

Machine vision system for weed detection using image filtering in vegetables crops

Sistema de visión de máquina para detectar maleza usando filtros de imágenes en cultivos de hortalizas

Camilo Andrés Pulido-Rojas*, Manuel Alejandro Molina-Villa, Leonardo Enrique Solaque-Guzmán

Grupo de investigación GIDAM, Facultad de Ingeniería, Universidad Militar Nueva Granada. Carrera 11 # 101-80. C. P. 11011. Bogotá, Colombia.

ARTICLE INFO

Received September 11, 2015

Accepted May 27, 2016

KEYWORDS

Weed detection, machine vision, weed removal, discriminate crop

Detección de maleza, visión de máquina, remoción de maleza, clasificación cultivo

ABSTRACT: This work presents a machine vision system for weed detection in vegetable crops using outdoor images, avoiding lighting and sharpness problems during acquisition step. This development will be a module for a weed removal mobile robot with *camera obscura* (Latin for "dark room") for lighting controlled conditions. The purpose of this paper is to develop a useful algorithm to discriminate weed, using image filtering to extract color and area features, then, a process to label each object in the scene is implemented, finally, a classification based on area is proposed, including sensitivity, specificity, positive and negative predicted values in order to evaluate algorithm performance.

RESUMEN: El presente trabajo expone un sistema de visión de máquina para la detección de maleza en cultivos de hortalizas, usando imágenes exteriores, evadiendo problemas de iluminación y nitidez durante la etapa de adquisición, ya que el presente desarrollo será un módulo para un robot móvil removedor de maleza con una cámara oscura para controlar las condiciones de luz. El enfoque de diseño se enmarca en el desarrollo de un algoritmo útil para discriminar maleza usando filtros en la imagen extrayendo características de color y área, luego se implementa un proceso de etiquetado para cada objeto en la escena. Finalmente, una clasificación basada en área es propuesta incluyendo el cálculo de los índices de sensibilidad, especificidad, valores predictivos positivos y negativos con el fin de evaluar el rendimiento del algoritmo.

1. Introduction

The agricultural sector in Colombia is focused on establishing environmental requirements that ensure quality and safety of crops production. This approach seeks to make use of technological development to support the processes of soil preparation, planting and weed identification. Nowadays, agricultural workers in a medium and low Colombian farms, remove weeds using herbicides or making a hand labor. The first alternative covers whole crop, causing environmental and economic concerns because it is not a selective treatment. The second way is a painstaking work and given the size of the crop, it may be necessary to sample the population to identify plagues and weed, with a high cost due to time spent on weed removal. Other similar works have developed machine vision applications for agriculture. In [1], a real time system using fast Fourier transform and Gray level co-occurrence matrix for image processing is integrated into the mechanical structure of spraying system for weed removal in Oil palm plantation,

test results obtain between 70.4 and 72.5% classification accuracy. Another work involves two processes: image segmentation and decision [2]. Source image is binarized applying a threshold from RGB format; with plant pixels, crop row direction is estimated; next, a Support Vector Machine classifier is trained and implemented to decide presence of weed offering an optimum performance in terms of computational cost.

In [3], a work to identify weeds using an *off-line* training from RGB images of corn, these images are segmented to differentiate vegetation and soil, estimating a probability density function and partitioning the image into cells formed detecting from culture lines using Hough transform. In operation, the system calculates the cells of a new image and applies Bayes' theorem to make the decision process of belonging to the *Weed* class. The authors in [4] use a JAI AD-130 machine vision camera capable of simultaneously capturing visible and near-infrared light spectrums; they characterized soil, green and dry plants using difference between red grey level and near infrared grey level (index NIR-R) in crop images, achieving a clear threshold for separate and segment objects in the scene; the results indicate that 0 to 50 grey level for dry plants and 150 to 250 for green plants is a good way to classify vegetation, suggesting that plants and soil can be separated using a threshold from histogram. Another work about weed

* Corresponding author: Camilo Andrés Pulido Rojas

e-mail: camilopulidorojas@gmail.com

ISSN 0120-6230

e-ISSN 2422-2844

detection is presented in [5]; first, image acquisition store 24-bit color images with 5 MP resolution in RGB color space, next excessive green algorithm [6] removes unwanted information as well as segment and separate crop images into vegetation and non-vegetation. Furthermore, median filtering is implemented to remove noise, then a labeling algorithm is used to connect components in image, then, size based features are extracted by using Mathematical Morphology; Area, Perimeter and Longest chord are estimated. Finally, with a threshold and crop masking made, weed is detected.

The main objective of the research presented in this paper is to develop a weed identification system using machine vision as a starting point to build an autonomous mobile robot for weed removal to be used in Colombian vegetable farms. The system consists of an 8MP camera to take images of crop and a laptop with 2.50 GHz processor and 6 GB RAM to data processing using algorithms about image filtering, optimal thresholding, labeling and area descriptor. The idea of the present work is to evaluate weed classification accuracy, further work implies parallel programming on embedded system for online processing. Results of detection will be transferred to a mechanical structure pulled by a tractor with an end effector responsible of pulling out weed plants. Current study was limited by a periodical weed removal task whereby an early age weed detection is carried out with a classification based on area. Tests were performed in a Horticulture crops of Nueva Granada Military University Campus in Cajicá, Colombia. The development of this project provides a technological tool to support Colombian farmers in the maintenance of vegetables crops.

This paper has been divided into five parts. The first part is about Introduction, chapter 2 describes Weed detection system and its stages: image acquisition, plant detection, feature extraction, labeling and classification. Results and discussion are explained in chapter 3. Then, conclusions, future work and acknowledgments are presented.

2. Weed detection system

The approach of the project is achieving a baseline method for developing a real time weed detection system through binary classification when vegetation is detected, that is, to separate soil and plants, then, to apply a feature extraction for discriminating weed. First, Green plant detection algorithm is implemented to remove soil from image such that image information is reduced. The next steps of algorithm focus only on vegetation, then, median filtering removes noise as "salt and pepper" with advantage of preserving edges. Third, the previous output is converted to binary; at this point, small objects are removed in order to avoid outliers. After, the pixels connected around their neighborhood are labeled, thus, all objects in the image are identified. Finally, area calculation for each object is done. With the values obtained, we set a threshold to differentiate weed from crop, such that the method is a feature extraction criterion based on size. Figure 1 shows the flow chart corresponding to the process described above. The algorithm was performed using MATLAB R2015.

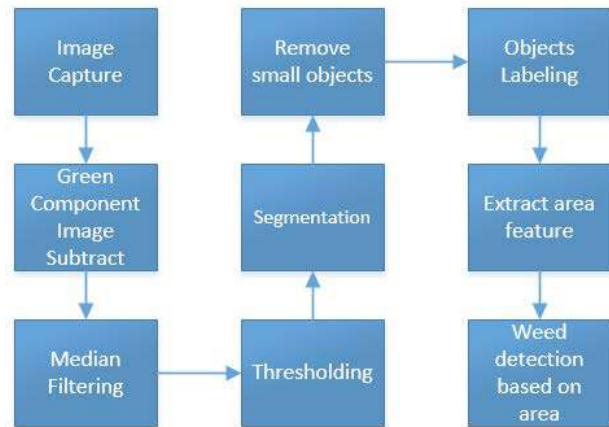


Figure 1 Vision system flow chart

2.1. Image acquisition

The digital images were captured in outdoor light conditions with perspective projection over crop (See Figure 2). Captures have 8MP resolution in RGB color space with JPG extension. The main idea of crop images acquisition is to avoid lighting and sharpness problems, therefore, color changes about vegetation are reduced. In this way, the accuracy of the first step of the plant classification algorithm increases provided that green color over objects is kept.



Figure 2 Input image

2.2. Green plant detection algorithm

A method for the segmentation of green plants and separation from the background is to use green component of RGB color model to get an image only with vegetation information. Previous studies have based their criteria for selection on an Index that stands out green component of source image; Excess Green Index [7, 8] and Normalized Difference Vegetation Index [9-11] are some methods that use this approach, however, they are aimed to perform on different sunlight and background conditions. In this project, the module for the weed remover robot will have a camera obscura and lamps in order to maintain uniform illumination. Then, it results appropriate to subtract green component from the original image.

At this point, source image is converted to grayscale intensity whereby the hue and saturation information is eliminated while retaining the luminance. All components in XY space corresponding to green value on image are subtracted from the corresponding element in grayscale array [1]. This method is one of the most practical ways of separating plants from soil.

$$I_{Plant}(x_{pixel}, y_{pixel}) = I_{Green_Source}(x_{pixel}, y_{pixel}, G) - I_{Gray}(x_{pixel}, y_{pixel}) \quad (1)$$

The response obtained using the method described above is shown in Figure 3:



Figure 3 Image subtracted

2.3. Median filtering

Median filtering is used for noise suppression in images subtracted, preserving edges whereby the relevant image information is conserved and tends to produce regions of constant or nearly constant intensity [12]. This filtering works using pixel values around 3-by-3 neighborhood mask. This window is moved over the points of an image, then, the value at the mask center is replaced with the calculation of the median from source image values within the window [13]. The output of this filter is shown in Figure 4.

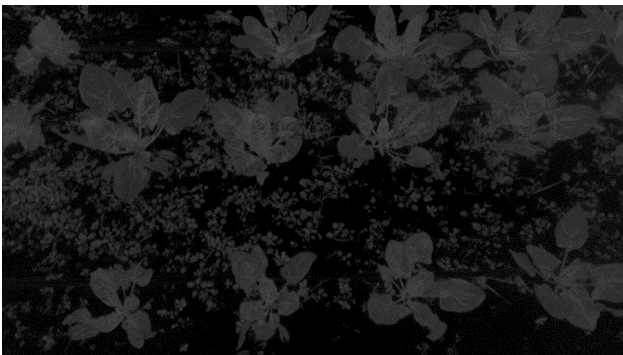


Figure 4 Median filtering

2.4. Feature extraction

The objects of interest result from the stage described above, therefore, it is necessary to segment the image in order to locate plants, assigning a label to each pixel and highlighting the similarity of the features used for detection of plants. In the present work, *color* and *area* serves as descriptors for a threshold classifier.

Threshold segmentation

Threshold segmentation method is appropriate because elements of interest have been highlighted above, having a clear difference of objects to be removed from the scene. Segmentation is accomplished by using the following expression [2]:

$$I_{bin}(x, y) = \begin{cases} 0, & I_{Median}(x, y) < t \\ 1, & I_{Median}(x, y) \geq t \end{cases} \quad (2)$$

The selection of *threshold* may not be obvious for human eye just to see the image. To deal with this, the parametric and unsupervised most famous and practical method is used. The *Otsu method* [14], estimates optimal threshold to segment through an exhaustive search of maximum variance between classes in gray levels. The selection of an appropriate threshold is carried out with image histogram, taking the value of the valley formed by two peaks, corresponding to the light and dark areas. Figure 5 shows the intensity level on the horizontal axis up to 50; however, the total value for the intensity level is 0-255.

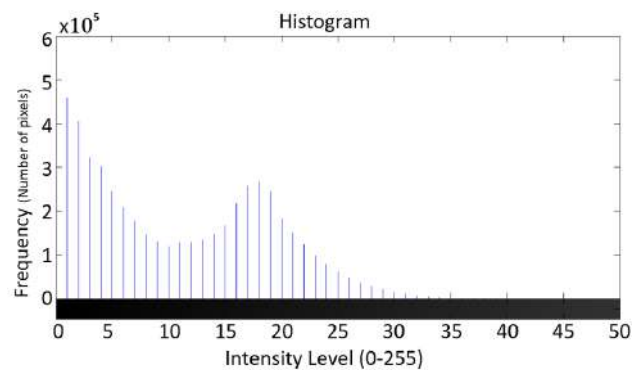


Figure 5 Image histogram

Then, the optimal value divides the two peaks in the graph above. Applying Otsu's method, the threshold found is 8. The results of the segmentation are shown in Figure 6.

2.5. Fill image holes

Because of feature extraction based on area, it is appropriate to fill in the holes in the image. Thus, evaluation in the next step is enhanced since compact objects are obtained. For this purpose, an algorithm based on morphological

reconstruction is used [15]. This method uses 4 or 8 connected neighborhood pixels to evaluate the resulting image [16]. The algorithm calculates a marked image stemming from source image borders using (3):

$$F_{mark}(x,y) = \begin{cases} 1 - I_{source}(x,y), & \\ (x,y \text{ is on the border of } I_{source}) & \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

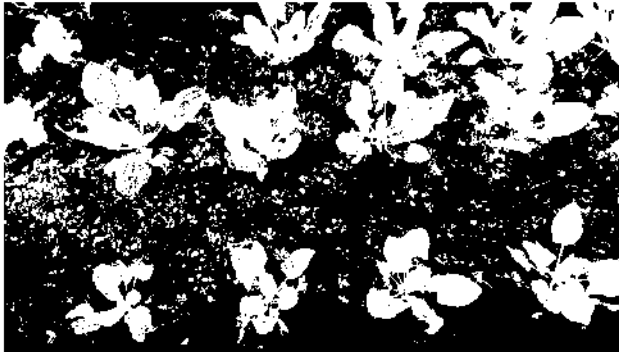


Figure 6 Threshold segmentation

Afterwards, the method iterates until it reaches all image positions (4), where B, a 3 x 3 matrix of ones, $H_0 = F$, and G, image to fill in the holes.

$$H_k = (H_{k-1} \otimes B) \cap G \quad (4)$$

Then, H_k , is a binary image with holes filled in. This result is shown in Figure 7.

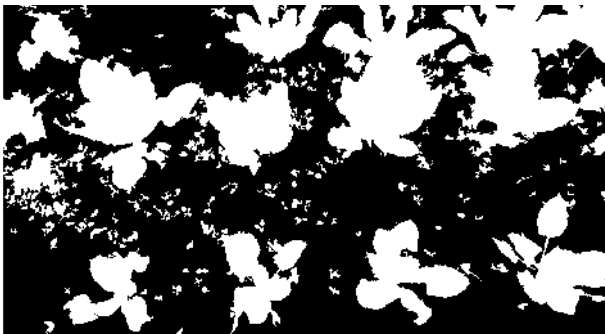


Figure 7 Image with holes filled

2.6. Labeling

Identifying objects in the scene, requires to label each element as a plant, getting a region description in order to extract features in the next step. Therefore, an algorithm based on connected components is used [17, 18]. The region labeling stage evaluates each pixel with a 4 neighbor-connectivity, using a heuristic stated on pixel values according to predecessor labels at north and west position.

2.7. Classification based on area

Once objects on the scene are labeled, the next step is to extract area features from each element to discriminate weed and crop. The algorithm presented, defines an area counting the number of pixels in the object region; then, the value is stored for all items. The elements are sorted according to the area values in descending order. When a difference with the next object evaluation is greater than 50%, the average of the previous elements is calculated. This value is the threshold for weed detection. The classification is shown in Figure 8.



Figure 8 Weeds detected

3. Results

The algorithm of weed detection system was tested using photos taken perpendicularly to crop lines, avoiding illumination disturbances in spinach and chard crops of Horticulture Technology, a dependency of Nueva Granada Military University Campus in Cajicá, Colombia. The images were labeled manually based on random behavior of weed and the expertise of crops manager, in order to compare and evaluate the performance of the proposed approach to weed detection.

The classes assigned to define the specificity and sensitivity are shown as follows:

- True positive (TP): Number of plants detected as weed correctly.
- True negative (TN): Number of plants detected as crop correctly.
- False positive (FP): Number of crop plants detected as weed.
- False negative (FN): Number of weed plants detected as crop.

The following images (see Figures 9, 10 and 11) and Tables 1, 2 and 3 report the results of the algorithm proposed. Each image is accompanied with a table that indicates sensitivity and specificity to group information about correct weed detection. Additionally, positive and negative values are calculated to highlight the percentage of true positive and true negative detections with respect to all classifications. The indices are calculated as follows:

- Sensitivity = $TP / (TP + FN)$
- Specificity = $TN / (FP + TN)$
- Positive predictive value (PPV) = $TP / (TP + FP)$
- Negative predictive value (NPV) = $TN / (FN + TN)$



Figure 9 Weeds plant detected in spinach crop 1



Figure 10 Weeds plant detected in spinach crop 2

Table 1 Sensitivity and specificity values spinach crop 1

Weed Plants (Confirmed manually)			
	Condition Positive	Condition Negative	
Plants detected as Weed	True Positive (TP) 20	False Positive (FP) 4	PPV 0.83
Plants detected as crop	False Negative (FN) 2	True Negative (TN) 9	NPV 0.818
	Sensitivity 0.909	Specificity 0.692	

Table 2 Sensitivity and specificity values spinach crop 2

Weed Plants (Confirmed manually)			
	Condition Positive	Condition Negative	
Plants detected as Weed	True Positive (TP) 80	False Positive (FP) 1	PPV 0.98
Plants detected as crop	False Negative (FN) 5	True Negative (TN) 6	NPV 0.545
	Sensitivity 0.94	Specificity 0.857	

Table 3 Sensitivity and specificity values chard crop

Weed Plants (Confirmed manually)			
	Condition Positive	Condition Negative	
Plants detected as Weed	True Positive (TP) 146	False Positive (FP) 1	PPV 0.99
Plants detected as crop	False Negative (FN) 12	True Negative (TN) 13	NPV 0.52
	Sensitivity 0.924	Specificity 0.92	



Figure 11 Weeds plant detected in chard crop

The sensitivity values shown in this section are around 0.90, which indicate a good performance of the proposed algorithm for detecting plants as weed correctly. Similarly, the specificity values near to 1 represent the ability of the system to detect correctly plants as crop. The positive predicted values are greater than 80 percent, whereby, most of the cases identified as weed are true positive. In contrast, the negative predicted values are near to 0.7, which means a 30 percent of false negatives obtained when weed size is greater than or near the crop size; this behavior is presented in Figures 10 and 11.

4. Conclusion and future work

This paper has shown a practical way for weed detection using an area-based feature to discriminate from crop. This study was limited to periodical weed removal tasks, whereby weed size is smaller than the crop size to carry out the approach reported, achieving a high sensitivity, specificity, PPV and NPV values. The use of low level features as color and area results appropriate in periodical treatment conditions in contrast to use texture or shape features that could have similitude between weed and crop. The algorithm presented in this paper leads on unsupervised learning, hence, additional computational cost is reduced compared to the algorithms that require training to identify weed with these descriptors. Then, color and area approaches result effective to recognize weed, avoiding a large database processing to discriminate crop. Additionally, the obtained results are focused on crop size; therefore, weed features are not relevant, which provide versatility for the application in the identification of different vegetable crops. This feature is important due to wide variety of weed plants. At color detection stage, the histogram serves as a basis to validate binary segmentation, bearing out a way to separate plants from soil using green component, which is the main objective in image acquisition the plant colors preservation. It is also remarkable that environment light controlled proposed like a camera obscura in a weed remover robot will help in the elimination of significant changes to parameterize green plants from image source because of variation in illumination. Thereby, other algorithms to try outdoor conditions images are not necessary and the computational cost is also reduced.

The results of this study suggest that the proposed algorithm for weed detection using low level features

and thresholding classifier has a high performance and accuracy validated with sensitivity and specificity indices above 90%. Likewise, high PPV values mean that most of plants detected as weed are *True Positive*, fulfilling the main objective of the presented work.

The use of technological tools to support Colombian farmers helps to reduce the cost per hour of work required for inspecting weeds in crops; for instance, three workers take between 20 and 60 minutes for inspection of crop area defined by 2x45m.

Future work on the current topic is aimed to use crop rows to estimate a grid, in which weed detection includes plants position, offering to the system other descriptor to discriminate crops, due to random behavior of weeds and include shape features to complement classification stage, likewise, the implementation on NVIDIA JETSON TK1 using CUDA programming for online processing and finally, camera calibration respect robot module, in order to concede coordinates about weed plants position to removal effector, giving a relative displacement of robot because it never stops to perform the task.

5. Acknowledgements

This work is supported by the project ING-1758 namely "Sistema de detección de malezas, suelo compacto, suelo seco, plagas y piedras en los campos de cultivo Fase 1", funded by the research vice rector of the Nueva Granada Military University in Bogotá, Colombia.

6. References

1. M. Mustafa, A. Hussain, K. Ghazali and S. Riyadi, "Implementation of Image Processing Technique in Real Time Vision System for Automatic Weeding Strategy", in *IEEE International Symposium on Signal Processing and Information Technology*, Giza, Egypt, 2007, pp. 632-635.
2. A. Tellaache, G. Pajares, X. Burgos and A. Ribeiro, "A computer vision approach for weeds identification through Support Vector Machines", *Applied Soft Computing*, vol. 11, pp. 908-915, 2011.
3. A. Tellaache, X. Burgos, G. Pajares and A. Ribeiro, "A vision-based method for weeds identification through the Bayesian decision theory", *Pattern Recognition*, vol. 41, pp. 521-530, 2008.
4. H. Liu, S. Lee and C. Saunders, "Development of a machine vision system for weed detection during both of off-season and in-season in broadacre no-tillage cropping lands", *American Journal of Agricultural and Biological Sciences*, vol. 9, pp. 174-193, 2014.
5. A. Shinde and M. Shukla, "Crop detection by machine vision for weed management", *International Journal of Advances in Engineering & Technology*, vol. 7, pp. 818-826, 2014.
6. H. Jeon, L. Tian and H. Zhu, "Robust Crop and Weed Segmentation under Uncontrolled Outdoor

- Illumination", *Sensors*, vol. 11, pp. 6270-6283, 2011.
7. D. Wobbecke, G. Meyer, K. Von and D. Mortensen, "Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions", *Transactions of the ASAE*, vol. 38, pp. 259-269, 1995.
 8. A. Muangkasem, S. Thainimit, R. Keinprasit and T. Isshiki, "Weed Detection over Between-Row of Sugarcane Fields Using Machine Vision with Shadow Robustness Technique for Variable Rate Herbicide Applicator", *Energy Research Journal*, vol. 1, no. 2, pp. 141-145, 2010.
 9. D. Wobbecke, G. Meyer, K. Von and D. Mortensen, "Plant species identification, size, and enumeration using machine vision techniques on near-binary images", in *SPIE Conference on Optics in Agriculture and Forestry*, Boston, USA, 1993, pp. 208-219.
 10. A. Perez, F. Lopez, J. Benlloch and S. Christensen, "Colour and shape analysis techniques for weed detection in cereal fields", *Computers and Electronics in Agriculture*, vol. 25, pp. 197-212, 2000.
 11. L. Wiles, "Software to quantify and map vegetative cover in fallow fields for weed management decisions", *Computers and electronics in agriculture*, vol. 78, pp. 106-115, 2011.
 12. I. Pitas and A. Venetsanopoulos, *Nonlinear digital filters: principles and applications*, 1st ed. New York, USA: Springer, 1990.
 13. B. Justusson, "Median Filtering: Statistical Properties", in *Two-Dimensional Digital Signal Processing II*, T. Huang (ed). Berlin, Germany: Springer, 2006, pp. 161-196.
 14. N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man , and Cybernetics*, vol. 9, no.1, pp. 62-66, 1979.
 15. P. Soille, *Morphological image analysis. Principles and applications*, 2nd ed. Berlin, Germany: Springer, 2004.
 16. J. Semmlow and B. Griffel, *Biosignal and medical image processing*, 3rd ed. Boca Raton, USA: CRC Press, 2014
 17. L. Shapiro and M. Haralick, *Computer and Robot Vision*, 1st ed. Boston, USA: Addison-Wesley, 1992.
 18. R. Kumar, K. Ramareddy and B. Rao, "A Simple Region Descriptor based on Object Area per Scan Line", *International Journal of Computer Applications*, vol. 3, no. 7, pp. 24-27, 2010.