

基于图像处理技术的四种苜蓿叶部病害的识别

秦丰¹ 刘东霞² 孙炳达³ 阮柳¹ 马占鸿¹ 王海光^{1*}

(1. 中国农业大学 植物保护学院, 北京 100193;

2. 河北北方学院 农林科技学院, 河北 张家口 075000;

3. 中国科学院 微生物研究所, 北京 100101)

摘要 基于图像处理技术,对4种苜蓿叶部病害进行识别研究。利用结合K中值聚类算法和线性判别分析的分割方法对病斑图像作分割,获得了较好的分割效果。结果表明:该分割方法在由4种病害图像数据集整合成的汇总图像数据集上综合得分的平均值和中值分别为0.877 1和0.899 7;召回率的平均值和中值分别为0.829 4和0.851 4;准确率的平均值和中值分别为0.924 9和0.942 4。进一步提取病斑图像的颜色特征、形状特征和纹理特征共计129个,利用朴素贝叶斯方法和线性判别分析方法建立病害识别模型,并结合顺序前向选择方法实现特征筛选,分别获得最优特征子集;同时利用这2个最优特征子集,结合支持向量机(Support vector machine, SVM)建立病害识别模型。比较各模型的识别效果,发现利用所建线性判别分析模型下的最优特征子集,结合SVM建立的病害识别模型识别效果最好,训练集识别正确率为96.18%,测试集识别正确率为93.10%。由此可见,本研究所建基于图像处理技术的病害识别模型可用于识别上述4种苜蓿叶部病害,为苜蓿病害的诊断和鉴别提供了一定依据。

关键词 苜蓿;叶部病害;图像识别;图像分割;特征优选;支持向量机

中图分类号 S 126; S 431.9

文章编号 1007-4333(2016)10-0065-11

文献标志码 A

Recognition of four different alfalfa leaf diseases based on image processing technology

QIN Feng¹, LIU Dong-xia², SUN Bing-da³, RUAN Liu¹, MA Zhan-hong¹, WANG Hai-guang^{1*}

(1. College of Plant Protection, China Agricultural University, Beijing 100193, China;

2. College of Agriculture and Forestry Science and Technology, Hebei North University, Zhangjiakou 075000, China;

3. Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract Automatic recognition of four different alfalfa leaf diseases was investigated based on image processing technology. The sub-images (lesion images) with one typical lesion or multiple typical lesions were obtained for further recognition by artificial cutting from the acquired digital disease images. Lesion segmentation was conducted by using a segmentation method integrating with *K*-median clustering algorithm and linear discriminant analysis. For the aggregated image dataset consisting of the lesion images of the four different alfalfa leaf diseases, the mean and the median of the scores were 0.877 1 and 0.899 7, respectively, those of the recalls were 0.829 4 and 0.851 4, respectively, and those of the accuracies reached 0.924 9 and 0.942 4, respectively. A total of 129 color, shape and texture features were extracted from the lesion images for further analysis. The Naive Bayes method and linear discriminant analysis were combined with sequential forward selection method to build disease recognition models, respectively. And the disease recognition models using support vector machine (SVM) were built based on the two optimal feature subsets consisting of the features optimized by using the two methods. Comparing the efficiency of those

收稿日期: 2016-02-02

基金项目: 公益性行业(农业)科研专项经费项目(201303057)

第一作者: 秦丰, 硕士研究生, E-mail: 15210590688@163.com

通讯作者: 王海光, 副教授, 主要从事植物病害流行病学和宏观植物病理学研究, E-mail: wanghaiguang@cau.edu.cn

four models, the disease recognition SVM model built based on the optimal feature subset obtained by linear discriminant analysis was found to be the best model for the image recognition of alfalfa leaf diseases. The recognition accuracy of the training set was 96.18% and that of the testing set was 93.10%. The results indicated that the image recognition of the four different alfalfa leaf diseases could be implemented with high accuracy by using the model proposed.

Keywords alfalfa; leaf disease; image recognition; image segmentation; feature optimization; support vector machine

苜蓿被称为“牧草之王”，是我国最重要的栽培牧草，对畜牧业有重要价值^[1,2]。苜蓿褐斑病(病原 *Pseudopeziza medicaginis*)、锈病(病原 *Uromyces striatus*)、小光亮叶斑病(病原 *Pleosphaerulina briosiana*)和尾孢菌叶斑病(病原 *Cercospora medicaginis*)是4种常见的苜蓿叶部病害，影响植株的光合作用和呼吸作用，造成可溶性糖和粗蛋白等含量下降，粗纤维含量上升，适口性下降，严重时可导致干草和种子减产，制约苜蓿产业的健康发展^[3]。苜蓿叶部病害种类多，病斑类型繁杂，不同病害病斑的颜色、形状等特征复杂多样，而且会随着病斑的扩展而发生改变。病害的准确诊断需要借助农业专家和农业技术人员的经验。病害的误诊容易导致防治措施的失败，不仅造成人力和物力的浪费，而且延误最佳防治时机，可能引发病害流行，造成严重的经济损失。因此，需要探索一种快速、智能的苜蓿叶部病害诊断方法，以便及时准确地诊断病害，及早制定和采取正确的防治措施，防止病害蔓延，减少病害损失。

随着科学技术的快速发展，图像获取已变得十分容易，相应的图像获取设备也更加多样化，如具有拍照功能的手机、各种图像传感器等，并且所获得图像分辨率越来越高。植物病害图像识别技术方便、快速，可以实现自动化，已在多种植物病害识别^[4-12]和评估^[13-15]研究中得到了应用。

准确进行病斑图像分割处理是植物病害图像识别的关键步骤之一。基于聚类算法的图像分割方法已广泛用于植物病害病斑图像的分割^[16-19]。然而，不同植物叶片表面的颜色往往复杂多变，病斑区域与健康区域在颜色空间上可能无明显的聚类，若仅利用基于聚类算法的图像分割方法分割颜色复杂的植物病害图像会导致病斑的过度分割或欠分割问题。监督分类是依据典型样本推断出用于分类的函数方程的一种技术。已有研究表明，利用监督分类方法可以较好地实现植物病害图像分割^[20]。然而，单独使用监督分类方法进行病斑图像分割不能有针对性

对性地自动获取某一病害图像中典型病斑区域和典型健康区域的特征。将聚类算法和监督分类算法结合在一起，可较好地实现图像的自动分割^[21,22]。

为实现植物病害的图像识别，提取获得的病斑图像颜色特征、纹理特征和形状特征等往往有几十个，甚至上百个，过多的特征会增加识别的复杂性，降低识别方法的准确性和应用性，因此，需要对提取的特征进行有效筛选，以便简化识别方法和提高识别准确率。

利用病斑图像的颜色特征、纹理特征和形状特征等实现植物病害图像识别的研究较多。Camargo等^[23]基于获得的由细菌性角斑病(病原 *Xanthomonas campestris*)、茎枯病(病原 *Ascochyta gossypii*)和稻绿蝽(*Nezara viridula*)造成的棉花叶部损害图像，提取图像的颜色特征、形状特征和纹理特征，利用支持向量机(Support vector machine, SVM)实现了图像识别。Xu等^[24]将获得的缺失氮素和钾素的番茄叶片图像由RGB颜色空间转换至 $L^*a^*b^*$ 颜色空间，分别利用差分算子、傅里叶变换方法和小波包分解方法提取 b^* 分量图像的纹理特征，利用遗传算法选择最优特征组合，结合模糊K最近邻法，实现了2种缺素症的识别。李冠林等^[10]通过对小麦条锈病和小麦叶锈病图像裁剪获得包含病斑的子图像，经中值滤波去噪后，利用K_means硬聚类算法完成病斑图像分割，提取病斑图像的形状特征、颜色特征和纹理特征，并从中优选26个特征参数利用SVM实现了对这2种病害的识别。Phadikar等^[11]基于获得的水稻胡麻斑病(病原 *Bipolaris oryzae*)、稻瘟病(病原 *Magnaporthe oryzae*)、水稻叶鞘腐败病(病原 *Sarocladium oryzae*)和水稻白叶枯病(病原 *X. oryzae*)的叶部病害图像，利用一种基于费米能的分割方法实现了病斑图像的分割，提取了病斑图像的颜色特征、形状特征和侵染位置特征，应用粗糙集理论选择了关键特征，结合一种基于规则的分类器实现了病害图像的识别。Omrani等^[18]根据提取的图像颜色特征及基

于小波和灰度共生矩阵的特征,利用以径向基函数为核函数的支持向量回归机实现了 3 种苹果叶部病虫害的识别。但是,目前利用图像识别技术诊断苜蓿叶部病害的研究尚鲜有报道。

本研究基于获得的苜蓿褐斑病、锈病、小光壳叶斑病和尾孢菌叶斑病病害图像,结合聚类算法(K 中值聚类算法)和监督分类方法(线性判别分析)进行病斑图像分割,提取病斑图像的颜色特征、纹理特征和形状特征,分别利用朴素贝叶斯方法和线性判别分析方法建立病害识别模型,并结合顺序前向选择方法实现特征筛选^[25],分别获得最优特征子集,根据最优特征子集建立病害识别 SVM 模型,比较各模型的识别效果,并确定最优模型,旨在为这 4 种苜蓿叶部病害提供一种快速、准确的无损识别方法。

1 材料与方法

1.1 图像获取

本研究使用的苜蓿病害图像共计 899 张,包括 76 张苜蓿褐斑病图像、136 张苜蓿锈病图像、231 张苜蓿小光壳叶斑病图像、456 张苜蓿尾孢菌叶斑病图像。将田间采集到的具有典型症状病斑的发病叶片带回室内,置于白色背景下拍照。拍照时,尽可能将叶片展平,保证相机镜头与叶片所在平面平行。获取的病害图像分辨率为 $4\ 256 \times 2\ 832$ 像素。

通过人工裁剪的方法从每一张病害图像中裁剪出包含 1 个或多个典型症状病斑的子图像,用于构建典型病斑图像库,并建立病害识别模型。子图像的大小取决于典型病斑的大小和包含的病斑数量。

1.2 病斑图像分割

本研究利用结合 K 中值聚类算法和线性判别分析的分割方法进行病斑图像的分割。

首先将裁剪获得的子图像由 RGB 颜色空间分别转换至 HSV 颜色空间和 $L^* a^* b^*$ 颜色空间。以图像中每个像素点的 a^* 分量值和 b^* 分量值作为该像素点的颜色特征,利用 K 中值聚类算法将所有像素点聚成 10 类。 K 中值聚类算法利用 MATLAB R2013b 编程实现,通过随机选择方法获得初始聚类种子,距离度量选择欧氏距离,算法最大迭代次数为 1 000 次,算法重复次数为 3 次,以类内距离之和最小作为 K 中值聚类准则。

利用 K 中值聚类算法将病斑图像中所有像素点聚为 10 类后,计算每一类中所有像素点的 H 分量平均值,并按 H 分量平均值由小到大的顺序排

序。与绿色健康叶片相比,4 种病害病斑的 H 分量值均较小,因此,选择 H 分量平均值最小的类,以该类中的像素点作为典型病斑样本。典型症状病斑部分与典型健康部分之间存在过渡区域,该区域的 H 分量值通常介于两者之间,因此,选择 H 分量平均值最大的 7 类,以这些类中的像素点作为典型健康样本,剩余 2 类中的样本作为待定样本,待定样本不参与分类模型的建立。

将典型病斑样本标记为正例,典型健康样本标记为负例,并组成训练集,以每个像素点的 a^* 分量值和 b^* 分量值作为特征,利用线性判别分析方法建立分类模型,并对图像中所有像素点进行分类。将分类模型判定为病斑的像素点赋值为 1,其余像素点赋值为 0,由此得到初始二值化的病斑分割图像。为避免原始子图像中可能存在的白色背景的干扰和影响,需要判断原始子图像中是否存在蓝色分量值高于 200 的像素点,若不存在,则对初始二值化病斑分割图像进行空洞填充,并计算所有连通区域的面积,去除小于最大连通区域面积 $1/16$ 的连通区域,得到最终的病斑分割图像;若原始子图像中存在蓝色分量值高于 200 的像素点,则在初始二值化病斑分割图像中将这此像素点赋值为 0,其余像素点赋值为 1,得到一幅二值化图像,将其与初始二值化病斑分割图像相乘,得到新的二值化图像,对该图像进行空洞填充,并计算所有连通区域的面积,去除小于最大连通区域面积 $1/16$ 的连通区域,得到最终的病斑分割图像。

在病斑分割时,图像中的所有像素点均被分为 2 类,即病斑像素点和非病斑像素点,因此,病斑分割等同于模式识别领域中的二分类问题,病斑分割效果评价方法可借鉴二分类模型的评价方法。利用 Adobe Photoshop CC 软件对子图像进行手动病斑分割,确定图像中所有像素点的真实类别。将结合 K 中值聚类算法和线性判别分析的病斑自动分割结果与手动分割结果进行比较,依据召回率、准确率和综合得分评价病斑自动分割的效果。

召回率和准确率是模式识别领域常用的 2 种分类模型效果评价指标^[26],在本研究中,其计算公式分别为:

$$\text{召回率} = \frac{N_1}{N_2} \quad (1)$$

$$\text{准确率} = \frac{N_1}{N_3} \quad (2)$$

式中： N_1 为被自动分割方法正确分类的病斑像素点总数； N_2 为手动分割方法下病斑像素点总数； N_3 为图像中的像素点总数。召回率和准确率的取值均为 0~1，值越大说明病斑自动分割效果越好。综合得分综合考虑了召回率和准确率 2 种评价指标的结果，本研究以召回率和准确率的平均值作为自动分割结果的综合得分，其计算公式为：

$$\text{综合得分} = \frac{\text{召回率} + \text{准确率}}{2} \quad (3)$$

该得分的取值为 0~1，其值越高说明自动分割效果越好。

1.3 病斑图像特征提取

为了充分挖掘 4 种苜蓿叶部病害病斑图像的特征信息，本研究提取了病斑图像的纹理特征、颜色特征和形状特征共计 129 个，其中，纹理特征包括 RGB、HSV 和 $L^*a^*b^*$ 3 个颜色空间共 9 个分量灰度图像的 Hu 不变矩 (63 个)、对比度 (9 个)、能量 (9 个) 和同质性 (9 个)，总计 90 个纹理特征；颜色特征包括 RGB、HSV 和 $L^*a^*b^*$ 3 个颜色空间共 9 个分量灰度图像的一阶矩 (9 个)、二阶矩 (9 个) 和三阶矩 (9 个) 以及 R、G、B 分量的颜色比值 (3 个)，总计 30 个颜色特征；形状特征包括病斑圆度、复杂度和二值化病斑图像的 Hu 不变矩 (7 个)，总计 9 个形状特征。

Hu 不变矩是描述图像纹理特征的方法之一，

具有平移、旋转和尺度缩放的不变性，7 个 Hu 不变矩的具体计算方法见参考文献 [27]。

对比度是一个像素在整个图像上与其邻居间的灰度对比的度量；能量是图像一致性的度量；同质性是共生矩阵中元素对角线分布的空间紧密度的度量 [28]。它们的计算公式见参考文献 [28]。

一阶矩、二阶矩、三阶矩的计算公式见参考文献 [29]。

R、G、B 3 个颜色分量的颜色比值计算公式分别为：

$$r = \frac{R}{R + G + B} \quad (4)$$

$$g = \frac{G}{R + G + B} \quad (5)$$

$$b = \frac{B}{R + G + B} \quad (6)$$

式中： r 、 g 、 b 分别为 R、G、B 3 个颜色分量的颜色比值。

圆度是指病斑区域接近圆形的程度，其值越大，说明病斑区域越接近圆形。复杂度指病斑区域的复杂性和离散程度，其值越大，复杂性越高，离散程度越大 [10]。圆度和复杂度的计算公式见参考文献 [10]。

为了方便展示特征选择结果，本研究为提取的每一个特征赋予 1 个 ID 号，结果如表 1 所示。其

表 1 提取的图像特征名称及其 ID 号

Table 1 The list of image features extracted and their ID numbers

特征 Feature	ID 号 Identity number	特征 Feature	ID 号 Identity number	特征 Feature	ID 号 Identity number	特征 Feature	ID 号 Identity number
φ Lab_L1	1	φ RGB_G6	34	φ shape2	67	三阶矩 Lab_L The third moment Lab_L	100
φ Lab_L2	2	φ RGB_G7	35	φ shape3	68	三阶矩 Lab_b The third moment Lab_b	101
φ Lab_L3	3	φ RGB_B1	36	φ shape4	69	三阶矩 Lab_a The third moment Lab_a	102
φ Lab_L4	4	φ RGB_B2	37	φ shape5	70	对比度 RGB_R Contrast RGB_R	103
φ Lab_L5	5	φ RGB_B3	38	φ shape6	71	能量 RGB_R Energy RGB_R	104
φ Lab_L6	6	φ RGB_B4	39	φ shape7	72	同质性 RGB_R Homogeneity RGB_R	105
φ Lab_L7	7	φ RGB_B5	40	一阶矩 RGB_R The first moment RGB_R	73	对比度 RGB_G Contrast RGB_G	106
φ Lab_a1	8	φ RGB_B6	41	一阶矩 RGB_G The first moment RGB_G	74	能量 RGB_G Energy RGB_G	107

表 1(续)

特征 Feature	ID 号 Identity number	特征 Feature	ID 号 Identity number	特征 Feature	ID 号 Identity number	特征 Feature	ID 号 Identity number
φ Lab_a2	9	φ RGB_B7	42	一阶矩 RGB_B The first moment RGB_B	75	同质性 RGB_G Homogeneity RGB_G	108
φ Lab_a3	10	φ HSV_H1	43	颜色比值 RGB_R Color ratio RGB_R	76	对比度 RGB_B Contrast RGB_B	109
φ Lab_a4	11	φ HSV_H2	44	颜色比值 RGB_G Color ratio RGB_G	77	能量 RGB_B Energy RGB_B	110
φ Lab_a5	12	φ HSV_H3	45	颜色比值 RGB_B Color ratio RGB_B	78	同质性 RGB_B Homogeneity RGB_B	111
φ Lab_a6	13	φ HSV_H4	46	一阶矩 HSV_H The first moment HSV_H	79	对比度 HSV_H Contrast HSV_H	112
φ Lab_a7	14	φ HSV_H5	47	一阶矩 HSV_S The first moment HSV_S	80	能量 HSV_H Energy HSV_H	113
φ Lab_b1	15	φ HSV_H6	48	一阶矩 HSV_V The first moment HSV_V	81	同质性 HSV_H Homogeneity HSV_H	114
φ Lab_b2	16	φ HSV_H7	49	一阶矩 Lab_L The first moment Lab_L	82	对比度 HSV_S Contrast HSV_S	115
φ Lab_b3	17	φ HSV_S1	50	一阶矩 Lab_b The first moment Lab_b	83	能量 HSV_S Energy HSV_S	116
φ Lab_b4	18	φ HSV_S2	51	一阶矩 Lab_a The first moment Lab_a	84	同质性 HSV_S Homogeneity HSV_S	117
φ Lab_b5	19	φ HSV_S3	52	二阶矩 RGB_R The second moment RGB_R	85	对比度 HSV_V Contrast HSV_V	118
φ Lab_b6	20	φ HSV_S4	53	二阶矩 RGB_G The second moment RGB_G	86	能量 HSV_V Energy HSV_V	119
φ Lab_b7	21	φ HSV_S5	54	二阶矩 RGB_B The second moment RGB_B	87	同质性 HSV_V Homogeneity HSV_V	120
φ RGB_R1	22	φ HSV_S6	55	二阶矩 HSV_H The second moment HSV_H	88	对比度 Lab_L Contrast Lab_L	121
φ RGB_R2	23	φ HSV_S7	56	二阶矩 HSV_S The second moment HSV_S	89	能量 Lab_L Energy Lab_L	122
φ RGB_R3	24	φ HSV_V1	57	二阶矩 HSV_V The second moment HSV_V	90	同质性 Lab_L Homogeneity Lab_L	123
φ RGB_R4	25	φ HSV_V2	58	二阶矩 Lab_L The second moment Lab_L	91	对比度 Lab_a Contrast Lab_a	124
φ RGB_R5	26	φ HSV_V3	59	二阶矩 Lab_b The second moment Lab_b	92	能量 Lab_a Energy Lab_a	125
φ RGB_R6	27	φ HSV_V4	60	二阶矩 Lab_a The second moment Lab_a	93	同质性 Lab_a Homogeneity Lab_a	126
φ RGB_R7	28	φ HSV_V5	61	三阶矩 RGB_R The third moment RGB_R	94	对比度 Lab_b Contrast Lab_b	127
φ RGB_G1	29	φ HSV_V6	62	三阶矩 RGB_G The third moment RGB_G	95	能量 Lab_b Energy Lab_b	128
φ RGB_G2	30	φ HSV_V7	63	三阶矩 RGB_B The third moment RGB_B	96	同质性 Lab_b Homogeneity Lab_b	129
φ RGB_G3	31	圆度 Circularity	64	三阶矩 HSV_H The third moment HSV_H	97		
φ RGB_G4	32	复杂度 Complexity	65	三阶矩 HSV_S The third moment HSV_S	98		
φ RGB_G5	33	φ shape1	66	三阶矩 HSV_V The third moment HSV_V	99		

中, $\varphi_{\text{Lab_L1}}$ 指 $L^* a^* b^*$ 颜色空间下, 病斑 L^* 分量灰度图像的第一 Hu 不变矩; φ_{shape1} 指二值化病斑图像的第一 Hu 不变矩, 反映了病斑的形状特征; 一阶矩 RGB_R 指 RGB 颜色空间下的 R 分量一阶矩; 颜色比值 RGB_R 指 RGB 颜色空间下的 R 分量颜色比值; 对比度 RGB_R 指 RGB 颜色空间下的 R 分量灰度图像对比度; 能量 RGB_R 指 RGB 颜色空间下的 R 分量灰度图像能量; 同质性 RGB_R 指 RGB 颜色空间下的 R 分量灰度图像同质性; 其余特征含义依此类推。

1.4 特征优选与病害识别模型的构建

通过图像分割得到 4 种苜蓿叶部病害的典型病斑图像共 1 651 张, 按 2:1 的比例将所有样本划分为训练集和测试集, 训练集共 1 100 张图像(其中, 褐斑病图像 111 张、锈病图像 267 张、小光壳叶斑病图像 371 张、尾孢菌叶斑病图像 351 张), 测试集共 551 张图像(其中, 褐斑病图像 56 张、锈病图像 133 张、小光壳叶斑病图像 185 张、尾孢菌叶斑病图像 177 张)。

由于不同特征的取值范围差异较大, 为避免这种差异对病害识别模型准确性的影响, 将所有特征的取值范围归一化至 $[0, 1]$ 范围内, 具体计算公式如下:

$$X_{\text{norm}}^i = \frac{X^i - X_{\text{min}}^i}{X_{\text{max}}^i - X_{\text{min}}^i} \quad (7)$$

式中: X_{norm}^i 为归一化后的第 i 个特征; X^i 为归一化前的第 i 个特征; X_{min}^i 和 X_{max}^i 分别为训练集中第 i 个特征的最小值和最大值。

分别利用朴素贝叶斯方法和线性判别分析方法建立病害识别模型, 并且在建模时利用顺序前向选择方法实现特征优选。假设共有 m 个特征, 备选特征集合为 U , 最优特征子集为 S 。初始的 U 即包含了全部 m 个特征的集合, 而初始的 S 集合为空集, 顺序前向选择方法的具体步骤如下:

第 1 步, 依次以集合 U 中的第 i 个特征 $X_i (i=1, 2, \dots, m)$ 为输入, 建立病害识别模型, 计算并记录识别正确率。假设该过程中所有模型的识别正确率最大值为 C_{max} , 将识别正确率最大的模型所用的特征放入集合 S 中, 假设该特征为 X_1 。将 X_1 从备选特征集合 U 中剔除。

第 2 步, 依次将集合 U 中的第 k 个特征 $X_k (k=2, 3, \dots, m)$ 与集合 S 中的特征组成新的特征子集, 以该特征子集为输入, 建立病害识别模型, 计算

并记录识别正确率。假设该过程中所有模型的识别正确率最大值为 C_{maxnew} , 如果 $C_{\text{maxnew}} \leq C_{\text{max}}$, 则停止算法; 否则, 令 $C_{\text{max}} = C_{\text{maxnew}}$, $S = \emptyset (\emptyset$ 为空集), 并将识别正确率最大的模型所用的特征子集放入集合 S 中。假设该特征子集为 $\{X_1, X_2\}$ 。将 X_2 从集合 U 中剔除。

重复第 2 步, 直到算法停止或集合 U 为空集。

SVM 可以较好解决小样本、非线性、高维数和局部极小点等问题^[30]。为了与基于朴素贝叶斯方法和线性判别分析方法所建病害识别模型进行比较, 分别利用 2 种方法(朴素贝叶斯方法和线性判别分析方法)下的最优特征子集, 结合 SVM 方法建立病害识别模型, 比较各模型的识别效果, 并确定最优模型。SVM 方法通过台湾林智仁研发团队编写的 LIBSVM 软件包中的 C-SVM 实现^[31], 以径向基函数作为核函数, 在 $2^{-10} \sim 2^{10}$ 范围内, 利用网格搜索算法搜索惩罚参数 C 和核函数参数 g , 搜索步距均为 0.4, 遍历网格内所有点针对训练集计算 3 次交叉验证识别正确率, 选择识别正确率最大时的 C 和 g 值作为模型参数, 分别记为 C_{best} 和 g_{best} 。以 C_{best} 和 g_{best} 为模型参数建立病害识别模型, 计算训练集和测试集的识别正确率, 并以此作为模型识别效果的评价依据。

2 结果与分析

2.1 病斑图像分割结果分析

利用结合 K 中值聚类算法和线性判别分析的病斑图像分割方法对 4 种苜蓿叶部病害病斑图像的分割结果如表 2 所示。对于苜蓿褐斑病图像, 分割方法的综合得分的平均值为 0.856 2, 中值为 0.881 0; 召回率的平均值为 0.790 5, 中值为 0.819 9; 准确率的平均值为 0.922 0, 中值为 0.936 6。对于苜蓿锈病图像, 分割方法的综合得分的平均值为 0.906 1, 中值为 0.913 7; 召回率的平均值为 0.851 6, 中值为 0.859 6; 准确率的平均值为 0.960 6, 中值为 0.967 1。对于苜蓿小光壳叶斑病图像, 分割方法的综合得分的平均值为 0.946 2, 中值为 0.958 3; 召回率的平均值为 0.928 7, 中值为 0.949 5; 准确率的平均值为 0.963 6, 中值为 0.969 0。对于苜蓿尾孢菌叶斑病图像, 分割方法的综合得分的平均值为 0.836 9, 中值为 0.849 6; 召回率的平均值为 0.778 6, 中值为 0.793 8; 准确率的平均值为 0.895 1, 中值为 0.910 9。对于全体病害图像, 分割方法的综

合得分的平均值为 0.877 1, 中值为 0.899 7; 召回率的平均值为 0.829 4, 中值为 0.851 4; 准确率的平均值为 0.924 9, 中值为 0.942 4。结果表明, 结合 K 中值聚类算法和线性判别分析的分割方法对病斑图

像的分割效果较好。该分割方法对 4 种病害病斑图像的分割结果如图 1 所示, 原始子图像中的所有病斑均被较好地分割出来, 表明该分割方法可以较好地实现病斑图像的自动分割。

表 2 结合 K 中值聚类算法和线性判别分析的病斑图像分割方法的分割效果

Table 2 Characterization of the effects of lesion segmentation on the images of the four diseases by combining the K -median clustering algorithm and linear discriminant analysis

图像数据集 Image dataset	召回率 Recall		准确率 Accuracy		综合得分 Score	
	平均值	中值	平均值	中值	平均值	中值
	Mean	Median	Mean	Median	Mean	Median
苜蓿褐斑病 Alfalfa common leaf spot	0.790 5	0.819 9	0.922 0	0.936 6	0.856 2	0.881 0
苜蓿锈病 Alfalfa rust	0.851 6	0.859 6	0.960 6	0.967 1	0.906 1	0.913 7
苜蓿小光亮叶斑病 Alfalfa Leptosphaerulina leaf spot	0.928 7	0.949 5	0.963 6	0.969 0	0.946 2	0.958 3
苜蓿尾孢菌叶斑病 Alfalfa Cercospora leaf spot	0.778 6	0.793 8	0.895 1	0.910 9	0.836 9	0.849 6
汇总图像数据集 Aggregated image dataset	0.829 4	0.851 4	0.924 9	0.942 4	0.877 1	0.899 7

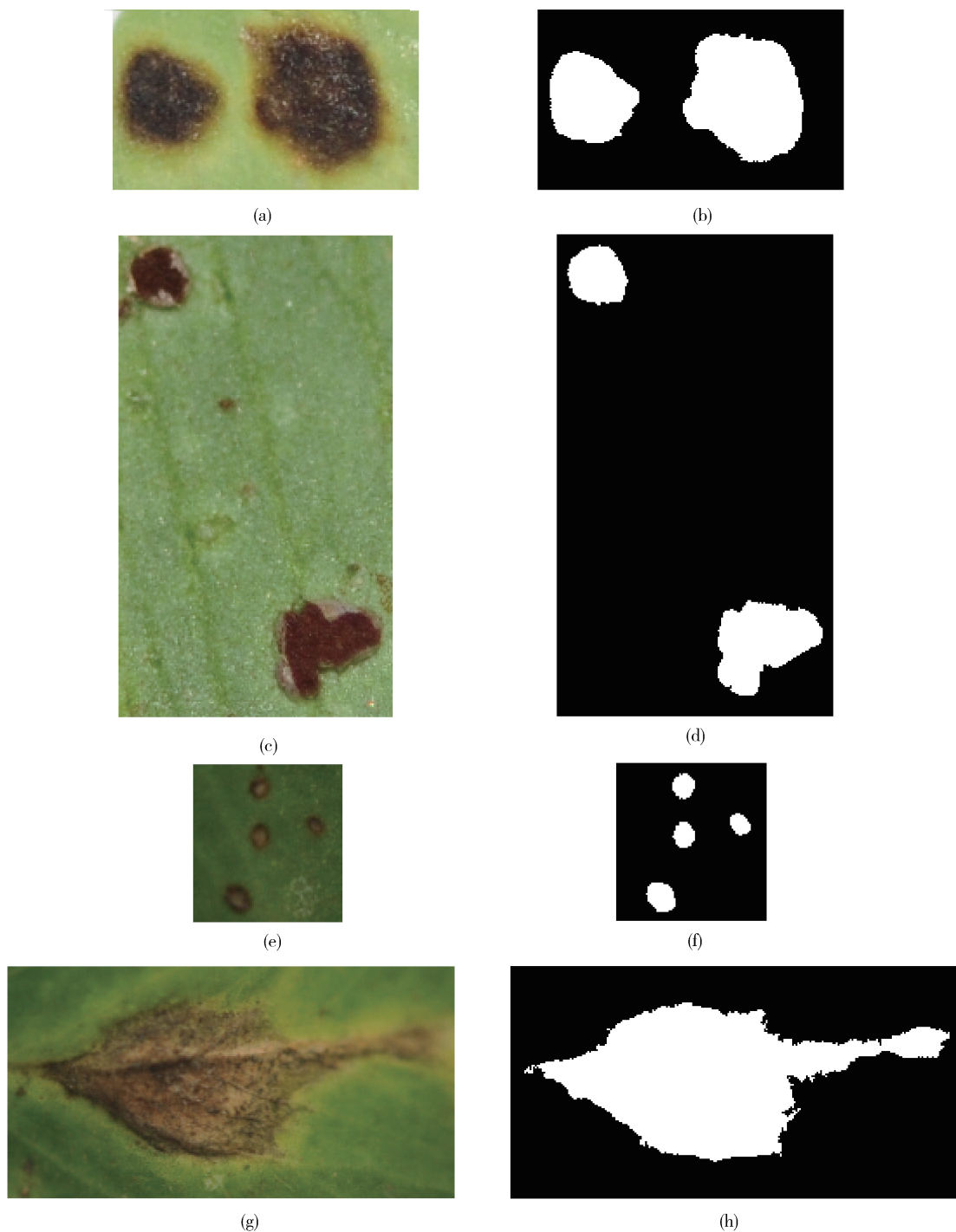
2.2 苜蓿叶部病害识别模型的建立与识别结果比较

分别利用朴素贝叶斯方法和线性判别分析方法建立病害识别模型, 并结合顺序前向选择法筛选特征, 2 种方法所建模型的识别结果如表 3 所示。朴素贝叶斯模型共选择了 8 个特征, 分别为圆度、一阶矩 RGB_G、一阶矩 Lab_b、三阶矩 HSV_H、三阶矩 HSV_S、能量 RGB_R、能量 RGB_B、能量 Lab_b, 将由这 8 个特征组成的最优特征子集命名为特征子集 1。对于所建朴素贝叶斯模型, 训练集识别正确率为 85.00%, 测试集识别正确率为 82.40%。线性判别分析模型共选择了 18 个特征, 分别为 φ Lab_L5、 φ Lab_b1、 φ RGB_R5、 φ HSV_H1、圆度、复杂度、一阶矩 HSV_V、二阶矩 RGB_R、二阶矩 RGB_G、二阶矩 RGB_B、二阶矩 HSV_H、对比度 RGB_G、能量 RGB_B、同质性 RGB_B、能量 HSV_H、同质性 HSV_H、能量 HSV_S、同质性 Lab_L, 将由这 18 个特征组成的最优特征子集命名为特征子集 2。对于

所建线性判别分析模型, 训练集识别正确率和测试集识别正确率均为 92.25%。

分别利用特征子集 1 和特征子集 2, 建立病害识别 SVM 模型 1 和 SVM 模型 2, 模型识别结果如表 4 所示。对于病害识别 SVM 模型 1, 训练集识别正确率为 92.27%, 测试集识别正确率为 88.38%, 模型的最优参数 C_{best} 和 g_{best} 分别为 48.503 和 0.574。对于病害识别 SVM 模型 2, 训练集识别正确率为 96.18%, 测试集识别正确率为 93.10%, 模型的最优参数 C_{best} 和 g_{best} 分别为 9.190 和 1.000。

基于特征子集 1 建立的 SVM 模型识别效果优于所建朴素贝叶斯模型, 其训练集识别正确率提高了 7.27%, 测试集识别正确率提高了 5.98%。基于特征子集 2 建立的 SVM 模型识别效果略优于所建线性判别分析模型, 其训练集识别正确率提高了 3.93%, 测试集识别正确率提高了 0.85%。上述结果表明, 所建病害识别 SVM 模型优于所建朴素贝叶斯模型和线性判别分析模型。



(a) 苜蓿褐斑病病斑图像 Lesion image of alfalfa common leaf spot; (b) 苜蓿褐斑病病斑图像分割结果 Image after lesion segmentation of alfalfa common leaf spot; (c) 苜蓿锈病病斑图像 Lesion image of alfalfa rust; (d) 苜蓿锈病病斑图像分割结果 Image after lesion segmentation of alfalfa rust; (e) 苜蓿小光壳叶斑病病斑图像 Lesion image of alfalfa *Leptosphaerulina* leaf spot; (f) 苜蓿小光壳叶斑病病斑图像分割结果 Image after lesion segmentation of alfalfa *Leptosphaerulina* leaf spot; (g) 苜蓿尾孢菌叶斑病病斑图像 Lesion image of alfalfa *Cercospora* leaf spot; (h) 苜蓿尾孢菌叶斑病病斑图像分割结果 Image after lesion segmentation of alfalfa *Cercospora* leaf spot

图 1 结合 K 中值聚类算法和线性判别分析的病斑图像分割方法的自动分割结果

Fig. 1 Automatic segmentation of lesion images using the segmentation method by combining the K _median clustering algorithm and linear discriminant analysis

表 3 基于朴素贝叶斯方法和线性判别分析方法所建病害识别模型的识别结果

Table 3 Recognition of the four alfalfa diseases using Naive Bayes model and linear discriminant analysis model, respectively

模型 Model	入选特征 ID 号 ID of the selected feature	特征子集 Feature subset	训练集识别正确率/% Recognition accuracy of the training set	测试集识别正确率/% Recognition accuracy of the testing set
朴素贝叶斯模型 Naive Bayes model	64,74,83,97,98,104,110,128	特征子集 1	85.00	82.40
线性判别分析模型 Linear discriminant analysis model	5,15,26,43,64,65,81,85,86, 87,88,106,110,111,113,114, 116,123	特征子集 2	92.25	92.25

表 4 分别基于特征子集 1 和特征子集 2 所建 SVM 病害识别模型识别结果

Table 4 Recognition of the four alfalfa diseases using the SVM models based on the optimal feature subsets obtained by the feature selecting methods

模型 Model	建模所用特征子集 Feature subset used for modeling	C_{best}	g_{best}	训练集识别正确率/% Recognition accuracy of the training set	测试集识别正确率/% Recognition accuracy of the testing set
SVM 模型 1 SVM model 1	特征子集 1	48.503	0.574	92.27	88.38
SVM 模型 2 SVM model 2	特征子集 2	9.190	1.000	96.18	93.10

3 结论与讨论

本研究以苜蓿褐斑病、锈病、小光亮叶斑病和尾孢菌叶斑病为研究对象,利用结合 K 中值聚类算法和线性判别分析的分割方法进行了病斑图像分割,病斑分割效果较好。本研究提取了病斑图像的颜色特征、形状特征和纹理特征共 129 个,利用朴素贝叶斯模型、线性判别分析模型和 SVM 模型对苜蓿褐斑病、锈病、小光亮叶斑病和尾孢菌叶斑病病害图像进行了识别。利用 φ_{Lab_L5} 、 φ_{Lab_b1} 、 φ_{RGB_R5} 、 φ_{HSV_H1} 、圆度、复杂度、一阶矩 HSV_V、二阶矩 RGB_R、二阶矩 RGB_G、二阶矩 RGB_B、二阶矩 HSV_H、对比度 RGB_G、能量 RGB_B、同质性 RGB_B、能量 HSV_H、同质性 HSV_H、能量 HSV_S 和同质性 Lab_L 共 18 个特征组成的最优特征子集,结合 SVM 方法所建模型的识别效果最好,训练集和测试集识别正确率分别为 96.18% 和 93.10%。因此,本研究建议以该模型作为 4 种苜蓿叶部病害的图像识别模型。利用上述 18 个特征,结合线性判别分析方法所建病害识别模型效果次之,

其训练集和测试集识别正确率均为 92.25%。结果表明利用本研究所提方法识别 4 种苜蓿叶部病害是可行的。本研究为苜蓿叶部病害的诊断和鉴别提供了一种方法,为解决苜蓿叶部病害识别难的问题提供了一种途径。

本研究所获取的病害图像分辨率为 $4\ 256 \times 2\ 832$ 像素,但是由于苜蓿病害种类不同,分割获得的病斑图像大小存在一定差异,考虑到本研究所提取的多数病斑图像特征受图像尺寸变化的影响较小,因此,研究中未进行病斑图像大小对病害图像识别影响的分析,针对这一问题需进一步研究和探讨。另外,本研究使用的是具有典型病害症状的图像,但是,病害病斑的形成一般具有不断扩展和变化的过程,病害不同发展阶段的图像识别问题也需进一步研究。

移动互联等现代信息技术是“互联网+”现代农业的重要支撑,对于实现农业现代化非常重要。随着移动互联等技术的发展和进步,基于移动终端的植物病害诊断识别系统和 App 得到迅速发展^[32-33],使得植物病害的诊断与识别更加便捷。苜蓿叶部病

害种类多,特别是叶斑类病害症状相似,导致难以准确诊断与识别。本研究为进一步开发包括更多种类苜蓿叶部病害的病害识别系统奠定了一定基础。在进一步的研究中,我们将采集更多种类的苜蓿叶部病害图片,探索适于移动终端的苜蓿叶部病害图像识别方法,从而促进苜蓿病害的有效防治和苜蓿产业的健康发展。

致谢 感谢中国农业科学院北京畜牧兽医研究所袁庆华研究员、王瑜老师和苗丽宏老师在研究中给予的大力支持和帮助。

参 考 文 献

- [1] 何峰,韩冬梅,万里强,李向林.我国主产区紫花苜蓿营养状况分析[J].植物营养与肥料学报,2014,20(2):503-509
He F, Han D M, Wan L Q, Li X L. The nutrient situations in the major alfalfa producing areas of China [J]. *Journal of Plant Nutrition and Fertilizer*, 2014, 20 (2): 503-509 (in Chinese)
- [2] 南丽丽,师尚礼,张建华.不同根型苜蓿根系发育能力研究[J].草业学报,2014,23(2):117-124
Nan L L, Shi S L, Zhang J H. Study on root system development ability of different root-type alfalfa [J]. *Acta Prataculturae Sinica*, 2014, 23(2):117-124 (in Chinese)
- [3] 李彦忠,南志标.牧草病害诊断调查与损失评定方法[M].南京:江苏凤凰科学技术出版社,2015
Li Y Z, Nan Z B. *The Methods of Diagnose, Investigation and Loss Evaluation for Forage Diseases* [M]. Nanjing: Phoenix Science Press, 2015 (in Chinese)
- [4] Pydipati R, Burks T F, Lee W S. Identification of citrus disease using color texture features and discriminant analysis [J]. *Computers and Electronics in Agriculture*, 2006, 52(1/2): 49-59
- [5] Arivazhagan S, Shebiah R N, Ananthi S, Varthini S V. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features [J]. *Agricultural Engineering International: CIGR Journal*, 2013, 15(1): 211-217
- [6] López-García F, Andreu-García G, Blasco J, Aleixos N, Valiente, J M. Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach [J]. *Computers and Electronics in Agriculture*, 2010, 71(2): 189-197
- [7] Barbedo J G A. An automatic method to detect and measure leaf disease symptoms using digital image processing [J]. *Plant Disease*, 2014, 98(12): 1709-1716
- [8] Sankaran S, Mishra A, Ehsani R, Davis C. A review of advanced techniques for detecting plant diseases [J]. *Computers and Electronics in Agriculture*, 2010, 72(1): 1-13
- [9] Patil J K, Kumar R. Advances in image processing for detection of plant diseases [J]. *Journal of Advanced Bioinformatics Applications and Research*, 2011, 2(2): 135-141
- [10] 李冠林,马占鸿,王海光.基于支持向量机的小麦条锈病和叶锈病图像识别[J].中国农业大学学报,2012,17(2):72-79
Li G L, Ma Z H, Wang H G. Image recognition of wheat stripe rust and wheat leaf rust based on support vector machine [J]. *Journal of China Agricultural University*, 2012, 17(2): 72-79 (in Chinese)
- [11] 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
- [12] Leiva-Valenzuela G A, Aguilera J M. Automatic detection of orientation and diseases in blueberries using image analysis to improve their postharvest storage quality [J]. *Food Control*, 2013, 33(1): 166-173
- [13] 陈占良,张长利,沈维政,陈晓霞.基于图像处理的叶斑病分级方法的研究[J].农机化研究,2008,30(11):73-75,80
Chen Z L, Zhang C L, Shen W Z, Chen X X. Grading method of leaf spot disease based on image processing [J]. *Journal of Agricultural Mechanization Research*, 2008, 30(11): 73-75, 80 (in Chinese)
- [14] 关辉,张长利,张春媛.基于图像处理的黄瓜叶片病斑分级方法的研究[J].农机化研究,2010,32(3):94-97
Guan H, Zhang C L, Zhang C Y. Grading method of cucumber leaf spot disease based on image processing [J]. *Journal of Agricultural Mechanization Research*, 2010, 32(3): 94-97 (in Chinese)
- [15] 李冠林,马占鸿,王海光.基于图像处理的葡萄霜霉病单叶严重度自动分级方法[J].中国农业大学学报,2011,16(6):88-93
Li G L, Ma Z H, Wang H G. An automatic grading method of severity of single leaf infected with grape downy mildew based on image processing [J]. *Journal of China Agricultural University*, 2011, 16(6): 88-93 (in Chinese)
- [16] 毛罕平,张艳诚,胡波.基于模糊C均值聚类的作物病害叶片图像分割方法研究[J].农业工程学报,2008,24(9):136-140
Mao H P, Zhang Y C, Hu B. Segmentation of crop disease leaf images using fuzzy C-means clustering algorithm [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2008, 24(9): 136-140 (in Chinese)
- [17] 李冠林,马占鸿,黄冲,迟永伟,王海光.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(增刊2):32-37
Li G L, Ma Z H, Huang C, Chi Y W, Wang H G. Segmentation of color images of grape diseases using K_means clustering algorithm [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2010, 26(S2): 32-37 (in Chinese)
- [18] Omrani E, Khoshnevisan B, Shamshirband S, Saboohi H, Anuar N B, Nasir M H N M. Potential of radial basis function-

- based support vector regression for apple disease detection[J]. *Measurement*, 2014, 55: 512-519
- [19] Dubey S R, Jalal A S. Fusing color and texture cues to identify the fruit diseases using images[J]. *International Journal of Computer Vision and Image Processing*, 2014, 4(2): 52-67
- [20] Zhang M, Meng Q G. Automatic citrus canker detection from leaf images captured in field[J]. *Pattern Recognition Letters*, 2011, 32(15): 2036-2046
- [21] 刘路, 王太勇, 蒋永翔, 支劲章. 结合 FCM 和 SVM 的纹理分割算法[J]. 计算机工程与应用, 2008, (33): 32-33
Liu L, Wang T Y, Jiang Y X, Zhi J Z. Algorithm of texture segmentation combining FCM and SVM [J]. *Computer Engineering and Applications*, 2008, 44 (33): 32-33 (in Chinese)
- [22] Wang X Y, Zhang X J, Yang H Y, Bu J. A pixel-based color image segmentation using support vector machine and fuzzy C-means[J]. *Neural Networks*, 2012, 33: 148-159
- [23] 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
- [24] Xu G L, Zhang F L, Shah S G, Ye Y Q, Mao H P. Use of leaf color images to identify nitrogen and potassium deficient tomatoes[J]. *Pattern Recognition Letters*, 2011, 32(11): 1584-1590
- [25] Theodoridis S, Koutroumbas K. *Pattern Recognition* [M]. 4th ed. Boston: Academic Press, 2009
- [26] Powers D M W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation [J]. *Journal of Machine Learning Technologies*, 2011, 2(1): 37-63
- [27] Gonzalez R C, Woods R E, Eddins S L. 数字图像处理 (MATLAB 版) [M]. 北京: 电子工业出版社, 2005
- Gonzalez R C, Woods R E, Eddins S L. *Digital Image Processing Using MATLAB* [M]. Beijing: Publishing House of Electronics Industry, 2005 (in Chinese)
- [28] Gonzalez R C, Woods R E. 数字图像处理 [M]. 3 版. 北京: 电子工业出版社, 2011
Gonzalez R C, Woods R E. *Digital Image Processing* [M]. 3rd ed. Beijing: Publishing House of Electronics Industry, 2011 (in Chinese)
- [29] Stricker M A, Orengo M. Similarity of color images [J]. *Proceedings of the International Society for Optical Engineering-Storage and Retrieval for Image and Video Databases III*, 1995, 2420: 381-392
- [30] Cortes C, Vapnik V. Support-vector networks [J]. *Machine Learning*, 1995, 20(3): 273-297
- [31] Chang C C, Lin C J. LIBSVM: A library for support vector machines [J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3): 1-27
- [32] 夏永泉, 王会敏, 曾莎. 基于 Android 的植物叶片图像病害检测 [J]. 郑州轻工业学院学报: 自然科学版, 2014, 29(2): 71-74
Xia Y Q, Wang H M, Zeng S. Plant leaf image disease detection based on Android [J]. *Journal of Zhengzhou University of Light Industry: Natural Science Edition*, 2014, 29(2): 71-74 (in Chinese)
- [33] 屈赞, 陶晔, 王政嘉, 王树桐. 基于 Android 的苹果叶部病害识别系统设计 [J]. 河北农业大学学报, 2015, 38(6): 102-106
Qu Y, Tao B, Wang Z J, Wang S T. Design of apple leaf disease recognition system based on Android [J]. *Journal of Agricultural University of Hebei*, 2015, 38(6): 102-106 (in Chinese)

责任编辑: 杨爱东