Automatic Detection of the Damaged Leaf Area in Digital Images of Soybean

A. C. Nazaré-JR., D. Menotti and J. M. R. Neves
Department of Computing
Federal University of Ouro Preto
Ouro Preto, Brazil

T. Sediyma
Department of Plant Science
Federal University of Viçosa
Viçosa, Brazil

Abstract— In agriculture, there are many difficulties involved when handling pests in the tillage. Among the types of pest attacks are the ones done directly on plants foliage. In this case, accurate detection of damaged leaf area is essential to the determination of pest control measures. The damaged leaf area detection is often done by a human, in a time-consuming and inaccurate way. In order to improve this process, here we propose a method for automatically detecting/quantifying the damaged leaf area using digital images. Experiments show that the proposed method improves in 40% the absolute error generated by other method proposed in the literature, and it is much more precise than the visual evaluation performed by a human specialist.

Damaged leaf area; soybean; shadow removing, extremity reconstruction.

I. INTRODUCTION

In agriculture, the determination of leaf area is of great importance for the application of techniques such as pruning, fertilization and planting density [1]. A feature that can be extracted by determining the leaf area is the quantification of damage caused by pests and diseases [2]. Such damage can be detected through the study of damaged leaf area by pests [2]. Detecting the precise amount of damaged leaf area is essential to determine control actions such as application of pesticides, since a small damaged leaf area may dispense control measures.

There are several methods for measuring leaf area, however, in practice, it is used mainly three: the human evaluation, the method of leaf dimensions and the methods which use devices such as planimeter and area integrator. Nevertheless, these methods require extensive work and are time-consuming. Moreover they have some degree of inaccuracy. And, the measurement techniques are not performed in the most cases by a farmer, but by an expert (agronomist), which delays the diagnosis [2].

With the advances in computing, especially in the graphics processing, it is possible to develop alternative methods for determining the damaged leaf area. Wilcken *et. al.* [2] demonstrated the feasibility of using digital image acquired by scanners as an alternative method for detecting the leaf area consumed by insects. Einhardt & Ferreira [3] concluded that computational method can be used to determine the leaf area, since it showed similar results to traditional methods. Mura *et. al.* [4] showed that the use of a computational method, also using images acquired by a scanner, is better to determine the percentage of soybean leaves affected by caterpillar (*Anticarsia gemmatalis*), with respect to the evaluation of an agronomist.

However, the method proposed in [4] has some disadvantages. To properly work, the method requires that the image of the leaf sample is free of impurities, such as specks and small pieces that disintegrate from the leaf during its handling on the scanner. In other words, the images should be composed of only the leaf sample and a white background, which is impractical in an agricultural environment. It is not performed any image preprocessing to remove the shadow that is incorporated into the leaf during the process of optical scanning for image acquisition, which leads to an inaccurate detection of the leaf area. Finally, the system does not enable automatic reconstruction of damaged extremity of the leaf. Then the damaged area of leaves with heavy damage at the border is detected in an unsatisfactory manner.

This paper proposes a new method of automatic detection and automatic quantification of the damaged leaf area of the soybeans using techniques of digital image processing and computational geometry that deals with the problems presented in the literature as shadow removal, elimination of undesirable objects from image acquisition and reconstruction of compromised extremities of the leaf[1]. The proposed method is also compared with the human expert evaluations and with the method of Mura *et. al*. [4].

II. METHODOLOGY

The proposed method consists of several steps of techniques of image processing and computational geometry applied to the color image of the leaf sample. These steps are described as follow.

A. *Image Acquisition*

Initially, the leaf is captured by an image scanner (flatbed) in *RGB* color space (*R* - red, *G* - green and *B* - blue) generating a color image of the sample, as illustrated in Figure 1(a).

B. *Shadow Removing*

This step aims at eliminating the shadow of the leaf sample. Firstly, the image is converted from *RGB* color space to *HSV* space (*H* - hue, *S* - saturation and *V* - value). According to [5], in the *HSV* color space the pixels in the shadow region have the following characteristics: high hue (*H*), high saturation (*S*) and low intensity (*V*). In this work, only the high tone (hue) was

[1] An extended version of this work, in portuguese, is published in [6].

sufficient for detecting the shadow in the leaf sample image. In *HSV* space, the hue represents the color itself (*e.g.*, green, red, blue, orange, etc.), and it is defined as an angle in the interval *[0º, 360º)*. By studying the mean histogram of the channel *H* of *150* leaf sample images, it was noticed that it presents a bimodal behavior. We are interested in the class of pixels with hue values less than *108º*. These values represent the leaf sample. The pixels with values greater than *108º* belong to the small area of shadow. In order to reduce the volume of information being processed, the pixels classified as shadow are subtracted from the initial image and the resulting is converted to a grayscale image (*i.e.*, *8* bits). Figure 1(b) illustrates the result of this preprocessing in a portion of Figure 1(a).

C. Filtering

In this step, it is applied the median filter [7] to decrease the noise from the digitization/acquisition process and also imperfections such as small leaf spots. This facilitates the process of segmentation that will be presented in Section II-D. The median filter replaces the value of a pixel by the median value of the sampling space, which consists of a 3×3 centered window on the considered pixel. The image in Figure 1(c) illustrates the result of this process using as input the resulting image of the shadow removing process.

D. Segmentation

In this context, segmenting an image consists of separating the object (leaf sample) from the background of the image. The segmentation was performed by the Otsu's automatic threshold algorithm [8]. This technique considers that the histogram of an image is composed of two classes, *i.e.*, the pixels of the object (leaf) and the pixels of the background. What simply distinguishes these two classes is a global threshold, which is automatically calculated by of Otsu'se algorithm, which maximizes the variance between the object and background classes. Thus, all pixels with values less than this threshold are classified as object and the other ones are classified as background, resulting in a binary image. The use of the median filter was important for this step, since it facilitates the identification of an optimal threshold for the image. Figure 1(d) illustrates the result of the segmentation of the filtered image in Figure 1(c).

E. Elimination of Undesirable Objects

After the segmentation is performed, the image may have other objects that are not members of the leaf sample, as specks and small pieces that disintegrate from the main sample. To deal with this problem, it was applied a labeling algorithm [7], which identifies all the connected objects in the image. It is considered objects the regions that have pixel values *0* (black). After this identification, the objects, except the one with largest size, which is a leaf sample to be analyzed, are excluded. An example of the presence of undesirable objects after segmentation is shown in Figure 1(d). These undesirable objects, if present in the image, can affect the accurate quantification of the damaged leaf area. The resulting image of this step (Figure 1(e)), from now on I_{SEG}, will be used as a baseline for the next steps and also for the computation of the damaged leaf area.

F. Recovering the Leaf Internal Damage

In this step, it is performed an intermediary detection of the damaged sample. That is, the internal damages of the leaf are reconstructed such that the leaf sample forms a single solid object, as shown in Figure 1(f). To accomplish this task, it is also used the labeling algorithm, cited in Section II-E. However, in this step, it detects regions with pixel values equal to *1* (white). The area outside the leaf is disregarded.

G. Reconstruction of Extremities

For detecting the damaged area present at the extremity of the leaf (*i.e.*, concave regions), it is required to reconstruct the broken edges. This operation is performed by two steps which are described as follows:

1) Corners Detection: Firstly, the corner points of the leaf are detected. A corner is a point on the curve that has a high curvature. This operation is performed using the method proposed by [9]. Figure 1(g) shows the points of corners, in red, detected in the image. These points are possible candidates of break points of the leaf contour. In other words, they are points where the natural curvature of the edge of the leaf sample is broken and they will be used for rebuilding the damaged extremities.

2) Reconstruction of Edge: Let C be the set of the detected corner points, *i.e.*, $C = \{P_1;\ P_2;\ P_3;\ \dots\ ;\ P_n\}$. Let R be the set of line segments generated from a combination of all elements of C, taken two at a time, *i.e.*, $R = \{\overline{P_iP_{i+1}};\ \overline{P_iP_{i+2}};\ \dots\ ;\ \overline{P_{n-1}P_n}\}$, where $i = 1, 2, 3, \dots, n-1$ and $\overline{P_iP_j}$ stands for the line segment between the corner points P_i and P_j. Once R is computed, this set of line segments is drawn into the image. The line segments were drawn by using the Bresenham algorithm [10]. One example of the outcome of this process is shown in Figure 1(h). Then, it is applied again the process presented in Section II-F in order to recover the new damaged inner regions of the sample. This is the final reconstructed image (Figure 1(i)), from now on I_{FINAL}. This image, along with I_{SEG}, will be used for quantificating the damaged area. Note that, in this step, we were not concerning to generate only the line segments outside the leaf sample, since the number of segments generated is easily computable.

H. Quantification

The determination of the damaged leaf area is defined as. Let $I_{DAMAGED}$ be the subtraction of I_{SEG} from $I_{FINAL,}$ *i.e.*,

$$I_{DAMAGED} = I_{FINAL} - I_{SEGe} \quad (1)$$

The amount, in pixels, of the damaged leaf area is determined by $I_{DAMAGED}(0)$, which corresponds to the number of pixel values *0* (black) in the $I_{DAMAGED}$ image. Therefore, the percentage of damage can be computed as:

$$P = \frac{I_{DAMAGED}\ (0)}{I_{FINAL}\ (0)} \quad (2)$$

where $I_{FINAL}(0)$ is the number of black pixels in I_{FINAL}:

III. COMPUTATIONAL EXPERIMENTS

In order to validate the proposed method, it was constructed a database of *185* leaf samples from Soybean (*Glycine max (L.) Merrill*) taken randomly from different positions of the canopy of plants in experimental fields of the Department of Plant Science, Federal University of Viçosa, Brazil. For image acquisition, it was used a scanner *HP*, model Scanjet *G2410*, coupled to a computer. The images of the samples were stored in the *BMP* format (Bitmap) – no compression, with a resolution of *200* dpi (dots per inch) and *24* bit per pixel (color).

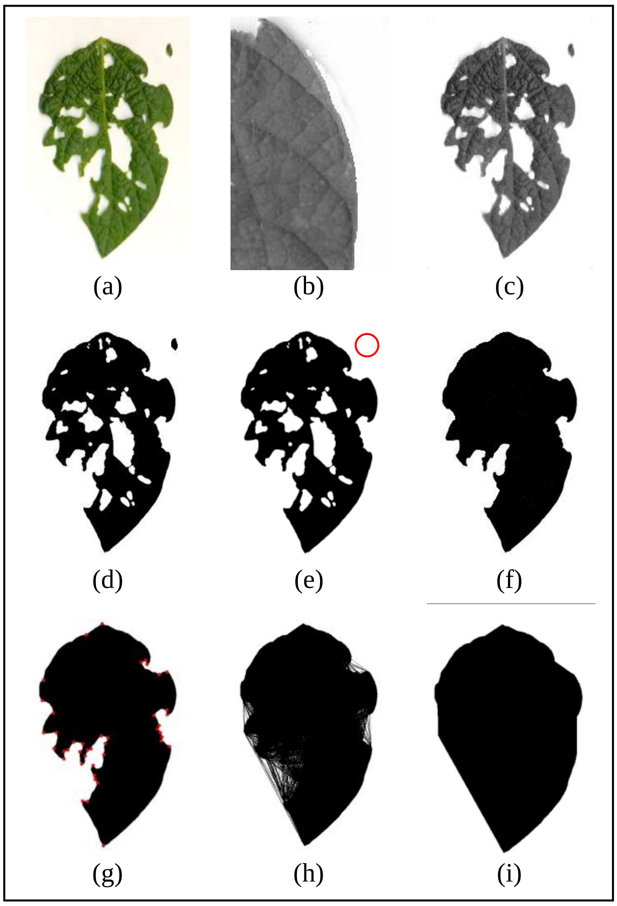

Figure 1. A leaf sample illustrating the steps of the proposed method.

The values of the damaged areas detected in leaves (*i.e.*, evaluation) by the proposed method (PROPOSED) were compared with those obtained by the method proposed in [4] (MURA-07). It is noteworthy that these two methods use digital images. Moreover, the values obtained by these methods were compared with human evaluations (HUMAN).

The PROPOSED and MURA-07 methods were implemented in Matlab. In the tests conducted in [4], images were acquired using as background a sheet of white paper in order to eliminate impurities, since they can affect the result obtained by the method. Then, for a fair comparison, in the MURA-07 method, it was included our step of elimination of undesired objects.

The results of the evaluation of methods using digital images are the percentages of damaged leaf area estimated for each sample. The third method consisted of an evaluation conducted by undergraduate students of the Federal University of Viçosa, already trained to perform a damaged leaf area analysis of a culture. Thus, it was developed a web system [11] where each leaf sample (image) was presented to four students to perform their separate evaluation by assigning percentage values of damaged leaf area. The final estimation of each sample was computed as the arithmetic mean of the four ratings. It is important to note that the choice of which leaf sample is given for each student was carefully done, such that none evaluation intersection is verified. That is, two different students never evaluated two same samples.

In order to generate references for the evaluation, the images of the samples were subjected to a process of manual segmentation and reconstruction (supervised by an expert in the area of Plant Science). In the process of manual segmentation, for each sample, it was generated two images of references (ground-truth): the damaged and rebuilt images. By using these images and Equation 2 to quantify the damaged area, it is possible to establish the ideal percentage of damage, *i.e.*, P_{REF}.

It was computed the absolute ($\mathcal{E}A_x$) and relative ($\mathcal{E}R_x$) errors of the method x for every sample, *i.e.*,

$$\mathcal{E}A_x = |P_{REF} - P_x| \,, \tag{3}$$

$$\mathcal{E}Rx = \begin{cases} \frac{|P_{REF} - P_x|}{P_{REF}}, & se\ P_{REF} \geq 1\% \\ |P_{REF} - P_x|, & se\ P_{REF} < 1\% \end{cases}, \tag{4}$$

where P_x is the result estimated by the method x and x stands for PROPOSED, MURA-07 and HUMAN.

The conditional test on P_{REF} in Equation 4 is justified by the fact that for samples with a damaged area less than *1*%, the $\mathcal{E}R_x$ would be biased, *i.e.*,

$$\lim_{P_{REF} \to 0^+} \frac{|P_{REF} - P_X|}{P_{REF}} = +\infty. \tag{5}$$

According to [1], control measures should be taken when it is detected a damaged leaf area by pests in adult soybean plants greater than or equal to *15*%. This analysis is of paramount importance to the farmer, because an incorrect indication may lead to unnecessary expenses and / or loss of production. For this reason, it was conducted an evaluation of methods in relation to indications of control measures. For each sample, it was took the decision according to the method indication (*i.e.*, the damaged leaf area percentage) and compared with the decision indicated by the true ideal percentage of. The numerical results, the leaf sample images, and the source code which implements the method presented here are available at [11].

IV. RESULTS AND DISCUSSIONS

Tables I and II present the values of mean absolute error (*MAE*) and mean relative error (*MRE*), respectively, of the methods evaluated. These values are arranged in the following format: $\mu \pm \sigma$, where μ is the average and σ is the standard deviation. The values were separated into groups, divided into bins of ideal percentage of damaged area, to allow a detailed analysis of the ratings.

The lower the value of *MAE*, the greater the accuracy on estimating the damaged leaf area. With respect to this criterion, it was observed that the PROPOSED method has better result because the *MAE* generated by it was a little less than half of *MAE* generated by the MURA-07 method and seven times smaller than the *MAE* generated by the HUMAN method.

TABLE I. MAE OF THE METHODS EVALUATED FOR EACH RANGE OF PERCENTAGE OF DAMAGED LEAF AREA.

Group	#	HUMAN	MURA-07	PROPOSED
0 |--- 5	60	2,99 ± 2,41	0,51 ± 0,48	0,37 ± 0,56
5 |--- 10	53	5.38 ± 3.41	1.15 ± 0.78	0.57 ± 0.48
10 |--- 15	43	7.01 ± 4.53	1.75 ± 1.80	0.93 ± 1.06
15 |--- 20	17	8.67 ± 5.10	3.53 ± 3.51	1.80 ± 1.95
20 |--- 99	12	13.76 ± 10.17	5.67 ± 3.79	2.38 ± 1.95
TOTAL	**185**	**5.83 ± 5.11**	**1.60 ± 2.19**	**0.82 ± 1.15**

TABLE II. MRE OF THE METHODS EVALUATED FOR EACH RANGE OF PERCENTAGE OF DAMAGED LEAF AREA.

Group	#	HUMAN	MURA-07	PROPOSED
0 |--- 5	60	89.99 ± 92.46	16.71 ± 19.72	12.10 ± 24.36
5 |--- 10	53	71.42 ± 41.97	15.39 ± 9.72	7.73 ± 6.18
10 |--- 15	43	57.94 ± 38.05	14.32 ± 15.05	7.55 ± 8.79
15 |--- 20	17	53.36 ± 32.47	21.55 ± 21.91	11.01 ± 12.20
20 |--- 99	12	51.86 ± 34.53	20.67 ± 10.85	8.55 ± 4.77
TOTAL	**185**	**71.38 ± 62.79**	**16.48 ± 16.03**	**9.46 ± 15.41**

In order to confirm the better accuracy of the PROPOSED method over the MURA-07 and HUMAN methods, it was performed a test of statistical significance. In other words, this test shows that the MAE of the PROPOSED method is smaller than the MAE of the other two methods compared. For such aim, it was applied the Student's *t*-test [12] with a significance level of 0.0005 having as the alternative hypothesis, H_1, *i.e*., the MAE of the PROPOSED method is smaller than the MAE generated by the other two methods, and as null hypothesis, H_0, otherwise. That is, the MAE of the PROPOSED method is greater than or equal to the MAE of the others. It was computed the values of Student's *t*-test between the PROPOSED method and the HUMAN ($t_1 = -11.987$) and MURA-07 ($t_2 = -6.799$) methods. In order to accept the null hypothesis, with significance level of 0.0005, t_1 and t_2 must be greater than –3.291. As this statement is false, the null hypothesis is rejected. Thus, we conclude that the PROPOSED method statistically estimates the damaged leaf area better than the other two methods.

Table III presents the maximum error generated by each method, which was observed in the same sample. The figures show that the maximum error of the PROPOSED method is quite lower than the two other methods, which means a high degree of reliability.

TABLE III. MAXIMUM ERROR RECHEAD BY EACH METHOD.

Method	Real Value	Estimated Value	Maximum Error
HUMAN	39.54 %	76.25 %	36.71 %
MURA-07	39.54 %	25.84 %	13.70 %
PROPOSED	39.54 %	31.64 %	7.90 %

It was also observed the capacity for the diagnosis of each method. Table IV shows the number of correct indications of control action performed by each method. The results showed that the PROPOSED method has a total accuracy rate of 95.68% ((21 + 156) / (29 + 156)), higher than the MURA-07 and HUMAN methods, which have rates of 92.97% ((16 + 156) / (29 + 156)) and 73.51% ((29 + 107) / (29 + 156)), respectively.

TABLE IV. CORRECT INDICATIONS FOR EACH METHOD IN RELATION TO THE NEED FOR CONTROL IN THE TILLAGE.

Indication	Reference	HUMAN	MURA-07	PROPOSED
Yes	29	29	16	21
No	156	107	156	156

V. CONCLUSION AND FUTURE WORKS

In this work, we proposed a new method for the detection of the damaged leaf area, which deals with noise removal (*e.g.*, shadows, objects, effects, etc.) and reconstruction of the leaf contour (using the concepts of computational geometry). The experiments showed that the proposed method quantifies the damaged leaf area with an accuracy that is in average seven times larger than human evaluation, and two times higher than the method proposed by [4].

As a proposal for future work, we plan to improve the proposed method with an interpolation algorithm for the reconstruction of curved edges instead of using line segments. We believe that with this extension, our method will achieve higher accuracy in the diagnosis of management of the tillage.

REFERENCES

[1] Picanço, F. L. Fernandes, E. G. F. Morais, M. R. Campos, and V. M. Xavier, *Tecnologias de Produção e Usos da Soja. Mecenas*, 209, ch. *Manejo Integrado das Pragas*, pp. 119–132.

[2] C. F. Wilcken, R. C. B. Moraes, M. L. Haddad, and J. R. P. Parra, "*Técnica de medição de área foliar consumida por insetos através de digitalizador de imagens*," Scientia Agrícola, vol. 55, no. 2, pp. 218–221, 1998.

[3] P.M. Einhardt and O.G.L. Ferreira, "*Validação de um método digital para determinação de área foliar*," in XVII *Congresso de Iniciaçao Científica* e X *Encontro Pós-Grad*., 2008, pp. 1–4.

[4] W. D. Mura, A. L. Oliveira, E. M. Sgarbi, and L. G. Sachs, "*Detecção automática da área foliar da soja danificada pela lagarta utilizando processamento digital de imagens*," in WUW-SIBGRABI 2007, 2007, pp. 1–4.

[5] W. Shugen, G. Zejin, and L. Deren, "Shadow detection and compensation for color aerial images," Geo-Spatial Information Science, vol. 6, no. 3, pp. 20–24, 2003.

[6] A. C. Nazaré-JR, D. Menotti, J. M. R. Neves, and T. Sedyiama,, "*Detecção Automática da Área Foliar Danificada da Soma através de Imagens Digitais*," Revista Eletrônica de Iniciação Científica, *Revista Eletrônica de Iniciação Científica* (REIC), SBC, submitted, 2010, pp. 1-17.

[7] R. C. Gonzales and R. E. Woods, Digital Image Processing, 3rd ed. Prentice Hall, 2008.

[8] N. Otsu, "A threshold selection method from gray-level histograms," in IEEE Transactions on Systems, Man and Cybernetics, vol. 9, no. 1, 1979, pp. 62–66.

[9] X. Chen-He and N. H. C. Yung, "Corner detector based on global and local curvature properties," Optical Engineering, vol. 47, no. 5, pp. 057 008 1–12, 2008.

[10] J.E. Bresenham, "Algorithm for computer control of a digital plotter," IBM Systems Journal, vol. 4, no. 1, 1965, pp 25-30.

[11] Web system, www.iceb.ufop.br/decom/lapdi/soja, 2010.

[12] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability form Engineers, 4th ed, John Wiley & Sons, 2006.