

基于径向基支持向量机的棉花虫害识别^{*}

张建华¹ 冀荣华² 袁雪¹ 李慧¹ 祁力钧¹

(1. 中国农业大学工学院, 北京 100083; 2. 中国农业大学信息与电气工程学院, 北京 100083)

【摘要】 针对棉花受棉蚜、棉叶螨、棉盲蝽、斜纹夜蛾和烟粉虱等害虫为害后叶片表面出现不同症状,利用计算机视觉技术识别棉花虫害。通过获取受害棉花叶片图像,预处理后转换至 $2G-R-B$ 空间,结合 Otsu 算法实现色斑分割,提取色斑图像 R 变量、 $(R+G+B)/3$ 变量的一阶矩、二阶矩和三阶矩为颜色特征,提取非色斑图像拓扑描述子和 Hu 不变矩为形状特征,提取 2 层双树复小波变换的细节图像均值和方差为纹理特征,并应用径向基支持向量机识别棉花棉蚜、棉叶螨、棉盲蝽、斜纹夜蛾、烟粉虱等虫害和正常叶片。试验结果表明,当径向基参数 σ 为 3 时,棉花虫害识别正确率达 88.1%。

关键词: 计算机视觉 棉花 虫害 支持向量机

中图分类号: TP391.41; S435.622 文献标识码: A 文章编号: 1000-1298(2011)08-0178-06

Recognition of Pest Damage for Cotton Leaf Based on RBF-SVM Algorithm

Zhang Jianhua¹ Ji Ronghua² Yuan Xue¹ Li Hui¹ Qi Lijun¹

(1. College of Engineering, China Agricultural University, Beijing 100083, China

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract

Based on different symptoms on pest damaged cotton leaf including cotton aphid, cotton spider mites, cotton plant bugs, cotton leafworm and whitefly, the recognition system of pest damage for cotton leaf was presented. After collecting cotton images, the mottling areas with cotton spider mites, cotton plant bugs and whitefly were segmented by Otsu method in $2G-R-B$ color space. The mean value, variance value and skewness value of mottling areas were extracted on the R and $(R+G+B)/3$ bands as color features if mottling areas appear, and topological descriptors and Hu invariant moments were extracted as shape features. Two layers dual-tree complex wavelet was used to evaluate the texture features of cotton leaf. A support vector machine (SVM) classifier with radial basis function were employed to classify cotton aphid, cotton spider mites, cotton plant bugs, cotton leafworm, whitefly and normal cotton leaf. Experiment results showed that the classification accuracy was 88.1% when σ was 3.

Key words Computer vision, Cotton, Pest, SVM

引言

棉花在生长过程中极易受害虫的侵入,害虫对棉花的生长和产量有较大影响,不同的棉花害虫需要不同的防治方法。很多农业生产者缺乏农业知识,无法判别虫害原因,而延误了害虫的除治,导致棉花的严重减产。因此,及时发现和识别棉花害虫

具有重要意义^[1-7]。

棉蚜、棉叶螨、棉盲蝽、斜纹夜蛾和烟粉虱等害虫为害田间棉花时大多栖息在叶片背面而不易察觉,且在棉花生长期棉花会受到不同害虫的危害,而在较全面对棉花虫害识别与分类方面,目前尚未见报道。因此,本文针对棉蚜、棉叶螨、棉盲蝽、斜纹夜蛾和烟粉虱等害虫为害棉花叶片的发生症状,对

收稿日期: 2010-11-04 修回日期: 2010-12-20

^{*} 高等学校博士点专项科研基金资助项目(20110008110010)和中国农业大学研究生科研创新专项资助项目(KYCX2011072)

作者简介: 张建华,博士生,主要从事植物病虫害检测及施药技术研究,E-mail: zjhua.2001@163.com

通讯作者: 祁力钧,教授,博士生导师,主要从事植物病虫害检测及施药技术研究,E-mail: qilijun@cau.edu.cn

其为害叶片进行图像分析,提取为害棉花叶片的形状、颜色和纹理特征,并利用支持向量机对棉花虫害进行识别。

1 图像采集

1.1 棉花叶片获取

2010 年 5 ~ 9 月在中国农业大学农业试验基地的棉田中选取被棉蚜、棉叶螨、棉盲蝽、斜纹夜蛾和烟粉虱为害后的叶片。其中,棉蚜主要发生在棉花生长早期,样本采集时间在 5 ~ 6 月,棉叶螨、棉盲蝽和斜纹夜蛾样本采集时间在 6 ~ 8 月,烟粉虱发生在棉花生长后期,在 8 ~ 9 月进行样本采集。每次收集的受害棉花叶片均放入保存箱中,带回实验室进行图像采集,如图 1 所示。

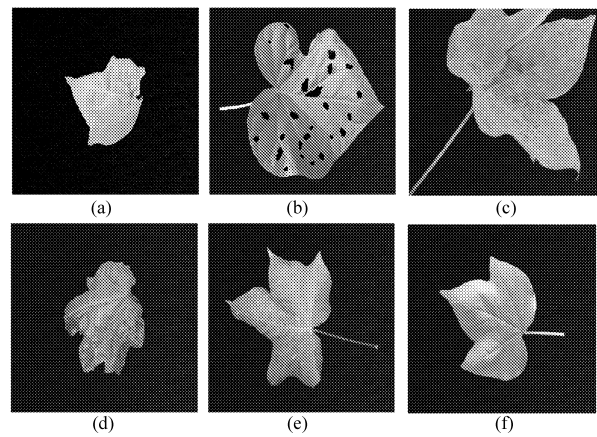


图 1 受害棉花叶片图

Fig. 1 Cotton leaf images damaged by pest

(a) 棉蚜 (b) 斜纹夜蛾 (c) 棉叶螨
(d) 棉盲蝽 (e) 烟粉虱 (f) 正常棉叶

1.2 图像采集系统

实验所用图像采集系统包含光照箱、光源、CCD 数码相机和计算机。其中,光照箱是自制木板立方箱体,体积为 50 cm × 50 cm × 50 cm;光源为 4 盏 30 W 白炽灯,分别固定在光照箱 4 个角;CCD 数码相机 (DSC - W320, Sony, Japan),分辨率为 1 024 × 768;计算机 (CPU AMD Athlon(tm) 5200 + /2.71 GHz, 内存 1 GB);图像采集卡 (Meter, Matrox Inc., Canada) 用于获取棉花图像;由于特征提取时用到双树复小波变换,图像长度、宽度必须相等且是 2 的倍数,因此采集图像大小保存为 512 像素 × 512 像素;Matlab R2008a 作为棉花叶片图像处理和分析平台。

2 图像预处理与色斑分割

2.1 图像预处理

图像预处理的快速中值滤波、背景分割、形态学开运算和逻辑减运算共 4 部分组成。快速中值滤波

可以有效消除微小噪声点,优化图像质量;在可控的均匀照明环境中,背景分割采用了全局阈值法,可以快速将棉花叶片与背景进行分割;形态学开运算,即先进行腐蚀运算,再进行膨胀运算,经试验发现,选择半径为 5 像素的圆盘型结构元素可以有效平滑棉花叶片边缘,并去除叶柄;将开运算后的二值图像与快速中值滤波后图像进行逻辑减运算,最终获得光滑边缘和均匀背景的图像,完成图像预处理。

2.2 色斑分割

色斑分割是病虫害识别过程中重要一项,其分割精度直接决定病虫害识别正确率,但色斑区域与非色斑区域有时难以区分。从棉花受害症状分析出发,发现受棉叶螨、棉盲蝽和烟粉虱为害棉叶会出现明显的色斑,且色斑的颜色和形状各不相同,而受棉蚜、斜纹夜蛾为害和正常棉叶则没有色斑。因此,本文采用如下方法对棉花虫害色斑进行分割:

(1) 棉花虫害色斑为红黄色、黑色或黄色,而叶片非色斑区域一般为绿色,因此将图像转换至 $2G - R - B$ 颜色空间。

(2) 通过加运算,对色斑区域与非色斑区域进行区别。

(3) 通过 Otsu 阈值分割算法确定一个最佳阈值,使图像经二值化处理后,目标与背景两个像素类间方差最大^[8],实现色斑区域分割。

(4) 利用 Otsu 阈值分割区域后图像作为掩模,提取预处理后图像中色斑区域。色斑分割结果如图 2 所示。

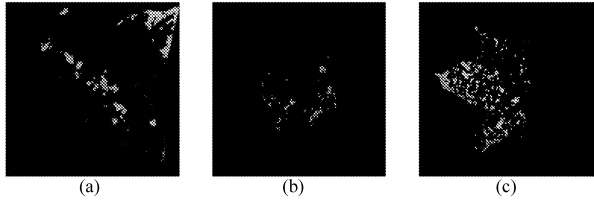


图 2 棉花叶片色斑分割结果

Fig. 2 Cotton leaf mottling segmentation images

(a) 棉叶螨 (b) 棉盲蝽 (c) 烟粉虱

3 特征提取与选择

特征提取和选择是模式识别中的一个关键问题,特征选择的好坏会影响到分类器的设计及性能^[9]。提取色斑的颜色作为特征变量。同时,受棉蚜为害的棉花叶片会发生不同程度的卷曲,故利用二层双树复小波变换提取棉叶的纹理特征,利用 Hu 不变矩提取棉叶的形状特征,并对特征向量进行选择。

3.1 色斑颜色特征提取

对色斑区域的直方图进行统计,从 R 变量和

$(R + G + B)/3$ 变量直方图(图 3)可以看出,棉叶螨、棉盲蝽和烟粉虱色斑区域的 R 变量和 $(R + G + B)/3$ 变量峰值相隔距离较大,且棉叶螨和棉盲蝽分布较集中,灰度分布区间分别为 150 ~ 230 和 50 ~ 150,而烟粉虱分布较广,其灰度分布区间为 50 ~ 225。因此,棉叶螨、棉盲蝽和烟粉虱为害棉花叶片色斑区域的 R 变量和 $(R + G + B)/3$ 变量具有很好的区分性。Stricker 和 Orengo 提出颜色矩表示图像的颜色分布^[10]。由于颜色分布信息主要集中在低阶矩中,因此提取色斑区域 B 变量和 $(R + G + B)/3$

变量的一阶矩、二阶矩和三阶矩作为颜色特征。其公式分别为

$$\mu_1 = \frac{1}{M + N} \sum_{x=1}^M \sum_{y=1}^N |I_i(x, y)| \tag{1}$$

$$\mu_2 = \frac{1}{M + N} \sum_{x=1}^M \sum_{y=1}^N |I_i(x, y) - \mu_i|^2 \tag{2}$$

$$\mu_3 = \frac{1}{M + N} \sum_{x=1}^M \sum_{y=1}^N |I_i(x, y) - \mu_i|^3 \tag{3}$$

式中 $I_i(x, y)$ ——图像灰度
 M, N ——图像长度和宽度

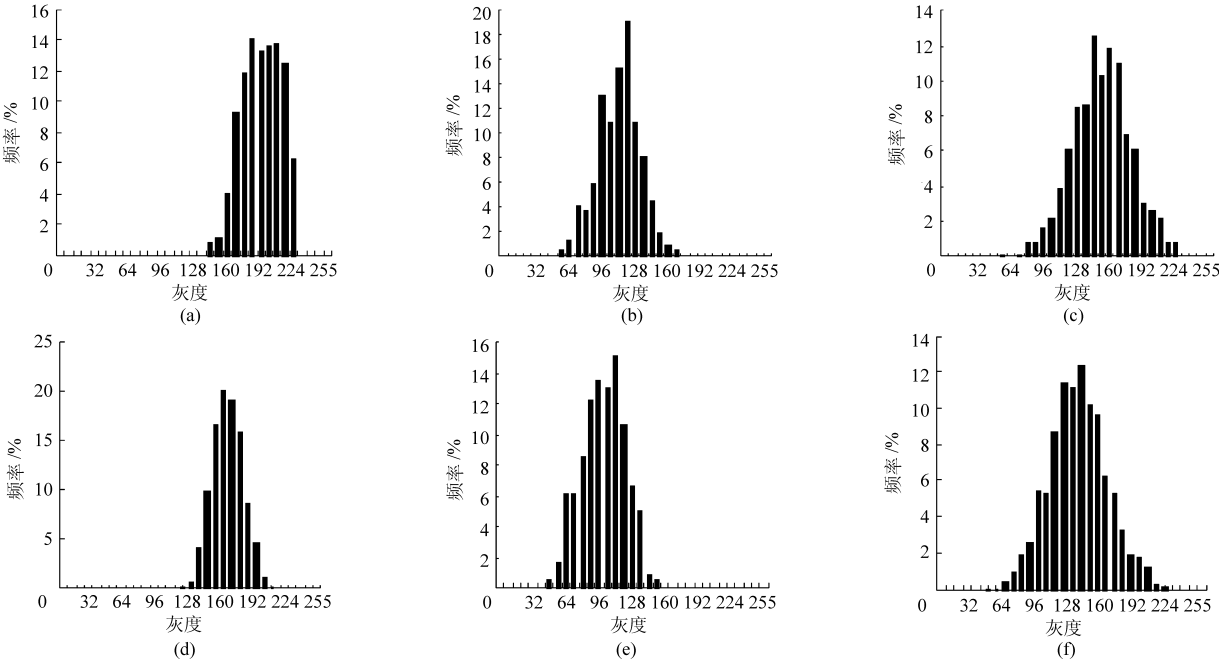


图 3 不同虫害色斑及变量直方图

Fig. 3 Histograms of mottling in R and $(R + G + B)/3$

(a) 棉叶螨, R 变量 (b) 棉盲蝽, R 变量 (c) 烟粉虱, R 变量

(d) 棉叶螨, $(R + G + B)/3$ 变量 (e) 棉盲蝽, $(R + G + B)/3$ 变量 (f) 烟粉虱, $(R + G + B)/3$ 变量

3.2 叶片形状特征提取

3.2.1 拓扑描述子

拓扑描述子是一个拓扑特性分量数^[11],在计算图像连通分量方面优势明显。因此本文采用该方法提取棉花叶片内孔个数。计算公式为

$$E = C - H \tag{4}$$

式中 H ——区域内孔个数
 C ——该区域的连通分量

计算步骤如下:

- (1)将预处理后图像(图 4a)进行灰度化处理,采用固定阈值 254 对其进行图像分割(图 4b)。
- (2)消除孔洞(图 4c)。
- (3)将图 4c 减去图 4b(图 4d),并统计连通分量即区域内孔的个数 H 。
- (4)根据式(4)计算出 E 。

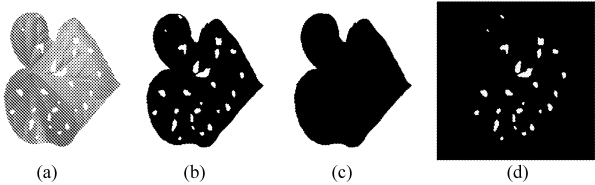


图 4 拓扑描述子提取

Fig. 4 Demonstration images of extracting topological descriptors

(a) 预处理后图像 (b) 阈值分割 (c) 消除孔洞 (d) 4e - 4b 图

3.2.2 Hu 不变矩

Hu 在矩概念的基础上提出了 7 个不变矩^[12],由于不变矩具有平移、比例和旋转等特性的不变性质,且能够描述图像的整体性质,从而在目标识别中得到了广泛的应用。由于棉叶螨、棉盲蝽和烟粉虱为害棉花叶片会出现明显的不同颜色色斑,但叶片形状不会发生明显变化,形状特征不能作为特征变

量用以区分棉叶螨、棉盲蝽和烟粉虱。而受棉蚜为害的棉花叶片会发生不同程度的卷曲,与斜纹夜蛾和正常叶片形状明显不同,因此,只对棉蚜、斜纹夜蛾与正常叶片通过提取 7 不变矩(φ_1 、 φ_2 、 φ_3 、 φ_4 、 φ_5 、 φ_6 、 φ_7)获得其形状因子并作为识别棉蚜叶片特征参数,从表 1 可见,棉蚜与斜纹夜蛾、正常叶片的 φ_3 、 φ_4 、 φ_5 、 φ_6 、 φ_7 变量有较大区别,但 φ_6 、 φ_7 与 φ_4 、 φ_5 有一定相关性。因此,本文使用 φ_3 、 φ_4 、 φ_5 提取

表 1 Hu 不变矩提取结果

Tab.1 Extracting results of Hu invariant moments							
叶片类型	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7
棉蚜	7.22	14.45	29.56	31.67	62.98	44.67	62.47
斜纹夜蛾	7.04	14.09	25.13	27.12	54.58	38.21	53.31
正常叶片	7.20	14.41	25.19	27.54	54.10	38.65	54.41

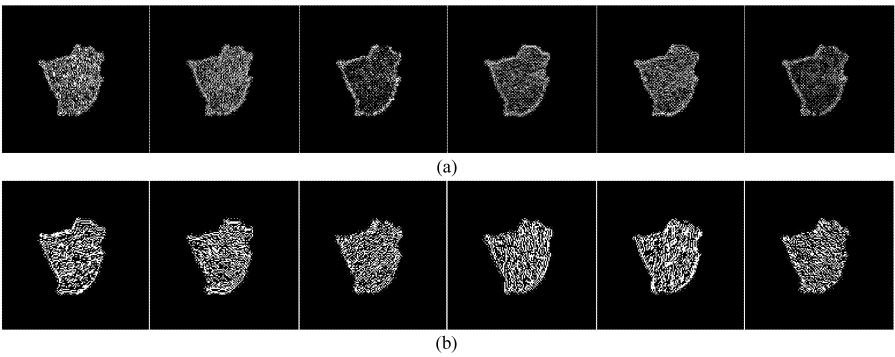


图 5 双树复小波变换 2 层分解细节子图像

Fig.5 Detail sub-band images after two layers dual-tree complex wavelet
(a) 第 1 层分解各实部细节子图像 (b) 第 2 层分解各实部细节子图像

本文分别对这两类图像双树复小波变换 2 层分解得到的 6 个实部细节子图像进行统计,并求取细节系数矩阵的均值和方差。由 2 层 12 个细节子图像的均值和方差组成 24 维的特征向量,作为图像的纹理特征。均值和方差的计算公式分别为

$$\bar{\mu}_i = \frac{1}{M+N} \sum_{x=1}^M \sum_{y=1}^N |I_i(x,y)| \tag{5}$$

$$\sigma_i = \frac{1}{M+N} \sum_{x=1}^M \sum_{y=1}^N |I_i(x,y) - \bar{\mu}_i|^2 \tag{6}$$

4 模式识别

支持向量机 (support vector machine, 简称 SVM) 在解决小样本学习、非线性以及高维等问题中表现出许多特有的优势,在模式识别中得到广泛应用^[14]。SVM 核函数包括线性、多项式、径向基 (RBF) 以及 Sigmoid 核函数。由于 RBF 核函数的计算复杂度不随参数的变化而变化,且在全部参数空间都满足 Mercer 条件,是 SVM 方法中最常用的核函数。因此本文采用径向基核函数

形状特征。
3.3 叶片纹理特征提取

双树复小波变换 (dual-tree complex wavelet transform, 简称 DT-CWT) 具有近似移不变以及更多方向选择性的优点,对于分析纹理特征更有优势^[13]。二维双树复小波变换包含两个平行的小波树,即树 A 和树 B 两个分支。其中上部树 A 的叠加滤波器组表示复小波变换的实部,下部树 B 的叠加滤波器组表示复小波变换的虚部。二维双树复小波变换产生了 6 个方向的高频子图像,6 个小波子带主方向分别为 15°、45°、75°、-15°、-45°、-75°,提取棉花图像 6 个方向上的细节系数矩阵,可实现纹理特征提取。利用二维双树复小波变换进行 2 层分解的分解图像如图 5 所示。

$$K(x_i,y_i) = \exp \left(- \frac{\|x_i - y_i\|^2}{2\sigma_i^2} \right)$$

式中 x_i ——输入样本
 y_i ——高斯函数的中心
 σ_i ——高斯函数的方差

基于径向基支持向量机的棉花虫害识别算法流程图如图 6 所示,该算法由图像预处理、色斑分割、特征提取和模式识别 4 部分组成。
具体步骤如下:

- (1)通过图像预处理去除叶柄,并使背景统一。
- (2)在 2G-R-B 空间,实现色斑区域与棉花叶片的分割。
- (3)判断是否有色斑,如果有色斑,则提取色斑区域的 B 变量和 (R+G+B)/3 变量的一阶矩、二阶矩和三阶矩作为颜色特征;否则应用 Hu 不变矩和拓扑描述子提取叶片的形状特征,并利用双树复小波变换提取叶片的纹理特征。
- (4)通过提取的特征向量和径向基支持向量机实现虫害的识别。

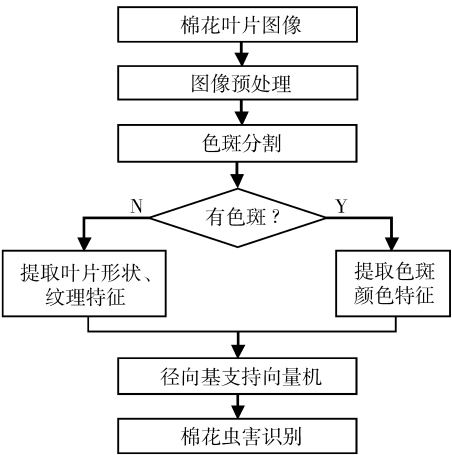


图 6 棉花虫害识别流程图

Fig. 6 Flow chart of cotton pest recognition

5 实验结果

本研究采用 LibSVM 工具箱作为分类器,通过图像采集系统采集棉花虫害图像共 316 幅。在虫害识别试验中,任意选取 158 幅棉花虫害图像作为训练样本,另外 158 幅作为测试样本,将特征变量数据转换成该分类器的输入格式,并分别利用 SVMtrain 函数和 SVMpredict 函数进行样本的训练和测试,惩罚因子 C 为 200, ξ 为 0.01,选取不同 σ 值的径向基核函数对棉花虫害进行识别试验。试验结果如表 2 所示。

表 2 不同 σ 值的径向基核函数棉花虫害识别试验结果

Tab.2 Recognition results of cotton pest with different σ

σ 值	平均识别正确率/%	σ 值	平均识别正确率/%
1	81.2	4	86.8
2	85.9	5	86.8
3	88.1		

由表 2 所示,当 $\sigma = 1$ 时棉花虫害平均识别正确率较低,为 81.2%;当 $\sigma = 3$ 时其平均识别正确率最高,为 88.1%。由表 3 可见,31 幅棉蚜图像中,由于其中 2 幅棉叶上有黑色污点色斑,分割后其颜色正好与棉盲蝽特征相吻合,被错误识别为棉盲蝽,1 幅棉蚜为害初期棉叶没有发生明显卷曲,错误识别为正常叶片,棉蚜识别正确率为 90.3%。28 幅斜纹夜蛾中 2 幅由于枯萎在纹理特征提取与棉蚜相同被错误识别,另有 2 幅棉叶边缘受到斜纹夜蛾为害,拓扑描述子提取形状参数失效,被错误识别为正常叶片,斜纹夜蛾识别正确率为 85.7%。25 幅棉叶螨中,3 幅因失水色斑由红色变成了黄色,错误识别为

烟粉虱,棉叶螨识别正确率为 88.0%。18 幅棉盲蝽中,3 幅被棉盲蝽为害的棉花嫩叶引起的色斑太小且棉花新叶本身卷曲被错误识别为棉蚜,棉盲蝽识别正确率为 83.3%。26 幅烟粉虱,3 幅色斑颜色与棉叶螨颜色相似棉叶错误识别为棉叶螨,1 幅色斑太小在色斑分割中被忽略,错误识别为正常叶片,烟粉虱识别正确率为 84.6%。因叶片上黑色污点,正常叶片 1 幅被识别为斜纹夜蛾,正常叶片识别正确率为 96.6%。试验结果表明:本文方法平均识别正确率达 88.1%,能较好地识别棉花虫害。

表 3 $\sigma = 3$ 时棉花虫害识别结果

Tab.3 Recognition results of cotton pest with $\sigma = 3$

叶片类型	识别类型					
	棉蚜	斜纹夜蛾	棉叶螨	棉盲蝽	烟粉虱	正常叶片
棉蚜	28	2	0	3	0	0
斜纹夜蛾	0	24	0	0	0	1
棉叶螨	0	0	22	0	3	0
棉盲蝽	2	0	0	15	0	0
烟粉虱	0	0	3	0	22	0
正常叶片	1	2	0	0	1	29
总数	31	28	25	18	26	30
识别正确率/%	90.3	85.7	88.0	83.3	84.6	96.6
平均识别正确率/%	88.1					

6 结论

- (1) 色斑分割算法是棉花虫害正确识别的关键,本文将图像转换为 $2G - B - R$ 颜色空间,并经过加运算使色斑与非色斑区别更加明显,利用 Otsu 算法确定的阈值能较好分割出色斑区域。
- (2) 通过分析色斑直方图分布特点,得出棉叶螨、棉盲蝽和烟粉虱为害的棉花叶片表面色斑在 R 变量和 $(R + G + B)/3$ 变量有较好区分性,分别提取色斑 2 个变量的一阶矩和二阶矩作为颜色特征,提取拓扑描述子和 Hu 不变矩作为叶片形状特征,并提取叶片 2 层双树复小波变换后 12 个细节图像的均值和方差作为纹理特征,并应用径向基支持向量机识别棉花棉蚜、棉叶螨、棉盲蝽、斜纹夜蛾和烟粉虱等虫害和正常叶片。
- (3) 对径向基核函数参数 σ 的不同取值进行了棉花虫害识别试验,当 $\sigma = 3$ 时平均识别正确率达到 88.1%,能较好识别棉花虫害,为棉花虫害远程诊断提供了理论基础。

参 考 文 献

- 1 陈月华, 杨方. 基于计算机视觉的害虫识别技术研究进展[J]. 农机化研究, 2008(1): 209 ~ 211.
Chen Yuehua, Yang Fang. Research advances of insect pest identification technology using machine vision[J]. Journal of Agricultural Mechanization Research, 2008(1): 209 ~ 211. (in Chinese)
- 2 陈佳娟, 纪寿文, 李娟, 等. 采用计算机视觉进行棉花虫害程度的自动测定[J]. 农业工程学报, 2001, 17(2): 157 ~ 160.
Chen Jiajuan, Ji Shouwen, Li Juan, et al. Automatic measurement of danger degree of cotton insect pests using computer vision[J]. Transactions of the Chinese Society of Agricultural Engineering, 2001, 17(2): 157 ~ 160. (in Chinese)
- 3 Panigrahi S. Background segmentation and dimensional measurement of corn germplasm[J]. Transactions of the ASAE, 1995, 38(1): 291 ~ 297.
- 4 Paul Biossard, Vincent Martin, Sabine Moisan. A cognitive vision approach to early pest detection in greenhouse crops [J]. Computers and Electronics in Agriculture, 2008, 62(2): 81 ~ 93.
- 5 冀荣华, 祁力钧, 傅泽田. 自动对靶施药系统中植物病害识别技术的研究[J]. 农业机械学报, 2007, 38(6): 190 ~ 192.
Ji Ronghua, Qi Lijun, Fu Zetian. Recognition of plant disease in automatic target spraying system[J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(6): 190 ~ 192. (in Chinese)
- 6 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.
- 7 王守志, 何东健, 李文, 等. 基于核 K-均值聚类算法的植物叶部病害识别[J]. 农业机械学报, 2009, 40(3): 152 ~ 155.
Wang Shouzhi, He Dongjian, Li Wen, et al. Plant leaf disease recognition based on kernel K-means clustering algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(3): 152 ~ 155. (in Chinese)
- 8 章毓晋. 图像工程(上册)图像处理[M]. 北京: 清华大学出版社, 2006.
- 9 张法全. 基于机器视觉和小波分析的农田害虫识别系统[D]. 郑州: 郑州大学, 2007: 61 ~ 63.
- 10 吴东洋, 业宁, 沈丽容, 等. 基于颜色矩的木材缺陷聚类识别[J]. 江南大学学报: 自然科学版, 2009, 8(5): 520 ~ 524.
Wu Dongyang, Ye Ning, Shen Lirong, et al. Clustering method for automatic timber defects detection based on the color moment[J]. Journal of Jiangnan University: Natural Science Edition, 2009, 8(5): 520 ~ 524. (in Chinese)
- 11 冈萨雷斯. 数字图像处理[M]. 阮秋琦, 等, 译. 北京: 电子工业出版社, 2005.
- 12 冀荣华. 变量施药系统中植物识别与定位方法研究[D]. 北京: 中国农业大学, 2008: 49 ~ 50.
- 13 练秋生, 尚燕, 陈书贞, 等. 基于 DT-CWT 和 SVM 的纹理分类算法[J]. 光电工程, 2007, 34(4): 110 ~ 113.
Lian Qiusheng, Shang Yan, Chen Shuzhen, et al. Texture classification algorithm based on DT-CWT and SVM[J]. Opto-Electronic Engineering, 2007, 34(4): 110 ~ 113. (in Chinese)
- 14 张淑雅, 赵一鸣, 李均利. 基于 SVM 的图像分类算法与实现[J]. 计算机工程与应用, 2007, 43(25): 41 ~ 43.
Zhang Shuya, Zhao Yiming, Li Junli, et al. Algorithm and implementation of image classification based on SVM [J]. Computer Engineering and Applications, 2007, 43(25): 41 ~ 43. (in Chinese)
- 15 关海鸥, 许少华, 谭峰. 基于遗传模糊神经网络的植物病斑区域图像分割模型[J]. 农业机械学报, 2010, 41(11): 163 ~ 167.
Guan Haiou, Xu Shaohua, Tan Feng. Image segmentation model of plant lesion based on genetic algorithm and fuzzy neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(11): 163 ~ 167. (in Chinese)