Bulgarian Journal of Agricultural Science, 23 (No 3) 2017, 491–497
Agricultural Academy

CHOOSING CLASSIFIER FOR WEED IDENTIFICATION IN SUGARCANE FIELDS THROUGH IMAGES TAKEN BY UAV

INACIO HENRIQUE YANO[1,2]; WESLEY ESDRAS SANTIAGO[3]; JOSE RICARDO ALVES[4]; LIA TOLEDO MOREIRA MOTA[4]; BARBARA TERUEL[1]

[1] *University of Campinas, School of Agricultural Engineering, Campinas – SP, 13083-875, SP, Brazil*
[2] *Embrapa Agriculture Informatics, Campinas – SP, 13083-886, SP, Brazil*
[3] *Federal University of the Jequitinhonha and Mucuri Valleys,Unai Cachoeira CEP 38610-000, MG, Brazil*
[4] *Pontifical Catholic University, Electrical Engineering, Faculty, Campinas – SP, 13086-900, Brazil*

Abstract

Yano, I. H., W. E. Santiago, J. R. Alves, L. T. M. Mota and B. Teruel, 2017. Choosing classifier for weed identification in sugarcane fields through images taken by UAV. *Bulg. J. Agric. Sci.*, 23 (3): 491–497

Sugarcane is the main raw material in the world production of sugar and ethanol. The weeds can cause 90% loss in sugarcane production. Thus the weed control is very important and usually made by herbicides application. The estimation of herbicide type and its dosage is in general done by sampling because sugarcane occupies extensive areas. This procedure causes problems of misapplication of herbicide, since the weed species and the level of its infestations could not be uniform in whole field. There are some solutions based on remote sense, using satellite image analysis which covers the whole field, that could solve the problems of the applications of herbicides by sampling, but this solution have problems with image resolution, and can only be used on high weed infestation and in the absence of clouds for good results. This work proposed and tested a process for weed surveying, based on pattern recognition in images taken by an UAV (Unmanned Aerial Vehicle). The UAV can take images very close to the plants, so the plants pattern recognition can be done in lower infestation levels than in images taken by satellites and also is not affected by the presence of clouds. In preliminary testes, three classifiers were tested; the best classifier was an Artificial Neural Network, which achieved an overall accuracy of 91.67% and a kappa coefficient of 0.8958.

Key words: images; machine learning; pattern recognition; sugarcane; UAV; weed

Introduction

Sugarcane is the main raw material in the world production of sugar and ethanol (Costa et al., 2013). The presence of weeds can interfere in the sugarcane development, since these plants compete for environment resources, especially water, light and nutrients, release allelopathic substances that can damage the sugarcane, can be host of pests and diseases common to culture and can turn the harvesting process more difficult (Kuva et al., 2003). There are about a thousand species of weeds, which can cause loss in sugarcane yield by 20-90% (Kaur et al., 2016). Interference on the growth of sugarcane is higher in the initial phase of development of sugarcane. The weeds control is also easier at the beginning of its appearance, so it is important to identify possible infestations as soon as possible (Mcmahon et al., 2000).

The weeds control is usually made by herbicides because of its practical advantages in the application, especially considering that the areas cultivated with sugarcane are quite extensive. Precisely because of the great extent of the plantations, the choice and dosage of herbicides have been carried out by sampling and based on the perception of observers to

**E-mail:* inacio.yano@feagri.unicamp.br, wesley.santiago@ufvjm.edu.br, ricardoalves@puc-campinas.edu.br, lia.mota@puc-campinas.edu.br, barbara.teruel@feagri.unicamp.br

identify weeds in crop parts, with special attention to species more aggressively present. Despite the identification of weed performed by human analysis is more efficient than performed by machines (Borregaard et al., 2000), sampling procedure causes misapplication of herbicides, since the degree of infestation and the species present in the crop is not the same in the whole field, causing inefficiencies in weeds control, high costs and environmental damage (Ahmad et al., 2011).

In order to avoid waste in the application of herbicides, which may result from the use of the sampling procedure, several works were developed for the identification of plant categories in satellites images (Cavalli et al., 2009). The satellite imagery mapping has the advantage of covering large areas, with the disadvantages: the resolution is usually not enough for plant identification; and particularly the problem of the existence of clouds that prevent good visualization (Johnson and Trout, 2012).

Another alternative to identify weeds in images is the use of Unmanned Aerial Vehicle (UAV) for mapping the whole field, which can take images without cloud interference and with higher resolution. Peña et al. (2013) describe the identification of weeds, which are the lines of maize crop, using a multispectral camera in an UAV.

In this study, a process for weed identification is proposed and tested, which is based on pattern recognition in RGB images (Teena et al., 2016) taken by an UAV. RGB cameras are lighter and cheaper than multispectral ones, so this solution is affordable for a larger number of users (Fehr et al., 2016) and had already been used in UAV Systems (Yun et al., 2012). In this work the weed surveying was performed not only between the lines, but also in any inch of the sugarcane field, and an Artificial Neural Network (ANN) was chosen as the weed identification classifier because of its better performance, when compared to Random Forest (RF) and to k-Nearest Neighbors (kNN), and because ANN have been used before in many remote sensing studies (da Silva Junior et al., 2016; Burks et al., 2000).

This section described the importance of weed detection in sugarcane fields, Section 2 describes the material and methods used in this work; Section 3 discuss the results and discussion, and Section 4 presents the conclusion of this study.

Materials and Methods

In this section, the material and methods used to take images and the process of weed identification will be described. Figure 1 presents the Weed Identification Process, which has six steps, which are presented in subsections 2.1 to 2.6. The step three uses classifiers libraries of Weka (Free Software of University of Waikato, Hamilton, New Zealand) to generate models for weed identification. The computer used for data processing was an Intel Core I5 of 3.4 GHz, 8 GB RAM, based on Microsoft Windows 10.

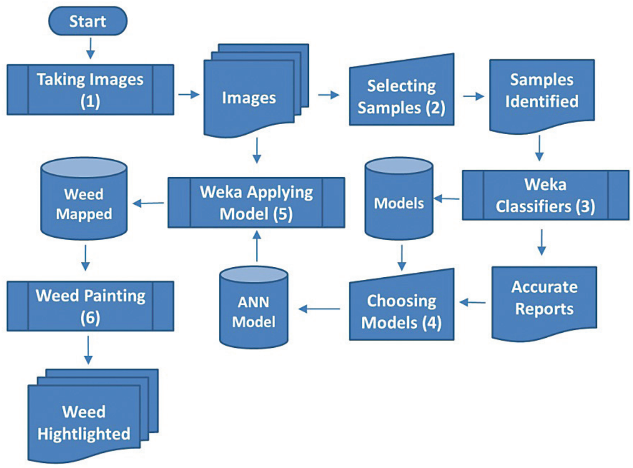

Fig. 1. Weed identification process flowchart

Taking Images

The first step of the Weed Identification Process is taking images using a RGB camera in an UAV. A DJI F450 Quadcopter (Figure 2a) was the UAV used in this work and the RGB camera was a GoPro Hero3 Silver Edition 10 MP (Figure 2b). The images were taken in a sugarcane experi-

a)

b)

Fig. 2a. DJI F450 quadcopter;
2b. GoPro Hero3 silver edition

mental field of the FEAGRI/UNICAMP, in Campinas-SP (lat 22 48’57”S, long 47°03’33”W), where the spontaneous emergence of weeds occurred.

Selecting Samples

The weed identification was done by a supervised machine learning process. After the images taken in step 1, the second step is selecting samples from these images to be used in Weka Classifiers (step 3) for models creation. The Selecting Samples process is a manual activity done by specialists, which will identify the weeds in the images and select samples of them for classifiers training. Figure 3 illustrates the Selecting Samples process in a part of an image taken by UAV with four species of plants identified, two narrow leaf species (sugarcane and signalgrass), two broad leaf species (peppergrass and shoo-fly plant) and soil, before (Figure 3a) and after the samples selected (Figure 3b).

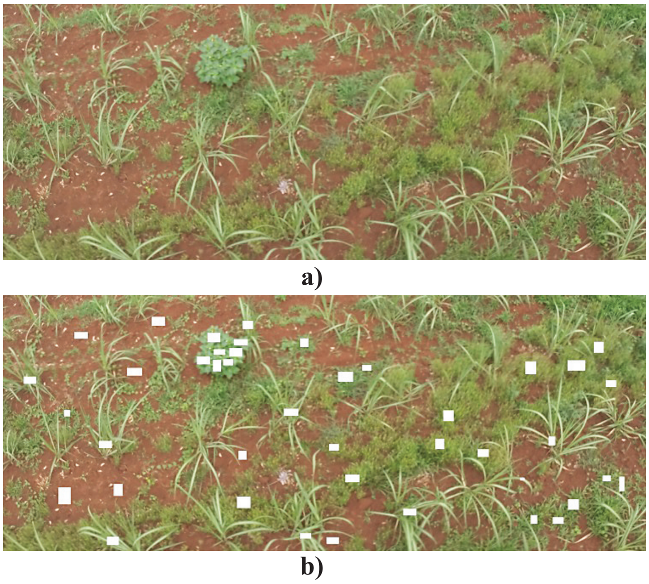

Fig. 3. Image of four species of plants identified (sugarcane, peppergrass, shoo-fly plant and signalgrass) and soil, before (a) and after (b) the samples selected

Figure 4 shows the four species in closer images (a) peppergrass, (b) sugarcane (c) shoo-fly plant and (d) signal grass) taken at ground level for a better seeing and easier specie identification.

Weka Classifiers

The third step is the Weka Classifiers process, which generates models and accurate reports. The data used for the classifiers training process were the statistical descriptors of pixels group (sub-image) of the plants samples selected from the images in step 2.

Fig. 4. Image of the four species identified in this work (a) peppergrass (*Lepidium virginicum*), (b) sugarcane (*Saccharum spp*), (c) shoo-fly plant (*Nicandra physalodes*) and (d) signalgrass (*Brachiaria decumbens*)

The use of these descriptors instead of pixel values is because even though a group of pixels belongs to the same plant, the values of them usually have some variation, so the statistical descriptors of a pixel group could be a better attribute to be used in plant recognition. In this work, eight statistical descriptors (average, average deviation, standard deviation, variance, kurtosis, asymmetry, maximum and minimum values) for each channel of the RGB image were used. Thus, there is a version of red, green and blue for each descriptor, totaling 24 descriptors.

The descriptors were obtained by dividing the samples into small sub-images, like a grid (Christensen et al., 2009). After that, the 24 descriptors were calculated for each sub-image. The size of the sub-image will depend on the image resolution and the level of weed infestation on the crop, higher weed infestation allow the use of larger sub-images with good pattern recognition. In this work, the grid unit or sub-image has 100 pixels, with good overall accuracy and Kappa coefficient (Viera and Garrett, 2005).

In this work, three classifiers were tested: Artificial Neural Network (ANN), Random Forest (RF) and k-Nearest Neighbors (kNN).

Artificial Neural Network (ANN)

The ANN used in this work was a Backpropagation (BP) because is mostly used for supervised learning, and used before with good results in weed identification from images (Barrero et al., 2016). The Backpropagation ANN requires input data with known output values, and like others ANNs is composed by input nodes, intermediate nodes and output nodes (Figure 5).

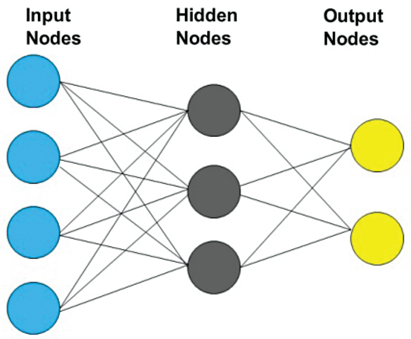

Fig. 5. Example of an artificial neural network

The input nodes are associated with the independent variables; each independent variable (statistical descriptor) is an input node. The intermediate nodes are used to enlarge the possibilities of the weights (multipliers) and, finally, there are the output nodes from which it gets the value of the dependent variable (plant previous identified). The ANN works by multiplying the values of network nodes and then summing up the results of multiplications that are accumulated in the output nodes. The difference between the dependent variable and the output node value is the error, which the iterative process tries to make smaller. When the error value is very low, it is considered that the network has converged and the model approaches the reality (Beale and Jackson, 1990).

Random Forest (RF)

Random Forest (Belgiu and Drăguț, 2016) is a technique based on decision trees sets (Safavian and Landgrebe, 1991). In the classification process, each sample is tested on each tree of the forest. The result of each test is a vote for a particular class. Completed all the tests, the sample are classified as belonging to the most voted class (Gislason et al., 2006). Figure 6 is an example of a RF that will answer “Yes” about some problem, because more than a half (60%) of the votes answered “Yes”.

k-Nearest Neighbors (kNN)

Another technique used to pattern recognition is the k-Nearest Neighbor (kNN) (Chen et al., 2011; Ma et al., 2010).

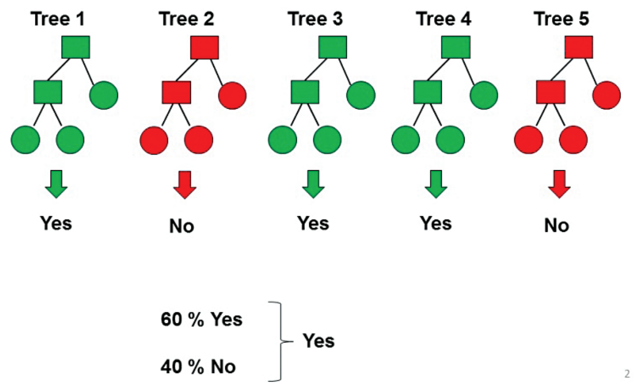

Fig. 6. Example of a random forest with five decision trees

In kNN, the closest elements are considered belonging to the same class. The value of k is the number of nearest neighbors of the sample or element that needs to be classified. The most common class among its k nearest neighbors will be assigned to the sample. To calculate the distance of the k closest elements, the Euclidean distance is the most used metrics, but the Manhattan distance and Minkowski distance are also used. In Figure 7, the X element will be classified as blue square instead of red circle, because there are more neighbors of this class near it.

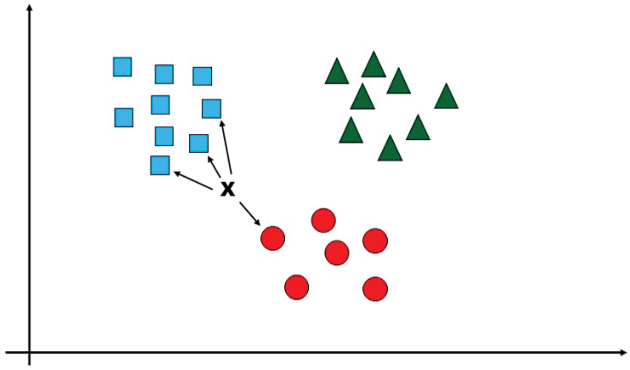

Fig. 7. Example of kNN with three classes

Choosing Model

Choosing Model is made by comparison of accurate reports generated in step 3. The Weka reports presents the overall accuracy, the most frequently used index, calculated by divididng the total number of correct identifications by the total number of instances, and Kappa statistics that uses all cells of the confusion matrix, making some compensation for the random hits, for this reason is widely used for the evaluation of classifiers (Xiao et al., 2006; Foody, 2002). In this work, the ANN achieved the best results, so the ANN Model was applied to the descriptors of the images in step 5.

Weka Applying Model

The ANN Model selected in step 5 was used to identify the plants in all images of sugarcane fields taken by UAV. Each sub-image of the images was identified as one of the four plants or soil used in the training of the classifiers process.

In this phase, the entire images taken by UAV were divided into sub-images, in the same process used to calculate the statistical descriptors before the classifiers training process. After that, each sub-image had their statistical descriptors calculated to be used as the input of Weka Software with the model selected (ANN) in step 4. The output of this step is the definition of which of the four plant species or soil is present on each sub-image.

Weed Painting

The Weed Painting is the last process of this work, in which a program written in Java language will read the output of the step 5 and draw an image with the four species of plants identified by a color-coding for a better visualization (Figure 8), Figure 8a is the original part of an image taken by UAV and Figure 8b is the same image after the four species of plants painted by a color-coding (yellow for peppergrass, red for sugarcane, purple for shoo-fly plant and blue for signalgrass). Even though there is some wrong plant identification in the Figure 8b, especially identification the sugarcane instead of signal grass, most of the plants were correctly identified.

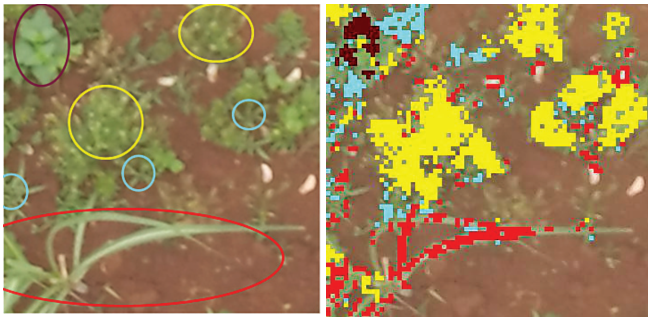

Fig. 8. Weed mapping with the four species identified in this work (a) original part of an image, (b) image with four species identified by a color-coding, where the color yellow was used for peppergrass, red for sugarcane, purple for shoo-fly plant and blue for signalgrass

Results and Discussions

This section presents the results and discussions of this work, based on the overall accuracy and Kappa coefficient, which were obtained from the Weka Classifiers process, and the weed mapping produced in the Weed Painting process.

Observing the statistical results, Artificial Neural Network had better results in this work with 91.67% of correctly classified instances and 0.8958 of Kappa coefficient, Random Forest was the second best classifier with 90.93% of correctly classified instances and 0.8866 of Kappa coefficient. One can read the results of Kappa coefficient of these two classifiers as almost perfect concordance accordingly with Table 1. The worst classifier was k-Nearest Neighbor with 81.11% of correctly classified instances and 0.7639 of Kappa coefficient and can be classified as substantial concordance, which was also a good result.

Table 1
Kappa coefficient interpretation accordingly with Landis and Koch (1977)

Kappa coefficient	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

Table 2 presents the confusion matrix of Artificial Neural Network (ANN) execution, so as the Table 3 for Random Forest (RF) and Table 4 for k-Nearest Neighbor (kNN). The columns of the Tables 2, 3, and 4 represent the results of the classifiers and the rows represent the selected samples classified by the specialists. In this research there were 108 samples of each class (plants or soil) totaling 540 samples. The letter "a" represents peppergrass, so as the letter "b" for sugarcane, letter "c" for shoo-fly plant, letter "d" for signal grass, and letter "e" for 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.

Table 2
Confusion matrix of Artificial Neural Network execution

Classified as ->	a	b	c	d	e	Row total
a = peppergrass	105	0	2	0	1	108
b = sugarcane	6	90	11	1	0	108
c = shoo-fly plant	2	13	89	4	0	108
d = signalgrass	0	1	4	103	0	108
e = soil	0	0	0	0	108	108
Column total	113	104	106	108	109	540

Table 3
Confusion matrix of Random Forest execution

Classified as ->	a	b	c	d	e	Row total
a = peppergrass	103	1	3	0	1	108
b = sugarcane	7	89	8	4	0	108
c = shoo-fly plant	1	9	94	4	0	108
d = signalgrass	0	2	8	98	0	108
e = soil	0	1	0	0	107	108
Column total	111	102	113	106	108	540

Table 4
Confusion matrix of k-Nearest Neighbor execution

Classified as ->	a	b	c	d	e	Row total
a = peppergrass	88	6	14	0	0	108
b = sugarcane	7	81	14	6	0	108
c = shoo-fly plant	3	21	76	8	0	108
d = signalgrass	0	10	13	85	0	108
e = soil	0	0	0	0	108	108
Column total	98	118	117	99	108	540

The ANN classifier obtained the best results for peppergrass, sugarcane and signalgrass, for shoo-fly plant, the best result was achieved by the RF classifier and for soil, kNN was the best classifier. From Tables 2, 3 and 4 is possible to observe that the soil was the class with the best accurate rate achieved by all the classifiers, because this class is quite different from the others, and so, easier to identify. Considering the set of results from all confusion matrixes the ANN model was chosen at the Choosing Model process to be applied in the Weka Appling Model to all images taken at the Taking Images process.

Another form of the results can be seen, in practical terms, by the images produced by the Weed Painting process. The Figure 8b presents the plants chosen in this study for identification painted in a color-coding for a better visualization. Even though the plants were mostly correctly identified, there are some mistakes in plants identification, this is because of the color difference between the samples selected during the Selecting Samples process and the plants present in the sub-images in the Weka Applying Model process. The color difference between plants of the same species happens due to several factors such as different stage of growth, brightness, shadows, soil composition and moisture, dust and injuries caused by diseases and pests.

Conclusions

In this work, three classifiers were tested for weed identification in images taken by UAV and the best of them (ANN) achieved 0.8958 of Kappa coefficient, which, according to Landis and Koch (1977), can be interpreted as almost perfect agreement. RF also achieved almost perfect agreement, demonstrating that this both classifiers were good for the plants identification selected in this study.

Although the results of Weed Painting process produced some wrong plant identification, most of the plants were correctly identified. To reduce this problem, future studies will be made with more variety of samples.

This work worked with spontaneous weed growth and will be continued with a creation of an experimental field with the seeding of the most harmful weed, in order to verify if the Weed Identification Process, proposed in this work, will have good results with these harmful weeds.

Acknowledgments

The authors thank the Instituto Plantarum de Estudos da Flora Ltda, Nova Odessa, SP – Brazil, for the valuable contribution on plant identification and Dr. José Teixeira Filho, professor at the School of Agricultural Engineering, University of Campinas – UNICAMP, Campinas, SP – Brazil, for the access to the experimental field. The authors also thank the Pontifical Catholic University of Campinas for the scholarship support related to this work.

References

Ahmad, I., M. H. Siddiqi, I. Fatima, S. Lee and Y. K. Lee, 2011. Weed classification based on Haar wavelet transform via k-nearesjt neighbor (k-NN) for real-time automatic sprayer control system. In: Proceedings of the 5th Int. Conf. on Ubiquitous Inf. Mgmt and Comm. ACM, p. 17.

Barrero, O., D. Rojas, C. Gonzalez and S. Perdomo, 2016. Weed detection in rice fields using aerial images and neural networks. In: Signal Processing, Images and Artificial Vision (STSIVA), 2016 XXI Symposium on. IEEE, pp. 1-4.

Beale, R. and T. Jackson, 1990. Neural Computing – an Introduction. Institute of Physics Publishing, *CRC Press*.

Belgiu, M. and L. Drăguț, 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*, 114: 24-31.

Borregaard, T., H. Nielsen, L. Nørgaard and H. Have, 2000. Crop-weed discrimination by line imaging spectroscopy. *J. Agric. Engr. Res.*, **75** (4): 389-400.

Burks, T. F., S. A. Shearer, R. S. Gates and K. D. Donohue. 2000. Backpropagation neural network design and evaluation for classifying weed species using color image texture. *Trans. ASAE*, **43** (4): 1029.

Cavalli, R. M., G. Laneve, L. Fusilli, S. Pignatti and F. Santini. 2009. Remote sensing water observation for supporting Lake Victoria weed management. *J. Environ. Manage.,* **90** (7): 2199-2211.

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.*, **9** (3): 220-227.

Christensen, S., H. T. Søgaard, P. Kudsk, M. Nørremark, I. Lund, E. S. Nadimi and R. Jørgensen, 2009. Site-specific weed control technologies. *Weed Res.*, **49** (3): 233-241.

Costa, C. D. S., E. M. Pedrosa, M. M. Rolim, H. R. Santos and A. T. Cordeiro Neto, 2013. Effects of vinasse application under the physical attributes of soil covered with sugarcane straw. *Eng. Agrícola*, **33** (4): 636-646.

da Silva Junior, C. A., M. R. Nanni, P. E. Teodoro, G. F. C Silva, M. G. de Lima and M. Eri, 2016. Comparison of mapping soybean areas in Brazil through perceptron neural networks and vegetation indices. *Afr. J. Agric. Res.*, **11** (43): 4413-4424.

Fehr, D., W. J. Beksi, D. Zermas and N. Papanikolopoulos. 2016. Covariance based point cloud descriptors for object detection and recognition. *Comput. Vis. Image Und.*, 142: 80-93.

Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sens. Environ.*, **80** (1): 185-201.

Gislason, P. O., J. A. Benediktsson and J. R. Sveinsson. 2006. Random forests for land cover classification. *Pattern Recognition Letters*, **27** (4): 294-300.

Johnson, L. F. and T. J. Trout, 2012. Satellite NDVI assisted monitoring of vegetable crop evapotranspiration in California's San Joaquin Valley. *Remote Sens.*, **4** (2): 439-455.

Kaur, N., M. S. Bhullar and G. Gill. 2016. Weed management in sugarcane-canola intercropping systems in northern India. *Field Crops Res.*, 188: 1-9.

Kuva, M. A., R. Gravena, R. A. Pitelli, P. J. Christoffoleti and P. L. C. A. Alves. 2003. Períodos de interferência das plantas daninhas na cultura da cana de açúcar. III-Capim-braquiária (*Brachiaria decumbens*) e capim-colonião (*Panicum maximum*). *Planta Daninha*, **21**(1): 37-44.

Landis, J. R. and G. G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33: 159-174.

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.

Mcmahon, G., P. Lawrenceand and T. O'grady. 2000. Weed Control in Sugarcane. Manual of Cane Growing. *Bureau of Sugar Experiment Stations*, Indooroopilly, pp. 241-261.

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

Safavian, S. R. and D. Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.*, **21**(3): 660-674.

Teena, M., A. Manickavasagan, A. M. Al-Sadi, R. Al-Yahyai, M. L. Deadman and A. Al-Ismaili, 2016. RGB color imaging to detect Aspergillus flavus infection in dates. *Emir. J. Food Agric.,* **28** (10): 683-688.

Viera, A. J. and J. M Garrett, 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med*, **37** (5): 360-363.

Xiao, J., Y. Shen, J. Ge, R. Tateishi, C. Tang, Y. Liang and Z. Huang. 2006. Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and Urban Planning*, **75** (1): 69-80.

Yun, M., J. Kim, D. Seo, J. Lee and C. Choi. 2012. Application possibility of smartphone as payload for photogrammetric UAV system. International Archives of the Photogrammetry. *Remote Sens. Spatial Inf. Sci.*, **39** (B4): 349-352.

Received April, 4, 2017; accepted for printing May, 2, 2017