



An Image Processing Method Based on Features Selection for Crop Plants and Weeds Discrimination Using RGB Images

Ali Ahmad¹, Rémy Guyonneau¹, Franck Mercier¹, and Étienne Belin^{1,2}(✉)

¹ Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS),
Université d'Angers, 49000 Angers, France

etienne.belin@univ-angers.fr

² ImHorPhen, SFR 4207 QUASAV, IRHS, UMR1345, Université d'Angers,
49045 Angers, France

Abstract. In the context of computer vision applied to precision agriculture, this paper presents an imaging system based on shape and intensity features, extracted from RGB images, for the discrimination between crop plants and weeds. A segmentation method with many constraints to overcome light acquisition conditions is used and coupled with morphological filtering suitable for denoising segmented images. A SVMs classifier based on a polynomial kernel function is implemented and a k-folds cross validation process is used to evaluate the performance of the SVMs classifier usable in 2 different configurations. On a training dataset, these 2 configurations are evaluated for the performance of classification in terms of true and false positive rates, according to ROC curves and area under curves. On a test dataset, these 2 configurations are exploited, giving both a relevant classification rate.

Keywords: Computer vision · RGB images · Features selection
Precision agriculture

1 Introduction

Weed control is a significant issue in agricultural crop production as weeds compete with crop plants and can have a negative impact on crop yields. Weeding is therefore a critical operation to maintain crop yields, and is mostly done by spraying herbicides over the entire surface of the field. As herbicides have harmful effects, weed management practices using chemical tools suggest to locally apply herbicides within the strictly necessary dosage. At a field scale, this requires tools capable to identify and locate weeds among crop plants in inter-row or in intra-row [1]. Machine vision and image processing are widely used in vegetal applications [2,3]. The detection and precise location of weeds can be obtained by computer vision and various image processing methods have been applied to this

purpose [4]. Popular methods for discrimination between plant crops and weeds are based on features extraction. Depending on applications, the features extraction is of important variability [5–7]. Other methods exploit frequency content of data, using the Gabor wavelet transform [8] or the Haar decomposition [9].

In this article, we detail an image processing method based on features selections capable to discriminate crop plants and weeds from RGB images. We first describe the image acquisition protocol, done under real outdoor conditions and present a specific algorithm implemented to efficiently separate green vegetation objects from the background. We then train and test a polynomial kernel-based SVMs classifier within shape and intensity features extracted from segmented objects, and we especially assess the performances of the classification regarding different configuration for the SVMs classifier. As such image processing method is intended to be used in concrete application framework, the study presented here is carried out within the framework of the discrimination between lamb’s lettuce plants and weeds.

2 Image Acquisition and Processing for the Selection of Vegetation Objects

Image acquisition takes place in real conditions, under a greenhouse of lamb’s lettuce plants, with various illumination conditions. RGB images are captured by a camera of 20 Mpixels with a spatial resolution of 5120×3840 pixels, mounted with a 35 mm objective. The camera is positioned vertically on an automated moving robot. A typical distance between the camera and the crop row is of 120 cm. We illustrate examples of acquired images in panel A of Fig. 1.

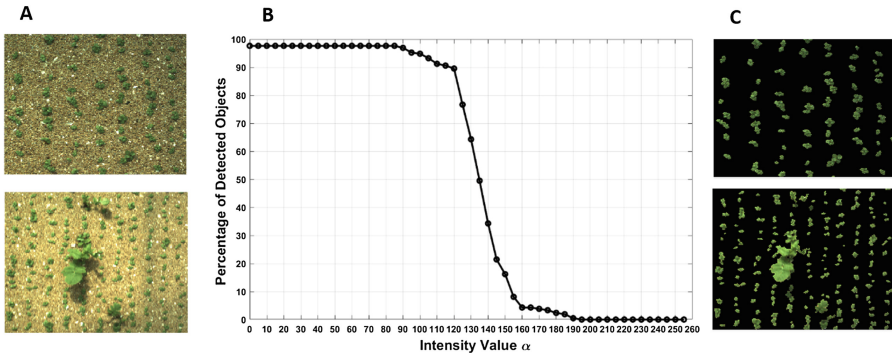


Fig. 1. Panel A illustrates 2 examples of acquired images in 2 different light conditions. These 2 images contain a total amount of 210 vegetation objects. Panel B represents the evolution of the percentage of detected vegetation objects in the binary images according to α values by comparing within the total amount of vegetation objects counted from the ground truth. Panel C shows the selection of vegetation objects obtained on the images of panel A.

Several methods have already been used in the segmentation of RGB images for the selection of vegetation. We implement here a specific algorithm, called modified excess green (MExG) described in [7], that will favor the green component of objects by combining, in a grey-level intensity image MExG, the non-normalized R, G and B grey-level intensity channels with the formula $MExG = 2 \times G - R - B$ within constraints. Such algorithm, independant from the camera and its resolution, presents some interest as it limitates image acquisition artifacts like brightness, saturation and blurring due to robot moving. Applied to our study case, the MExG algorithm is set by constraints defined as $G < R$ and $G < B$ and $G < \alpha$ with α representing an intensity value between 0 and 255 to be determined. If one pixel satisfied such constraints, then it is set to zero in the MExG image. Binarized images are then calculated by thresholding to zero. A morphological filtering step based on erosion and dilation is performed on binarized objects to eliminate artefacts and to smooth edges due to blurring. The size of the structural element of the filter is evaluated according to the size of objects in images. Experimentally, the radius size of the filter is set to 5 pixels and, to evaluate α , two images representing two different acquisition conditions with annotated ground truth giving number of vegetation objects to be detected are used. We compare the number of detected objects in the binary images with the ground truth, according to α values. As shown in panel B of Fig. 1, a chosen value for α between 0 and 80 is relevant for a good binarization of vegetation objects visible in panel C of Fig. 1.

3 Support Vector Machines (SVMs) Classification Method

Based on extracted features, classifier algorithm is used on training data in order to establish a discrimination model. According to our method which aims at being exploited in real time in further applications, the classifier algorithm must then be fast and requires a minimal memory to store the knowledge obtained from the learning process. We choose to use a SVMs for learning and classification steps, that is well-adapted to the discrimination of vegetation objects [10]. A SVMs algorithm employs an additional kernel function H to project data from the original features space to an alternative higher dimensional features space, and is then able to classify non linearly two or more classes from alternative space via hyper planes defined from the kernel function H . In our study, the chosen kernel function H is a polynomial kernel of degree d that constitutes an adjustable parameter to be tested if accuracy of the classification has to be improved. A 10-folds cross validation approach is computed. A training dataset composed of 10 images, acquired in various light conditions and containing a total amount of 4500 vegetation objects is used. These images have been annotated by experts, defining one crop plants class and one weeds class, to establish ground truth data. For the feature selection, in a first approach, we choose to use a small set of features, only composed with a shape feature and a color feature: the area of vegetation object as shape feature f_1 and the average intensity pixels

of vegetation objects as color feature f_2 . Features (f_1, f_2) are extracted for each vegetation object and coded as a row vector of a matrix of 4500×2 . We estimate the performance 10 times. For each time, the 4500 vectors will be split into 9-folds for training and 1 for validation. A classical approach for evaluating a classifier performance is based on the use of receiving operating characteristic (ROC) curves. These ROC curves are computed as average of 10 estimations in our case, and then the area under curve (AUC) is also estimated for each ROC curve. By definition, the AUC is the probability that a positive event is classified as positive by the test over the extent of the possible decision threshold values. In our case, the positive event is the weed detection and false positive event is when a crop plant is detected as a weed. Usually, the performance of classification is determined by AUC values. If $AUC > 0.9$ then classification is considered as excellent whereas if $AUC < 0.5$ then classification is considered as poor.

4 Performance of the SVMs Classification Method Evaluated on a Training Dataset

In the following, we propose to evaluate performance of the SVMs classification method on a training dataset. We assess the performance with a polynomial kernel of degree $d = 1$ and features (f_1, f_2) . In this configuration, the AUC is of 0.88 as illustrated in Panel A of Fig. 2. This is a satisfactory result that may be improved by adjusting the parameters. On one hand, to improve performance, we choose to increase the degree d of the polynomial function. The results obtained for the AUC are reported in Fig. 2, where the degree d increases up to the order 3. As visible, always with the 2 features only, there is a beneficial effect to increase the degree of the polynomial. On the other hand, another improvement consists in maintaining degree $d = 1$ and add other shape and color features. Complementary to already extracted features (f_1, f_2) , and commonly used in literature [11], a set of 8 additional shape features are defined (perimeter, convex area, compactness, solidity, diameter of a circle with the same area as the region,

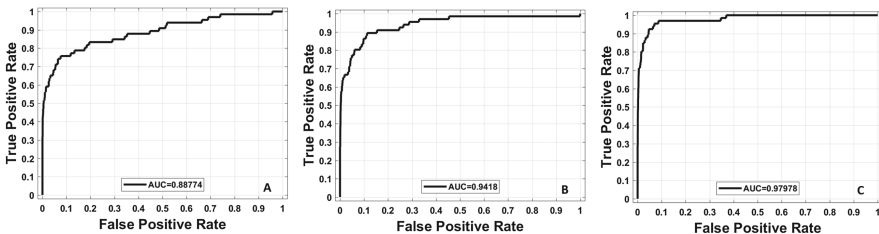


Fig. 2. Evolution of ROC curves and AUC of the classification method according to the polynomial kernel of degree d within 2 features (f_1, f_2) extracted. Panel A: $d = 1$ and AUC is of 0.88, panel B: $d = 2$ and AUC is of 0.94, panel C: $d = 3$ and AUC is of 0.98.

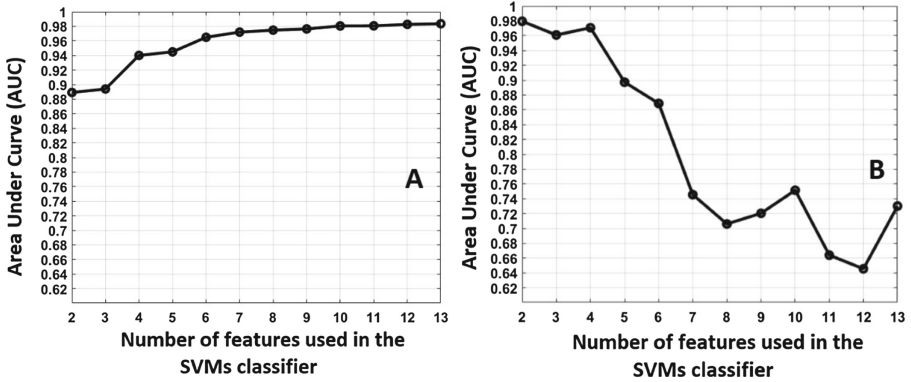


Fig. 3. Evolution of AUC values when features are added in the SVMs classifier. Panel A: $d = 1$ and maximum AUC upper than 0.98 is obtained for the set of 13 features. Panel B: for $d = 3$ AUC remains maximum with only 2 features.

eccentricity, minor and major axis lengths) and 3 additional color features are introduced (standard deviation, maximum and minimum of intensities). The contribution of added features is evaluated according to the evolution of the AUC each time a feature is added. This evolution is plotted in Fig. 3. As illustrated in Panel A of Fig. 3, there is a gain of the performance when the number of features increases. Thus, with the set of 13 features, the AUC is greater than 0.98 for a degree $d = 1$. This is a satisfactory result for the classification task envisaged here. To go further, it is interesting to note that the addition of features reduces performance when the degree is greater than 1. This is illustrated with Panel B of Fig. 3 for $d = 3$ where additional features degrade the performance.

5 Application of the SVMs Classification Method on a Test Dataset

The performance of the classification is therefore satisfactory under 2 different parametrical configurations for the SVMs classifier, with a set of 13 features and a polynomial kernel of degree $d = 1$ or, within a polynomial kernel of higher degree such as $d = 3$ but with a reduced set of features. Considering these 2 parametrical configurations, we now exploit the SVMs classifier on a test dataset composed of 4 images containing 1762 vegetation objects illustrated in the Panel A of Fig. 4. According to the ground truth showed in Panel B of Fig. 4, a correct classification rate can be calculated. This rate is of 0.9739 in the configuration within degree $d = 1$ and the set of 13 features showed in Panel C of Fig. 4. In the second configuration within degree $d = 3$ and the set of 2 features, the rate is of 0.9888, showed in Panel D of Fig. 4. In both cases, classification is very good and computation time is quite relevant. However, the accuracy of the classification can be discussed for the 2 configurations tested here. As visible in Panels C and D of Fig. 4, there are more some false positive

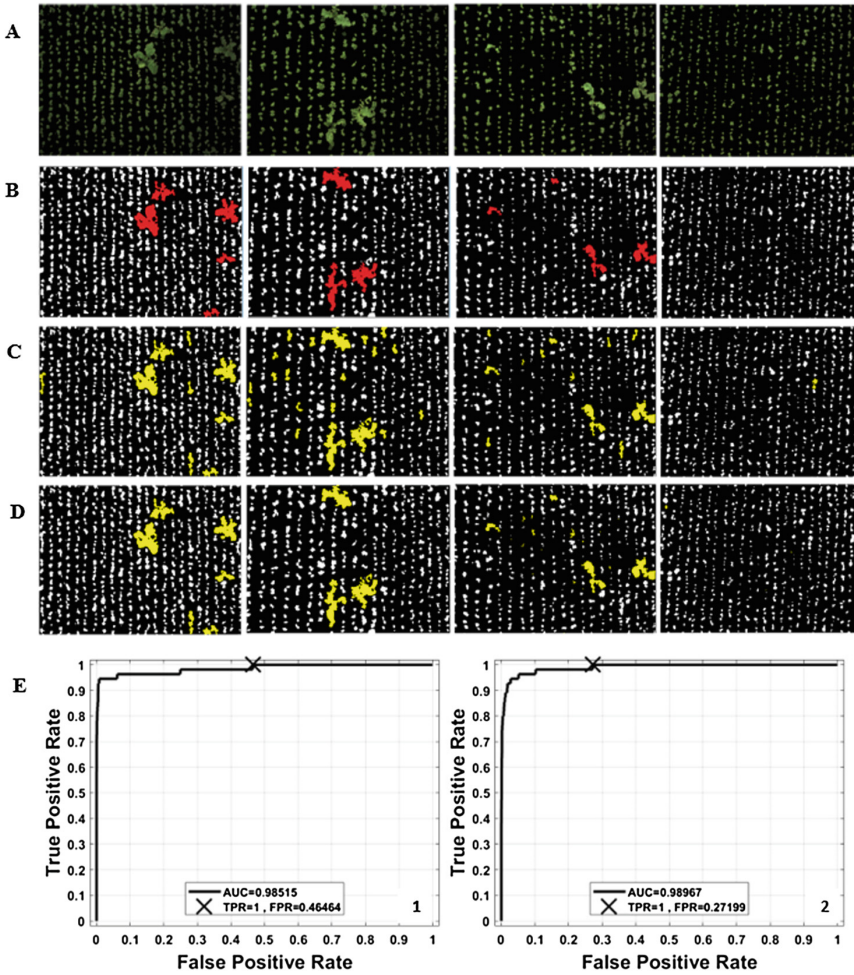


Fig. 4. Results of SVMs classifier applied to 4 images with known ground truth to estimate the accuracy of the classification. Panel A: initial images. Panel B: ground truth images where red objects correspond to weeds. Panel C: result of classification using the set of 13 features within degree $d = 1$. Panel D: result of classification using the set of 2 features within degree $d = 3$. Yellow objects correspond to object classified as weeds in panel C and D. Panels E1 and E2 show the evolution of ROC curves for the assesment of SVMs classifier accuracy on the test dataset. The false positive rates (FPR) are calculated when all weeds are detected (true positive rate TPR=1). Panel E1: configuration with degree $d = 1$ using 13 features. AUC is of 0.985 and FPR of 0.46. Panel E2: configuration with degree $d = 3$ using 2 features. AUC is of 0.989 and FPR of 0.27. (Color figure online)

events in the case of configuration with $d = 1$ and 13 features than the other one. This is due to redundancy caused by increasing the number of features. In Panels E1 and E2 of Fig. 4, we illustrate with ROC curves the accuracy of these 2 configurations. A complete weeds detection ($TPR = 1$) is accomplished in both cases. Nevertheless, regarding accuracy, in configuration with $d = 1$ using 13 features, the false positive rate (FPR) is quite important as it equals to 0.46 whereas the FPR equals 0.27 in other configuration. It leads to consider this second configuration as more adapted and recommended to the classification task envisaged here.

6 Conclusion

In this study, we proposed a computer vision tools for crop plants and weeds discrimination. The segmentation method using a constraints based-modified excess green protocol have been applied to the RGB images acquired outdoor in fields to overcome the variability of light conditions. A study based-cross validation method operating with ROC curves and AUC values as criteria was conducted to assess and improve the classification of the employed SVMs classifier used with a polynomial kernel function. Different parametrical configurations concerning both the degree d of the polynomial kernel and the number of shape and color features to take into account have been tested to assess and overcome the initial performance of the SVMs classifier. Firstly, a configuration was considered for the classification task within degree $d = 1$ and only 2 features, the area and the average intensity of vegetation objects. Secondly, this classification performance is improved by two ways. On one hand, by increasing the number of features from 2 to 13 and by maintaining the degree of polynomial function to $d = 1$. In this case, we have shown that the probability of redundancy increases explaining an important false positive rate for a complete weeds detection. On the other hand, we have chosen to increase polynomial degree from $d = 1$ to $d = 3$ and to keep only the 2 initial features. Again, a correct classification rate was obtained for this configuration but with a lower false positive rate than for the first configuration. This makes the second configuration a better solution for increasing the accuracy of our classification method. Such an approach with this computer vision system for image acquisition and image processing seems efficient and robust for the intended application here.

In future works, in the framework of our study-case, some focus could be done on the classification performance and accuracy. For instance, impact of other kernel function could be studied, such as the radial basis function kernel instead of polynomial function. In addition, within the framework of the discrimination between lamb's lettuce plants and weeds, the overlap of vegetation objects is a real challenge in image processing. For a relevant classification task in dense culture, some extend of this study could focus in solving the problem of overlapping to separate vegetation objects by sift or texture image processing methods [12].

Acknowledgement. This work received support from the Fonds Unique Interministériel (FUI) in the framework of the project PUMAGri.

References

1. Slaughter, D., Giles, D., Downey, D.: Autonomous robotic weed control systems: a review. *Comput. Electron. Agric.* **6**, 63–78 (2008)
2. Belin, E., Rousseau, D., Boureau, T., Caffier, V.: Thermography versus chlorophyll fluorescence imaging for detection and quantification of apple scab. *Comput. Electron. Agric.* **90**, 159–163 (2013)
3. Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, E., Chapeau-Blondeau, F.: On the use of depth camera for 3D phenotyping of entire plants. *Comput. Electron. Agric.* **82**, 122–127 (2012)
4. Desai, R., Desai, K., Desai, S., Solanki, Z., Patel, D., Patel, V.: Removal of weeds using image processing: a technical review. *Int. J. Adv. Comput. Technol. (IJACT)* **4**, 27–31 (2015)
5. Astrand, B., Baerveldt, A.: An agricultural mobil robot with vision-based perception for mechanical weed control. *Auton. Robot* **13**, 21–35 (2002)
6. Blasco, J., Alexios, N., Roger, J., Rabatel, G., Molto, E.: Robotic weed control using machine vision. *Biosyst. Eng.* **83**, 149–157 (2002)
7. Aitkenhead, M., Dalgetty, I., Mullins, C., Strachan, N.: Weed and crop discrimination using image analysis and artificial intelligence methods. *Comput. Electron. Agric.* **39**, 157–171 (2003)
8. Tang, L., Tian, L., Steward, B.L.: Classification of broadleaf and grass weeds using gabor wavelets and an artificial neural network. *Am. Soc. Agric. Eng.* **46**, 1247–1254 (2003)
9. Bakhshipour, A., Jafari, A., Nassiri, S., Zare, D.: Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst. Eng.* **157**, 1–12 (2017)
10. Siddiqi, M., Lee, S., Khan, A.: Weed image classification using wavelet transform, stepwise linear discriminant analysis, and support vector machines for an automatic spray control system. *J. Inf. Sci. Eng.* **30**, 53–64 (2014)
11. Du, J., Wang, X., Zhang, G.: Leaf shape-based plant species recognition. *Appl. Math. Comput.* **185**, 883–893 (2007)
12. Rasti, P., Ahmad, A., Belin, E., Rousseau, D.: Learning on deep network without the hot air by scattering transform. Application to weed detection in dense culture. In: 6th International Workshop on Image Analysis Methods for the Plant Sciences (IAMPS), Nottingham, 22–23 January 2018