

基于 K 均值聚类的小麦腥黑穗病菌冬孢子图像分割

邓继忠¹, 李 敏¹, 袁之报², 王 张³, 李 娇¹

(1 华南农业大学 工程学院, 广东 广州, 510642; 2 海南出入境检验检疫局 热带植物隔离检疫中心, 海南海口, 570311; 3 海南出入境检验检疫局 检验检疫技术中心, 海南海口, 570311)

摘要: 图像分析技术能够实现对小麦腥黑穗病菌冬孢子特征的定量分析, 冬孢子区域的分割则是分析的前提. 针对小麦腥黑穗病菌图像的特点, 考虑到传统分割方法的局限性, 提出了一种基于彩色图像的分割算法. 研究以病菌彩色图像 B 分量为聚类对象, 以 R 、 G 、 B 分量值之和不变为迭代终止条件, 利用 K 均值聚类的方法分割病菌图像, 使类内像素均值的距离和取得局部极小值. 与其他分割方法进行比较的结果表明, 该分割算法既分割出亮度不均匀的背景, 对噪声的敏感度较小, 又减少了分割后的冬孢子粘连, 分割出的冬孢子数目增加.

关键词: 小麦腥黑穗病菌; 图像分割; K 均值聚类; 冬孢子

中图分类号: TP391 文献标志码: A 文章编号: 1001-411X(2012)02-0266-04

Image Segmentation of Wheat Bunt Teliospores Based on K -Means Clustering

DENG Ji-zhong¹, LI Min¹, YUAN Zhi-bao², WANG Zhang³, LI Jiao¹

(1 College of Engineering, South China Agricultural University, Guangzhou 510642, China; 2 Post-Entry Quarantine Station for Tropical Plant, Hainan Entry-Exit Inspection and Quarantine Bureau, Haikou, 570311, China; 3 Inspection and Quarantine Technology Center, Hainan Entry-Exit Inspection and Quarantine Bureau, Haikou, 570311, China)

Abstract: In order to study the features of wheat bunt teliospores to be researched quantitatively through image analysis technique, the region of teliospores should be segmented beforehand. In view of the limitation of traditional segmentation methods, a kind of K -means clustering algorithm for colorized image segmentation was proposed according to the characteristics of wheat bunt. B component was chosen as segmentation clustering features, and the termination condition of iteration was the invariance of the sum of R , G and B components for the sum of distances between intra-class pixels to reach its local minimal. An experiment was conducted for wheat bunt teliospores segmentation and the results showed that the proposed algorithm was superior to other traditional methods since it was capable of filtering asymmetrical background or impurities, reducing teliospore adherences and increasing the number of segmented teliospores.

Key words: wheat bunt; image segment; K -means clustering; teliospores

小麦腥黑穗病是毁灭性的检疫对象, 一般发病田块小麦可减产 10% ~ 20%, 严重的减产 50% 以上, 甚至绝收^[1]. 小麦印度腥、小麦矮腥和小麦网腥 3 种黑穗病常常混入进口小麦原粮中, 而且它们的冬孢子形态特征比较相似^[2], 给检疫造成困难. 目前, 国内主要通过显微镜观察的途径来对其进行检测和识别, 其准确程度与分析人员的经验有关, 难以保证检测的稳定性和客观性.

利用图像分析技术对小麦冬孢子特征进行分析, 是进行检测与识别的有效手段, 而冬孢子图像的分割是对病菌特征量化分析的重要前提^[3]. 但是, 显微图像存在几个特点: (1) 背景亮度不均匀造成灰度变化复杂; (2) 冬孢子核边界复杂且不规则; (3) 载玻片残留杂质; (4) 切片制作时的刀口或抓痕; (5)

冬孢子重叠或粘连. 这些特点都给冬孢子的分割带来了困难,简单的阈值分割等分割方法已难以满足需求.

近年来,为解决显微图像的分割难题,研究者们对各种分割算法进行了深入研究. 目前常用的显微图像分割方法有阈值法,区域增长法和聚类分析法^[4-7]. 陈宇等^[4]把自适应局部阈值分割算法应用于复杂背景下的细胞图像;Shirin Nasr-Isfahani 等^[5]在选择的分离区边界点使用分水岭算法来进行细胞分割. 但由于细胞图像的特殊性,实际分割结果受对比度不明显、背景复杂等影响,分割效果也因病害图像而异. 本文针对小麦腥黑穗病,通过多次试验,确定利用K均值聚类的方法分割小麦病菌图像. 并将该方法与阈值法、区域增长法^[6]等分割方法进行了比较和分析.

1 K均值聚类分割的原理与算法

1.1 K均值聚类的基本原理

K均值聚类(K-Means clustering)算法是一种不断迭代调整K个聚类质心(即类别)的算法^[7].

1.1.1 数据集X划分^[8] 设数据集 $X = \{X_1, X_2, \dots, X_n\}$ 是聚类分析对象的全体(称为论域). X中的每个对象 $X_k (k=1, 2, \dots, n)$ (称为样本)用有限个参数值来刻画,每个参数值刻画 X_k 的某个特征. 聚类分析就是分析论域X中n个样本所对应的模式向量的相似性,按照各样本间的亲疏关系,产生X的K划分. 数据集X的K划分得到的K个子集 $X_1, X_2, \dots, X_K$,如果满足下列条件,则称之为X的硬K划分:
 $\bigcup_{i=1}^K X_i = X$; 当 $i \neq j$ 时, $X_i \cap X_j = \emptyset$; 对所有 i , 有 $\emptyset \subset X_i \subset X$ 成立.

1.1.2 K均值聚类 K均值聚类算法是一种已知聚类类别数的聚类算法. 其基本思想是:指定类别数为K,对样本集合进行聚类,聚类的结果由K个聚类中心来表达,基于给定的聚类目标函数(或者说是聚类效果判别准则),算法采用迭代更新的方法,每一次迭代过程都是向目标函数值减小的方向进行,最终的聚类结果使目标函数值(E)取得极小值,达到较优的聚类效果.

$$E = \sum_{i=1}^k \sum_{x \in Q_j^i} |f(x) - \mu_j^{i+1}|^2, \quad (1),$$

式中, $f(x)$ 代表点x的灰度值; Q_j^i 代表第i次迭代后赋给第j类的像素集合; μ_j^{i+1} 代表第i+1次迭代后第j类的均值. 该函数给出了每个像素与其对应的类均值的距离和.

1.2 基于K均值聚类的小麦病菌图像分割算法

病菌图像分割实质就是把冬孢子从图像中提取

出来,即可以看作是一个两分类问题:一类为需要提取的冬孢子图像,一类为背景和其他无关的噪声. 即 $k=2$.

1.2.1 基于灰度图像的K均值聚类分割 基于灰度图的K均值聚类分割的算法如下:

1) 任意设定2个初始质心即分割灰度值 fAverage1 和 fAverage2;

2) 把待分割图像的每一个像素点p分配到灰度值与之相距最近的那个质心所代表的聚类中,背景赋值为255,目标赋值为0;

3) 根据公式 $\mu_j^{i+1} = \frac{1}{N_j} \sum_{x \in Q_j^i} g(x)$ 计算分配后的新的质心并赋给 fAverage1 和 fAverage2. 其中, N_j 是第j类中的像素个数; $g(x)$ 表示类内像素点p的灰度值, $j=1, 2$.

4) 重复步骤2)和3)直到 fAverage1 和 fAverage2 不再改变.

1.2.2 基于彩色图像的K均值聚类分割 基于彩色图像的K均值聚类分割的算法如下:

1) 任意设定2个初始质心向量 fAverage1 和 fAverage2,以R、G、B分量值作为其分量;

2) 以B分量值为判别标准,判断每一个像素点与初始质心的距离:

① 取点 (x, y) ,其中, $x \in [0, \text{Width}]$, $y \in [0, \text{Height}]$;

② 判断点 (x, y) 与初始质心B分量值的距离,并把该点分配到距离较近的质心所属的类中,设定背景类灰度值为255,目标类灰度值为0;

③ 根据公式

$$\mu_j^{i+1} = \frac{1}{N_j} [\mu_j^i (N_j - 1) + \text{new}]$$

$i \in [0, \text{Width} \times \text{Height}]$, $j=1, 2$, 计算分配后的新的质心并分别赋给 fAverage1 和 fAverage2;其中, N_j 是第j类中的像素个数; μ_j^i 表示第j类中本次循环前质心的R、G、B分量值, new 表示本次循环新分配像素点的灰度值.

3) 重复步骤2)直到质心向量的3个分量值之和不再改变.

2 结果与分析

2.1 试验样本图像

研究所用样本图像由海南出入境检验检疫局提供,均为显微镜下获取的 2048×1536 pixel的JPG格式的小麦腥黑穗病菌图像,包括16幅矮腥黑穗病菌 *Tilletia controversa* Kühn 病菌图像,17幅网腥黑穗病菌 *Tilletia caries* (DC.) Tul. 病菌图像和5幅印度腥黑穗

病 *Tilletia indica* Mitra 病菌图像. 图 1 所示为矮腥黑穗病菌样本图像.

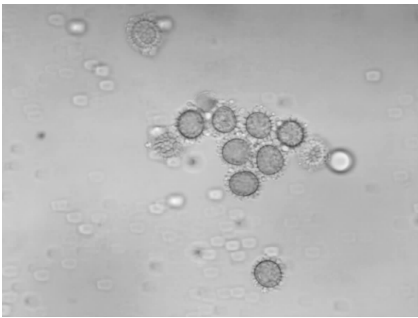


图 1 小麦矮腥黑穗病菌孢子图像

Fig. 1 The image of *Tilletia controversa* Kühn spore

2.2 几种算法的对比结果

利用 Visual C ++ 6.0 编程,将基于 K 均值聚类的图像分割算法与直方图阈值分割法、基于最小均方错分误差法、最小类内方差法、最佳阈值法及区域生长法等算法进行分割的对比试验,分割效果好坏的判据是该算法能否直接完整提取各图像的冬孢子特征区域,或者完整保留冬孢子区域的轮廓以利于后续的进一步提取处理.

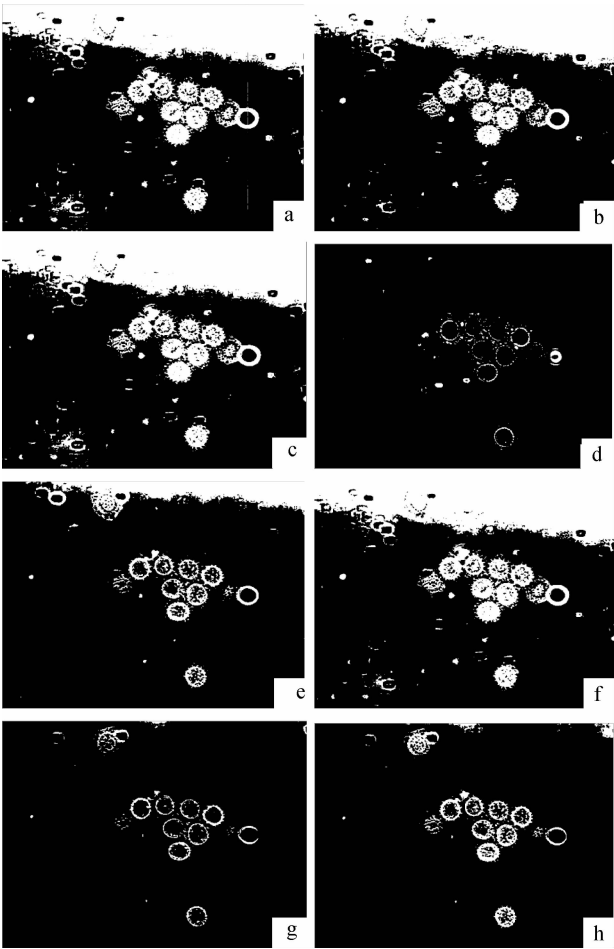
对包括小麦矮腥黑穗病、小麦网腥黑穗病及小麦印度腥黑穗病共 38 幅样本图像进行了各算法的对比试验.

图 2 所示为各算法对其中 1 幅矮腥黑穗病菌图像分割得到的试验结果. 可以看出,基于均方错分误差最小法、最小类内方差、最佳阈值分割等自动阈值分割方法,虽能将冬孢子区域从背景中提取出来,但各冬孢子特征区域有粘连,且不能完全剔除背景;而区域生长法分割过度,后续处理无法完整提取特征区域;直方图阈值分割法的效果相对较好,但需要人机交互.

K 均值聚类对背景的剔除效果优于其他分割方法,但基于灰度的分割有可能使特征区域的边界信息不完整,将在后续处理中导致冬孢子特征区域的丢失,如图 3a 所示. 而基于本文算法的彩色图像分割在保证背景分割效果的同时,更完整地分割出冬孢子.

图 3 分别是对图 2g、图 2h 进行形态学处理后的图像. 图 4 是以图 3b 为基础对原始图像进行逻辑运算,最终分离出的单个病菌孢子图像.

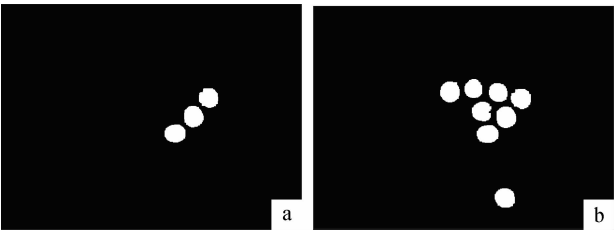
试验结果表明,各算法处理速度无明显优劣,但从冬孢子区域提取是否完整这一效果来看,基于彩色的 K 均值聚类分割的效果要优于其他算法,采用该算法只有 2 幅图像分割不完整,分析原因是拍摄的图像太模糊,而其他算法也无法分割出该图像的冬孢子区域.



a: 最小均方错分误差分割; b: 最小类内方差分割; c: 最大类间方差分割; d: 区域生长法分割; e: 直方图阈值分割; f: 最佳阈值; g: 基于灰度图 K 均值聚类分割; h: 基于彩色图 K 均值聚类分割.

图 2 不同算法分割结果

Fig. 2 Segmentation results of different methods



a: 基于灰度图像的 K 均值聚类分割; b: 基于彩色图像的 K 均值聚类分割.

图 3 形态学处理的图像

Fig. 3 Images after morphology processing

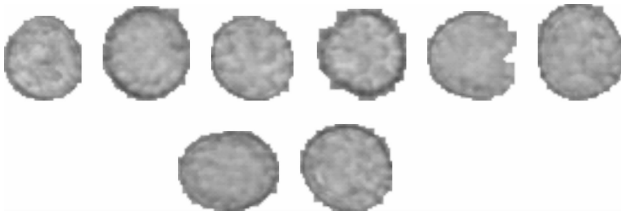


图 4 分离出的单个病菌孢子图像

Fig. 4 The image of single spore after separation

3 小结

病菌显微图像的复杂性使得其分割成为一项困难但却很有意义的课题. 随着信息化的发展,应用图像处理技术研究显微图像的领域越来越多,中外学者也提出了许多分割方法,但始终没有一种通用的分割方法. 而图像分割的误差导致许多病菌分析还是由人工检测完成.

针对小麦病害显微图像的特点,本文提出以病害彩色图像B分量为聚类对象,以R、G、B分量值之和不变为迭代终止条件,利用K均值聚类的方法分割冬孢子图像,使类内像素均值的距离和取得局部极小值. 与基于灰度图的K均值聚类分割和其他分割方法比较,结果表明该分割算法能较好分割亮度不均匀的背景,对噪声的敏感度也较小,减少了冬孢子粘连,分割出的病菌孢子数目也有所增加.

由于采光不均匀和载玻片上存在杂质等原因,使得病菌图像较复杂. 由于本文是基于颜色分量的聚类,光照不均匀仍可能引起某些冬孢子边缘不完整,一些亮度与冬孢子差异很小的噪声也会分割出来. 因此,考虑用纹理分析的方法,减弱颜色对图像分割的影响将是下一步研究的方向.

参考文献:

[1] 中国农业网. 小麦腥黑穗病防治技巧[EB/OL][2010-04-10]. <http://www.zgny.com.cn/ifm/tech/2010-4-10/>

(上接第265页)

本新种与产于福建的紫宽带步甲 *C. obscurus* Xie et Yu(谢为平等^[5]; 陈元洪等^[9])相似,但其前胸背板为盾形,形状与后者明显不同,前半部宽而阔,侧缘中部不呈角状突起,后角钝而阔;此外,新种触角第1节与第3节近等长(在 *C. obscurus* 中则第1节略短),上唇、上颚为黄色而非黑色,鞘翅两侧于中央处近平行,前斑位于第6~10行距而非6~9行距.

与同分布于海南省的非宽带步甲 *C. philippinus* Jedlička, 1939 相比,本种体形稍长,前胸背板不如后者横,侧缘中央不明显膨突,触角仅第1~3黑褐色(后者全黑)、口须黄色(后者口须黑色),且鞘翅前斑起自第6行距(后者起自第5行距),易于区别.

参考文献:

[1] CSIKI E. Carabidae: Harpalinae: III [M]//JUNK W. Coleopterorum Catalogus. Berlin: W Junk, 1929:348-527.
[2] WU C F. Catalogus insectorum sinensium: III: The fan memorial institute of biology [M]. Peiping: The Fan Memorial Intitute of Biology, 1937:1312.
[3] JEDLIČKA A. Neue Carabiden aus Ostasien: XII: Teil.

101504. shtml.
[2] 梁再群,郭翼奋,朱颖初,等. 根据统计分析冬孢子形态特性区分小麦矮腥黑穗病和网腥黑穗病的方法[J]. 植物保护学报,1982,9(4):243-250.
[3] NATTKEMPER T W, TWELLMANN T, RITTER H, et al. Human vs. machine: Evaluation of fluorescence micrographs[J]. Computers in Biology and Medicine, 2003, 33(1):31-43.
[4] 陈宇,范影乐,庞全. 复杂背景下的细胞图像分割技术研究[J]. 计算机工程与应用, 2005, 7(14):42-59.
[5] NASR-ISFAHANI S, MIRSAFIAN A, MASOUDI-NEJAD A. A new approach for touching cells segmentation [C]// [Anon]. Proceedings of the International Conference on Biomedical and Informatics. Sanya: IEEE Computer SOC, 2008:816-820.
[6] 赵进辉,罗锡文,周志艳,等. 基于颜色与形状特征的甘蔗病害图像分割方法[J]. 农业机械学报, 2008, 39(9):100-103.
[7] 魏伟一, 蔺想红. 基于最大区域互信息量的生物细胞图像分割[J]. 西北师范大学学报, 2010, 46(3):44-46.
[8] PAL N R, PAL S K. A review on image segmentation techniques[J]. Pattern Recognition, 1993, 26(9):1277-1294.
[9] 马奎斯德萨 J P. 模式识别:原理、方法及应用[M]. 吴逸飞,译. 北京:清华大学出版社, 2002:66-69.
[10] 刘豪,潘中良. 电路板图像分割的K均值聚类算法研究[J]. 自动化与信息工程, 2009(2):1-3.

【责任编辑 周志红】

[M]. Prague: [S. N.], 1939: 1-8.
[4] JEDLIČKA A. Monographie der Tribus Panagaeini aus Ostasien (Coleoptera: Carabidae) [J]. Annotationes Zoologicae et Botanicae, 1965, 12:1-15.
[5] 谢为平, 虞佩玉. 中国偏须步甲族昆虫的分类研究(鞘翅目:步行虫科) [M] // 张广学. 系统进化动物学论文集. 北京: 科学出版社, 1991:151-172.
[6] KIRSCHENHOFER E. Neue und wenig bekannte Panagaeini deröstlichen Paläarktis sowie der Orientalis (Coleoptera, Carabidae) [J]. Entomofauna, 2000, 21(29):321-372.
[7] BAEHR M. Carabidae: Panaegeini [M] // LÖBL I, SMETANA A. Catalogue of Palaearctic Coleoptera: I: Archostemata-Myxophaga-Adephaga. Stenstrup: Apollo Books, 2003:446-448.
[8] 虞佩玉,唐真正. 鞘翅目:步甲科[M]//黄复生. 海南森林昆虫. 北京:科学出版社, 2002:305-310.
[9] 陈元洪,陈玉妹,胡奇勇,等. 鞘翅目:步甲科 [M]//黄邦侃. 福建昆虫志:VI. 福州:福建科学技术出版社, 2002:11-62.

【责任编辑 周志红】