续的目标识别奠定了基础。

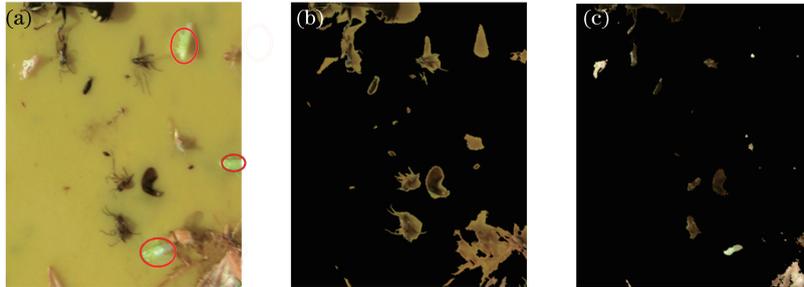


图 4 不同算法的分割结果。(a)原图;(b)传统 Ostu 算法;(c)  $K\_means$  聚类算法

Fig. 4 Segmentation results using different algorithms. (a) Original image; (b) traditional Ostu algorithm; (c)  $K\_means$  clustering algorithm

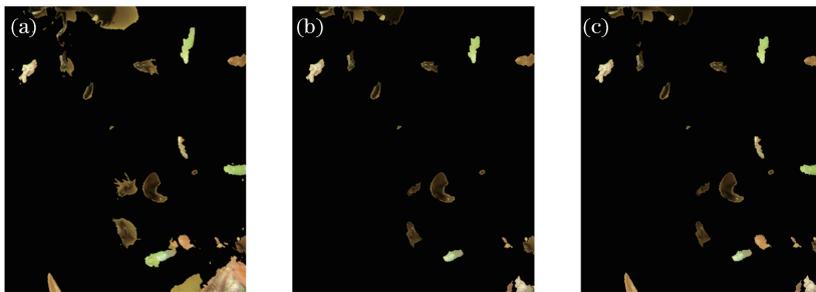


图 5 DBSCAN 聚类结果。(a)阈值取 8;(b)阈值取 12;(c)多聚类融合图像

Fig. 5 DBSCAN clustering results. (a) Setting threshold to 8; (b) setting threshold to 12; (c) clustering fusion result

表 1 不同算法分割出的茶小绿叶蝉个数

Table 1 Number of *Empoasca flavescens* separated by different algorithms

Total number	Number		
	Ostu	$K\_means$	SLIC+DBSCAN8+DBSCAN12
204	164	187	200

### 3.2 茶小绿叶蝉识别效果的评价

为验证所提方法的有效性,对不同天气和不同光照强度下拍摄的 151 幅图像进行试验。从这 151 幅图中分割出的目标图像中提取了 200 个茶小绿叶蝉以及 3000 个其他类害虫,并从中各随机抽取 1/2 (茶小绿叶蝉 100 个,其他类害虫 1500 个)作为训练样本,剩余的作为测试样本。用 LSSVM 对训练样本进行训练,将训练好的模型对茶小绿叶蝉进行识别能力检测,随机抽取不同的训练样本和测试样本共 10 次,分别建立识别模型,得到测试样本的平均识别结果如表 2 所示。其中,整体识别精度和 TNR

分别高达 98.81% 和 99.45%,而茶小绿叶蝉的查全率和查准率分别为 86.0% 和 91.29%,同时两个综合指标 G-mean 和 F-measure 分别为 92.47% 和 88.53%。这是因为相比其他害虫来说,茶小绿叶蝉的个数非常少,训练 LSSVM 模型的样本数量不均衡导致分类超平面偏移。即使茶小绿叶蝉误识别的个数较多,整体识别精度仍会处于一个很高的水平,而其他的精度测量标准却表现得不太理想。

因此,使用改进的 SMOTE 算法对错分样本进行 4 倍生成,以提高关键性样本(分类超平面附近样本)的识别精度。从表 2 可以看出,少数类样本的 TPR 以及综合指标 G-mean、F-measure 分别为 98.2%、98.63% 和 92.56%,相比于未使用 SMOTE 算法有非常明显的提升,但同时 TNR 和 Precision 的值有所下降。

为解决 TNR 和 Precision 这两个指标下降的问题,在改进 SMOTE 算法的基础上,利用 KS 算法选取具有代表性的多数类样本来减少样本之间的冗余

信息以及缩小与少数类之间的数量差异。从表 2 可以看出,用 KS 算法选择 1000 个多数类样本后, Precision 由 87.54% 提升到 91.76%, F-measure 由 92.56% 提升到 94.54%。可见,所提算法解决了数据集分类不均衡时,超平面偏移使少数类样本识别率以及评价分类器性能指标较低的问题。

图 6 所示为不同算法选择不同个数样本的精度变化情况。图 6(a) 给出了 KS 算法选择不同个数的多数类样本时各精度指标的变化。由图 6(a) 可以看出,当选择个数从 600 增加到 1400 时, Percent\_test、TNR、Precision 以及 F-measure 这四项指标的精度值先上升后下降,在选择个数为 1000 时达到最大值;总体识别精度 Percent\_test 和

表 2 不同训练方法得到的测试精度

Table 2 The test accuracy obtained by different training methods

Algorithm	Percent_test	TPR	TNR	Precision	G-mean	F-measure	%
LSSVM	98.81	86.0	99.45	91.29	92.47	88.53	
LSSVM+SMOTE	99.01	98.2	99.07	87.54	98.63	92.56	
LSSVM+SMOTE+KS	99.30	97.5	99.42	91.76	98.46	94.54	

随机选择不同个数的多数类样本时各精度指标的变化如图 6(b) 所示。当选择个数从 600 变化到 1400 时, Percent\_test、TNR、Precision 以及 F-measure 的精度逐渐增大, TPR 略有下降的趋势, G-mean 无明显变化。这是因为随机选择只是简单地改变了训练样本数量间的差异,选择的样本无法

TNR 变化得较为平缓, Precision 和 F-measure 变化得较为剧烈; TPR 和 G-mean 这两项指标精度值先降后升,在选择个数为 1000 时达到最小值。Percent\_test、TNR、Precision 的变化趋势从侧面验证了多数类样本中包含有冗余信息,并且 KS 算法能够有效减少冗余信息。当选择样本数为 1000 时,效果最为显著,随着选择样本数量减少,信息冗余性降低。但因为选择的样本数较少而无法全面代表多数类样本,随着选择样本数增加,冗余信息也随之增加。在上述两种情形中都获取不到较为完整和有效的多数类样本的信息,同时这两种情形都会导致 TPR 和 G-mean 在选择的样本数为 1000 时反而较小。

代表全部的样本信息,而随着选择样本个数增加,其包含的有用信息越多。因此 Percent\_test、TNR、Precision 以及 F-measure 呈现逐步上升的趋势,而选择的多数类样本数越少,分类超平面附近的样本就更容易被分类成少数类样本,导致 TPR 的值随选择样本数增加而略有下降的趋势。

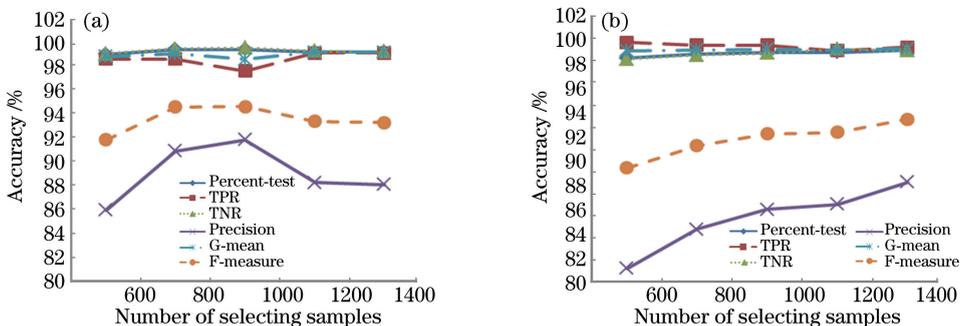


图 6 不同算法精度与选择样本个数的关系。(a) KS 算法;(b)随机选择

Fig. 6 Accuracy versus number of selecting samples for different algorithms. (a) KS algorithm; (b) random selection

茶小绿叶蝉在不同天气和光强下的部分识别结果如图 7 所示,图中的方框、圆圈分别表示茶小绿叶蝉被正确识别和漏识别,方框和圆圈的组合表示茶小绿叶蝉被误识别。图 7(a) 显示了茶小绿叶蝉被正确识别的情况。图 7(b) 显示了当光线较暗时某些阴影部分被错误识别的情况。图 7(c) 显示了在光照较强时,因茶小绿叶蝉的翅膀呈白色遮挡了其身体的颜色而导致的漏识别的情况。因此,如何减

少和消除复杂环境下茶小绿叶蝉的错误识别和漏识别还有待进一步研究。

## 4 结 论

用机器视觉的方法实现了茶园开放环境下黄板上茶小绿叶蝉的识别。由于光照条件和背景具有不确定性,阈值分割的方法并不能达到理想的分割效果。在超像素分割的基础上引入多聚类图像融合的

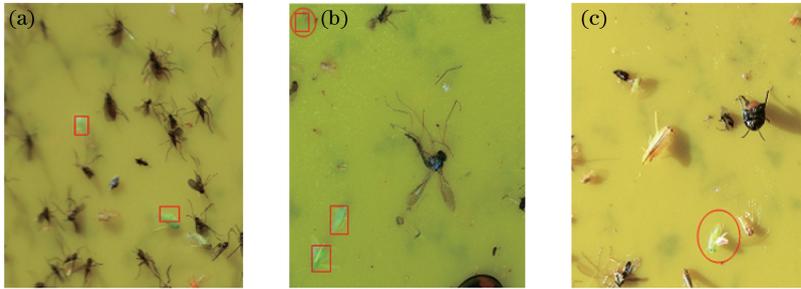


图7 部分识别结果。(a)正确识别;(b)错误识别;(c)漏识别

Fig. 7 Partial recognition results. (a) Correct recognition; (b) misrecognition; (c) leakage recognition

分割算法可有效解决复杂环境下茶小绿叶蝉被漏分割的问题,同时保证了目标图像分割的准确性。在用LSSVM进行识别时,针对茶小绿叶蝉和其他害虫数量不均衡的问题,采用改进SMOTE算法和KS算法相结合的方法提高了茶小绿叶蝉的识别准确率,整体精度为99.30%,茶小绿叶蝉的查准率为91.76%。但是该识别方法也存在少量错误识别和漏识别的情况,因此寻找稳健性更高的特征以提高茶园中茶小绿叶蝉的识别精度将是下一步研究的内容。

### 参 考 文 献

- [1] 彭林, 孙晓琳, 陈宗懋, 等. 茶树主要病虫害防治手册[M]. 杭州: 国家产业技术体系病虫害功能研究室, 2009: 5-6.
- [2] Wu L Y, Jin S, Wu M Y, *et al.* Advances in tea leafhopper integrated pest management[J]. *Guizhou Agricultural Sciences*, 2014, 42(3): 64-68.  
吴亮宇, 金珊, 吴命燕, 等. 茶树假眼小绿叶蝉防治技术的研究进展[J]. *贵州农业科学*, 2014, 42(3): 64-68.
- [3] Wang Q S, Wang D F, Wu G Y. Research advances on *Empoasca vitis* (Gothe) in tea trees in China[J]. *Fujian Journal of Agricultural Sciences*, 2013, 28(6): 615-623.  
王庆森, 王定锋, 吴光远. 我国茶树假眼小绿叶蝉研究进展[J]. *福建农业学报*, 2013, 28(6): 615-623.
- [4] Yao Y J, Wang J L, He L, *et al.* Research on trap of *Empoasca vitis* Gothe by yellow board[J]. *Journal of Tea*, 2010, 36(2): 90-92.  
姚雍静, 王家伦, 何莲, 等. 黄色诱虫板对茶小绿叶蝉的诱捕效果研究[J]. *茶叶*, 2010, 36(2): 90-92.
- [5] Lu Z Y, Liu S C. Drought identification based on multi feature fusion for early maize image[J]. *Laser & Optoelectronics Progress*, 2017, 54(8): 081005.  
路志英, 刘书辰. 基于多特征融合的玉米前期图像早情识别[J]. *激光与光电子学进展*, 2017, 54(8): 081005.
- [6] Xu L F, Xu X B, Hu M, *et al.* Corn leaf disease identification based on multiple classifiers fusion[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(14): 194-201.  
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.
- [7] Ni J, Yao X, Tian Y C, *et al.* Design and experiments of portable apparatus for plant growth monitoring and diagnosis [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(6): 150-156.  
倪军, 姚霞, 田永超, 等. 便携式作物生长监测诊断仪的设计与试验[J]. *农业工程学报*, 2013, 29(6): 150-156.
- [8] Cao L P. The research progress on machine recognition of plant diseases and insect pests [J]. *Chinese Agricultural Science Bulletin*, 2015, 31(20): 244-249.  
曹乐平. 基于机器视觉的植物病虫害实时识别方法[J]. *中国农学通报*, 2015, 31(20): 244-249.
- [9] Ling C D, Chen H, Yang X, *et al.* Fundus image hard exudates detection based on SLIC superpixels and DBSCAN clustering [J]. *Journal of Huaqiao University*, 2015, 36(4): 399-405.  
凌朝东, 陈虎, 杨骁, 等. 结合SLIC超像素和DBSCAN聚类的眼底图像硬性渗出检测方法[J]. *华侨大学学报*, 2015, 36(4): 399-405.
- [10] Sun Z, Song Q, Zhu X, *et al.* A novel ensemble method for classifying imbalanced data [J]. *Pattern Recognition*, 2015, 48(5): 1623-1637.
- [11] Zhang Y, Li X, Gao X, *et al.* A simple algorithm of superpixel segmentation with boundary constraint [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(7): 1502-1514.
- [12] Mou L, Zhang X W, Zhang Z, *et al.* Saliency

- detection optimization method in natural scene [J]. *Laser & Optoelectronics Progress*, 2016, 53(12): 121501.
- 牟丽, 张学武, 张卓, 等. 自然场景下的显著性检测优化方法[J]. *激光与光电子学进展*, 2016, 53(12): 121501.
- [13] Wang X. VCells: Simple and efficient superpixels using edge-weighted centroidal voronoi tessellations [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(6): 1241-1247.
- [14] Zhou J, Zhu J R, Wang M J. Pear orchard scene segmentation based on conditional random fields[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(2): 8-13.
- 周俊, 朱金荣, 王明军. 基于条件随机场的梨园场景图像分割方法[J]. *农业机械学报*, 2015, 46(2): 8-13.
- [15] Achanta R, Shaji A, Smith K, *et al.* SLIC superpixels compared to state-of-the-art superpixel method[J]. *IEEE Transation on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.
- [16] Zhou S G, Zhou A Y, Jin W, *et al.* A fast DBSCAN algorithm[J]. *Journal of Software*, 2000, 11(6): 735-744.
- 周水庚, 周傲英, 金文, 等. 一种快速 DBSCAN 算法[J]. *软件学报*, 2000, 11(6): 735-744.
- [17] Li G L, Ma Z H, Huang C, *et al.* Segmentation of color images of grape diseases using  $K$ \_means clustering algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2010, 26(14): 32-37.
- 李冠林, 马占鸿, 黄冲, 等. 基于  $K$ \_means 硬聚类算法的葡萄病害彩色图像分割方法[J]. *农业工程学报*, 2010, 26(14): 32-37.
- [18] Vapnik V. *Statistical learning theory* [M]. New York: Wiley, 1998.
- [19] Khanmohammadi M, Karami F, Mir-Marqués A, *et al.* Classification of persimmon fruit origin by near infrared spectrometry and least squares-support vector machines [J]. *Food Engineering*, 2014, 142(6): 17-22.
- [20] Sun T, Wu H F, Liang Z G, *et al.* Application of SMOTE arithmetic for unbalanced data [J]. *Beijing Biomedical Engineering*, 2012, 31(5): 528-530.
- 孙涛, 吴海丰, 梁志刚, 等. SMOTE 算法在不平衡数据中的应用[J]. *北京生物工程医学工程*, 2012, 31(5): 528-530.
- [21] Zheng W C, Chen S Y, Wang X Q. Imbalanced datasets based SMOTE-SVM-AID algorithm [J]. *Journal of Wuhan University of Technology*, 2012, 34(11): 58-62.
- 郑文昌, 陈淑燕, 王宣强. 面向不平衡数据集的 SMOTE-SVM 交通事件检测算法[J]. *武汉理工大学学报*, 2012, 34(11): 58-62.
- [22] Yi B H, Zhu J J, Li J. Imbalanced data classification on micro-credit company customer credit risk assessment using improved SMOTE support vector machine [J]. *Chinese Journal of Management Science*, 2016, 24(3): 24-30.
- 衣柏衡, 朱建军, 李杰. 基于改进 SMOTE 的小额贷款公司客户信用风险非均衡 SVM 分类[J]. *中国管理科学*, 2016, 24(3): 24-30.
- [23] Zhang X D. Initial center selection method for  $k$ -medoids clustering based on KS sampling[J]. *Journal of Inner Mongolia University of Technology*, 2016, 35(4): 297-302.
- 张晓东. 基于 KS 抽样的  $k$ -medoids 聚类初始中心选取法[J]. *内蒙古工业大学学报*, 2016, 35(4): 297-302.