

Identification of weed species, kikuyu grass (pennisetums) in crops using machine vision

BEATRIZ NATHALIA SERRATO PANQUEBA, CARLOS ARTURO CASTILLO MEDINA

Electronic and Telecommunications Engineering

Universidad Católica de Colombia

Diagonal 46 A # 15 B-10

BOGOTÁ - COLOMBIA

bserrato@ucatolica.edu.co, cacm.castillo@gmail.com

Abstract: This article shows the results of applying image processing to estimate the presence of weed species Kikuyu grass in crops. A study was conducted in field (in site), within more than 30 locations in the area Cundinamarca - Boyaca in Colombia; an analysis of classical methods of image classification is performed to determine the best way to identify arvense species. It is concluded that at present, the benefits presented by analysis with neural networks, allow ensure a correct classification despite the difficulties that arise with environmental variables. This document is to disseminate the results obtained from the analysis of images taken in the field in order to estimate automatically the presence of weed species, Kikuyu grass, by using computer vision and traditional approaches of classification.

Key- words: machine vision, weed identification, classical classifier, neuronal networks, unwanted weed, weed in crops.

1 Introduction

Precision Agriculture is now a promising alternative to improve agricultural processes and maximize crop production. Farming system driven by information gathered from specific technological devices improve production processes by controlling crop-exogenous variables that guarantee maximum production with minimum environmental impact. It is also necessary to adjust parameters such as weed control rates, fertilizer use, specific watering patterns, and the proper types of pesticides and herbicides.

Colombia is a purely agricultural country; most crops have made production from manual labor which provides an opportunity for improvement to be competitive in world markets. In particular, the identification and removal of unwanted weeds in crops is done manually, thus creating an environment free of competencies (in crops) but at a high production cost. Additionally, eradication of weeds occurs from the use of chemicals that maintain the balance in production but these are also applied from manual techniques which creates an environment conducive to the future destruction of the soil.

2 Problem Formulation

The presence of weeds in crops is an important variable in the time to analyze the productivity of a given crop, since the management of these can prevent secondary impacts such as pests, diseases that bring consequences that can be detrimental to the profitability expected. In particular, unwanted herbs growing in crops must be treated in order to maintain a balance, because if they do not reach control measures, could lose about 40% of world agricultural production [1]. Chemical control can be done by applying pre-emergent herbicides immediately after sowing or at an early stage [2]; when these products are applied, the soil should have very good moisture content so that no erosion occurs which is a permanent damage to soils. This is how computer vision can be an option for the management of weed species, to process and analyze images taken on field [3], using classical vision algorithms and key machine learning techniques as a primary resource.

Considering the above, we define strategies to detect the presence of weed species in crops, in order that in the future a controlled and sectioned application, to eradicate weeds, is performed by reducing the environmental impact caused by the use of agricultural chemicals.

3 Problem Solution

In the process conducted it was included the images acquisition in field, which is pre-processing, segmentation and use of conventional techniques for the classification among the species. The description of the step by step of each of the stage developed during the research is detailed below:

3.1 Image Acquisition

Different crops were randomly chosen, in an early stage of sowing in order to ensure that the crop was not in a level of very advanced development since weed control is performed in initial stages of the crop. In particular, some images were taken of the crops such as corn, potatoes, beans, onions, strawberries, beets, cabbage and chard; characterized because its production process is carried out at ground level and currently have the presence of weed species in different proportions.

Specifically, 305 photographs were taken in the area of Cundinamarca - Boyaca in Colombia, ensuring that lighting levels were similar. The time slot intended for making images was early morning



Fig.1 Growing bean with presence of weed

It was intended that the images taken had presence of weed species very close to the crop so that it could conduct an initial recognition of the herbs that were in competition, and thus determine some factors which could become discriminating among species in the screening or classification stage.

3.2 Processing and segmentation

In this step, a sequence of morphological [4] transformations was prepared to allow the enlistment of the image, so that it has the characteristics required in the process of segmentation, intended to discriminate among weed species and cultivation of other elements that are present in the image; for this purpose tools offered by OpenCV [5] by Python were useful.

Initially, the images were analyzed on color spaces, RGB and HSV, to identify which of them provided much information as could extract only the areas of interest. Why it was determined that the HSV space, specifically the S channel provided a further definition of the boundaries of the areas of interest.

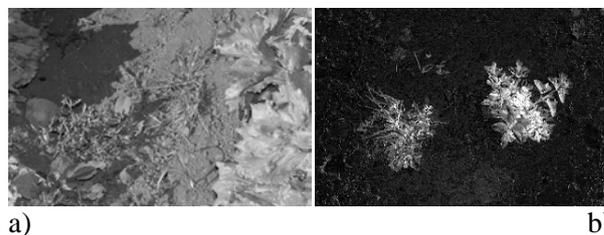


Fig.2 a). Lettuce crop, with presence of weed species in gray scale RGB color space. b) Strawberry crops with presence of weed species in the S channel of the space HSV

Figure 2 (a), evidences that visibly the differentiation between crop and area with presence of weed species is very low, since the gray levels obtained do not have a high degree of separation by the color in the image. Moreover in (b), the shadow that is generated in the presence of land has low saturation levels, which can differentiate very well between crops and planting land. This is taken as the basis for a segmentation process in which only areas of interest are available, which in our case is vegetative; masks were created to overlap the processed image with the original.

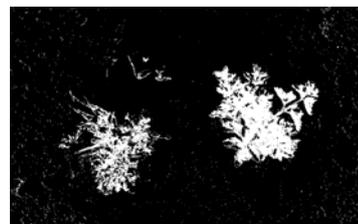


Fig.3 Segmentation of strawberry crop in the presence of weed species

In the segmentation process, they are obtained some regions with presence of vegetable species, but also the related areas are those where the light reflectance determines mathematical values to obtain the threshold values of the segmentation; making it necessary to define algorithms to remove unwanted pixels in the images in order to obtain a cleaner source of information. To do this a treatment operation is applied, such as erosion and dilation to maintain the zones with larger area and eliminate only those generating noise. After obtaining such clean image, it proceeds to mask the image in RGB

scale with the one segmented and thus the areas of interest are extracted.



Fig.4 Segmentation mask with the image in RGB color scale

3.3 Classification

After obtaining only the areas with vegetable presence in the picture, it carried out a random seed irrigation that allowed extracting color characteristics in different RGB channels to estimate values differentiating between crops and weeds species. 400 samples were selected as training data and 150 such as test data, all of them organized so they could be analyzed in a manner supervised by typical classification techniques.

Comparisons of different channels (RGB) where the existing relations between crop and weed species were performed. Particularly we analyzed the various options such as leaving fixed values that come from the R channel of arvense species and compare it with the data provided by the R, G and B channels of the crop. To differentiate between species four classifiers were implemented, hoping to get the lowest possible error in classification.

3.3.1 Classifier K-NN [6]

This classification method is supervised, nonparametric, which estimates the value of the probability density function of a particular item to belonging to or not to a defined class. In the process of learning any belongings it is not supposed since the distribution of the predictor variables is based on the supervised data. Data obtained from this classification process are those listed in Table 1.

Arvense Analyzed Channel	Wrong Predictions	Test error (%)
R	168	42
G	89	22
B	196	49

Table 1 Results of nearest neighbour classification

It is possible to identify that the comparison with the values provided by the G channel of weeds enables discriminate a clearer distinction between the other two options (A and B of arvense species); however, testing has an error of 22%. It is very important to

use other classifiers in order to find one that provides a lower percentage of error.

3.3.2 Bayesian classifier

This classifier analyzes the sample for discrimination, considering functions of lower cost and higher probability [7]. Considering Bayes' theorem is required a priori probability of each class to be analyzed and the probability of the attributes of each class, in order to calculate the posterior probability. The results obtained after applying this classifier are as show in Table 2.

Arvense Analyzed Channel	Wrong Predictions	Test error (%)
R	122	30
G	69	17
B	211	52

Table 2 Results of Bayesian classification

According to the values presented in the classification through this method it is again observable that channel which provides better discriminant relationship between crop and weed species is the G channel, resulting in an error test 17%; percent lower, than the one presented by the nearest neighbour classifier.

3.3.3 Support Vector Machines SVM

This classification methodology obtains information from the samples that are taken from different channels; and accompanied by labels that relate the classes belonging; thus building a predictive model that identifies the correspondence of entering new values to determine the class to which it belongs to [8]. The results obtained by this method of classification are shown in Table 3.

Arvense Analyzed Channel	Wrong Predictions	Test error (%)
R	156	39
G	70	17.5
B	204	51

Table 3 Results of classification by support vector machines

3.3.4 Artificial Neural Networks RNA

Artificial neural networks are systems enable the processing of structured information that are inspired by the behaviour of biological neural networks; neurons are organized by layers that are randomly assigned different weights and which are modified as the ratio of importance are related to the sample values and evolution of the algorithm [9]. These systems are adaptive learning by experience provided by training models.

It is of great importance to perform the validation process with the support of the training process [10], in order to detect undesirable developments in the learning stage; among them is the overtraining which affects the results due to excessive refinement in setting parameters for the network [11]. This problem occurs with increasing error in the curve validation and training decrease.

Arvenses Analyzed Channel	Wrong Predictions	Test error (%)
R	156	39
G	70	17.5
B	204	51

Table 4 Results of classification using artificial neural networks

This neural network was developed in several steps, in which it was possible to demonstrate the problems of overtraining and the problem of excessive neurons in the hidden layer as increasing neurons also increased training time and in turn increased the network error; if instead, very few neurons caused an imbalance in the weights, which did not allow solving the problem adequately. For this reason, we chose to have in the hidden layer 7 neurons with sigmoid activation function and re-propagation training algorithm. The minimal error got for the dataset was 17.5% compared to the evaluation values which place this method of classification as the one providing lower test error values.

4 Conclusions

Image processing is critical because when performing features extraction, data should also be provided in order to build a suitable dataset, since it is this information which assigns values of analysis to the classification systems; it is expected that the characteristics that predominate in one type of crop, also represents in another similarly on condition that everyone must be in an early stage.

Crops that are analyzed at a time of day that generate shadow of too high values, do not meet the conditions necessary for the algorithms, maintain a single decision-making pattern. However most of the samples analyzed, despite the differences in illumination, were not affected so that the algorithm defined initially fulfilled the specific features.

After analyzing the percentage results of test errors, it is concluded that the classifier that yielded better results of classification is by means of artificial

neural networks, with 7 neurons in the hidden layer with sigmoid activation function; however, the results are not too far from the results obtained by two classifiers tested in this research. Particularly the Bayesian classifier obtained values very close since the effectiveness was maintained thanks to the fulfilment of independence between the analyzed classes.

4.1 Future Works

This analysis is the initial stage of a proposal that allows the controlled application of herbicides. This requires refining image processing algorithms that allow fieldwork even though high lightness levels available, which in turn provide shadow in crops. Additionally, there is the interest to work alternative methodologies for detection of unwanted herbs in crops to supplement the detection performed by the computer vision system.

References:

- [1] E. C. Oerke, «Crop losses to pest,» Journal of agricultural science, vol. 144, pp. 31-43, 2006.
- [2] C. C. R. M. T. J. Perez L, «Control químico preemergente de la maleza de tomate en cáscara,» Interciencia, vol. 39, n° 6, pp. 422-427, 2014.
- [3] C.-J. Z. y. Y.-S. S. Guo-Quan Jiang, «A machine vision based crop rows detection for agricultural robots,» Wavelet analysis and pattern recognition, pp. 114-118, 2010.
- [4] Y. W. S. M. Weyrich M, «Quality assessment of row crop plants by using a machine vision system,» de 39th annual conference of the IEEE IECON2013, 2013.
- [5] U. F. F. F. Revelo D, «Reconstrucción 3D de escenas mediante un sistema de visión estéreo basado en extracción de características y desarrollado en OpenCV,» 2012.
- [6] S. S. Cost S, «A weighted nearest neighbour algorithm for learning with symbolic features,» Machine Learning, vol. 10, pp. 57-78.
- [7] G. D. G. M. Friedman N, «Bayesian network classifiers,» Machine learning, vol. 29, pp. 131-163.
- [8] S. D., «Comparación de máquinas de soporte vectorial vs. regresión logística. ¿Cuál es más recomendable para discriminar?,» Escuela de estadística, Universidad Nacional de Colombia, 2012.
- [9] B. C., «Neural networks for pattern recognition,» Oxford University Press, 1995.

- [10] C. S, «Sistema electrónico para la clasificación de datos de fragancia de café utilizando redes neuronales artificiales,» México, 2013, p. 41.
- [11] C. CYTED, Aplicaciones de las redes de neuronas en supervisión, diagnosos y control de procesos, Caracas: EQUINOCCIO, Ediciones de la Universidad Simón Bolívar, 1999.