

向日葵叶部病害图像分割方法

李艳雪, 吕芳

(内蒙古工业大学 信息学院, 内蒙古 呼和浩特 010000)

摘要: 为了有效提高向日葵的产量,除了需要掌握向日葵的栽培技术外,研究可以模拟人甚至超越人的视觉功能的图像识别成为研究图像处理的一大关键。为了实现向日葵叶部病害图像的准确识别,需要用分辨率较高的相机在自然光照下采集,之后需要先进行图像预处理与病斑分割,从而达到病害的精确高识别率。采集向日葵病害图像之后,图像预处理部分使用平滑滤波方法进行去噪和增强,选用合适的邻域模板达到良好的处理效果,对预处理之后的病害图像进行分割,病斑分割时将 K-means 聚类方法与 DBSCAN 算法进行对比,最终选择 DBSCAN 算法进行分割。针对随机选取初始聚类中心的问题,DBSCAN 算法主要是改进 K-means 聚类算法和结果对初始聚类中心的敏感依赖程度,主要目的是清晰地分割出病斑部分。

关键词: 采集;图像预处理;颜色空间;图像分割;聚类算法;K-means 聚类;DBSCAN 算法

中图分类号:TN911

文献标识码:A

文章编号:1673-629X(2019)05-0148-04

doi:10.3969/j.issn.1673-629X.2019.05.031

Image Segmentation Method of Sunflower Leaf Disease Image

LI Yan-xue, LYU Fang

(School of Information, Inner Mongolia University of Technology, Hohhot 010000, China)

Abstract: In order to effectively improve the yield of sunflower, besides mastering the sunflower cultivation techniques, researching a image recognition which can simulate people beyond the visual function has become a key to the research of image processing. In order to accurately identify the disease image of sunflower leaves, a higher resolution camera is required to collect in natural light when collecting. Then, image pretreatment and spot segmentation are needed to achieve high recognition rate of disease. Image preprocessing part using smoothing filter to mania and strengthen, choose the appropriate template to achieve well treatment effect. After the pretreatment of image segmentation, in the segmentation of disease spots, K-means clustering method is used to compare the DBSCAN algorithm, and finally DBSCAN algorithm is selected for segmentation. The DBSCAN algorithm is to improve the sensitivity and dependence of K-means clustering algorithm and results on the initial clustering center, aiming at the problem of randomly selecting the initial clustering center. Its main purpose is to clearly segment the disease spots.

Key words: acquisition; image preprocessing; color space; image segmentation; clustering algorithm; K-means clustering; DBSCAN algorithm

0 引言

向日葵是人们生活中必不可少的一部分,并且为农业经济做出了很大的贡献^[1]。但在实际生长中,向日葵病害的种类有很多种,例如:细菌性叶斑病、霜霉病、黑斑病……这些病害的存在很大程度上影响了向日葵的产量。随着国民经济的快速发展和人们对健康生活认知的提高,向日葵的种植面积逐步扩大,向日葵病害越来越不容忽视^[2]。向日葵叶部病害识别也亟待解决。图像处理技术广泛应用于农业领域,如识别黄

瓜叶片病害等。图像预处理与图像分割是图像识别的两大重要步骤。很多技术已经在农作物上得以运用,得到了很好的效果。

1 图像采集

在图像处理之前,图像采集是一个很重要的步骤,关乎之后的处理与识别准确性。通过病害图像可以快速有效地观察向日葵叶部病害信息。显然,早年间黑白相机拍摄得到的图像像素较低,并且设备存储容量

收稿日期:2018-05-17

修回日期:2018-09-25

网络出版时间:2018-12-21

基金项目:国家自然科学基金(61561038);内蒙古科技计划项目

作者简介:李艳雪(1993-),女,硕士研究生,研究方向为图像处理;吕芳,教授,研究方向为图像处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20181221.1445.018.html>

较小,拍摄效果已经不能满足图像处理的需求。在日益发展的今天,高分辨率数码相机已经成为拍摄的重要工具,它不仅存储容量大,还可直接将模拟图像转换为计算机能够识别处理的数字图像,这就避免了转化过程中一些噪声的加入。向日葵病害采集工作都是在田间进行,所以需要设备具有易携带特性,并且因为数码相机的众多优势,文中采用一台 CCD 照相机在自然光照下进行向日葵叶部病害图像的采集。采集得到的图像以 JPEG 格式存储。

2 图像预处理

通过上述操作,采集到的向日葵叶部病害图像为真彩色图像,具有红(R)、绿(G)、蓝(B)三原色。采集过程中因自然光照等各种噪声的影响,采集到的图像存在失真、模糊的情况。如果不去除图像中一些干扰信息,增强有用的可靠信息,会对之后的识别产生很大的影响,因此图像的增强去噪处理是一关键步骤,用来得到最接近病害本身的图像(下面以向日葵叶片叶斑病为例)^[3]。

图像增强技术是在一定标准下,处理后的图像比原图像效果更好。图像增强方法^[4]分为两种:空间域增强和频域增强,传统的图像增强技术大多是基于空间域对图像进行处理。空间域的图像增强技术主要有灰度变换方法、直方图方法和滤波法等。采用这几种图像增强技术对采集到的田间向日葵叶部病害图像进行预处理操作,处理分析如下。

2.1 灰度变换增强方法

该方法不改变原图像中像素的位置,只改变像素

的灰度值,并逐点进行^[5]。

图1为灰度变换增强图像。

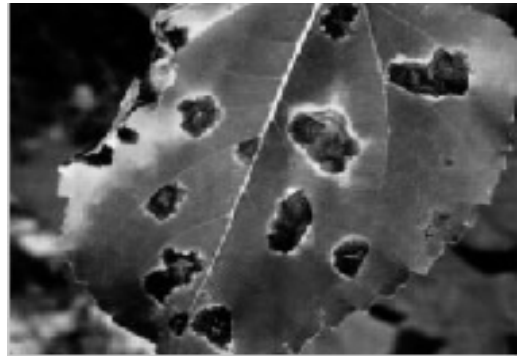


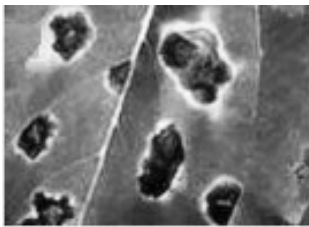
图1 灰度变换增强图像

2.2 直方图增强方法

图像灰度直方图表示灰度图像中具有每种灰度像素的个数,反映了图像中每种灰度级出现的频率,是图像的基本统计特征之一,也是图像对比度增强的最常用方法^[5]。直方图均衡化和直方图规定化是直方图增强常用的两种方法。

2.3 空域滤波法

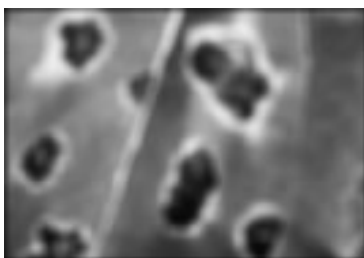
在图像增强中,空域滤波法很常用,并且处理效果较好,可以很好地改善图像质量,增强图像中有用信息,削弱不必要的信息特征^[6]。空域滤波是通过邻域操作完成的,需要用到模板,每一个模板就是一个空域滤波器,模板系数的不同直接影响了去噪增强的效果。常用的滤波器有平滑滤波器、中值滤波器、锐化滤波器等,除了空域滤波增强之外,还有频域滤波增强,这里不做介绍。采用模板进行滤波,对图像的去噪增强处理效果,如图2所示。



(a) 原始图像



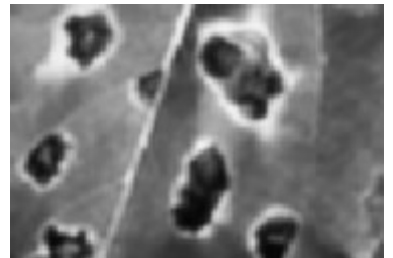
(b) 添加高斯噪声图像



(c) 5*5模板邻域去噪



(d) 3*3模板中值滤波



(e) 3*3模板邻域去噪

图2 图像去噪增强效果

为了与实际契合,图2(b)中人为加入了高斯噪声,并且分别采用不同方法不同模板系数的去噪增强

效果。可以看到,邻域大小直接影响去噪增强的效果,邻域越大,平滑效果越好,但过大会导致处理之后的图

像越发模糊,综合考虑选择比较合适的邻域系数,使用 3×3 模板滤波器。通过观察向日葵原始彩色病害图像与去噪之后的图像,可以看到处理后图像消除了噪声,留下了想要的部分,达到了滤波目的,并且为之后精确的分割做了充分的准备。

3 聚类算法分割对比

3.1 颜色空间的选取和相互转换

颜色模型常用的有 RGB 颜色模型^[7] (CRT, 电视)、CMY 颜色模型(印刷行业)、HSV 颜色模型(人眼模型)、CIE $L^*a^*b^*$ 颜色模型。因为 Lab 颜色模型与人眼所看到的颜色较接近且不受设备影响,同时弥补了 RGB 和 CMYK 两种颜色模型的不足,所以研究采用 Lab 颜色模型^[8-9]。采用 Lab 颜色空间进行向日葵叶部病害彩色图像的分割^[10],因为预处理之后的图像是 RGB 彩色图像,需要进行颜色空间转换。转换公式为:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.240 & 479 & -1.537 & 150 & -0.498 & 535 \\ -0.969 & 256 & 1.875 & 992 & 0.041 & 556 \\ 0.055 & 648 & -0.204 & 043 & 1.057 & 311 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412 & 453 & 0.357 & 580 & 0.180 & 423 \\ 0.212 & 674 & 0.715 & 160 & 0.072 & 169 \\ 0.019 & 334 & 0.119 & 193 & 0.950 & 227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$L^* = 116f(Y/Y_n) - 16$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)]$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)]$$

3.2 K-means 聚类分割原理及 DBSCAN 聚类算法原理

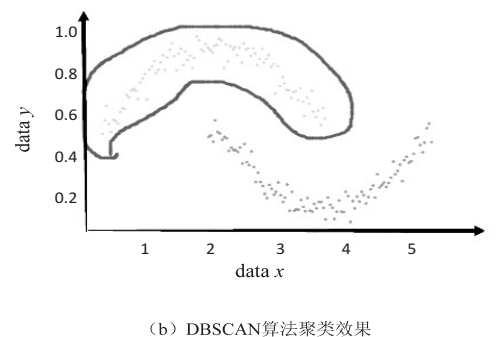
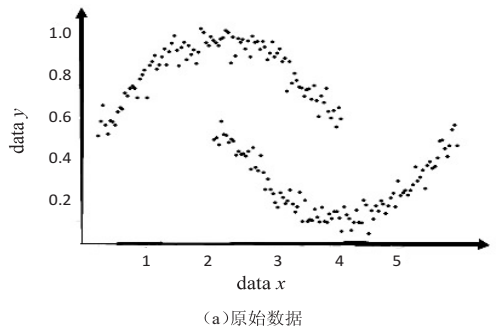
由于计算速度快、聚类效果较好等优点,K-means 聚类算法是最常用的算法之一,也是数据挖掘的重要手段,广泛应用于机器学习领域。K-means 聚类算法^[11-13]以图像的像素为数据点,按照指定的簇数进行聚类,最终将每个像素点以其对应的聚类中心替代,重新组成图像^[14]。K-means 聚类分割^[15]算法简单有效,但是其聚类结果对初始中心值选取较为敏感,从而一定程度上导致聚类结果并不是特别准确^[16]。因此,文中利用 K-means 聚类算法与 DBSCAN 聚类算法分别分割病斑并进行对比。基于密度的 DBSCAN 算法主要针对算法中随机选取初始聚类中心的问题,改善聚类算法和结果对初始聚类中心的敏感依赖程度。因图像数据对象中分散比较随机,可能存在部分数据对象较分散的情况,在选取聚类中心时,可以较好地去除分散的密度低的数据对象并且去除噪声点。这个过程中

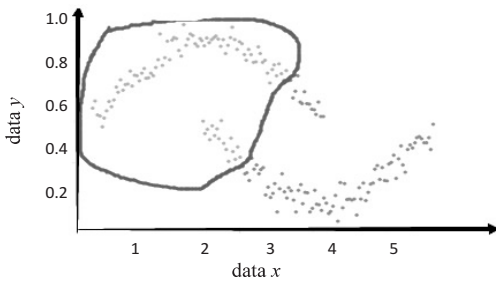
不再使用数据对象的平均值为中心点,而是考虑每簇中居于中心位置的中心点,通过迭代来选择最佳聚类效果。K-means 聚类算法是针对所有对象,所以对距离较远的异常点特别敏感,基于密度的 DBSCAN 算法会把这些异常点当作噪声进行丢弃,不再基于模型做聚类,而是基于密度^[14],这为聚类算法初始中心的选取做了很好的铺垫。

3.3 DBSCAN 算法步骤及仿真结果

DBSCAN 聚类算法^[17] (density-based spatial clustering of applications with noise) 是一种比较有代表性的基于密度的聚类算法。与划分和层次聚类方法不同,它将簇定义为密度相连的点的最大集合,能够把具有足够高密度的区域划分为簇,并可在噪声的空间数据库中发现任意形状的聚类。DBSCAN 聚类算法步骤为:首先需要标准化一组新的测试样本,这些样本应该遵循每个样本的属性及其规范化,将样本归化,归化后的数值进行具体测试;找到中心质点,并且找到最大的半径样本,使用两点之间距离公式,找到最大的价值;将所有样本分配给各自所在的半径范围,这样就可以建立良好的初始聚类中心,并且依次重复此操作(选取最优的聚类中心),最后进行类合并,也就是所谓的聚类。这样的操作使得最后选取的初始聚类中心更加合适,有利于以后的聚类分割,避免了随机选取初始聚类中心,结果较受该中心选取的影响。该算法增强了聚类的稳定性和精确性。

通过举例运行仿真程序,可以发现 K-means 算法与 DBSCAN 算法之间的效果差别(如图 3 所示)。可以看到,通过 K-means 算法和基于密度的 DBSCAN 算法都把数据集分成了鲜明的两个类别。





(c) K-means聚类算法效果

图3 聚类算法结果对比

由图(b)可以看到,轮廓内外部分分别代表了不同的数据对象,分割明显,效果较好;但图(c)中两类数据对象并未明显分开,数据对象产生交叉现象。因此,DBSCAN 算法的效果要比 K-means 算法好很多。

3.4 分割病斑及总结

上节中清晰地看到把原始数据集分成了两类,并且 DBSCAN 聚类效果远优于 K-means 算法,效果很好。由此,现在把两种算法运用到实际中分割病斑。以向日葵叶斑病为例,在 Matlab 软件仿真病害图像结果如图4所示。

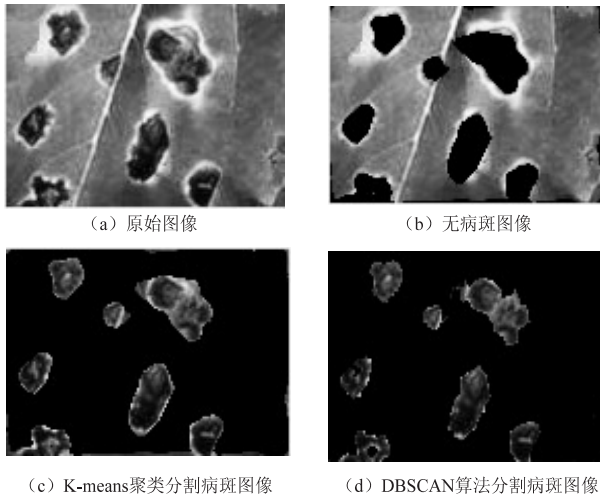


图4 对比聚类分割效果

由图4可看到,基于密度的 DBSCAN 聚类分割效果较好,更贴近病斑原来的形状,病斑部分与无病斑部分几乎完全实现分离,解决了 K-means 算法的不足。综上所述,通过观察结果并考虑其优缺点,研究选择 CIE L^*a^*b 颜色空间,并使用基于密度的 DBSCAN 聚类算法实现向日葵叶部病斑的分离,能够得到高质量的聚类分割效果。

4 结束语

文中研究了向日葵叶部病斑图像,以向日葵叶斑病为例,进行图像采集、图像预处理和病斑分割操作。为了达到图像质量要求并符合田间实际情况,采用高万方数据

分辨率、高像素相机在自然光照下拍摄,采集过程中掺入的噪声用平滑滤波方法进行去噪增强处理,选择了 L^*a^*b 颜色模型,更符合人眼观察效果,采用基于密度的 DBSCAN 聚类方法实现向日葵叶部病斑的有效分割,仿真结果证明了该方法的有效性。

参考文献:

- [1] 徐 箭. 食用向日葵的发展趋向[J]. 现代农业,2012(8): 47.
- [2] 贾永红,孟瑞霞,黄志强,等. 内蒙古河套地区向日葵主要害虫及天敌种类的调查和鉴定[J]. 内蒙古农业大学学报:自然科学版,2009,30(3):74-78.
- [3] 张铮等. 数字图像处理与机器视觉[M]. 北京:人民邮电出版社,2010.
- [4] 高彦平. 图像增强方法的研究与实现[D]. 济南:山东科技大学,2005.
- [5] 杨 丹,赵海滨,龙 哲,等. Matlab 图像处理实例详解[M]. 北京:清华大学出版社,2015.
- [6] 冈萨雷斯. 数字图像处理[M]. 第2版. 阮秋奇,译. 北京:电子工业出版社,2003.
- [7] ZHANG Shichao. Shell-neighbor method and its application in missing data imputation[J]. Applied Intelligence, 2011, 35(1):123-133.
- [8] LIU J Q, YANG Y H. Multiresolution color image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(7):689-700.
- [9] 贾 渊,姬长英. 农产品自动检测中的常见颜色模型[J]. 农机化研究,2004,36(4):205-208.
- [10] CHEN J M, LU H Y. Color image segmentation based on secondary watershed and growcut algorithm[J]. Advanced Materials Research, 2014, 989-994:4032-4037.
- [11] KRISHNA K, MURTY M N. Genetic k-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics: Part B, 1999, 29(3):433-439.
- [12] 熊开玲,彭俊杰,杨晓飞,等. 基于核密度估计的 K-means 聚类优化[J]. 计算机技术与发展,2017,27(2):1-5.
- [13] 程艳云,周鹏. 动态分配聚类中心的改进 K 均值聚类算法[J]. 计算机技术与发展,2017,27(2):33-36.
- [14] ABUBAKER M, ASHOUR W. Efficient data clustering algorithms: improvements over k-means[J]. International Journal of Intelligent Systems & Applications, 2013, 15(3):37-49.
- [15] PANIGRAHI S. Background segmentation and dimensional measurement of corn germplasm[J]. Transactions of the ASAE, 1995, 38(1):291-297.
- [16] 孙士保,秦克云. 改进的 K-平均聚类算法研究[J]. 计算机工程,2007,33(13):200-201.
- [17] 傅德胜,周 辰. 基于密度的改进的 K 均值聚类算法及实现[J]. 计算机应用,2011,31(2):432-434.