

基于 K-means 聚类和椭圆拟合方法的白粉虱计数算法

王志彬^{1,2}, 王开义^{1,2}, 张水发^{1,2}, 刘忠强^{1,2}, 穆翠霞³

(1. 北京农业信息技术研究中心, 北京 100097; 2. 国家农业信息化工程技术研究中心, 北京 100097;
3. 中华女子学院计算机系, 北京 100101)

摘要:为了能够对害虫的准确计数, 该文以白粉虱为例, 提出了一种基于 K-means 聚类和椭圆拟合方法的白粉虱计数算法。该方法首先利用 K-means 聚类算法对白粉虱图像进行分割, 使白粉虱从背景图像中分离, 然后利用基于最小二乘法的椭圆拟合方法对分割结果进行椭圆拟合, 统计椭圆的个数, 提取椭圆中心点的颜色特征值, 将其作为新的分类中心, 重新对白粉虱图像进行分割和椭圆个数的统计, 最后将算法收敛时的椭圆个数作为当前白粉虱的个数。对辣椒、黄瓜、番茄和茄子 4 种作物叶片上附着的白粉虱进行了计数试验, 该算法在这 4 种作物上的平均计数错误率依次为 2.80%, 8.51%, 5.00%, 1.56%, 并且分别比阈值化方法和 K-means 聚类方法的平均计数错误率降低了 11.65% 和 70.18%。试验结果表明: 所提方法能够实现对不同作物上白粉虱的准确计数, 且算法具有很好的泛化性。该研究结果可为虫害的检测以及采取正确的防治措施提供重要依据。

关键词:机器视觉, 虫害控制, 算法, 自动计数, K-means 聚类, 椭圆拟合, 白粉虱

doi: 10.3969/j.issn.1002-6819.2014.01.014

中图分类号: S652

文献标志码: A

文章编号: 1002-6819(2014)-01-0105-08

王志彬, 王开义, 张水发, 等. 基于 K-means 聚类和椭圆拟合方法的白粉虱计数算法[J]. 农业工程学报, 2014, 30(1): 105 - 112.

Wang Zhibin, Wang Kaiyi, Zhang Shuifa, et al. Whiteflies counting with K-means clustering and ellipse fitting[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(1): 105 - 112. (in Chinese with English abstract)

0 引言

虫害是导致农作物减产的一个重要因素。及时地发现和控制虫害, 对确保农作物的产量和质量具有十分重要的作用^[1-2]。其中, 对作物叶片上害虫的准确计数是检测虫害信息以及采取正确防治措施的一个重要依据。目前, 常用的计数方法主要分为人工计数法和基于机器视觉的计数方法 2 大类^[3-5]。

基于机器视觉的计数方法主要利用图像处理、模式识别等技术对农作物虫害图像进行自动分析, 有效地识别出害虫的种类及数量。该方法具有速度快、精度高、客观性强等特点, 克服了人工计数方法中的劳动量大、效率低、计数结果不准确等缺点, 成为了农作物病虫害防治领域的研究热点^[6-10]。沈佐锐等用了基于熵的分割算法对白粉虱图像进行分割, 然后对分割后的二值图像利用区域标记算法得

到白粉虱的数量^[11]。翁桂荣则利用数学形态学及分水岭分割算法实现对飞行状态的蝗虫种群密度监测^[12]。邱白晶等通过建立 G 分量阈值来区分黄瓜蚜虫区域和非蚜虫区域, 然后将图像中连通域的个数标记为蚜虫的个数^[13]。Boissard 等利用认知视觉技术实现对玫瑰叶片上白粉虱的计数^[14]。Bechar 等首先对虫害图像进行 RGB - 灰度的线性变换, 然后对虫害区域进行建模, 从而实现了对害虫的识别与计数^[15]。Shariff 等则通过模糊逻辑数字图像处理算法实现了对稻田病虫害图像的识别和个数的统计^[16]。

在上述方法中, 主要采用了图像分割技术^[17-19], 将目标图像从背景图像中分割出来, 然后再进行个数的统计。但是该技术主要存在以下问题: 1) 如何确定图像分割阈值。在图像分割时往往采用固定阈值或者固定参数, 这使得算法在对虫害图像分割时效果不理想; 2) 大多数计数算法主要针对一种作物进行学习和试验, 当算法应用于其他作物时, 算法的移植性差、计数精度不高。因此, 如何提高计数算法的鲁棒性和准确性, 是基于机器视觉的计数方法的一个重要研究方向。

白粉虱 *Trialeurodes vaporariorum* (Westwood), 又称小白蛾子, 是温室、大棚内种植作物的主要虫害之一^[11]。该虫以群集在作物叶背面吸取汁液, 使

收稿日期: 2013-07-30 修订日期: 2013-11-18

基金项目: 农业部公益性行业科研专项项目 (201203026)

作者简介: 王志彬 (1983 -), 男, 博士, 主要从事农作物病虫害方面的研究。北京 北京农业信息技术研究中心, 100097。

Email: wangzb@nercita.org.cn

通信作者: 王开义 (1974 -), 男, 博士, 副研究员, 主要从事农产品流通信息化方面的研究。北京 北京农业信息技术研究中心, 100097。
Email: wangky@nercita.org.cn

受害叶片褪绿变黄、萎缩，甚至枯死，同时还能导致煤污病的发生，严重影响了种植作物的产量和质量。本文以白粉虱为对象，提出了一种基于 K-means 聚类和椭圆拟合方法的白粉虱计数算法，以期为实现田间农作物病虫个数的在线计数提供新的技术和方法。

1 材料与方法

1.1 白粉虱图像

在大田开放环境下，本文采集了北京市小汤山国家精准农业研究示范基地中的辣椒、黄瓜、番茄和茄子 4 种作物叶片上附着的白粉虱（包括烟白粉虱和温室白粉虱）图像。

本文采用了人工手持数码相机拍摄的方式进行白粉虱图像的采集。所用数码相机为 SONYDSC-W35（分辨率为 2048×1536 ）。拍摄图像时，相机设置为自动调节焦距和光圈，自动白平衡，关闭闪光灯，采取遮阴（避免阳光直接照射）和避开刮风、降雨天气拍摄病虫图像。为获取清晰的病虫图像，消除运动模糊的干扰，拍摄时将植株固定，减少抖动。所采集的部分白粉虱图像如图 1 所示。



a. 辣椒
a. Pepper



b. 黄瓜
b. Cucumber

图 1 辣椒和黄瓜叶片上的白粉虱图像

Fig.1 Images of pepper and cucumber leaves with whitefly

从图 1 可以看出，所采集的白粉虱图像具有以

下 2 个显著特点：1) 白粉虱大小不一，且存在严重的粘连。白粉虱在不同生长期，其大小、颜色、形状均存在一定程度上的差异，且白粉虱虫体较小、粘连现象严重；2) 采集的作物、时间以及地点不同，白粉虱的背景图像不同。在农田环境下，由于光照条件、背景和叶片的不确定性，使得算法应用背景变得更加复杂、多变。上述白粉虱图像的特点为准确地分割、统计白粉虱的数量增加了难度。

1.2 研究方法

1.2.1 K-means 聚类算法

聚类是无监督学习的一种重要形式，其主要任务是将数据对象分为多个类或簇，同一簇内的数据对象相似度尽可能大，而不同簇间的数据对象相似度尽可能小^[20]。其中，K-means 聚类是解决聚类问题的一种经典算法，使用广泛^[21-23]。该算法具体如下：1) 随机从数据集中选取 K 个点作为初始聚类中心；2) 计算各数据到聚类中心的距离，把数据归到离它最近聚类中心所在的类；3) 计算新形成的每一个聚类数据的平均值，得到新的聚类中心；4) 若相邻 2 次的聚类中心没有任何变化，说明样本调整结束，聚类准则函数已经收敛。

由于图像分割可以看成聚类问题，即图像中像素点类别未知的前提下，根据像素点的特征值，将图像划分为若干个区域。因此，本文利用 K-means 聚类算法实现对白粉虱图像的分割，即将其分为 2 大类：白粉虱和背景。

1.2.2 基于最小二乘法的椭圆拟合

经图像分割后的白粉虱在外形上类似椭圆，因此本文采用椭圆拟合的方法实现对其个数的统计。

在笛卡尔坐标系下，任意位置上的一个椭圆可用 5 个参数来确定，由该参数确定的椭圆方程描述为^[24]

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + 1 = 0 \quad (1)$$

式中，各项的系数 A 、 B 、 C 、 D 、 E 可采用最小二乘法来进行拟合，即通过计算二值图像中边缘点到理想拟合椭圆的距离平方和最小^[25]。按照最小二乘法原理将式 (1) 转换为如下目标函数

$$F(A, B, C, D, E) = \sum_{i=1}^N (Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + 1)^2 \quad (2)$$

式中， N 为二值图像中边缘点的总个数； (x_i, y_i) 为待拟合椭圆边缘点的坐标。

按照极值原理，若要使 F 值最小，则可通过对各参数求偏导数，如式 (3) 所示

$$\frac{\partial F}{\partial A} = \frac{\partial F}{\partial B} = \frac{\partial F}{\partial C} = \frac{\partial F}{\partial D} = \frac{\partial F}{\partial E} = 0 \quad (3)$$

由此，得到椭圆拟合的一个线性方程组，然后

对该方程组采用矩阵求逆或高斯列主元消去等方法即可得到椭圆方程的 5 个参数 A 、 B 、 C 、 D 、 E 。

根据上述椭圆参数，可求解出平面坐标系中椭圆中心的位置 (x_c, y_c) ，长半轴 a 和短半轴 b ，计算公式如下

$$x_c = \frac{BE - 2CD}{4AC - B^2}, \quad y_c = \frac{BD - 2AE}{4AC - B^2} \quad (4)$$

$$a = 2 \times \sqrt{\frac{-2F}{A + C - \sqrt{B^2 + (\frac{A - C}{F})^2}}} \quad (5)$$

$$b = 2 \times \sqrt{\frac{-2F}{A + C + \sqrt{B^2 + (\frac{A - C}{F})^2}}} \quad (6)$$

根据式(4)、式(5)、式(6)，即可实现对椭圆的拟合，并根据椭圆长半轴和短半轴的大小筛选出符合要求的椭圆，实现白粉虱个数的准确统计。

1.2.3 白粉虱计数方法

本文将 K-means 聚类算法与椭圆拟合方法进行结合，提出了一种白粉虱计数方法。该方法首先采用 K-means 聚类算法对白粉虱图像进行分割，然后利用最小二乘法的椭圆拟合方法对分割结果进行椭圆拟合，并对椭圆的个数进行统计，以此作为白粉虱的个数。该算法具体如下：1) 设定白粉虱颜色特征值和背景图像的颜色特征值为 2 个分类中心；2) 对待计数图像分块，利用中间块图像更新白粉虱颜色特征值；3) 计算各像素到 2 个分类中心的距离，将像素归到距离最小的那个聚类中心所在的类；4) 对图像分割结果进行椭圆拟合；5) 根据椭圆的长、短半轴，对椭圆进行筛选和个数的统计；6) 若相邻 2 次的椭圆个数没有变化，则将椭圆个数作为白粉虱的个数输出，否则，更新分类中心，重新分割和计数。

与 K-mean 聚类算法相比，本文所提出的白粉虱计数算法主要做了以下 3 点改进：

1) 初始分类中心的选择

初始分类中心的选择对于实现白粉虱图像的准确分割具有十分重要的作用。若采用固定的分类中心必然存在算法的适应性问题。为此，本文采用了图像自学习的方法，利用待计数图像进行初始分类中心的学习与选择。

白粉虱图像的分割可以看成是背景和白粉虱的分类问题。由于白粉虱与背景图像在颜色上存在较大的差异，因此可以选择颜色特征作为分类的特征值。在背景颜色特征值计算方面，可以利用图像中的 R 、 G 、 B 各颜色分量的平均值作为背景图像的颜色特征或分类中心。由于白粉虱个体

小，在图像中占的比例较低，对背景颜色特征的计算影响较小，利用该方法求取背景的颜色特征具有较高的稳定性，且对于不同植物都具有很好的效果。

在白粉虱特征值计算方面，由于白粉虱在不同背景下、不同生长期内的颜色不同，且同一副图像中的白粉虱颜色也略有差异。为了准确地计算出白粉虱的颜色特征，本文设计了一种白粉虱颜色特征值的计算方法：首先根据白粉虱虫体的特点人为设定白粉虱的初始颜色特征值，然后对待计数图像进行分块，选择中间块图像进行分割、椭圆拟合，在此基础上，计算各椭圆中心点处颜色特征值的平均值，并将该值作为新的白粉虱分类中心。其中，对图像进行分块，可以避免整幅图像中白粉虱颜色的变化；利用中间块图像进行白粉虱颜色特征值的更新，可以提高算法的适应性，实现在不同作物上进行白粉虱的计数。

2) 椭圆拟合与白粉虱计数

形状是白粉虱虫体的一个重要视觉特征。在对图像进行分割后，通过基于最小二乘法的椭圆拟合方法即可实现对白粉虱虫体的椭圆拟合以及准确计数。椭圆拟合结果如图 2 所示。



注：椭圆是所提算法的拟合结果，圆点是拟合的椭圆中心点。

Note: Ellipse is the fitted result of the proposed method, and circle dot is the center of ellipse.

图 2 白粉虱椭圆拟合结果

Fig.2 Ellipse fitting results of whitefly

根据图 2 中白粉虱椭圆拟合的结果，经式(4)、式(5)、式(6)，即可求出拟合的椭圆中心点、长半轴、短半轴以及椭圆的面积。由此可计算出图像中白粉虱的大小，并筛选出符合要求的椭圆，去除噪声点的干扰，实现白粉虱的准确计数。但是，当图像中白粉虱存在粘连现象时，会使椭圆的面积大小不一，这样同一个椭圆所代表的白粉虱个数也就不同。为此，本文设计了一种简单的计数方法：首先根据中间块图像中白粉虱面积的平均值，去除当前图像中面积过大和过小的椭圆，然后统计剩余

椭圆的平均面积，当拟合的椭圆面积大于该平均面积的 1.5 倍时，则将其计数为 2 个白粉虱，否则计为 1 个白粉虱。利用该方法能够间接处理粘连问题，提高计数结果的准确度。

3) 算法收敛与分类中心更新

准确的图像分割结果是保证白粉虱计数结果精确的基础。为此，本文利用拟合的椭圆个数来衡量当前图像分割结果的准确性，即：若相邻 2 次的椭圆个数没有变化，则说明当前图像分割效果最好，算法已收敛，此时收敛的椭圆个数即为图像中最准确的白粉虱个数；否则，说明分割结果不理想，需要更新分类中心。

在更新白粉虱分类中心时，由于椭圆的中心往往就是白粉虱的中心（如图 2 所示），其颜色特征值具有代表性。为此，可提取图像中的多个椭圆中心，并求取均值作为图像白粉虱新的颜色特征值，重新对图像进行分类和个数的统计。

2 结果与分析

2.1 图像处理

在试验中，为便于对图像进行处理，本文利用双线性插值方法将采集的白粉虱图像缩放为 1300×1000 ，并分为 $5 \times 5 = 25$ 块的子图像。其中，所采集的白粉虱图像中间区域干扰因素较少，本文利用中间块图像进行算法的学习，如第 8、13、18 块，以获得图像中白粉虱的颜色、形状、大小等信息。在实际应用中，可根据采集图像的大小，选择合适的分块份数。分块结果如图 3 所示。

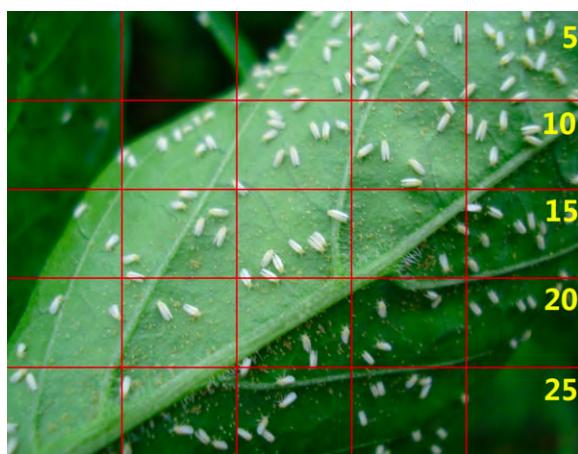


图 3 白粉虱图像分块结果

Fig.3 Partition results of whitefly image

2.2 图像分割与椭圆拟合

按照本文算法，对采集的白粉虱图像进行了分割、椭圆拟合，其中以图 3 中第 7 块图 4a 为例对试验结果进行说明，试验结果如图 4 所示。

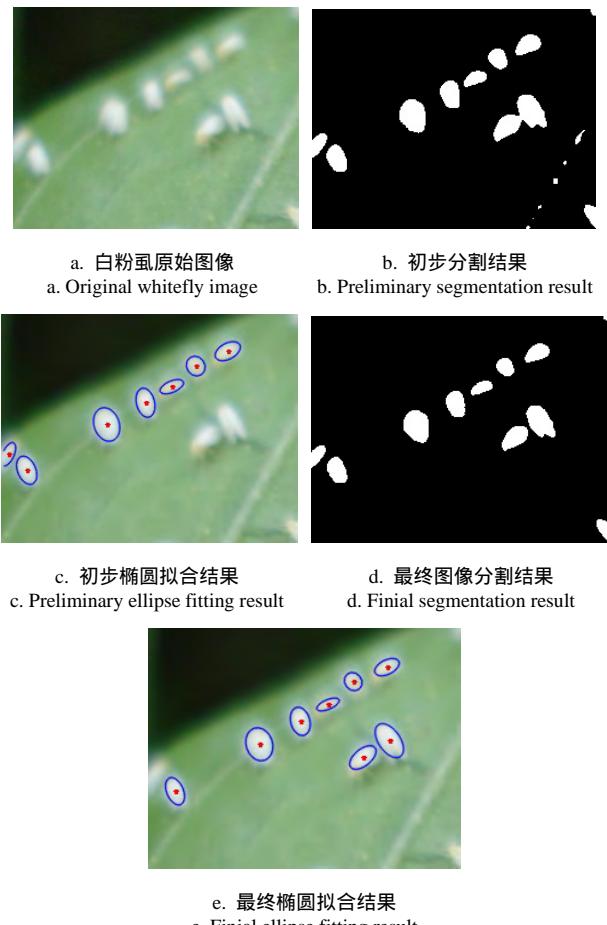


图 4 辣椒白粉虱图像分割与椭圆拟合结果

Fig.4 Segmentation and ellipse fitting results of pepper whitefly image

对图 4a 进行分割后的初步试验结果为图 4b。从图 4b 可以看出，图像的分割效果明显，但存在部分粘连和噪声点（如叶片脉络）。造成该问题的一个重要原因是因为白粉虱颜色特征值计算的不准确，即中间块和边缘块图像之间白粉虱颜色存在差异。对图 4b 的分割结果进行椭圆拟合，拟合结果为图 4c。从图 4c 可以看出，粘连部分和噪声点没有实现椭圆的拟合，其原因为利用中间块图像学习时，可以获得图像中白粉虱的大小、面积等信息，从而筛选出合适的椭圆。但由于粘连部分拟合椭圆过大，而噪声点部分拟合的椭圆过小，不符合图像中白粉虱的大小，从而均被排除。为提高图像分割质量和椭圆拟合的准确度，本文利用当前图像的拟合结果进行学习，更新白粉虱的分类中心。如图 4c 所示，本文提取了椭圆中心点处的颜色特征值，并取其均值作为新的白粉虱分类中心。按照该分类中心进行分割和椭圆拟合的结果如图 4d 和 4e 所示。从图 4d 可以看出，分割效果比图 4b 更明显，去除了粘连和噪声点。在此基础上，实现了更好的椭圆拟合结果，如图 4e 所示。但是，在图 4e 中图像边缘处的部分白

粉虱图像未能实现拟合，其原因为本文所提出的计数算法是建立在对整幅图像分块处理的基础之上，这就不可避免的造成对图像边缘处白粉虱完整性的破坏，从而影响了椭圆拟合的精度。

对各分块图像的椭圆拟合结果进行整合，则图像最终拟合结果如图 5 所示。

2.3 白粉虱计数结果

为验证所提算法的有效性，本文对采集的辣椒、黄瓜、番茄和茄子 4 种作物叶片上附着的白粉虱（包括烟白粉虱和温室白粉虱）进行了计数试验，重点以辣椒和黄瓜 2 种作物为主，并将所提算法与常用的图像分割方法如阈值化方法、K-means 聚类算法进行了对比。

在试验中，本文分别利用阈值化方法和 K-means 聚类算法对白粉虱图像进行分割，然后直接对分割结果中的连通区域进行计数，以此作为图像中白粉虱的个数。其中，阈值化方法中的分割阈值则采用了该算法在辣椒作物上的最优分割阈值 ($T=200$)；K-means 聚类算法的初始分类中心的背景分类中心采用了图像中 R 、 G 、 B 各颜色分量的平均值，白粉虱分类中心则根据虫体的颜色特征选择固定值 $R=209$ 、 $G=220$ 、 $B=205$ 。算法计数错误率的计算方法如式 (7) 所示；计数试验结果如表 1 所示。

$$\text{error rate} = \left(\sum_{i=1}^n |d_i - t_i| \right) / \sum_{i=1}^n t_i \quad (7)$$

式中， d_i 为算法识别的白粉虱个数； t_i 为人工计数的白粉虱个数。

表 1 本文算法与阈值化方法、K-means 聚类算法的计数精度比较

Table 1 Comparison of counting precision of proposed method respectively with that of thresholding algorithm and K-means clustering algorithm

类别 Category	人工计数 Manual counting	自动计数方法 Automatic counting methods					
		阈值化方法 Thresholding method		K-means 聚类方法 K-means clustering method		本文方法 Proposed method	
		识别数 Correct number	错误率 Error rate/%	识别数 Correct number	错误率 Error rate/%	识别数 Correct number	错误率 Error rate/%
辣椒 Pepper	130	85	34.62	76	41.54	128	1.54
辣椒 Pepper	105	100	4.76	128	21.90	102	2.86
辣椒 Pepper	125	123	1.60	89	28.80	120	4.00
茄子 Eggplant	64	54	15.63	32	50.00	63	1.56
番茄 Tomato	140	137	2.14	213	52.14	133	5.00
黄瓜 Cucumber	115	75	34.78	301	161.74	102	11.30
黄瓜 Cucumber	70	84	20.00	187	167.14	66	5.71
均值 Average			16.22		74.75		4.57

从表 1 可以看出，本文算法在辣椒、黄瓜、番茄和茄子 4 种作物上的计数错误率分别为：在辣椒作物上，错误率分别为 1.54%、2.86% 和 4.00%，平均错误率为 2.80%；在茄子上，错误率为 1.56%；在番茄上，错误率为 5.00%；在黄瓜上，错误率分



a. 辣椒
a. Pepper



b. 黄瓜
b. Cucumber

图 5 辣椒和黄瓜叶片上的白粉虱椭圆拟合结果
Fig.5 Ellipse fitting results of whitefly on pepper and cucumber leaves

别为 11.30% 和 5.71%，平均错误率为 8.51%。该算法在 4 种作物上均具有较高的白粉虱计数精度，但是在不同作物上，计数结果具有一定的差异性。在辣椒作物上计数精度较高，而在黄瓜作物上的计数精度略低。主要原因为白粉虱附着在作物的叶片

上，作物的种类不同，叶片特征不同。对于辣椒作物而言，叶片光滑而无太多的纹理，白粉虱附着在叶片上，清晰、干扰因素少，从而容易分割和计数；对于黄瓜作物而言，叶片大而薄、具有刺毛和气孔，白粉虱附着在叶片上时，会被刺毛覆盖，增加了白粉虱分割的难度，从而造成了计数结果的不准确。此外，不同作物上，白粉虱所处的生长期不同，如幼虫和成虫，也会造成计数结果的差异。

本文方法分别比阈值化方法和 K-means 聚类方法的平均计数错误率降低了 11.65% 和 70.18%。对于阈值化方法，由于算法的分割阈值采用了该算法在辣椒作物上的最优分割阈值，这就使得算法在辣椒作物部分图像上具有较好的分割和计数效果，而对于其他作物则图像分割和计数结果较不理想，说明了算法的泛化能力差；对于 K-means 聚类算法，由于白粉虱个体小，在图像中占的比例较低，构成了不平衡数据的分类问题，若直接采用 K-means 聚类算法对其进行分类，难以实现对白粉虱和背景图像的准确分类，图像分割效果差。此外，上述 2 种分割方法直接对分割结果中的连通区域进行计数，难以正确区分白粉虱虫体区域和非虫体区域，严重影响了算法的计数精度。

然而，本文所提出的计数算法则弥补了上述缺陷。该算法能够自适应的学习图像中白粉虱和背景图像的特征，避免了算法固定阈值或者固定参数的设定，增强了算法的适应性，且通过对分割结果进行椭圆拟合，并将收敛的椭圆个数作为白粉虱的个数，不但确保了图像分割的效果，而且能够正确的区分白粉虱虫体区域和非虫体区域，确保了计数结果的准确性。

3 讨论

在农田环境下，由于光照条件、背景和叶片的不确定性，使得基于机器视觉的计数方法还停留在实验室阶段，未能实现对田间农作物病虫个数的在线计数。因此，如何提高计数算法的鲁棒性和准确性，是该类方法的一个重要研究方向。已有的研究成果，如文献[13]通过 G 分量阈值来区分黄瓜蚜虫区域和非蚜虫区域，然后将图像中连通域的个数标记为蚜虫的个数，实现了对黄瓜蚜虫的准确计数。但是该方法在对农作物图像进行分割时，采用 G 分量阈值，这使得算法受应用环境影响较大，未能实现不同作物上蚜虫的计数。而文献[5]则采用了监督学习算法 SVM (support vector machine) 实现了对水稻灯诱虫的准确识别与计数。但是该类方法主要是针对一种作物进行学习和试验，当算法应用于其他作物时，需要重新采集数据样本进行学习，算法

的移植性差。

而本文所提出的白粉虱计数方法，是一种无监督学习算法。该方法具有以下 2 个特点：

1) 主动学习性：该方法能够自适应的学习待计数图像中白粉虱和背景图像的特征，避免了算法固定阈值或者固定参数的设定，且无需建立样本数据库，增强了算法的适应性，实现了对不同作物上白粉虱的计数；

2) 椭圆拟合：对白粉虱图像的分割结果进行椭圆拟合，并将收敛的椭圆个数作为白粉虱的个数，不但确保了图像分割的效果，而且能够正确的区分分割结果中的白粉虱虫体区域和非虫体区域，从而避免了噪声的干扰，进一步提高了计数结果的准确性。

由于本文所提出的计数方法是一种基于 K-mean 聚类的算法，这就不可避免的存在聚类中心的选择问题^[26-28]。虽然本文利用了中间块图像的颜色特征值作为聚类中心，但是对整幅图像分块后，不可避免的造成对图像边缘处白粉虱完整性的破坏，从而影响了算法的计数精度。此外，如何利用图像的多特征^[29-30]来提高算法的计数精度，也将是本文算法改进的方向。

4 结 论

本文提出了一种基于 K-means 聚类和椭圆拟合方法的白粉虱计数算法，在辣椒、黄瓜、番茄和茄子 4 种作物上进行了白粉虱计数试验，试验结果表明：

1) 所提方法在辣椒、黄瓜、番茄和茄子 4 种作物上均具有较高的白粉虱计数精度，平均计数错误率为 4.57%，比阈值化方法和 K-means 聚类方法分别降低了 11.65% 和 70.18%。

2) 所提方法能够自适应的学习待计数图像中白粉虱的颜色、形状、大小等信息，从而有利于算法在不同作物上实现白粉虱图像的分割和计数，提高了算法的适应性和泛化性。

3) 所提方法能够充分利用白粉虱的颜色和形状 2 个重要视觉特征，通过图像分割和椭圆拟合的方式实现了两者的融合，进一步提高了白粉虱计数结果的准确性。

如何选择聚类中心以及利用图像的多特征来提高算法的计数精度，将是本文算法改进的 2 个重要方面。此外，图像中白粉虱的粘连也是造成计数结果不准确的一个重要原因，改进算法有待进一步的研究。

[参 考 文 献]

- [1] 周志艳，罗锡文，张杨，等. 农作物虫害的机器检测与监测技术研究进展[J]. 昆虫学报，2010，53(1)：98 - 109.
Zhou Zhiyan, Luo Xiwen, Zhang Yang, et al. Machine

- based technologies for detecting and monitoring insect pests of crops: A review[J]. *Acta Entomologica Sinica*, 2010, 53(1): 98 - 109. (in Chinese with English abstract)
- [2] Huddar S R, Gowri S, Keerthana K, et al. Novel algorithm for segmentation and automatic identification of pests on plants using image processing[C]// Third International Conference on Computing Communication & Networking Technologies, Coimbatore, India, 2012, 1 - 5.
- [3] 姚青, 吕军, 杨保军, 等. 基于图像的昆虫自动识别与计数研究进展[J]. *中国农业科学*, 2011, 44(14): 2886 - 2899. Yao Qing, Lü Jun, Yang Baojun, et al. Progress in research on digital image processing technology for automatic insect identification and counting[J]. *Scientia Agricultura Sinica*, 2011, 44(14): 2886 - 2899. (in Chinese with English abstract)
- [4] Cho J, Choi J, Qiao M, et al. Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis[J]. *International Journal of Mathematics and Computers in Simulation*, 2007, 1(1): 46 - 53.
- [5] Yao, Q, Lü J, Liu Q, et al. An insect imaging system to automate rice Light-Trap pest identification[J]. *Journal of Integrative Agriculture*, 2012, 11(6): 978 - 985.
- [6] Wen C, Guyer D. Image-based orchard insect automated identification and classification method[J]. *Computers and Electronics in Agriculture*, 2012, 89:110 - 115.
- [7] Shen W Z, Zhang C L, Chen Z L. Research on automatic counting soybean leaf aphids system based on computer vision technology[C]// International Conference on Machine Learning and Cybernetics, Hong Kong, China, 2007, 1635 - 1638.
- [8] Singh C B, Jayas D S, Paliwal J, et al. Identification of insect-damaged wheat kernels using short-wave near-infrared hyper spectral and digital colour imaging[J]. *Computers and Electronics in Agriculture*, 2010, 73(2): 118 - 125.
- [9] Zou X, Ding W. Design of processing system for agricultural pests with digital signal processor[J]. *Journal of Information and Computational Science*, 2012, 9(15): 4575 - 4582.
- [10] Solis-Sanchez L O, Garcia-Escalante J J, Castaneda-Miranda R, et al. Machine vision algorithm for whiteflies (*Bemisia tabaci* Genn.) scouting under greenhouse environment[J]. *Journal of Applied Entomology*, 2009, 133(7): 546 - 552.
- [11] 沈佐锐, 于新文. 温室白粉虱自动计数技术研究初报[J]. *生态学报*, 2001, 21(1) : 94 - 99. Shen Zuorui, Yu Xinwen. Preliminary research on automated counting technology for *Triaeurodes vaporariorum* (Westwood)[J]. *Acta Ecologica Sinica*, 2001, 21(1): 94 - 99. (in Chinese with English abstract)
- [12] 翁桂荣. 数学形态学在害虫种群密度监测中的应用[J]. *农业工程学报*, 2008, 24(11) : 135 - 138. Weng Guirong. Monitoring population density of pests based on mathematical morphology[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2008, 24(11): 135 - 138. (in Chinese with English abstract)
- [13] 邱白晶, 王天波, 李娟娟, 等. 黄瓜蚜虫的图像识别与计数方法[J]. *农业机械学报*, 2010 ,41(8) :151 - 155. Qiu Baijing, Wang Tianbo, Li Juanjuan, et al. Image recognition and counting for glasshouse aphis gossypii[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(8): 151 - 155. (in Chinese with English abstract)
- [14] Boissard P, Martin V, Moisan S. A cognitive vision approach to early pest detection in greenhouse crops[J]. *Computers and Electronics in Agriculture*, 2008, 62(2): 81 - 93.
- [15] Bechar I, Moisan S, Thonnat M, et al. On-line video recognition and counting of harmful insects[C]// 19th International Conference on Pattern Recognition, Istanbul Turkey, 2010: 4068 - 4071.
- [16] Shariff A R M, Aik Y Y, Hong W T, et al. Automated identification and counting of pests in the paddy fields using image analysis[C]// 4th World Congress on Computers in Agriculture, Florida, USA, 2006: 759 - 764.
- [17] 刘艳, 赵英良. Otsu 多阈值快速求解算法[J]. *计算机应用*, 2011 , 31(12) : 3363 - 3365. Liu Yan, Zhao Yinglang. Quick approach of multi-threshold Otsu method for image segmentation[J]. *Journal of Computer Applications*, 2011, 31(12): 3363 - 3365. (in Chinese with English abstract)
- [18] Ning J, Zhang L, Zhang D, et al. Interactive image segmentation by maximal similarity based region merging[J]. *Pattern Recognition*, 2010, 43(2): 445 - 456.
- [19] 陈月华, 胡晓光, 张长利. 基于机器视觉的小麦害虫分割算法研究[J]. *农业工程学报*, 2007 , 23(12) : 187 - 191. Chen Yuehua, Hu Xiaoguang, Zhang Changli. Algorithm for segmentation of insect pest images from wheat leaves based on machine vision[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2007, 23(12): 187 - 191. (in Chinese with English abstract)
- [20] Duda R O, Hart P E, Stork D G. *Pattern Classification*[M]. Wiley-Interscience, 2001.
- [21] Yusoff I A, Matisa N A, Hasikin K. Automated two-dimensional K-means clustering algorithm for unsupervised image segmentation[J]. *Computers and Electrical Engineering*, 2013, 39(3): 907 - 917.
- [22] 李灿灿, 王宝, 王静, 等. 基于 K-means 聚类的植物叶片图像叶脉提取[J]. *农业工程学报*, 2012 , 28(17) : 157 - 162. Li Cancan, Wang Bao, Wang Jing, et al. Extracting vein of leaf image based on K-means clustering[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2012, 28(17): 157 - 162. (in Chinese with English abstract)
- [23] Yao H, Duan Q, Li D, et al. An improved K-means clustering algorithm for fish image segmentation[J]. *Mathematical and Computer Modelling*, 2013, 58(3/4): 790 - 798.
- [24] 闫蓓, 王斌, 李媛. 基于最小二乘法的椭圆拟合改进算法[J]. *北京航空航天大学学报*, 2008 , 34(3) : 295 - 298. Yan Bei, Wang Bin, Li Yuan. Optimal ellipse fitting method based on least-square principle[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2008, 34(3): 295 - 298. (in Chinese with English abstract)
- [25] Li L, Feng Z, He K. A randomized algorithm for detecting multiple ellipses based on least square approach[J]. *Opto-Electronics Review*, 2005, 13(1): 61 - 67.
- [26] Erisoglu M, Calis N, Sakallioğlu S. A new algorithm for initial cluster centers in k-means algorithm[J]. *Pattern Recognition Letters*, 2011, 32(14): 1701 - 1705.
- [27] Wu D, Zhang Y, Yang F, et al. Improved K-means algorithm based on optimizing initial cluster centers[J]. *ICIC Express Letters*, 2013, 7(3B): 991 - 996.

- [28] Cao F, Liang J, Jiang G. An initialization method for the K-Means algorithm using neighborhood model[J]. Computers & Mathematics with Applications, 2009, 58(3): 474 - 483.
- [29] Ai D, Duan G, Han X, et al. Multiple feature selection and fusion based on generalized N-dimensional independent component analysis[C]// 21st International Conference on Pattern Recognition, Tsukuba, Japan, 2012: 971 - 974.
- [30] Kong S, Wang X, Wang D, et al. Multiple feature fusion for face recognition[C]// 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, Shanghai, China, 2013: 1 - 7.

Whiteflies counting with K-means clustering and ellipse fitting

Wang Zhibin^{1,2}, Wang Kaiyl^{1,2}, Zhang Shuifa^{1,2}, Liu Zhongqiang^{1,2}, Mu Cuixia³

(1. Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China; 2. China National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; 3. Department of Computer Science, China Women's University, Beijing 100101, China)

Abstract: Insect pests are one of the important factors leading to crop loss. Accurate insect counts provide an important basis for pest detection, and for proper preventive measures to be taken. At present, the common counting methods are mainly based on computer vision, but this type of technology primarily has the following problems: 1) how to determine the threshold of image segmentation. The effects of the algorithms are unsatisfactory, as their thresholds or parameters are fixed when they are used to segment insect images. 2) Most counting algorithms are mainly aimed at one certain crop for learning and testing. If applied to other crops, their portability is poor, and the counting results are inaccurate. Therefore, how to improve the generalization and accuracy of counting algorithm is an important direction for research on a counting method based on machine vision.

To solve the above problems, a novel counting algorithm for whiteflies based on k-means clustering and ellipse fitting method was proposed in this paper. It combined k-means clustering algorithm with ellipse fitting and automatically learned the features of whiteflies and background to segment and count whitefly images accurately. First, whitefly image were segmented by a k-means clustering algorithm to separate the whiteflies from the background, and then the segmentation results were fitted using an ellipse fitting based on least square method and adding up the ellipse number. The color features of the ellipse centers were extracted as new centers of classes. The segmentation and counting was iterated until the difference between two continuous counts met the needs of the algorithm and the convergence ellipse count was output as the number of whiteflies. Moreover, to improve the adaptability of the algorithm to count whiteflies on various crops, the whitefly images to be counted were parted into blocks and the center block was used to learn the features of whiteflies such as color, size, and area. The learned result was set as the initial value of the algorithm. Thus, the accuracy and generalization of the algorithm was improved.

To verify the effectiveness of the proposed algorithm, the counting experiment was performed on whitefly images of cayenne peppers, cucumbers, tomatoes, and eggplants respectively. These images were captured in the open environment from Xiao Tang Shan field research and a demonstration base of national precision agriculture in Beijing. The experimental results compared to that of the threshold method and the K-means clustering method showed that: 1) The count results of the proposed method had a high accuracy in cayenne peppers, cucumbers, tomatoes, and eggplants. The error rates of the pepper were 1.54%, 2.86%, and 4.00%; eggplant, 1.56%; tomato, 5.00%; cucumber, 11.30% and 5.71%. 2) The proposed method had better image segmentation results and higher count accuracy, compared to the threshold method and the K-means clustering method. Moreover, the counting error rate was decreased by 12.46% and 70.18% respectively. 3) The adaptive method learns the features of whiteflies such as color, sharpness, and size in the image to be counted, which is propitious for the accurate segmentation and counting of whitefly images. 4) The method makes the most of two important visual features of whiteflies, color and shape, and combines them by image segmentation and ellipse fitting to further increase the accuracy of the count results.

Key words: computer vision, pest control, algorithms, automatic counting, K-means clustering, ellipse fitting, whitefly

(责任编辑：张俊芳)