Automatically Measuring Early and Late Leaf Spot Lesions in Peanut Plants Using Digital Image Processing

Jayme Garcia Arnal Barbedo

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

`jayme.barbedo@embrapa.br`

Abstract. This paper presents a method to measure the lesions originated by the Cercospora arachidicola *and* Cercosporidium personatum *fungi, which cause, respectively, the early and late leaf spots in peanut plants. The proposed method is based on a modified version of a previous proposal by the author, and uses mainly well-known image processing techniques, as well as specialist knowledge, to separate lesions from healthy tissue. The resulting tool provides good area estimates with minimum user interference and low computational burden.*

Abstract. Este artigo apresenta um método para medir lesões causadas pelos fungos Cercospora arachidicola *e* Cercosporidium personatum*, os quais causam, respectivamente, a cercosporiose e a mancha preta em plantas de amendoim. O método proposto é baseado numa versão modificada de uma proposta anterior feita pelo autor, e usa principalmente técnicas bem conhecidas de processamento de imagens, bem como conhecimento especialista, para separar lesões de tecido saudável. A ferramenta resultante fornece boas estimativas de área com mínima interferência por parte do usuário e baixa complexidade computacional.*

1. Introduction

One of the main goals of phytopathology is to investigate the behavior of the diseases so better forms of treatment and control can be devised and applied. In this context, measuring the ratio between the areas occupied by diseased and healthy tissues is paramount. Until recently, those measurements were done entirely by hand, using techniques that lack both accuracy and practicality [Barbedo 2013].

As technology evolved, new tools that reduce the intensity of manual labor have become available, like SigmaPro [SSPS 1998], Assess [Lamari 2002] and ImageJ [Abramoff et al. 2002]. Most of those tools still demand a certain degree of human intervention, particularly regarding the task of adjusting image processing parameters. Also, many of them are paid and proprietary. Due to those factors, manual area measurements are still commonplace.

In the last few years, a number of automatic methods for measuring lesions have been proposed [Weizheng et al. 2008], [Camargo and Smith 2009], [Zhang 2010], [Pang et al. 2011], [Patil and Bodhe 2011], [Sekulska-Nalewajko and Goclawski 2011], [Contreras-Medina et al. 2012]. Those methods generally work well under the conditions and situations to which they were trained, but lack generalization capabilities.

A more generic method was recently proposed by the author [Barbedo 2013]. Almost inevitably, this kind of approach has to deal with a tradeoff between generality and

accuracy. A generic method will most likely present lower accuracy for a given situation (for example, a specific plant species and disease) than a tool designed having specifically that situation in mind. In other words, a generic method is designed to provide satisfactory results over a large range of plants, diseases and conditions, while specific methods try to provide the most accurate results for a specific type of problem.

The method proposed in [Barbedo 2013], being generic, has good foundations but was not finely tuned to any situation. In this paper, such a method is modified to carefully meet the characteristics present in peanut leaves infected by early and late leaf spots, caused, respectively, by the Cercospora arachidicola and Cercosporidium personatum fungi. As will be seen along the paper, the method only uses well-known image processing techniques (specially mathematical morphology) and heuristics based on specialist knowledge and observation.

The paper is organized as follows. Section 2 presents the image database used in the development of the method and describes the proposed algorithm. Section 3 presents the experiment setup and the results achieved using the proposed method. Finally, Section 4 presents the conclusions and final remarks.

2. Material and Methods

2.1. Image Database

The database used in the development of the algorithm includes 124 images containing early leaf spots and 114 images containing late leaf spots. The images were captured in a laboratory after the samples were removed from the respective plants, using a flatbed scanner. They were stored in the RGB (Red-Green-Blue) format and quantized with 8 bits, and have dimensions of 834 x 935 pixels in average. Each image has multiple leaves, and the set comprises the whole range of disease severity levels, from totally healthy to heavily infested leaves. The background of the images is nearly white, and the leaves are presented still attached to the respective stems. Fig. 1a shows an example of a typical image.

2.2. The Method

The first step of the algorithm is to segment the image into leaves and background. Since the background is very bright, it has good contrast with respect to the leaves. As a result, the segmentation procedure is trivial – a mask is generated by considering only pixels whose absolute value is lesser than or equal to 200 in the grayscale version of the image (since a 8-bit quantization is being used, the maximum pixel value, representing white, is 255). This excludes almost all background pixels and, in some cases, a number of leaf pixels as well. To consider pixels that have been mistakenly removed, any holes inside the mask are filled. To remove background pixels that may remain, all objects that are smaller than 1% of the largest object are removed. The resulting mask image, which is used as reference to isolate the leaves and stems from the background, is shown in Fig. 1b.

The next step of the algorithm is dedicated to the removal of petioles and stems from the mask image, so only the leaves remain. First, the mask image is morphologically opened using as structuring element a disk whose radius is given by $\sqrt{l^2 + L^2}$, where l and L are, respectively, the shortest and longest sides of the image. The resulting image (see Fig. 1c) is subtracted from the original mask image, giving rise to Fig. 1d. As can be
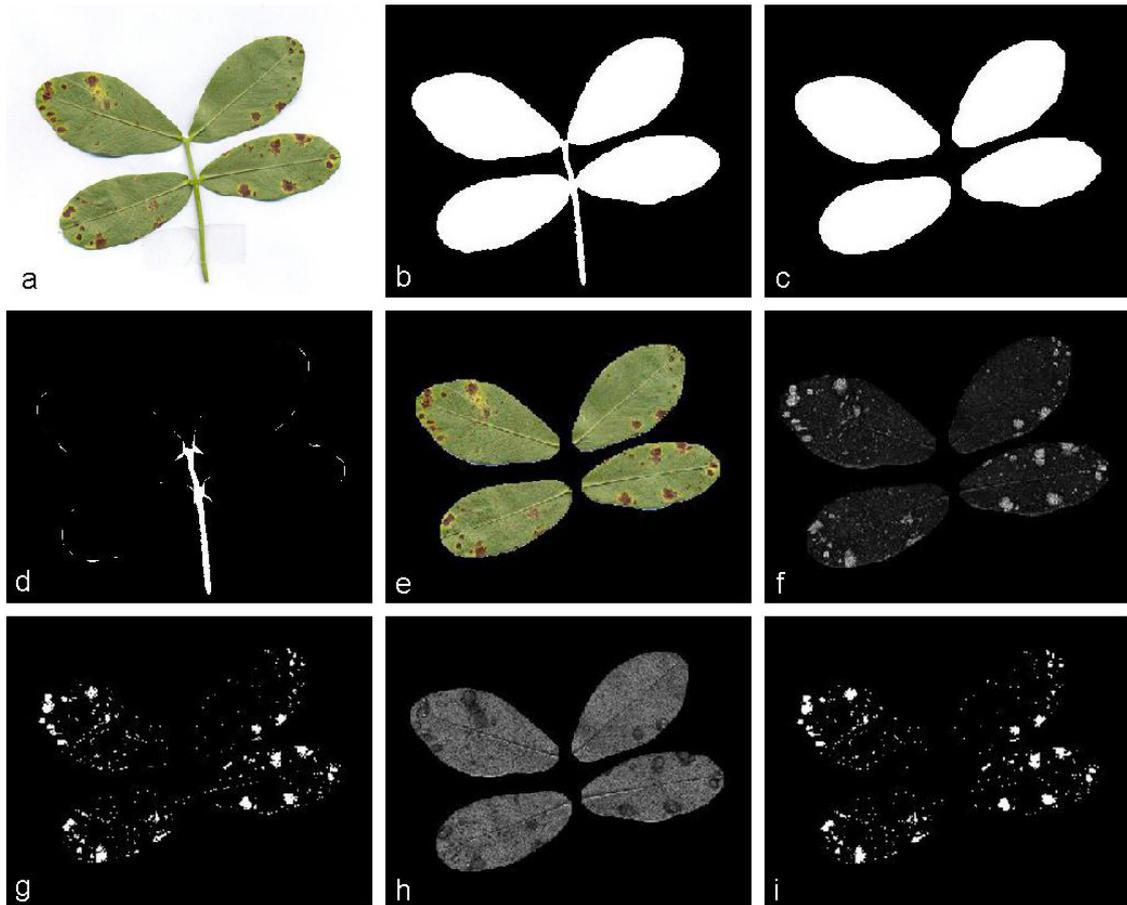
Figure 1. a) Typical image present in the database. b) Example of mask image prior to fine tuning. c) Example of morphologically opened image. d) Image containing the locations of stems and petioles prior to the purge of spurious objects. e) Image containing only the areas of interest. f) Representation of the image in the third channel (yellow) of the CMYK color model. g) First estimate for the lesions locations. h) Representation of the image in the second channel (magenta) of the CMYK color model. i) First estimate for the lesions locations.

seen, this image includes the exact location of stems and petioles, and also some spurious objects. Those spurious objects can be discarded simply by eliminating all objects that are smaller than half the size of the largest object. Finally, the stem and petiole pixels are removed from the original mask image, and the mask is applied to the original image (Fig. 1e).

The RGB color model is good for storage and visualization purposes, but it is usually not suitable for more sophisticated manipulations. After an extensive study considering the most widely used color models, the third channel (yellow) of the CMYK (cyan, magenta, yellow, and key) model showed the best characteristics in the context of the present problem. In particular, this channel greatly improves the contrast between lesions and healthy tissue, as can be seen in Fig. 1f.

After the color transformation, the pixel values now are in the range from 0 (black) to 1 (white). The resulting image is then binarized using a threshold of 0.25, that is, all pixels with values above this threshold are made white, and all others are made black. The resulting image (Fig. 1g) is a first estimate for the lesions locations.

At this point, only some minor adjustments have to be made. As can be seen in previous figures, the veins and the leaves themselves can cast dark shadows that, under the chosen color channel, can be very similar in shade to the lesions themselves, which can cause misidentifications. However, it was observed that, in the magenta channel of the CMYK model, these shadows appear very bright, while the lesions are dark (Fig. 1h). Therefore, all pixels detected as lesions in the yellow channel are discarded if they have a value above 0.5 in the magenta channel. Finally, all holes inside the detected objects (lesions) are filled, resulting in the final estimate shown in Fig. 1i.

3. Experimental Setup and Results

The method was trained and tuned using a single image of each type of spot (early and late). Both images were not used in the tests to be described next.

Prior to submitting the images to the program that implements the method, they were manually annotated into diseased and healthy regions. It is important to highlight that the manual segmentation between lesions and healthy tissue has a degree of subjectivity, as the limits between the regions are not abrupt, but gradual. Therefore, the manual annotations cannot be considered the ground-truth, but just a reference to the automatic segmentation. It is important to take this into account when analyzing the results.

Table 1 presents the overall accuracy achieved for the early and late leaf spots, with respect to the manually annotated images. The column "false hits" is given by f/a, and indicates the proportion of pixels mistakenly detected; the column "misses" is given by m/a, and indicates the proportion of lesion pixels that the method failed to detect; and the column "deviation" is given by $(A-a)/a$, and indicates how the estimated area compares with the manually annotated one. In the ratios above, f is the number of pixels wrongly detected as lesions, m is the number of undetected lesion pixels, a is the total number of manually annotated lesion pixels, and A is the total number of lesion pixels detected by the method.

As can be seen in Table 1, the estimates are less than 1% off the target values. Considering that the gradual transition from lesions to healthy tissue does not allow to

Table 1. Results using the method.

Lesion	False Hits	Misses	Deviation
Early Spot	1.1%	0.2%	0.9%
Late Spot	1.1%	0.1%	0.9%

Table 2. Comparison with other methods.

Method	False Hits	Misses	Deviation
Proposed	1.1%	0.2%	0.9%
Generic [Barbedo 2013]	3.1%	1.6%	1.7%
Assess [Lamari 2002]	0.7%	0.3%	0.5%
ImageJ [Abramoff et al. 2002]	0.8%	0.3%	0.6%

infer a real ground-truth, it can be argued that the performance of the proposed automatic method in terms of accuracy is as good as the manual estimate, while being much more practical and faster.

Table 2 presents a comparison with other methods proposed in the literature, considering both early and late spot together.

As can be seen in Table II, Assess [Lamari 2002] and ImageJ [Abramoff et al. 2002] have slightly better performance than the proposed method. However, they are semi-automatic, which means they depend on a close interaction with the user to work properly. Moreover, Assess is paid and proprietary. On the other hand, the method proposed in [Barbedo 2013] performs worse due to its generality capabilities, as discussed before.

4. Conclusion

This paper presented a new method for measuring the areas occupied by early and late leaf spots, which manifest in the leaves of peanut plants. The method is mostly based on well-known mathematical morphology operations and on heuristics resulting from specialist knowledge. Tests have shown that the proposed method performs as well as semi-automatic and manual methods, being, at the same time, much more practical and faster. Given the fact that the characteristics of the leaves and lesions tackled in this work are quite similar to other combinations of plants and diseases, it is reasonable to assume that the proposed method is suitable to be used in other kinds of situations. Future work will be dedicated to investigate this possibility, and also to find the best way to extend the method to situations currently outside its scope.

References

Abramoff, M., Magalhaes, P., and Ram, S. (2002). Multispectral inspection of citrus in real-time using machine vision and digital signal processors. *Computers and Electronics in Agriculture*, 33(2):121–137.

Barbedo, J. G. A. (2013). An automatic method to detect and measure leaf disease using digital image processing. *Computers and Electronics in Agriculture*, submitted.

Camargo, A. and Smith, J. S. (2009). Image pattern classification for the identification of disease causing agents in plants. *Computers and Electronics in Agriculture*, 66(2):121–125.

Contreras-Medina, L. M., Osornio-Rios, R. A., Torres-Pacheco, I., Romero-Troncoso, R. J., Guevara-González, R. G., and Millan-Almaraz, J. R. (2012). Smart sensor for real-time quantification of common symptoms present in unhealthy plants. *Sensors (Basel, Switzerland)*, 12(1):784–805.

Lamari, L. (2002). *Assess: Image Analysis Software for Plant Disease Quantification*. APS Press, St. Paul.

Pang, J., Bai, Z.-y., Lai, J.-c., and Li, S.-k. (2011). Automatic segmentation of crop leaf spot disease images by integrating local threshold and seeded region growing. In *2011 International Conference on Image Analysis and Signal Processing*, pages 590–594.

Patil, S. B. and Bodhe, S. K. (2011). Leaf disease severity measurement using image processing. *International Journal of Engineering and Technology*, 3(5):297–301.

Sekulska-Nalewajko, J. and Goclawski, J. (2011). A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering. In *VII International Conference on Perspective Technologies and Methods in MEMS Design*, pages 172–175.

SSPS (1998). *Sigma Scan Pro 5.0*. SPSS Science Marketing Dep., Chicago.

Weizheng, S., Yachun, W., Zhanliang, C., and Hongda, W. (2008). Grading Method of Leaf Spot Disease Based on Image Processing. In *2008 International Conference on Computer Science and Software Engineering*, pages 491–494, Wuhan.

Zhang, M. (2010). Citrus canker detection based on leaf images analysis. In *The 2nd International Conference on Information Science and Engineering*, pages 3584–3587.