Automatic Method for Counting and Measuring Whiteflies in Soybean Leaves Using Digital Image Processing

Jayme Garcia Arnal Barbedo

[1]Embrapa Agricultural Informatics, Campinas, SP, Brazil

jayme.barbedo@embrapa.br

Abstract. *This paper presents an automatic method based on digital image processing for analyzing the leaves of soybean plants hosting whiteflies. The method is capable not only of counting and measuring whitefly nymphs and adults, but it is also capable of counting and measuring empty whitefly exoskeletons, as well as lesions that may be present in the leaf. The approach used in the algorithm is very simple, employing color model transformations to isolate the elements of interest in the image, and mathematical morphology to fine tune the results. This approach provides very accurate estimates under the tested conditions, and preliminary tests have shown that the algorithm is flexible enough to be used in other situations with only a few minor adjustments.*

Abstract. *Este artigo apresenta um método automático baseado em processamento digital de sinais para analisar folhas de soja atacadas por moscas brancas. O método é capaz não somente de contar e medir ninfas e adultos da mosca branca, como também é capaz de contar e medir exoesqueletos vazios, bem como lesões que podem estar presentes nas folhas. A abordagem usada no algoritmo é muito simples, empregando transformações de modelo de cor para isolar os elementos de interesse na imagem, e morfologia matemática para ajuste fino dos resultados. Esta abordagem fornece estimativas acuradas sob as condições testadas, e testes preliminares mostraram que o algoritmo é suficientemente flexível para ser usado em outras situações, com apenas alguns ajustes.*

1. Introduction

Whiteflies are small sap-sucking insects that can be found in a wide variety of host plants [Flint 2002]. There are more than 1550 species of whiteflies identified so far [Martin and Mound 2007]. Whiteflies may cause losses in crops both by direct feeding on the plants' tissue, and also by transmitting a variety of diseases [Flint 2002].

As can be seen, the whitefly is one of the most important agricultural pests and, as such, efficient ways of control must be devised in order to cut the losses caused by the insect. The effectiveness of this control depends, among other factors, on the early detection of the infestation and on studies about the effectiveness of control strategies.

Early detection can be performed visually by humans, however it may be nearly impossible to continuously monitor large crops without having some kind of automated system. In this context, image processing-based tools capable of detecting the infestation using only the images provided by cameras strategically placed over the crop can be invaluable. In this context, some systems for automatic identification and counting of insects have been proposed in the literature. Those proposals are capable of identifying other kinds of insects besides the whitefly (e.g. aphids and thrips). Most of them perform the identification and counting using images from sticky traps [Cho et al. 2007], [Martin et al. 2011], [Bechar and Moisan 2010], while a few use images from the leaves themselves [Mundada and Gohokar 2013], [Huddar et al. 2012]. This last approach has the advantage of being able to detect insects in larval and nymph stages.

In the studies about the effectiveness of different control strategies, normally several samples of plants receiving different kinds of treatment have to be collected. Each of those samples contains a certain number of specimens that must be classified according to their life cycle stage, and then counted accordingly. Such a classification is often based on the size of the specimens. Therefore, an image processing-based tool capable of detecting and measuring each one of those specimens could be very useful, especially in terms of speed, as this kind of tool can process batches of images very quickly. To the author's knowledge, there is no method of this kind available in the literature.

The method proposed here can tackle both challenges described above, with focus on soybean plants. The algorithm is based on two main strategies: 1) the original images are modified according to some color transformations in order to emphasize the features of the objects of interest; 2) some mathematical morphology operations are applied in order to finely delimit the objects, following the same guidelines used in [Barbedo 2012b] and [Barbedo 2012a]. After that, the objects can easily be counted and measured.

As will be shown in Section 3, the method is effective in identifying the different stages of the whitefly life cycle, being accurate both in terms of counting and surface measurement. The method is also capable of identifying, counting and measuring lesions that may be present in the leaf. Finally, the algorithm can identify and count empty exoskeletons shed by whitefly nymphs when they need to grow.

2. Material and Methods

2.1. Image Database

The database used in the development of the algorithm has 475 images, each containing a number of whiteflies at different stages of their life cycles. The images were captured in laboratory using a camera coupled with a tube of 10 cm in diameter, whose function is to limit the area in which the specimens are to be counted and also to serve as a measure reference to convert from pixels to centimeters. The images were stored in the RGB (Red-Green-Blue) format and quantized with 8 bits, and have dimension of 3648 x 2736 pixels. Fig. 1a shows an example of a typical image.

2.2. Isolation of the Region of Interest (ROI)

As can be seen in Fig. 1, the region of interest is the area delimited by the plastic tube. In order to isolate the ROI, first the original RGB image is transformed to the CMYK (cyan, magenta, yellow, and key) color model, from which only the cyan (C) and yellow (Y) channels are used, as shown in Fig. 2a and b. Those channels are binarized using, respectively, the values 150 and 100 as thresholds, that is, all pixels above those values are made white, and all others are made black. The resulting binary masks are shown in Fig. 2c and d. In the following, the binary operation AND is performed between those

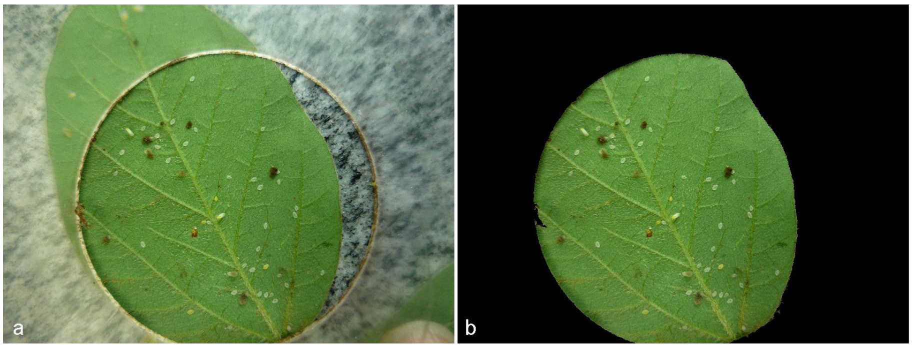

Figure 1. a) A typical image of a leaf containing whiteflies. b) Masked image.

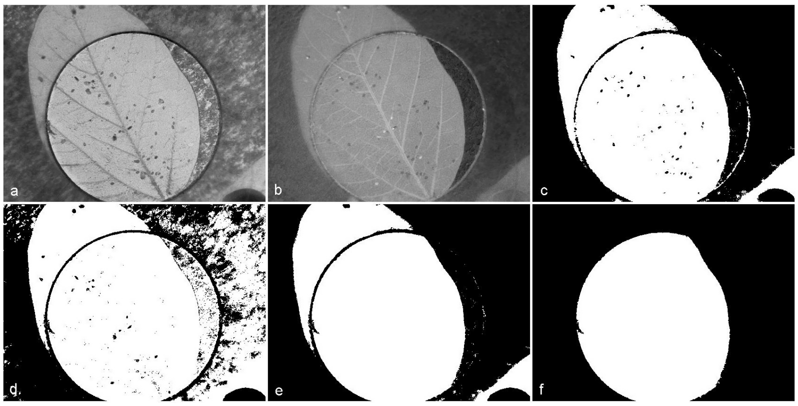

Figure 2. a) Representation of the image in the C channel of the CMYK color model. b) Representation of the image in the Y channel of the CMYK color model. c) Binary mask of the C channel. d) Binary mask of the Y channel. e) Combination (AND) of the two masks. f) Final mask defining the ROI.

binary masks, and all holes are filled (Fig. 2e). Finally, only the largest continuous object is kept, resulting in Fig. 2f. This mask is then applied to the original image (Fig. 1b).

2.3. Definition of the Color Schemes

As commented before, color manipulations are one of the two major strategies used in the paper. When the representation of an image is transformed from one color model to another, the channels of the original model are combined using certain equations, generating the new channels. In this work, three transformations were considered: RGB to CMYK, RGB to Lab (lightness and two color-opponent components) and RGB to XYZ (composed by three channels based on relative luminance). Additionally, a study was carried out about the effects of chaining up to three different color transformations. In this chaining, each intermediate result is treated as RGB. For instance, if the chain of transformations is RGB-CMYK-Lab, the equation applied to transform from CMYK to Lab will be the RGB-Lab one, and not the CMYK-Lab one. If this last one was used, the whole chain would be equivalent to a simple RGB-Lab transformation. Since CMYK has four channel and RGB has three, the K channel is discarded in the process. Fig. 3 illustrates the procedure expanding the example above.

After extensive tests, the following transformation chains yielded the best results:

- Whiteflies (all stages): RGB - Lab - XYZ (third channel - Z);

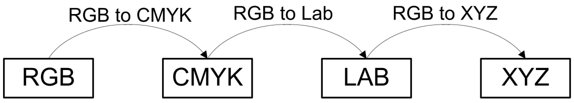

Figure 3. The color transformation process. The equations used to jump from one color model to another always consider the previous model as being RGB, no matter the actual presentation.

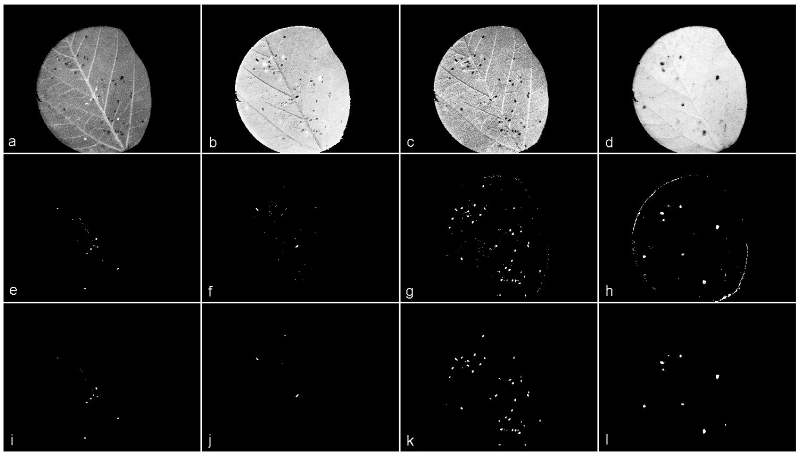

Figure 4. a-d) Resulting images after the selected color transformation chains. e-h) Binarized images. i-l) Binarized images after elimination of spurious objects.

- Whiteflies (adults): RGB - XYZ - XYZ - CMYK (first channel - C);

- Empty exoskeletons: RGB - CMYK - XYZ (third channel - Z);

- Lesions and mold: RGB - Lab - CMYK (second channel - M).

2.4. Delimitation of the Objects

After the color transformation chains are applied (Fig. 4a-d), the resulting images are binarized according to the following rules:

- Whiteflies (all stages): all pixels with value above 242 are made white (Fig. 4e);

- Whiteflies (adults): all pixels with value below 13 are made white (Fig. 4f);

- Empty exoskeletons: all pixels with value below 64 are made white (Fig. 4g);

- Lesions and mold: all pixels with value below 128 are made white (Fig. 4h).

The objects of interest are highlighted in the binary images, together with a large number of spurious objects. In order to eliminate those undesirable elements, all objects that are smaller than 10% of the largest object are removed (Fig. 4i-l). Objects that touch the borders of the ROI are also eliminated, because the border region often contain distortions that may lead to errors. After that, all connected objects are identified, counted and measured.

3. Experimental Setup and Results

The method was developed having only one image as reference. This image was not used in the tests to be described next. Also, approximately 20% of the images present in

Table 1. Overall results. The second, third and fourth columns are related to the counting accuracy, while the last column is related to the area measurement.

Object	False Hits	Misses	Deviation	Area Accuracy
Whiteflies (all)	6.2%	1.2%	5.3%	87%
Whiteflies (adults)	0.2%	0.2%	0.2%	95%
Exoskeletons	1.7%	5.3%	3.9%	83%
Lesions and Mold	0.8%	3.3%	2.6%	84%

the database described in Section 2.1 were removed from the tests due to poor lighting conditions.

All elements present in the images (whiteflies, exoskeletons and lesions) were counted and measured manually prior to the image capture. These manual estimates were used as references for the results yielded by the program. It is important to highlight that the manual measurements are subject to errors due to the number and variety of elements. Therefore, the manual annotations cannot be considered the ground-truth, but just a reference to the estimated lengths.

Table 1 shows the overall results for each type of object. The column "false hits" indicates the proportion of objects mistakenly detected; the column "misses" indicates the proportion of objects that the method failed to detect; the column "deviation" reveals the difference between manual and automatic counts; and the column "area accuracy" reveals the difference between manual and automatic measurements.

The following conclusions can be drawn from Table 1:

- Whiteflies (all): the number of misses is very low, with the number of false hits being considerably higher (although still low). Almost all false hits are due to the venations of the leaves, which sometimes can have regions with color tones very close to that of whiteflies. The area accuracy for the whitefly nymphs is very high (around 97%), however it is very low for adults (24%). This happens because most of the body of adult whiteflies is occluded by the wings, which have tones very close to those of empty exoskeletons, so they are not detected. This is not a major problem, as the program has a part dedicated exclusively to the detection and measurement of adult specimens.

- Whiteflies (adults): the algorithm is very accurate in all aspects.

- Exoskeletons: the method presents good accuracy in counting empty exoskeletons. The few false hits are due to debris in the leaf, and misses are mostly due to the high degree of transparency presented by some elements. The area measurement is not as accurate, again due to the transparency of certain parts of the objects, which prevents those regions to be detected and taken into account.

- Lesions and mold: the counting accuracy is very high, with a few misses due to subtlety of some lesions. The area accuracy drops due to the fuzzy boundaries presented by virtually all mold spots and also by some lesions.

A problem that was observed in some images is that when there are whiteflies at very different stages of their life cycle, that is, if there is a big difference in size, some of the small specimens may be discarded in the process of debris removal. This is still an open issue that should be addressed in the future.

Because some of the objects to be detected are small, the resolution has to be enough so those can be resolved. The ideal resolution depends on how the images were captured, but a good reference number would be 5 MPixels.

Finally, no comparison with other methods is presented because, as commented before, there are no equivalent proposals available in the literature.

4. Conclusion

This paper presented an automatic method for counting and measuring whiteflies, exoskeletons and lesions in soybean leaves. The method is based on two main procedures, color model transformations and mathematical morphology. The resulting algorithm can accurately estimate the number of objects. The area measurement is reasonably accurate, but it has some flaws that should be addressed in the future.

Besides the area measurement procedure, there are other aspects that deserve further investigation. First, the method was tested under ideal conditions of light and color, so its robustness under more challenging circumstances is still to be determined. Future research will also deal with the problem of missing small objects when much larger objects are present in the same image. Finally, since the whitefly affects many other cultures other than soybean, the extension of the method to other plant species is another objective worth pursuing.

References

Barbedo, J. G. A. (2012a). Method for automatic counting root nodules using digital images. In *Proc. Intl. Conf. Comp. Science and Its Applications*, pages 159–161.

Barbedo, J. G. A. (2012b). Method for counting microorganisms and colonies in microscopic images. In *Proc. Intl. Conf. Comp. Science and Its Applications*, pages 83–87.

Bechar, I. and Moisan, S. (2010). On-line counting of pests in a greenhouse using computer vision. In *Visual Observation and Analysis of Animal and Insect Behavior*.

Cho, J., Choi, J., Qiao, M., Ji, C.-W., Kim, H.-Y., Uhm, K.-B., and Chon, T.-S. (2007). Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis. *Journal of Mathematics and Computers in Simulation*, 1(1):46–53.

Flint, M. (2002). Whiteflies: Integrated pest management for home gardeners and professional landscapers. Technical Report Publication 7401, University of California.

Huddar, S., Gowri, S., Keerthana, K., Vasanthi, S., and Rupanagudi, S. (2012). Novel algorithm for segmentation and automatic identification of pests on plants using image processing. In *Intl. Conf. Comp. Comm. & Networking Technologies*, pages 1–5.

Martin, J. and Mound, L. (2007). An annotated check list of the world's whiteflies (Insecta: Hemiptera: Aleyrodidae). *Zootaxa*, 1492:1–84.

Martin, V., Moisan, S., Paris, B., and Nicolas, O. (2011). Towards a video camera network for early pest detection in greenhouses. In *Int. Conf. on Diversifying Crop Protection*, pages 12–15.

Mundada, R. and Gohokar, V. (2013). Detection and classification of pests in greenhouse using image processing. *IOSR Journal Electronics and Comm. Eng.*, 5(6):57–63.