Weed Identification in Sugarcane Plantation Through Images Taken from Remotely Piloted Aircraft (RPA) and kNN Classifier

Inacio Henrique Yano[1, 2], Nelson Felipe Oliveros Mesa[1], Wesley Esdras Santiago[3], Rosa Helena Aguiar[1], Barbara Teruel[1]

[1]Faculty of Agricultural Engineering, University of Campinas, Campinas, Brazil

[2]Embrapa Informatica Agropecuaria, Campinas, Brazil

[3]Institute of Agricultural Sciences, Federal University of the Jequitinhonha and Mucuri Valleys, Unai, Brazil

Email address:

inacio.yano@embrapa.br (I. H. Yano), nfoliveosm@gmail.com (N. F. O. Mesa), esdras.wes@gmail.com (W. E. Santiago), rosahel@feagri.unicamp.br (R. H. Aguiar), barbara.teruel@feagri.unicamp.br (B. Teruel)

Abstract: The sugarcane is one of the most important crops in Brazil, the world´s largest sugar producer and the second largest ethanol producer. The presence of weeds in the sugarcane plantation can cause losses up to 90% of the production, caused by the competition for light, water and nutrients, between the crop and the weeds. Usually sugarcane plantations occupy large fields, and due to this, the weeds control is mostly chemical, which is more practical and cheaper than mechanical control. In the chemical control, the dosage and type of herbicides has been calculated by sampling, which causes problems of waste and misapplication of herbicides, since the degree of infestation may be variant from one location to another, as well as the species presents in the plantation. In order to avoid unnecessary waste in the herbicides application, there are some studies about weed identification using images taken from satellites, solution that have proved to have the advantage of covering the whole plantation, solving the problems of sample surveying, nevertheless, this method its dependent of a high weed density to ensure a good pattern recognition and its affected by the influence of clouds in the imagery quality. This work proposes a system for weed identification based on pattern recognition in imagery taken from a Remotely Piloted Aircraft (RPA). The RPA is able to fly at low altitude, so it is possible to take images closer to the plants and make the weed identification even in low infestation levels. In an initial evaluation, the system reached an overall accuracy of 83.1% and kappa coefficient of 0.775, using k-Nearest Neighbors (kNN) classifier.

Keywords: Infestation, Machine Learning, Pattern Recognition, Remote Sensing

1. Introduction

Sugarcane is an important tropical crop, largely used for sugar and ethanol production [1], especially for Brazil, the world´s largest sugarcane producer [2]. The presence of weeds in the sugarcane plantation can cause losses up to 90% of the production [3], these plants compete with crop for nutrients, water and light, hamper harvesting, and also can reduce the longevity of sugarcane plantation [4].

The sugarcane plantation occupies large fields, and for this reason, the weed control is done by herbicides [5]. The herbicide dosage estimation is calculated by sampling, causing problems of herbicide misapplication, due to the growth spatial variability of several weed species in the plantation.

In order to achieve a better precision in the herbicide application, studies about weed surveying using satellite imagery have been performed [6], however, to obtain good results, this method depends on large spots of weeds, because of the low spatial resolution, and also due to the low temporal resolution [7].

Another way for weed surveying is the use of Remotely Piloted Aircraft (RPA) imagery. The RPA can be used in low altitude flights, take images very close to the plants and do not have the problem of cloud interference. Even though the RPA cannot be used on rainy or windy days, usually, it can provide images more frequently than images taken from satellites. In [8] there is an example of a map of three categories of weed coverage produced from images taken from an RPA, which achieved 86% of overall accuracy using a multispectral camera and Object Based Image Analysis (OBIA).

The proposition of this work was identify three previous selected species of plants, the crop, one large leaf weed and one narrow leaf weed. This aim was achieved by k-Nearest Neighbors (kNN) model generated by Weka Software, using RGB images taken from an RPA as input data and reached 83.1% of overall accuracy in preliminary tests. RGB cameras are lighter and cheaper than multispectral ones, resulting in much longer drone flight time, and also can be used for a larger number of farmers.

2. Material and Methods

Figure 1 presents the flowchart of the weed identification process, which was divided into four parts. The blue arrow refers to the model creation flowchart which, once concluded, permits the use of the model for all the images taken from RPA, this second flow is represented by the dashed red lines. The computer used for Image Processing, kNN Weka Model and the Weed Mapping was an Intel Core I5 de 3.4 GHz 8 GB RAM, with Windows 10.
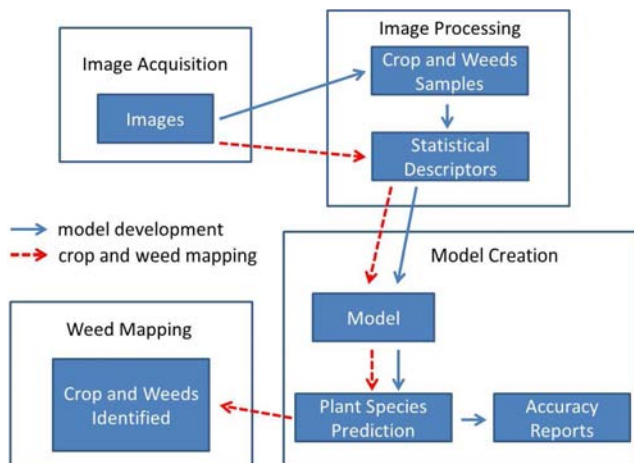

Figure 1. Weed Identification Process Flowchart.

2.1. Image Acquisition

The first step of the Weed Identification Process is Image Acquisition, which was done by a GoPro Hero3 Silver Edition 10 MP in a DJI Phantom (Figure 2). The flight occurred at the experimental field of Agricultural Engineering Faculty (FEAGRI) of Unicamp, latitude: 22°48" 57' south, longitude: 47° 03" 33' west. The images were

taken at heights of 3 to 4 meters.

Figure 2. DJI Phantom with a GoPro Hero3 Silver Edition.

2.2. Image Processing

The aim of this work was identify three species of plants, presented in Figure 3, sugarcane (crop), coco-grass (narrow leaf weed) and morning glory (large leaf weed), having this in mind, the kNN classifier was chosen, because of its supervised machine learning technique and has been used in others studies of weed identification before with good results [9]. The input data for the kNN model development were statistical descriptors of samples of crop and weeds.

(a)

(b)
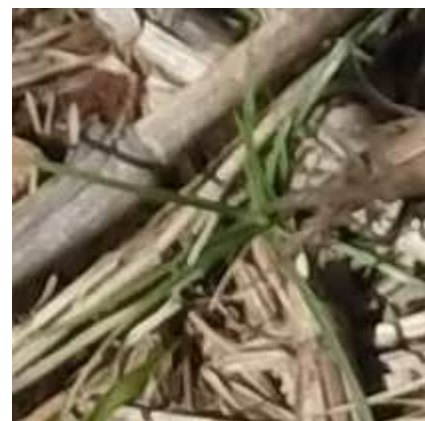
(c)

Figure 3. *Image of the Three Species Identified in This Work in Closer Images for a Better Visualization: (a) Sugarcane (Saccharum spp), (b) Narrow Leaf Weed (Cyperus Rotundus Commonly Known as Coco-grass) and (c) Large Leaf Weed (Ipomonea sp Known as Morning Glory).*

The extracting of crop and weed samples was a manual activity and the identification of the weed samples was performed by specialists [10]. Figure 4 shows in the left side the original parts of images taken from a RPA of sugarcane (a), large leaf weed (c) and narrow leaf weed (e), and in the right side the same parts of images with the samples extracted of sugarcane (b), large leaf weed (d) and narrow leaf weed (f), from these parts of images were also extracted samples of soil and straw.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4. *Original Part of Images Taken from RPA of Sugarcane (a), Large Leaf Weed (c) and Narrow Leaf Weed (e), and the Same Parts of Images with the Samples Extracted of Sugarcane (b), Large Leaf Weed (d) and Narrow Leaf Weed (f).*

The crop and weed samples were divided into small group of pixels (sub-images), like a grid [11]. A program written in Java language calculated the statistical descriptors for each sub-image. Statistical descriptors give more information than only the range of pixels values and this surplus of information can be useful for the pattern recognition process. In this work were used eight statistical descriptors (mean, average deviation, standard deviation, variance, kurtosis, skewness, maximum and minimum values) [12]. The average of pixels values of a sub-image is the mean and the measures of statistical dispersion around the mean are the variance, standard deviation and average deviation. The skewness, minimum and maximum pixel values gives the shape of the histogram and the kurtosis reflects the sharpness of the histogram [13].

These eight statistical descriptors were applied for all three channels of the RGB system, totalling 24 descriptors for each sub-image to be used in the machine learning training and in the identification process.

In this work, better results were achieved by 5x5 grid unit (pixel area), however the sub-image size depends on the image resolution and the weed infestation level, higher weed infestation are easier to identify, so the identification can be done in larger sub-image, in the opposite lower weed infestation requires smaller sub-image size.

The sub-images extracted from the samples were divided into two groups - training and validation sets. For a better training 240 sub-images of each class were used in the training set, if a class had less than 240 sub-images, some sub-images were used more than once to complete this number. The kNN Classifier achieved 83.1% of overall accuracy using 40 sub-images of each class to be identified in the validation set. In this preliminary training process four classes were chosen: the crop (sugarcane), a narrow leaf weed (coco-grass), a large leaf weed (morning glory) and straw and soil. The straw and soil was the fourth class, which was necessary to distinguish them from the plants.

2.3. Model Creation

The core of the identification process is the Weka, a Data Mining with Open Source Machine Learning Software, developed by the University of Waikato, Hamilton, New Zealand. The Weka kNN Classifier used the statistical data calculated from the samples of the images to generate a model, which can be used for three plant species prediction. The model performance was evaluated using the validation set, in order to calculate the overall accuracy and kappa coefficient [14].

In k-Nearest Neighbors (kNN), a predefined number of closest elements in distance define the class of the new element. The number of nearest neighbors of the new element that needs to be classified is the value of k. Each nearest neighbor is a vote, the most voted class among its k nearest neighbors will be assigned to the new element. The standard Euclidean distance is the most common choice to calculate the distance of the k closest elements [15]. In Figure 5, the X element will be classified as red circle instead of blue square, because there are more red circles neighbors than blue squares neighbors near it.
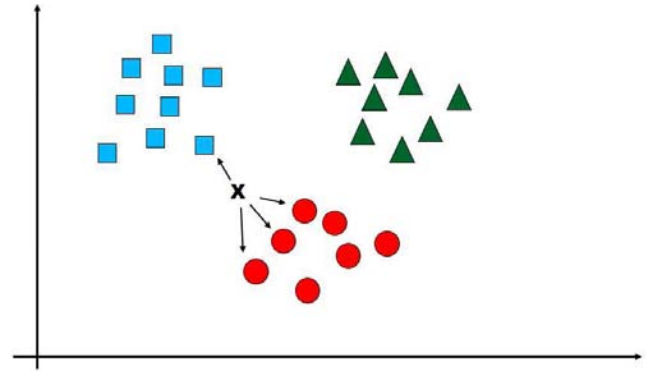

Figure 5. Example of a kNN with Three Classes.

2.4. Weed Mapping

The fourth and last step is Generating Map, where a program written in Java language will read the output of the Weka kNN Model and draw a weed map. In this map the crop and the weeds are identified by a color-coding, red for sugarcane, blue for narrow leaf weed and yellow for large leaf weed. From images like Figure 6a the statistical descriptors were calculated in the Image Processing activity (step 2), in the output of the step 2 the model generated in step 3 (Model Creation activity) were applied to identify each sub-image as crop, large leaf weed, narrow leaf weed, or straw or soil. Once identified a map can be drawn. Figure 6b is an example of a map in which the crop and the weeds were identified, there are several sub-images incorrectly identified, due to the variations of colors found in the leaves of the same plant, caused by difference in leaves age, sunlight exposure and nutrient deficiencies, but most of them were correctly identified.

(a)
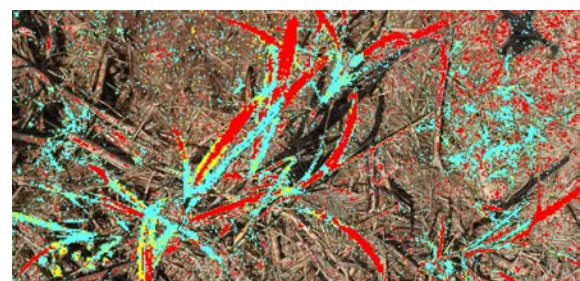
(b)

Figure 6. Image of Three Species Identified in This Work (a) Original Image, (b) Image with Three Species Marked by a Color-coding, Where the Color Red was Used for Sugarcane, Blue for Narrow leaf Weed and Yellow for Large Leaf Weed.

3. Results

Table 1 presents the confusion matrix of k-Nearest Neighbor execution. The letter "a" represents sugarcane, so as the letter "b" for coco-grass, letter "c" for morning-glory, and letter "d" for straw and soil. The corrected instances predicted by the classifiers are located in the diagonals of the tables, all the values outside the diagonals are wrong predictions made by the classifiers. The straw and soil was the class with the best accurate rate, sugarcane also had good results, for coco-grass and morning glory the kNN made some mistakes involving these two classes and both had the worst results.

Table 1. Confusion Matrix of K-Nearest Neighbor Execution.

Classified as -->	a	b	c	d	Row total
a: sugarcane	35	4	0	1	40
b: coco-grass	1	32	6	1	40
c: m. glory	2	11	26	1	40
d: straw & soil	0	0	0	40	40
Column total	38	47	32	43	160

a: sugarcane; b: coco-grass; c: m. glory; d: straw & soil.

The Weka Software generated for this execution an overall accuracy of 83.1% and kappa coefficient of 0.775, accordingly with Table 2 these results can be classified as substantial agreement. Although the map produced in [8] were about three categories of weed coverage and these results were about three species of plants (crop and two weeds), both studies mapped three classes and achieved similar accuracy assessment.

Table 2. Kappa Coefficient Interpretation Accordingly with [16].

Kappa coeficiente	Agreement Strength
<0	Poor
0-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

The weed identification process also produces a weed mapping, which allows crop and weed identification by a color-coding for a better visualization. The Figure 6b is an example of weed mapping, even though there were some wrong identification, most of the sub-images were correctly identified. In order to reduce the number of errors, a larger sample size will be tested in the Model Creation activity in future studies.

4. Conclusions

A weed mapping can be a useful tool for a more efficient application of herbicides in sugarcane plantations. In this work, a weed identification process was developed using RGB images taken from a RPA and kNN classifier. In preliminary tests, the overall accuracy achieved 83.1% and 0.775 of Kappa coefficient, which can be interpreted as substantial agreement.

These initial accuracy assessment and weed map with the plants represented by a color code demonstrating that kNN can be used for a weed map construction.

As a continuation of this work, studies for a map production with the weeds geolocation data will be made.

Acknowledgements

References

[1] D' Hont, A., Souza, G. M., Menossi, M., Vincentz, M., Van-Sluys, M. A., Glaszmann, J. C., & Ulian, E. (2008). Sugarcane: a major source of sweetness, alcohol, and bio-energy. In *Genomics of tropical crop plants* (pp. 483-513). Springer New York.

[2] Rudorff, B. F. T., Aguiar, D. A., Silva, W. F., Sugawara, L. M., Adami, M., & Moreira, M. A. (2010). Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. Remote sensing, 2(4), 1057-1076.

[3] Firehun, Y., & Tamado, T. (2006). Weed flora in the Rift Valley sugarcane plantations of Ethiopia as influenced by soil types and agronomic practises. Weed biology and management, 6(3), 139-150.

[4] Ferreira, E. A., Procópio, S. O., Galon, L., Franca, A. C., Concenço, G., Silva, A. A., & Rocha, P. R. R. (2010). Weed management in raw sugarcane. Planta Daninha, 28(4), 915-925.

[5] Cerdeira, A. L., Paraíba, L. C., Queiroz, S. C. N. D., Matallo, M. B., & Ferracini, V. L. (2015). Estimation of herbicide bioconcentration in sugarcane (Saccharum officinarum L.). Ciência Rural, 45(4), 591-597.

[6] Cavalli, R. M., Laneve, G., Fusilli, L., Pignatti, S., & Santini, F. (2009). Remote sensing water observation for supporting Lake Victoria weed management. Journal of environmental management, 90(7), 2199-2211.

[7] Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. International journal of remote sensing, 7(11), 1417-1434.

[8] Peña, J. M., Torres-Sánchez, J., de Castro, A. I., Kelly, M., & López-Granados, F. (2013). Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) pictures. PLoS One, 8(10), e77151.

[9] Chen, Y., P. Lin, Y. He and Z Xu. 2011. Classification of broadleaf weed images using Gabor wavelets and Lie group structure of region covariance on Riemannian manifolds. Biosystems engr., 09(3): 220-227.

[10] Rahman, M., Blackwell, B., Banerjee, N., & Saraswat, D. (2015). Smartphone-based hierarchical crowdsourcing for weed identification. Computers and Electronics in Agriculture, 113, 14-23.

[11] Christensen, S., Søgaard, H. T., Kudsk, P., Nørremark, M., Lund, I., Nadimi, E. S., & Jørgensen, R. (2009). Site‑specific weed control technologies. Weed Research, 49(3), 233-241.

[12] Metre, V., & Ghorpade, J. (2013). An overview of the research on texture based plant leaf classification. arXiv preprint arXiv: 1306.4345.

[13] Dixit, A., Hegde, N., & Hiremath, P. S. (2016). Cluster Analysis of Satellite (LISS-III) Pictures of Earth surface.

[14] Vieira, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. Fam Med, 37(5), 360-363.

[15] Ma, L., M. M. Crawford and J. Tian. 2010. Local manifold learning-based-nearest-neighbor for hyperspectral image classification. IEEE Trans. Geoscience Remote Sens. (TGRS), 48(11): 4099-4109.

[16] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. biometrics, 159-174.