

Performance Evaluation of Discriminant Analysis and Decision Tree, for Weed Classification of Potato Fields

¹Farshad Vesali, ²Masoud Gharibkhani and ³Mohmmad Hasan Komarizadeh

¹Department of Agriculture Machinery Engineering, University of Tehran, Karaj, Iran

²Department of Agricultural Machinery Engineering, Ege University, Bornova, Izmir, Turkey

³Department of Mechanic of Agricultural Machinery Engineering, University of Urmia, Urmia, Iran

Abstract: In present study we tried to recognizing weeds in potato fields to effective use from herbicides. As we know potato is one of the crops which is cultivated vastly all over the world and it is a major world food crop that is consumed by over one billion people world over, but it is treated by weed invade, because of row cropping system applied in potato tillage. Machine vision is used in this research for effective application of herbicides in field. About 300 color images from 3 potato farms of Qorveh city and 2 farms of Urmia University-Iran, was acquired. Images were acquired in different illumination condition from morning to evening in sunny and cloudy days. Because of overlap and shading of plants in farm condition it is hard to use morphologic parameters. In method used for classifying weeds and potato plants, primary color components of each plant were extracted and the relation between them was estimated for determining discriminant function and classifying plants using discrimination analysis. In addition the decision tree method was used to compare results with discriminant analysis. Three different classifications were applied: first, Classification was applied to discriminate potato plant from all other weeds (two groups), the rate of correct classification was 76.67% for discriminant analysis and 83.82% for decision tree; second classification was applied to discriminate potato plant from separate groups of each weed (6 groups), the rate of correct classification was 87%. And the third, Classification of potato plant versus weed species one by one. As the weeds were different, the results of classification were different in this composition. The decision tree in all conditions showed the better result than discriminant analysis.

Keywords: Color components, potato plant, three composition of classification, weed

INTRODUCTION

The potato (*Solanum tuberosum L.*) is a major world food crop. In world food production, potato is exceeded only by rice, wheat and maize. Potatoes are consumed by over one billion people world over; half of them are in the developing countries alone. Potato is the fifth farm product of the world and third of Iran (FAO, 2010). Eskandari *et al.* (2011) investigate some method like hand weeding; power machine weeding and weeding using herbicides on rice field and expressed that chemical treatment were the best treatments because in all sampling stages, the least dry weight of weed produced. This appropriate control has an important effect on yield while the most yield rate belonged to this treatment. Also they declared, if herbicides use just for weeds (not for soil or main plant) it will be most efficient. But weeds invading on potato farm, lead to increasingly use of selective and unselective herbicides which causes the land and environmental pollution. On the other hand, high price of selective herbicides makes them noneconomic for farmers. So if there is a machine that can spray the

herbicide directly on the weed, the used amount of it will reduce by 50 to 70%, which leads to money saving and reduction of environmental pollution. The objectives of this study were to represent a useful and instant method for segregating weeds from potato plant, using machine vision. The study were done to approach this aim is in two divisions: the first one is to recognize weed between planting row and second one is to recognize weed on planting row.

In the first division, all the plants between the planting row should be removed; no matter if they are weed or not. Bases on this principal Woebbeck *et al.* (1995) simply tried to separate plants from soil background, But Lee (1999) recognized tomato leaves from weeds and segregated them using some specifics of tomato plant and designed an herbicide sprayer. Recognizing weeds based on texture analysis is a powerful method to separate smooth surface from shaggy and non-uniform surface. Scarr (1998) used this method for separating onion, because the leaves of onion are strait up and it is not possible to use shape assessment methods. Using texture analysis method, the problems related to

shape assessment and overlapping leaves was almost solved. El-Faki *et al.* (2000), compared four index based on main color components of plants to evaluate the effective parameters on weed recognition. They surveyed the effect of some parameters like soil moisture, light intensity and image resolution on weed recognition accuracy. They also evaluated the efficient of this method in field condition. Astrand and Baerveldt (2003) used some combinations of color and shape features for sugar beet weed segmentation. They evaluated shape features single plants and showed that plant recognition based on color vision is feasible with three feature and a 5-nearest neighbor classifier. Color features could solely have up to 92% success rate in classification. This rate increased to 96% by adding two shape features. Yang *et al.* (2000) developed a data mining technique, decision-tree analysis and used it to classify multi-spectral data sets from 33 experimental plots containing various combinations of crop (corn or soybean) and weed species. Their results indicate that a reasonable degree of differentiation (0.85 ± 0.06) may be obtained for the most complex of the classification problems investigated, which was to classify the 33 plots into 11 plot categories. Also they reported that classification success may be mainly related to the input type and may or may not be related to the number of inputs (Yang *et al.*, 2000). Burgos-Artizzu *et al.* (2009), used Case-Based Reasoning (CBR) for choosing the best pattern for segregating new images based on previous segmenting experiments and were able to speedup weed recognition on the planting row suing this method, which showed increase in Pearson Correlation from 60.1 to 79.7%.

The main goal of this study is reducing usage of herbicides in potato fields by recognizing the weeds in potato fields and effective use from herbicides. To achieve this goal we used image processing method to detect the weeds from potato plant. After that the ability of decision trees and discriminant analysis for classifying weeds and potato plants, was evaluated.

MATERIALS AND METHODS

For image data base of this research, 300 color images from 3 potato farms of Qorveh city and 2 farms of Urmia University, was acquired in year 2010. The image acquisition system consisted of a digital camera (Sony CyberShot W200, Japon) and the image capture size was $2,048 \times 1,538$ pixel in the horizontal and vertical directions respectively. In most of these images we could see five species of most important weeds of potato fields: Convolvulus arvensis (Field Bindweed), Centaurea arvense, Cirsium arvense (Creeping Thistle), Chenopodium murale (salt-green or sowbane) and Chenopodium album (goosefoot). For algorithms

development, image processing tool from MATLAB version R2009a toolbox was used.

Classification of weeds is influenced by the weeding time of the farm. In farms that the growth of plant and weed are concurrent, the weeding process was done in first stages of plant growth and both plant and weed are small, the success probability of morphologic method is high, but for most of row crops like potato because of plant and weed leaves overlapping, it is almost impossible to apply morphological methods for classification. The overlapping of leaves makes it very hard to assign a specific and uniform shape for plant, in different images. There are some other factors like leaves deflexion or deformation of leaves under environmental condition like plant diseases or herbs invade and leaves shift by wind, that makes the morphological specifications assessment very hard. Though in this study color processing and texture recognition was evaluated and finally these traits were used for separating weeds from potato plant by means of a classifier.

Preprocessing operations: As images were acquired in field condition, it was necessary to do preprocessing operations on the images to reduce the effect of light variation on images. Therefore light intensity and contrast of images with low mean light intensity, were increased using coefficient of Gama Correction Function, so light variation in images were reduced. This effect was obvious in surveying some images before and after preprocessing operations.

Separating plant and soil background: At first it seems that because of the green color of plant, we can easily separate plant from field using valley of green component histogram for thresholding, but as the background (soil) color is not constant and there is always some differentiation in light intensity in different parts of image and between images, it is impossible to apply that method. So we used another method for this job. As the dominant color of the plant is green, the quantity of green component (G) should be more than other two components (R and B), which is explained by Eq. (1):

$$G > \frac{R+G}{2} \Rightarrow 2G - R - B > 0 \quad (1)$$

Using this equation and choosing zero for threshold quantity, it is possible to separate plant from the soil (Gonzalez and Woods, 2008). After applying Eq. (1) there were some noises remaining in images. To remove these noises, after converting the images to binary image, Opening and Closing function was applied to them and the as a result, complete segmentation of plant was obtained without any noise.

Table 1: Mean and standard deviation of three main components of potato plant and five types of weeds

	R		G		B	
	Mean	SD	Mean	SD	Mean	SD
Centaurea arvensis	136.73	44.03	166.42	45.39	135.05	53.08
Convolvulus arvensis	108.46	37.09	153.53	38.79	119.73	43.84
Cirsium arvense	96.12	35.58	142.07	35.73	97.18	38.76
Chenopodium murale	112.23	39.98	159.42	39.22	99.24	41.38
Chenopodium album	93.74	33.75	144.91	33.65	102.44	38.65
Potato	75.31	36.41	124.25	40.84	78.54	38.27
Weeds	105.18	39.19	151.62	38.33	106.44	43.19

Classifying potato plant and weeds: The classification between potato plant and weeds was done using two discriminant analysis methods and a decision tree method. Discriminant analysis uses training data to estimate the parameters of discriminant functions of the predictor variables. Discriminant functions determine boundaries in predictor space between various classes. The resulting classifier discriminates among the classes (the categorical levels of the response) based on the predictor data. So Discriminant Analysis (DA) is a technique used to build a predictive model of group membership based on observed characteristics of each case. Discriminant function is the linear or nonlinear combinations of the standardized independent variables, which yield the biggest mean differences between the groups. As mentioned before El-Faki *et al.* (2000) used color component combination, but by surveying color component indexes for potato plant and weeds, no useful difference was seen between them and because of negative quantities of some indexes, it led to wrong recognition. Though, three primary color components were used for discriminant analysis input. These values were chosen from 175 separated images of 300 acquisitions. As each of these components is obtained from a pixel, the quantity of input data for discriminant analysis was too much. There for, to reduce the quantity of imputed data and also to assimilating the effective chance of each type of weeds, the total amount of them for potato plant and weeds were reduced to 50,000; 10,000 data for potato and 8,000 data for each weed. Averages and standard deviation of this data for each group was showed in Table 1.

Discriminant analysis was used to determine the membership model of data. This procedure was done by means of Statistics Toolbox in MATLAB software.

The other method which was used to segregate potato plant and weeds is decision tree. A Decision Tree (DT) is a machine learning algorithm based on a sequential divide and conquers approach (Han and Kamber, 2001; Breiman *et al.*, 1984). The algorithm builds classification or

regression models in an unambiguous way by recursively partitioning data. It learns the predictor or response pattern from the training data and derives a series of decision rules to appropriately represent the pattern. Each rule divides the training data into subsets using a threshold. In most DT algorithms, the data is classified at a time into two subsets (binary tree). These decision rules are hierarchical and sequential in nature and can be presented in a flowchart like, top-down tree structure (Witten and Frank 2000; Liu and Paulsen, 2000). In this study, Univariate tree or binary tree was used in each group. In this type of decision tree which is considered as simple and quick decision trees, there is just one condition in each node. In both methods (DA and DT), 80% of data was chosen randomly for training and remained 20% data was used for testing the method.

In this study, three different composition of potato against weeds were considered as below:

- Classification of potato and weeds
- Classification of potato versus all weeds
- Classification of potato versus each weeds species one by one

In First composition, classification was done between 10,000 potato plants R, G, B (Red, Green, Blue) data and 40,000 data from all weed types as one group. In second composition classification was done between potato plant and weeds in six group, 10,000 data for potato and 8,000 for each kind of weeds. Finally potato plant was classified separately comparing with each of five weed types.

RESULTS AND DISCUSSION

For evaluating accuracy of separating plant from soil background with Eq. (1), 10 images were chosen randomly and the edge of plants were chosen and then separated accurately from soil background by hand using Adobe Photoshop CS4 software. The obtained image of this separation was subtracted from the image which was obtained of Eq. (1) in MATLAB. In ideal situation the result of subtraction should be equal to zero, this results to a completely black image. But it wouldn't be possible and there were some pixels remained. By dividing the number of these pixels to number of all pixels in image, the error percent of separated soil background, will be obtained. This procedure is shown in Fig. 1.

This work was done for other images to and the mean deviation of field wrong separation in these 10 images was 2.3%. Low error rate shows that combination of opening and closing operation in addition to preprocessing and Eq. (1) can separate soil background accurately. For comparison El-Faki *et al.* (2000) encounter a high error rate (about 15%) in separation soil

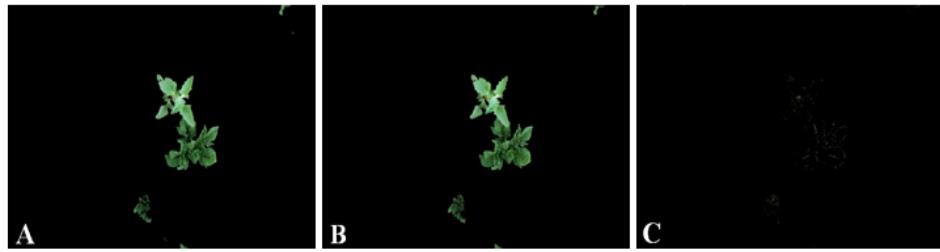


Fig. 1: Comparison between two segregation methods; (A) Segregation using equation No.2; (B) Segregation using Adobe Photoshop CS4 software; (C) Deviation between two methods

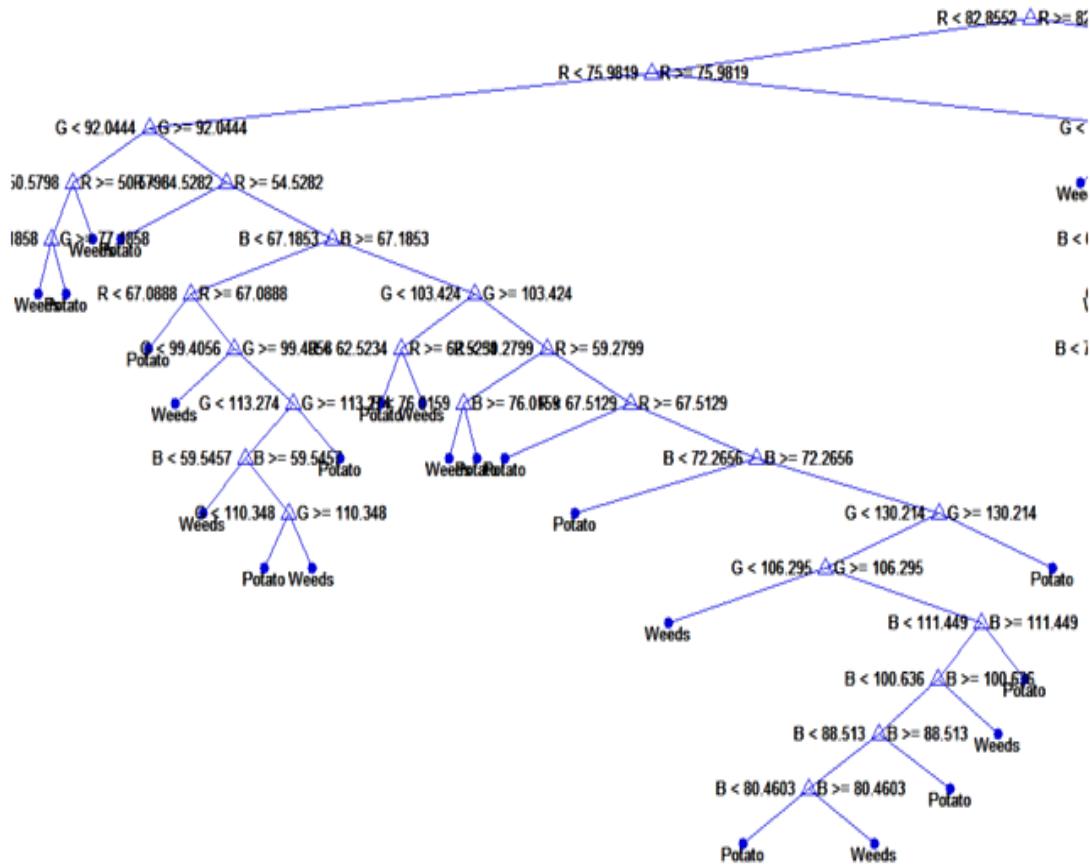


Fig. 2: Part of the decision tree that used to classify the weeds from potato

Table 2: The result of classifying potato and weeds in composition a

	Predicted group membership			
	Discriminant analysis		Decision tree	
	Potato%	Weeds%	Potato%	Weeds%
Potato	153	838	408	583
	15.44	84.56	41.17	58.83
Weeds	2995	1014	3608	401
	74.71	25.29	90.00	10.00
Correct classification rate	76.67		83.82	

background that maybe because of using Eq. (1) without any preprocessing before separation. Jafari *et al.* (2006) to reduce error in separation due to light change had to split their images to two group 2 (Images that were in light and Images that were in shadow) (Jafari *et al.*, 2006; Scar *et al.*, 1998).

It is obvious that when the potato plant was classified in condition that all weed types placed in one group (composition a), classification rate in both methods was higher than classifying potato plant and each kind of

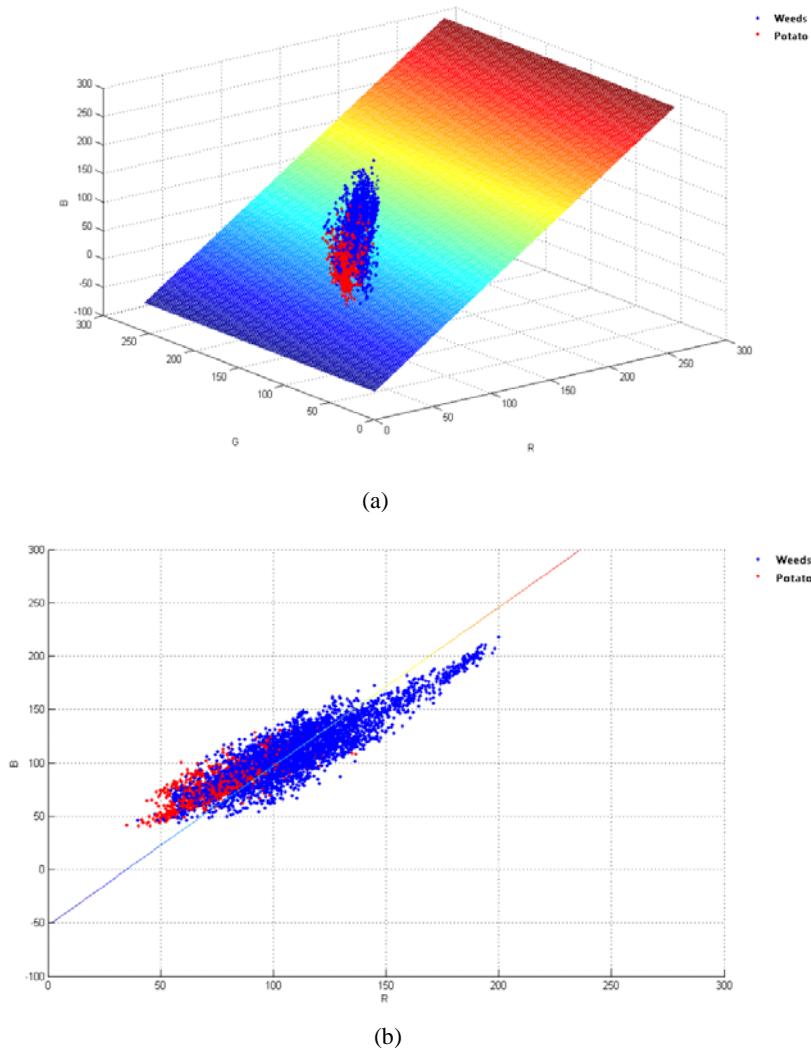


Fig. 3: 3D graph of discriminant function, classifying potato plant from weeds using R, G, B components amount. (a) 3D view; (b) View of R-B plate (blue dots stand for weeds and red dots stand for potato plant)

weeds separately (composition b). In composition c, depending on the type of weeds, different classification rate was assessed.

Table 2 shows the assumed membership by both methods in test stage in composition a.

The result of this comparison shows that in this status, the accuracy of decision tree in classifying potato plant and weeds, is higher than discriminant analysis. This decision tree has 1,200 sub branches and because of huge size of this tree, just part of it is placed in Fig. 2.

Discriminant analysis produces a discriminant function for each data set and the discriminant function for this case is as Eq. (2) the discriminant function is in fact assessed the constants of a plate that classified potato

plant from weeds based on previous training course of discriminant analysis.

$$DF_{(P\&W)} = -0.0530 \times R - 0.0015 \times G + 0.0014 \times B + 4.9774 \quad (2)$$

As mentioned before, the Eq. (2) is equation of a plate, that if value of each pixel put on this equation, the negative or positive result value of equation determines whether this pixel is related to potato or weed.

In Fig. 3, plate of discriminant function (2) with R, G and B of both groups is shown. Part (B) of this graph is produced by rotation of 3D graph and it helps to better understanding of how discriminant function classifies the

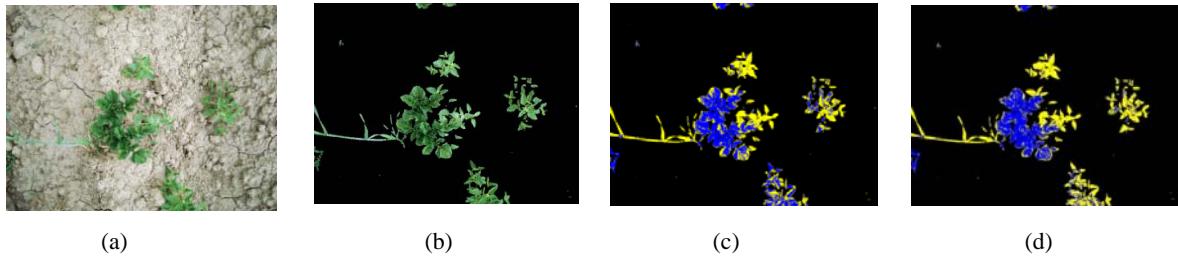


Fig. 4: Classification of weed from potato plant. (a) Primary image (b) The images after segmenting soil background. (c) Classified image weeds from potato with discriminant analysis (d) Classified image (weeds from potato) with decision tree (The yellow color is represent weeds and the blue represent potato plant, this colors chosen arbitrary)

Table 3: The result of classifying potato and weeds in composition b

Plant		Predicted group membership (%)					
		Potato	<i>Centaurea arvense</i>	<i>Convolvulus arvensis</i>	<i>Cirsium arvense</i>	<i>Chenopodium murale</i>	<i>Chenopodium album</i>
Potato	DT	60.04	1.92	5.54	11.60	6.86	14.03
	DA	65.97	0.81	4.04	10.16	6.99	12.03
<i>Centaurea arvense</i>	DT	1.49	76.58	8.09	4.98	6.22	2.61
	DA	2.04	82.45	8.08	4.36	6.71	0.11
<i>Convolvulus arvensis</i>	DT	6.65	9.53	53.07	11.42	5.52	13.80
	DA	8.36	13.99	47.08	6.93	5.15	18.5
<i>Cirsium arvense</i>	DT	14.02	5.17	12.30	45.01	8.61	14.88
	DA	20.8	7.54	10.88	18.99	18.30	23.50
<i>Chenopodium murale</i>	DT	5.58	3.24	3.37	7.79	74.16	5.84
	DA	5.99	10.48	1.89	8.73	67.81	5.11
<i>Chenopodium album</i>	DT	18.76	2.91	18.28	15.38	4.72	39.95
	DA	26.94	4.18	20.59	9.64	5.68	32.99
Total CCR	DT	58.12					
	DA	52.48					

data. Number of pixels for the weed was about 4 times more than pixels for potato plant so as it is shown in Fig. 2, the number of blue dots is more than red dots.

In composition b, that all weeds and potato plant are placed in separate classes, the decision tree showed the best accuracy. The CCR (Correct Classification Rate) was 58% comparing with discriminant analysis which showed (Correct Classification Rate) of 52% in this composition. So the high accuracy of decision tree is proofed. More details of classification with DA and DT in composition b are placed in Table 3.

Decision tree was able to classify all type of weeds from potato simultaneously and the results of it was better than discriminant analysis as showed in Table 3.

In composition C as the weeds were different, the results of classification were different. Though, by means of discriminant analysis the highest result was obtained by classifying potato plant and *Centaurea arvense* and this rate was 95% and the lowest result belonged to classification between potato plant and *Chenopodium album* by 73%. Decision tree resulted to better classification rates for each weed. The result of

classification using decision tree also showed that the highest rate belonged to separating potato plant from *Centaurea arvense* by 97% and the lowest rate belonged to classification between potato plant and *Chenopodium album* by 78%. Consequently, the decision tree showed better result for classification between potato plant and weeds in all cases, which shows the high capability of this method for classification.

In Fig. 4a, we can see two sample images of weed and potato plant. After applying Eq. (1) soil background was segmented (Fig. 4b) and the effect of discriminant function and decision tree that classified weeds from potato plant is shown in Fig. 4c and d. As a result weeds became yellow and potato plant became blue distinguishable.

CONCLUSION

Gama correction function was used to normalize the brightness and contrast of pictures. Afterward, for separating soil background, the Eq. (1) was used. This

equation separated the soil background and plants more accurately, when it used with opening and closing function. Ability of two methods to Classify Potato plants and five types of common potato weeds was evaluated in three compositions. In these three compositions, the result of decision tree was better than the discriminant analysis. For the example when classifying all types of weed as one group and potato as another group (Table 2), DT classification was 6% higher than DA (DT: 58.12%; DA: 52.48%). However it is important, which type of weed control device is considered to be used. For spot spraying with selective herbicides, where the main objective is to minimize the herbicide consumption, such classification rates are desirable. Furthermore when classification was done between one type of weed and the potato plant, CCR was high. In many cases, near a potato plant just one or two types of weed are exist in real condition, so this procedure would be able to classify weeds and potato accurately.

REFERENCES

- Astrand, B. and A. Baerveldt, 2003. An agricultural mobile robot with vision-based perception for mechanical weed control. *Auton. Robot.*, 13: 21-35.
- Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone, 1984. *Stone Classification and Regression Trees*. Belmont, Wadsworth, Cal.
- Burgos-Artizzu, X.P., A. Ribeiro, A. Tellaeche, G. Pajares and C. Fernández-Quintanilla, 2009. Improving weed pressure assessment using digital images from an experience-based reasoning approach. *Comput. Electr. Agric.*, 65: 1324-1333.
- El-Faki, M.S., N. Zhang and D.E. Peterson, 2000. Factors affecting color-based weed detection. *Trans. ASAE*, 43(2): 1001-1009.
- Eskandari, F.C., H. Bahrami and A. Asakereh, 2011. Evaluation of traditional, mechanical and chemical weed control methods in rice fields. *AJCS*, 5(8): 1007-1013.
- FAO, 2010. Statistic Food and Agriculture Organization of the United Nations. Retrieved from: <http://www.fao.org/corp/statistics/en/>.
- Gonzalez, R.C. and R.E. Woods, 2008. *Digital Image Processing*. Prentice Hall, New Jersey-Pearson.
- Han, J. and M. Kamber, 2001. *Data Mining: Concepts and Techniques*. Academic Press, San Diego, CA, USA.
- Jafari, A., S.S. Mohtasebi, H.E. Jahromi and M. Omid, 2006. Weed detection in sugar beet fields using machine vision. *Int. J. Agric. Biol.*, 8(5): 602-605.
- Lee, W.S., D.C. Slaughter and D. Geiles, 1999. Development of a machine vision system for weed control using precision chemical application. *Trans. ASAE*, 39(13): 220-227.
- Liu, J. and M.R. Paulsen, 2000. Corn whiteness measurement and classification using machine vision. *Trans. ASAE*, 43(6): 1669-1675.
- Scarr, M.R., C.C. Taylor and I.L. Dryden, 1998. Unsupervised texture segmentation using reversible jump Markov chain Monte Carlo methodology. University of Leeds, Statistics Tech. Report STAT.
- Witten, I.H. and E. Frank, 2000. *Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations*. Morgan Kaufmann Publishers, San Francisco, CA, USA.
- Woebbeck, D.M., G.E. Meyer, K. Von Bargen and D.A. Mortensen, 1995. Shape features for identifying young weeds using image analysis. *Trans. ASAE*, 38: 271-281.
- Yang, C.C., S.O. Prasher, J.A. Landry, H.S. Ramaswamy and A. DiTommaso, 2000. Application of artificial neural networks in image recognition and classification of crop and weeds. *Can. Agric. Eng.*, 42(3): 147-152.