

谢泽奇, 张会敏, 张善文, 等. 基于颜色特征和属性约简的黄瓜病害识别方法[J]. 江苏农业学报, 2015, 31(3): 526-530.
doi:10.3969/j.issn.1000-4440.2015.03.010

基于颜色特征和属性约简的黄瓜病害识别方法

谢泽奇, 张会敏, 张善文, 张云龙

(郑州大学西亚斯国际学院, 河南 郑州 451150)

摘要: 为了减少黄瓜叶部病害给农业生产带来的损失, 提高病害的识别率和精度, 提出了一种基于颜色特征和属性约简算法的黄瓜病害叶片分割与识别方法。该方法首先利用最大类间方差(Otsu)阈值法对黄瓜病害叶片图像进行病斑分割; 其次提取病斑图像的36个分类特征, 再利用基于区分矩阵的属性约简算法进行特征选择; 最后利用最近邻分类器进行病害识别。该方法在5种常见黄瓜病害叶片图像数据库上进行了病害识别试验, 结果表明, 识别率高达94.8%。说明, 该方法对作物病害叶片图像识别是有效可行的。

关键词: 颜色特征; 属性约简; 病斑分割; 病害识别

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-4440(2015)03-0526-05

Cucumber disease recognition based on color feature and attribute reduction

XIE Ze-qi, ZHANG Hui-min, ZHANG Shan-wen, ZHANG Yun-long

(SIAS International University, Zhengzhou University, Zhengzhou 451150, China)

Abstract: To reduce the loss caused by cucumber leaf disease, and improve disease recognition rate and accuracy, a leaf image segmentation and disease recognition method for cucumber was proposed based on color feature and attribute reduction algorithm. Firstly, the Otsu algorithm was applied to segment the cucumber diseased leaf image. Secondly, 36 diagnostic characters of disease lesion were extracted, and selected by using attribute reduction algorithm based on discernibility matrix. Finally, the crop diseases were recognized using the nearest neighbor classifier. As an effective and feasible approach for crop disease recognition, this method could recognize as high as 94.8% of five cucumber diseases.

Key words: color feature; attribute reduction; image segmentation; disease recognition

黄瓜病害对其产量和质量带来了巨大损失, 如何利用计算机图像处理技术在黄瓜种植期间实现病害及时检测识别和自动防治在农业物联网技术研究

领域具有重要意义。由于实际病害叶片形状、纹理和颜色的多样性, 使得叶片图像中提取的病害分类特征也各种各样, 因此, 需要进行特征选择^[1]。粗糙集(Rough sets, RS)是一种经典特征选择方法, 能直接从给定问题的描述集出发, 通过不可分辨关系和等价关系给出问题的近似域, 从而得到该问题的内在规律。多年来, RS已成功应用于人工智能、模式识别、信息处理以及决策分析等领域^[2-3]。属性约简是RS的一个重要应用^[4]。基于RS的属性约简是在保持信息系统分类或决策能力不变的前提下, 通过知识化简导出分类问题的分类规则和决策。基于区分矩阵的属性约简策略是建立在两两互异目标

收稿日期: 2014-11-19

基金项目: 国家自然科学基金项目(61272333); 河南省科技厅科技攻关项目(142102310518、142400410853、142300410309); 河南省教育厅科学技术研究重点项目(14B520064、15A520100); 郑州大学西亚斯国际学院校级科研项目(2014KYYB23)

作者简介: 谢泽奇(1981-), 男, 河南镇平人, 硕士, 讲师, 研究方向为计算机应用、图像处理。(E-mail) xzq0413@163.com

通讯作者: 张会敏, (E-mail) zhm0413@163.com

上属性集合的区分矩阵^[5]。区分矩阵的元素代表能够区分两个目标的属性组合。实际应用中可以将区分矩阵中出现次数多少作为属性重要性的判断依据,即出现次数多的属性重要性越大。

近年来,国内外专家学者对基于作物叶片的病害识别进行了广泛研究。耿长兴等分析黄瓜病害图像信息的分布特征,通过线性运算实现病害目标与复杂环境背景快速有效的分离^[6];田有文等利用黄瓜叶片的纹理特征向量结合支持向量机(Support vector machines, SVM)分类方法对黄瓜的霜霉病、白粉病进行识别^[7];张芳等利用颜色、形状、纹理等典型特征参数结合支持向量机的方法对黄瓜叶部病害进行自动识别^[8]。上述作物病害识别方法和系统都在某一方面取得了较好效果,但由于作物病害种类和病害叶片图像的多样性,使得仅利用病害叶片颜色、纹理和形状等进行病害识别的鲁棒性不高。本试验在研究作物病害叶片图像分割的基础上,将基于区分矩阵的属性约简策略应用于黄瓜病害叶片图像分类特征选择中,提出了一种有效的黄瓜叶片病害识别算法。该算法首先利用最大类间方差法对黄瓜叶片图像进行病斑分割;然后提取病斑图像的36个分类特征;再利用基于区分矩阵的属性约简算法对36个特征进行特征选择;最后用最近邻分类器进行病害识别,为其他作物病害识别与防治提供参考。

1 材料与方法

1.1 材料

本试验所使用的病害样本图像均是在陕西杨凌农业示范区科学园黄瓜温室中自然光照的非强光条件下采集得到黄瓜病害叶片图像,黄瓜品种为中农26号。

1.2 方法

1.2.1 作物叶片病斑图像特征提取 黄瓜叶片病斑图像的有效分割是黄瓜病害识别的前提条件^[9-10]。其中,最大类间方差(Otsu)阈值法因其分割精度高、适用范围广等优点成为目前广泛采用的一种图像阈值分割方法^[11]。该方法的工作原理是将病斑叶片图像像素分为病斑和正常两类,计算出两类病害的类间方差值,若类间方差值小于某一给定值,合并初始划分的两类;然后计算此时所有类的类间方差值和分离因素 F 值。若 F 值大于某个给定值,退出该算法;否则,就按顺序在已有类中继续

图像分割,最后得到叶片病斑图像。本试验算法采用Otsu阈值法分别将叶片图像的 H 、 I 、 S 分量进行病斑分割,再转换为二值化图像,分别提取 H 、 I 、 S 的36个分类特征,得到一个特征向量,然后生成包括条件属性集合和结论属性集合、满足粗糙集数据处理要求的二维关系规则表。

由于实际得到的病害叶片图像为红绿蓝(RGB)模式,该色彩模式对光照较敏感,而色饱和强度(HIS)颜色模型反映了人的视觉对色彩的感觉^[12],能克服采集过程中光照对病害识别率的影响,可将RGB模式转换HIS模式。其转换公式如下:

$$\begin{cases} I = (R+G+B)/3 \\ S = 1 - \frac{3}{R+G+B} [\min(R, G, B)] \\ H = \begin{cases} \theta, & \text{若 } G \geq B \\ \theta - 2\pi, & \text{若 } G < B \end{cases} \end{cases} \quad (1)$$

$$\text{式中, } \theta = \arccos \left\{ \frac{[(R-G)+(R-B)]/2}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\}。$$

为了得到能区分作物病害的颜色特征值,分别计算病害叶片图像 R 、 G 、 B 分量和分割后病斑图像的 H 、 I 、 S 分量的均值、方差、偏度、峰值、能量、熵共 $6 \times 6 = 36$ 个统计特征参数^[13]。计算公式如下:

$$\begin{cases} p(b) = h(b)/s \\ b = \sum_{b=1} b p(b) \\ b_K = 1/\delta^3 \sum (b-\bar{b})^3 p(b) \\ \delta^2 = \sum_{b=1} (b-\bar{b})^2 p(b) \\ b_F = 1/\delta^4 \sum_{b=1} (b-\bar{b})^4 p(b) - 3 \\ b_N = \sum_{b=1} [p(b)]^2 \\ b_E = \sum_{b=1} p(b) \lg[p(b)] \end{cases} \quad (2)$$

式②中, $p(b)$ 为灰度级, $h(b)$ 为直方图, s 为图像的面积, b 为灰度均值, \bar{b} 为 b 的均值、 b_K 为偏度, b_F 为峰值, b_N 为能量, b_E 为熵。

1.2.2 属性约简算法进行特征选择 若直接把提取的36个病害图像特征参数用于构建病害识别模型,识别效果的稳定性较差。因为各个特征对识别结果的贡献不同,有些特征甚至影响识别结果。因此,需要对得到的特征进行选择。其中,基于区分矩阵的属性约简方法可有效对特征属性进行约简^[14],其基本思想是将属性在区分矩阵中出现次数作为目标属性的重要性判断函数。计算各属性的重要性判

断函数 f , 由此可确定第一个重要属性, 此时重要属性的判断函数 f 值最大。在第一个重要属性简单求出后, 进一步快速简单地求出后续重要属性。设决策表 S 的核值集合为 $Core$, 其属性约简过程如下:

输入: S 的属性约简矩阵 RM ; 输出: S 的属性约简 $Redu$ 。

①初始化, 令 $Redu = Core, H = \emptyset$;

②令 $Q = \{RM(a_j) \mid a_j \in Redu \text{ 或 } AFI(a_j) = 0\}$, 则 $RM = RM - Q, B = A - Redu - H$;

③对 $a_j \in B$, 计算 $f(a_j) = \max\{a_k\}, (k=1, 2, \dots, m)$, 其中 $f(x)$ 为属性重要性判断函数;

④ $Redu \leftarrow Redu \cup \{a_j\}$;

⑤对每个 $a_j \in B$, 令 $H_2(a_j) = RM(a_j) \cap RM(a_k)$, $H_3(a_j) = |H_2(a_j)|$;

⑥对所有 $a_j \in B$, 令 $RM(a_j) \leftarrow RM(a_j) - H_2(a_j)$, $f(a_j) \leftarrow f(a_j) - H_3(a_j)$; 若 $f(a_k) = 0$, 则 $H_1 = \cup\{a_k\}$;

⑦重复步骤②~⑥, 直到判断函数 AFI 为零向量;

⑧输出 $Redu$ 。 $Redu$ 为信息表的一个属性约简。

本算法使用基于区分矩阵的属性约简方法对得到的 36 个特征进行选择。该方法是先将关系表中的每类特征属性进行归一化和离散化处理; 然后利用基于区分矩阵的属性约简方法对关系表进行属性约简, 依次消去可省略的属性(列)和合并重复的对象(行), 再对每一个对象进行简化, 消去冗余的属性值; 最后根据一定的评选准则选取有效识别规则的属性简化表, 最终获得优选的最简单判定规则。

1.2.3 最近邻分类器进行病斑识别 最近邻分类的思想是: 若待识别模式与样本 W_i 之间的距离最小, 而且 $x^k \in W_i$, 则判定 x 属于 W_i 类^[15]。将 x 与 W_i 类之间的距离定义为:

$$D_i(x) = \|x - x^k\|, x^k \in W_i, IH \quad (3)$$

则最近邻分类方法的决策规则为:

若 $D_i(x) \leq D_j(x), i=1, 2, L, C; j=1, 2, L, C; i \neq j$, 则 $x \in W_i$ 。

由于本算法只针对于黄瓜叶片病斑中最常见的 5 种病斑进行识别, 因此, 定义 D_1, D_2, D_3, D_4, D_5 分别表示待识别叶片病斑特征与细菌性角斑病、炭疽病、霜霉病、白粉病和白粉虱病 5 类病斑之间的距离。

1.3 数据统计分析

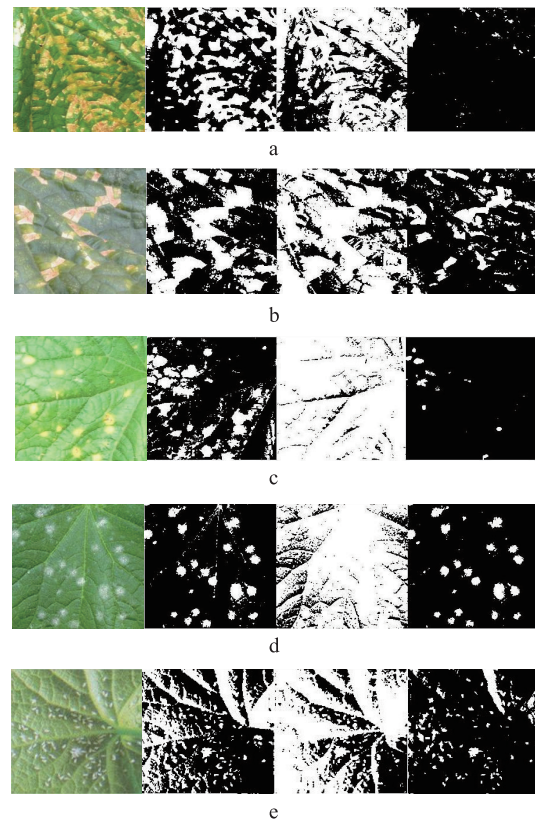
为了验证本试验方法的有效性, 采集温室黄瓜叶片最常见的细菌性角斑病、炭疽病、霜霉病、白粉病和白粉虱病的 5 种病斑叶片各 100 幅, 其中 50 幅图像

作为训练集, 其余 50 幅图像作为测试集。然后在 Windows XP 系统中利用 Matlab7.1 软件中自带的图像处理工具箱为图像处理和平台, 计算以上特征参数, 数据分析采用统计分析软件 SAS V9.2。

2 结果与分析

2.1 病斑分割

采用 Otsu 阈值法分别将病害叶片进行病斑分割, 分割定位待识别的黄瓜叶片病害部分病斑, 利用式(2)提取病害叶片病斑的 36 个分类特征。其中, 5 种黄瓜病害叶片及对应分割病斑图像的 H, I, S 分量如图 1 所示, 最左边图像为 5 种病害叶片原始图像, 右边二值图像为 5 种病斑叶片对应的分割结果, 依次为 H, I, S 分量。5 种病斑图像采用 Otsu 阈值法自动得到二值化图像的熵值、分割阈值和运行时间如表 1 所示。



a: 细菌性角斑病; b: 霜霉病; c: 炭疽病; d: 白粉病; e: 白粉虱病。

图 1 5 种黄瓜病害叶片分割

Fig. 1 Leaf image segmentation of five kinds of cucumber diseases

2.2 病斑识别

将得到的36个特征进行离散化,即把每个特征划分成有限区域(试验中为5个区域),使每个区域中对象的决策值相同。利用基于区分矩阵的属性约简方法对36个特征进行属性约简,得到6个重要属性,分别为病斑分量 H 的均值、方差、能量,分量 I 的均值,分量 S 的方差和能量。再将这6个属性组成一个特征向量来表示这个样本。最后利用最近邻分类器对病害种类进行分类,识别结果见表2,其中,每种病斑50幅图像用于测试,表2给出了基于全部36个特征和基于其他2种方法^[16-17]的识别结果。

表1 5种黄瓜病害叶片分割结果参数

Table 1 Segmentation result of five kinds of cucumber diseased leaves

参 数	细菌性角斑病	霜霉病	炭疽病	白粉病	白粉虱病
二值化图像熵值	0.975 7	0.975 7	0.965 2	0.924 0	0.964 2
分割阈值	148	156	181	131	149
运行时间(s)	1.484 5	1.594 0	2.265 0	1.395 7	1.172 0

表2 黄瓜病害叶片的识别结果

Table 2 Recognition rates of cucumber diseased leaves

病 害	识别率(%)			
	36特征识别法	文献[11]方法	文献[12]方法	本研究算法
细菌性角斑病	78.0	90.0	86.0	92.0
炭疽病	82.0	84.0	80.0	94.0
霜霉病	76.0	86.0	84.0	94.0
白粉病	76.0	82.0	78.0	96.0
白粉虱病	84.0	86.0	82.0	98.0
总计	79.2	85.6	82.0	94.8

从黄瓜叶片图像采集和识别中可得出以下结论:(1)36特征识别率低于本文算法采用的约简后的6特征识别法,因此,病害识别中并非特征越多越好,应对识别特征进行优化组合或特征选择;(2)关于采集的病斑叶片图像提取特征,同一病害样本在不同成像环境下拍摄,图像特征向量存在一定差异;(3)不同黄瓜品种、不同发病时期的病症表现差异较大。这些差异对病斑识别结果的影响可以通过制定图像的采集规范有效解决。

3 讨 论

本研究在结合颜色特征和属性约简算法的基础

可以看出,若只利用全部36个特征进行病害识别,其识别率仅为79.2%,因为36个特征中存在的一些冗余特征影响病害的识别率。与文献中其他算法相比,本研究提出的方法识别率最高,对于黄瓜最常见的病斑识别率可达94.8%,其中对白粉虱病的识别率最高可达98.0%,因为白粉虱病的分割效果较好,特征较明显,对细菌性角斑病识别率达到92.0%,表明该方法用于黄瓜叶部各种常见病害图像识别是有效可行的。

试验过程还发现,病斑分割与识别方法对于黄瓜病害中等程度的识别率略高于重度病害和轻度病害。

上,提出了一种黄瓜叶片病害识别算法。该算法利用最大类间方差阈值法对作物叶片病斑进行图像分割,然后提取病斑图像的36个特征,再利用粗糙集对提取特征进行属性约简,得到了6个重要代表特征。最后利用最近邻分类器对病害进行分类。结果表明该方法具有较好的有效性和可行性。由于病害叶片拍摄状态和光照等环境因素对颜色特征存在影响,发病程度及病斑的典型性可能对识别准确率也有影响,所以对拍摄照片中颜色特征的定量关系有待进一步研究。对于某些颜色、形状特征非常相似的其他黄瓜病症,还应进一步提取病斑的颜色、形状和纹理等特征,以实现病害的有效识别。

参考文献:

- [1] 郭 鹏,李乃祥. 黄瓜病害图像自动分割方法研究[J]. 农机化研究,2014,36(8):10-13,18.
- [2] 王国胤,姚一豫,于 洪. 粗糙集理论与应用研究综述[J]. 计算机学报,2009,32(7):1229-1246.
- [3] 陈 昊,杨俊安,庄镇泉. 变精度粗糙集的属性核和最小属性约简算法[J]. 计算机学报,2012,35(5):1011-1017.
- [4] 杨传健,葛 浩,汪志圣. 基于粗糙集的属性约简方法研究综述[J]. 计算机应用研究,2012,29(1):16-20.
- [5] 马 翔,张继福,杨海峰. 基于区分矩阵的启发式属性约简算法[J]. 计算机应用,2010,30(8):1999-2002,2037.
- [6] 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.
- [7] 田有文,李天来,张 琳,等. 高光谱图像技术诊断温室黄瓜病害的方法[J]. 农业工程学报,2010,26(5):202-206.
- [8] 张 芳,王璐,付立思,等. 基于支持向量机的黄瓜叶部病害的识别研究[J]. 沈阳农业大学学报,2014,45(4):457-462.
- [9] 袁 媛,李 森,陈 晟,等. 复杂背景黄瓜叶部病害图像分割方法[J]. 农业机械学报,2013,44(10):233-237.
- [10] 张 芳,王 璐,付立思,等. 复杂背景下黄瓜病害叶片的分割方法研究[J]. 浙江农业学报,2014,26(5):1346-1355.
- [11] 邹小林,冯国灿. 融合视觉模型和最大类间方差的阈值分割算法[J]. 计算机应用,2013,33(3):670-673,837.
- [12] 温长吉,王生生,于合龙,等. 基于改进蜂群算法优化神经网络的玉米病害图像分割[J]. 农业工程学报,2013,29(13):142-149.
- [13] REVATHI P, REVATHI R. Knowledge discovery in diagnose of crop diseases using machine learning techniques [J]. International Journal of Engineering Science and Technology, 2011, 3(9): 7187-7190.
- [14] 吕跃进,翁世洲,何朝丽. 基于布尔区分矩阵与关联规则挖掘的属性约简算法[J]. 计算机应用与软件,2012,29(10):40-43,116.
- [15] 王卫东,郑宇杰,杨静宇,等. 一种基于预分类的高效最近邻分类器算法[J]. 计算机科学,2007,2(34):198-200.
- [16] 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(4):115-121.
- [17] 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.

(责任编辑:袁 伟)