Weed Classification Based on Haar Wavelet Transform via k-Nearest Neighbor (k-NN) for Real-Time Automatic Sprayer Control System

Irshad Ahmad
College of Computer & Information Technology, Al Jouf University, Sakaka, Al Jouf, Saudi Arabia
iakhalil2003@yahoo.com

Muhammad Hameed Siddiqi
Ubiquitous Computing Lab, Kyung Hee University, Korea
siddiqi@oslab.khu.ac.kr

Iram Fatima
Ubiquitous Computing Lab, Kyung Hee University, Korea
iram.fatima@oslab.khu.ac.kr

Sungyoung Lee
Ubiquius Computing Lab, Kyung Hee University, Korea
sylee@khu.ac.kr

Young-Koo Lee
Ubiquitous Computing Lab, Kyung Hee University, Korea
yklee@oslab.khu.ac.kr

ABSTRACT
Weeds cause harm to crops by competing for water, light, nutrients and space, reducing crop yields and inhibiting the efficiency of machinery. Although there is a large volume of work on this topic, previous studies have lacked accuracy and efficiency. In this paper an algorithm has been developed and analyzed for real-time specific weeds discrimination that is employed Haar Wavelet Transform (HWT). There are three stages of this paper, segmentation stage, feature extraction stage, and classification stage. Based on the proposed algorithm, 200 highest and informative coefficients are extracted at feature extraction stage, then at the classification stage, a well known classifier k-NN (k-Nearest Neighbor, for which k = 2) has been applied. The proposed method was tested on the database of 200 samples of each category. Furthermore, the result of our proposed method improves the performance by 7-10% as compared with some of the existing techniques that used histogram maxima with threshold, and angular cross sectional intensities. In conclusion, the proposed technique is more accurate at weed leaf classification than the existing techniques when images are captured by a CCD camera. The overall accuracy and efficiency utilizing the proposed method using haar wavelet transform is 94% and 40 ms respectively.

General Terms
Algorithms, Performance, Reliability, Experimentation

Keywords
Weed, HWT, KNN, Herbicides, Machine vision

1. INTRODUCTION
A weed is defined as ”any plant growing in a place at the wrong time and doing more harm than good”, which competes with a crop for water, light, nutrients and space and therefore reduces crop yields and affects the efficiency of machinery. Thus, those plants which interfere with human activity in crop and non-crop areas are considered weeds.

Mechanical cultivation is commonly used method for weed control, but the removal of specific weeds from a field is a primary limitation with this method. To solve this limitation, agricultural chemicals (herbicides and fertilizer products) are widely used. In fact, the success of agriculture is attributable to the effective use of chemicals/herbicides.

A real-time weed leaf classification system or machine vision system is important for this purpose. The author of [1] has developed and tested a machine-vision-system-guided precision sprayer, the accuracy of which was 75% in weed zone. Thus weed recognition component became a critical form operation and can significantly affect crop yield.

Improved efficiency in chemical application would also increase profitability in the agricultural production sector. Some spraying systems exist, such as Selectively Spraying, Spot Spraying or Intermittent Spraying, which are attached to the herbicide applicators. Thus, farmers need alternatives for weed control in order to reduce the usage of chemicals and cost of production, as well as prevent time wastage (the time consumed by hand hoeing).

Herbicides are vitally important in weed control and high crop yield; however these chemicals often produce harmful effects [2]. Normally herbicides are applied uniformly because weeds are highly aggregated and tend to occur in clumps and/ or patches and also remain relatively stable in size and location from year to year [3].

Herbicides are applied to whole fields of weeds like a blanket, without taking into consideration the types of weeds in the field [4]. Furthermore, in comparison to the uniform application method, the reduction of herbicide not only provides an economic advantage, but it is also environmentally friendly. The author of [5] has reported that when real-time weed sensing such as remote sensing and machine vision is applied, then there is a possibility that the amount of herbicides in a control patch sprayer will be reduced. Both the systems essentially require image acquisition and image processing techniques. The authors of [6, 7] have found that the size of the image varies by an order of megabytes for which the elapsed time is 0.34s to 7s. But the elapsed time depends on image resolution and the type of weed, for which the algorithm uses hardware configurations.

Although there, large volumes of methods are developed for a real-time automatic sprayer control system, previous studies have lacked accuracy and efficiency. The objective of this paper is to develop an accurate and an efficient algorithm that can classify broad and narrow weed leaves, and then to compare the accuracy with some of the existing algorithms. After comparison our approach is the accurate one among the existing algorithms, whose accuracy is 94%. The average efficiency of the developed algorithm was 40 m sec.

The rest of the paper is organized as follows. First, we discuss the related work about this filed. Then in next two sections, we will discuss the overview of our methods for the classification of the real-time specific weed leaves and the results and the comparison with the existing. Finally we provide concluding remarks about the accuracy and efficiency of our algorithm and some future work.

2. RELATED WORK

Many researchers have examined strategies to reduce production costs, protect the environment and control weeds with less herbicide. Several methods and algorithms have been developed for real-time selective herbicide systems by [4, 9] that can classify, localize and recognize weed leaves. The author of [10] used PDA as a processing device and measured the Weed Coverage Rate to discriminate between narrow and broad leaves. A system that could use spatial distribution information in real-time and apply only the necessary amounts of herbicide to the weed-infested area would be much more efficient and minimize environmental damage. Therefore, a high spatial resolution, real-time weed molestation detection system seems to be the solution for site-specific weed management.

The spray threshold is limited by the fraction of background pixels that are misclassified as plant material. If the spray threshold is set too close to the background misclassification rate, then herbicide will be wasted spraying background. Therefore, a larger misclassification rate limits the smallest plant that can be detected without targeting the background for spray [11]. The author of [12] proposed a local edge detection algorithm based on both gradient magnitude threshold and gray level analysis. He modified Hough Transform to recognize the edge points generated by the new edge detector which worked well for most images and was fast enough for real-time applications.

Many methods have been developed for real-time automatic sprayer control system. In [13], they proposed a method for such type of system and they got the average accuracy about 87% of classification. There are several steps that are involved in this method, so the processing speed of this method is slow, because at the filtering and feature extraction stage, the two common filters low pass filter and high pass filter are applied. Based on these filters they found different continuity measure (means the neighborhood pixel values can be measured by checking its continuity of 3, 5, 7 or 10 with different angel), so by this way the feature vectors are obtained. Finally the weed categories were classified by using the linear formula $y=mx+c$, so all the entire stages take much time to classify an image for which the image size was 240 x 320. Also some of their works [14], they found that SIFT method was the robust one amongst all of them in-term of accuracy, but it takes lots of time to classify an image, because SIFT uses Gaussian filter for preprocessing (convert RGB image to gray), and subjected to the filtering technique of Difference of Gaussian (DoG) that is similar to Laplace of Gaussian technique, in which the image first smoothes by convolution with the Gaussian kernel. Due to which the SIFT-C (Scale Invariant Feature Transform by Convolution) feature vector has been constructed from the key-descriptor (key-points), and put them into histogram bins that involved the determination of every similar magnitude direction. The transformations of SIFT-C on key-descriptor produced SIFT local maximum and minimum that were represented by 3D array, which were used for the classification of an object. So to process these entire stages and 3D array takes much time to classify weed categories for which the image size was 240 x 320. Although there a large volume of methods were developed for a real-time automatic sprayer control system, previous studies have lacked accuracy and efficiency.

In our previous work [15], the specific weeds have been classified based on multilevel wavelet decomposition based on only four different types of sub-wavelets, i.e. Daubechies (db4), Symlet (sym4), Biorthogonal (bior3.3) and Reverse Biorthogonal (rbio3.3). The objective of this research was to develop an algorithm that can recognize the absence of weeds and differentiate the presence of broad and narrow weed leaves, which could also be used to recognize the type of weeds that is to be killed by the appropriate herbicides using the automatic sprayer control system. This method has been developed for real-time automatic sprayer control system that has been developed by [1], which included a CCD camera, Central Processing Unit (CPU), decision box and two dc pumps for spraying, the setup of which is shown in Figure 1.

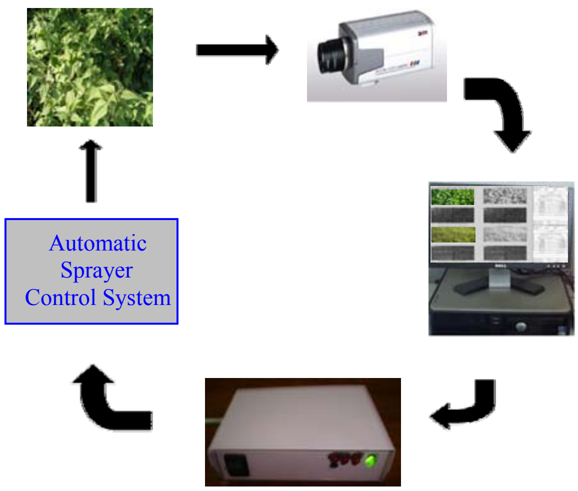

Figure 1. Automatic sprayer control system.

3. MATERIAL AND METHOD

In the proposed algorithm, haar wavelet decomposition is used. A set of 200 highest coefficients was extracted, that is used to classify the broad and marrow weed leaves. A well known classifier k-Nearest Neighbor (k-NN) is used to classify two weeds leaf categories. There are three stages in the proposed algorithm: pre-processing (image segmentation or image pruning), feature extraction, and classification.

3.1 Preprocessing (Image Segmentation)

The pre-processing stage is necessary to improve the quality of the images and made the feature extraction phase more reliable for the enhancement of broad and narrow weeds leaves. For this purpose, some common techniques like histogram equalization, morphological dilation were adopted for the removal of unnecessary and hidden details for fast and easy processing. The histogram equalization stage dealt with enhancing the contrast of suspicious areas in the image. In [16], the morphological dilation dealt with the image by using a defined structuring element (SE) that consisted of odd number of rows and columns. In the proposed algorithm the image preprocessing (image pruning) was applied for the purpose of removing the un-necessary information from the image. Hence the noise has been diminished from the images. The database of the images having resolution 240 pixel row and 320 pixel columns and almost 50% of the whole images compressed of the background with a lot of noise. In this stage a cropping operation is applied to remove the irrelevant and hidden details from the images and hence the images being enhanced.

3.2 Feature Extraction

In this stage the decomposition process has been applied by haar wavelet transform, for which the images were in gray scale. The reason for converting from RGB to gray scale was to improve the efficiency of the proposed algorithm. In one of previous work, we have extracted the feature by using watershed transformation [17].

An integer value communicated with each pixel of the image (as an index in an ordered table of colors) contained a matrix of integers. True color images often interacted with three matrices, for RGB coding. The wavelet decomposition could be interpreted as signal decomposition in a set of independent feature vector, spatially oriented frequency channels. Each vector consisted of sub-vectors as shown in Equation.1:

$$V_0^{2D} = V_0^{2D-1}, V_0^{2D-2}, V_0^{2D-3}, \ldots\ldots\ldots, V_0^{2D-n} \tag{1}$$

Where V represents the 2D feature vector. If we have a 2D image x which breaks up into orthogonal sub images corresponding to different visualization, so Equation.2 is generated as given below:

$$X = A_1 + D_1 \tag{2}$$

The Eq. 2 shows one level decomposition, where A_1 called approximation coefficient vector and D_1 is called detail coefficient vector. Detail coefficient vector further consist of three coefficient vectors, horizontal, vertical and diagonal coefficient vectors. In the proposed algorithm, during the feature extraction stage, the image has been decomposed up to one level of decomposition. Decomposition along three directions of detail spaces implies that in 2D as shown in Equation. 3.

$$X = A_1 + \left[(D_h)_1 + (D_v)_1 + (D_d)_1 \right] \tag{3}$$

Where D_h, D_v, and D_d are known as horizontal, vertical and diagonal coefficients at one level decomposition. It means that all the coefficient vectors are connected with each other like a chain. Figure. 2 shows the decomposition process and the chain model of all the training coefficients, when the image is decomposed up to multilevel. But in this algorithm the image is decomposed up to one level.

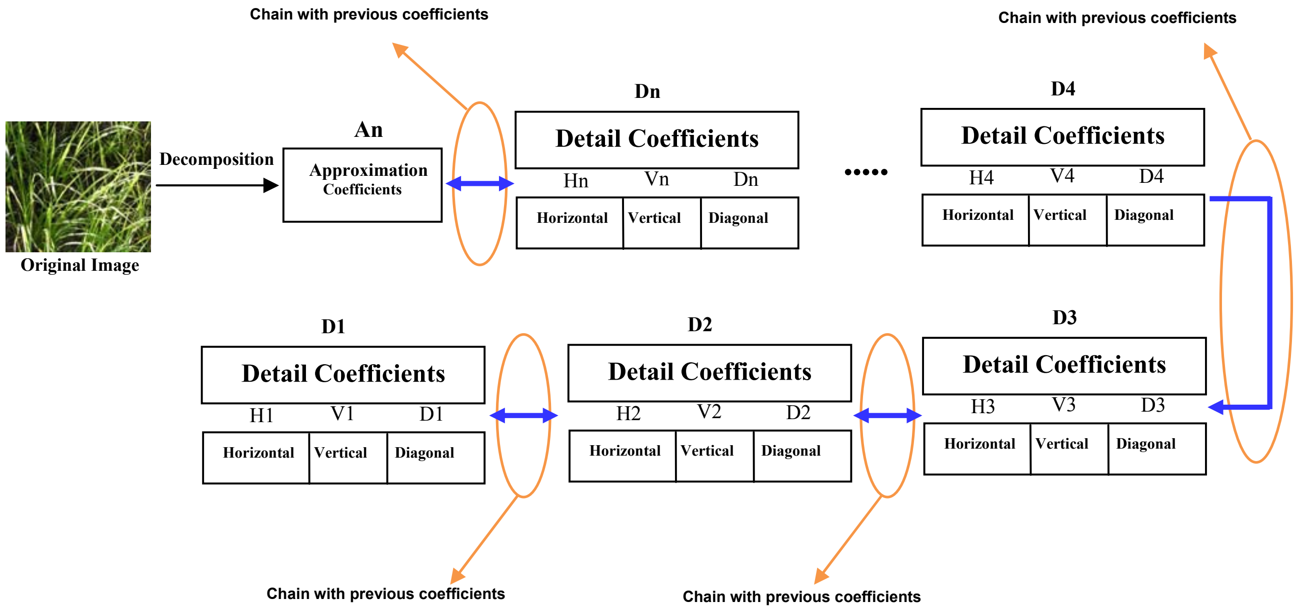

Figure 2. All the coefficients in the form of chain model.

During the decomposition process, the approximation coefficient vector and detail coefficient vectors are obtained. These vectors are obtained by convolving the low-pass filter for approximation coefficient vector and high-pass filter for detail coefficient vector. Figure. 3 shows the decomposition of the 2D image using haar wavelet transform.

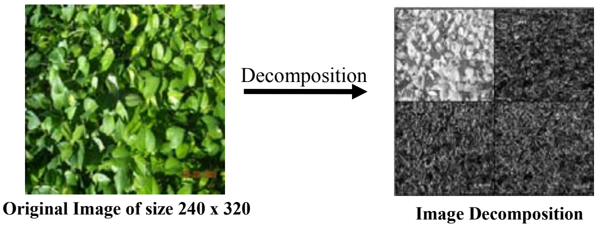

Figure 3. Broad weeds decomposition using Haar Wavelet Transform.

In this research, the haar wavelet transformation has been applied up to one level of decomposition. Using the wavelet decomposition the images were enhanced and some highest and informative features were extracted from these coefficients (approximation and detail). These features were produced during the process of the wavelet decomposition; thereby, making the classification of the real-time specific weeds possible.

3.3 Classification

During the feature extraction stage, 200 highest coefficients were selected randomly from the decomposition level and then the average of these coefficients is determined. Figure.4, 5, 6, and 7 show diagonal, horizontal, vertical, and approximation coefficients respectively.

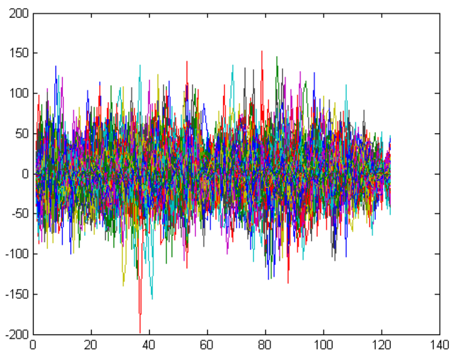

Figure 4. Diagonal coefficients.

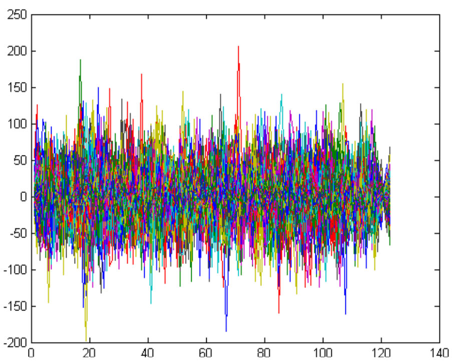

Figure 5. Horizontal coefficients.

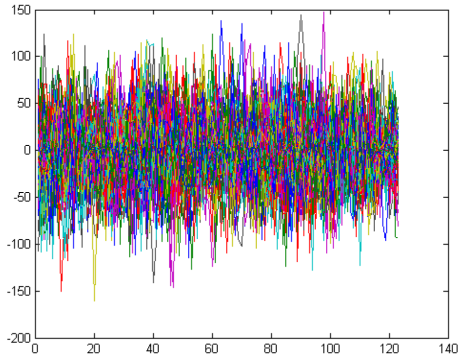

Figure 6. Vertical coefficients.

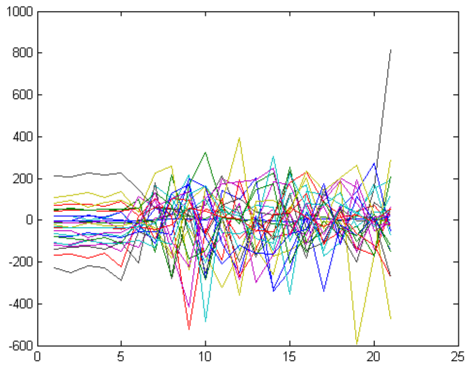

Figure 7. Approximation coefficients.

Feature vectors that have been obtained from these coefficients (extracted from the original image) are used to model and train the classifier. Figure 8 shows the extracted coefficients that were extracted from the first level of decomposition of the original image that is used for the purpose of modeling a training feature vector for each category.

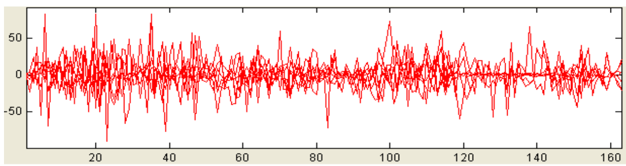

Figure 8. Representation of the extracted coefficients.

The class core vector for one image has been found by Equation.4.

$$V^{1}_{core} = V^{1}_{A} + V^{1}_{D_h} + V^{1}_{D_v} + V^{1}_{D_d} \quad (4)$$

For each class the core vector was the mean of 25% of the training vector (means 25% of each coefficient) that will be used as a training vector for classification, as shown in Equation. 5.

$$V^{j}_{train} = \frac{1}{N}\sum_{i=1}^{N} V^{i}_{core} \quad (5)$$

where N is the number of samples. A well known classifier k-Nearest Neighbor (k-NN) is used to design a classifier in order to recognize and differentiate between broad and narrow specific weed leaves for the real time automatic sprayer control system. In this algorithm there are two weed leaf categories, and each category has its own training vector, so the value of k is 2. For a new feature vector the distance between feature vector and the training vector is calculated using Equation. 6.

$$Dist = \sqrt{\sum_{i=1}^{k} (V^{i}_{train} - V^{i}_{test})^{2}} \quad (6)$$

The system automatically classified the feature vector in the class for which the distance was smallest.

4. RESULTS AND DISCUSSION

In this paper the proposed algorithm classifies a real-time specific weed leaf based on haar wavelet transform, and then compared with some of the existing algorithms. Among them our proposed algorithm produced the best result.

This algorithm enabled a real-time automatic sprayer control system to distinguish broad and narrow weed leaves according to their properties and then to classify them. By using this algorithm the right type of herbicides can be applied on the real-time specific weed leaves. To build the training feature vectors, 200 images are used for each class. The proposed algorithm used Haar Wavelet Transform (HWT) for the classification of real-time specific weed leaves. Table 1 shows the accuracy of classification using haar wavelet transform. The average accuracy of the proposed algorithm is 94%.

Table 1. Classification of different weeds using haar wavelet transform

Different types of weeds	Accuracy of classification (%)
Broad	94
Narrow	92
Little or no weeds	97

In the proposed algorithm, 200 highest coefficients were extracted at feature extraction stage, which helps the classifier at the classification stage. Figure 8 shows the classification of broad and narrow weed leaves using haar wavelet transform.

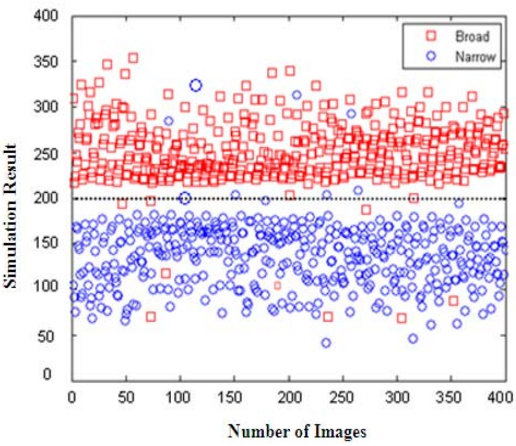

Figure 8. Results with the selection of 200 highest and informative coefficients

It is obvious from the Figure 8 that the classification based on the extracted coefficients (training feature vector) gives an accurate result using haar wavelet transform.

It is also noted from Figure 8 that the misclassification were found in images, which are consisted high intensity values, due to which they are overlapped on each other. As a result a narrow weed leaf is wrongly classified as a broad weed leaf category, and on the other hand the broad weed leaf is wrongly classified as a narrow weed leaf category. Further research is needed to overcome this problem. But still these findings indicate that the proposed feature extraction algorithm has great potential for feature vector representation of weed images in this classification task.

3.4 Comparison with existing algorithms

The developed algorithm was compared with some of the existing algorithms in term of accuracy. Among all of them the current algorithm is the most accurate of all tested. Figure 9 shows the comparison results between the developed algorithm and the existing algorithms using the database of broad and narrow weed leaves.

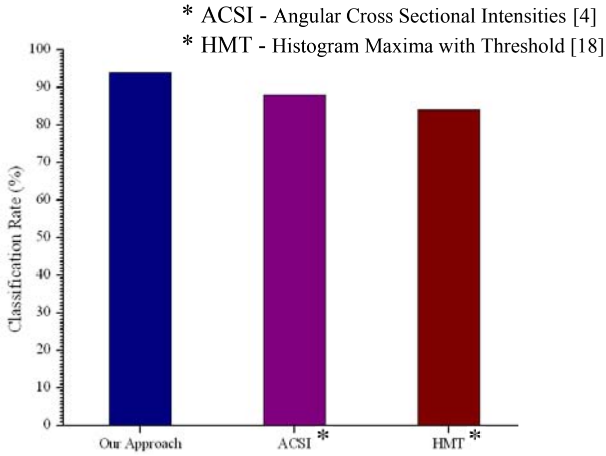

Figure 9. Comparison with existing algorithms

Figure 9 demonstrates that the decomposition based on haar wavelet transform is most accurate among the existing algorithms. This finding is significant when determining the most suitable algorithm for use in real-time specific leaves discrimination system, whose setup is shown in Figure 1.

5. CONCLUSION AND FUTURE WORK

Although there a large volume of methods were developed for a real-time automatic sprayer control system, previous studies have lacked accuracy and efficiency. In this paper, we proposed a complete approach for real-time specific weed leaf discrimination that is employing haar wavelet decomposition and then compared it with existing algorithms. The developed algorithm was successfully tested on the database of 200 samples (of each category) for weed leaf classification in order for selective spraying of herbicides using vision recognition system. This study has described a more effective pre-processing and post-processing technique for dealing with weed classification in order to improve the efficiency and accuracy of weeding strategies. The developed algorithm used Haar Wavelet Transform (HWT) that shows an effective and reliable classification of broad and narrow weed leaf category, for which the accuracy and efficiency were 94% and 40 ms respectively. In conclusion, the haar wavelet transform is the most accurate and efficient with images captured by a CCD camera.

Lighting conditions, wind and other natural environmental parameters degrade the performance of algorithm. Further research is needed to perform environmental adaptive weeds recognition to develop such classifier, which will classify the real-time specific weed leaves according to the environmental conditions.

6. ACKNOWLEDGMENT

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National Industry Promotion Agency) (NIPA-2010-(C1090-1021-0003)).

7. REFERENCES

[1] Tian, L., Reid, J. and Hummel, J. 1999. Development of a precision sprayer for site-specific weed management. *Trans. ASAE.* 42 (4), 893–900.

[2] Sunil, K. M., Weckler, P. R., Taylor, R. K., 2007. Effective spatial resolution for weed detection. *In: Annual International Meeting (Sponsored by (ASABE).* American Society of Agricultural and Biological Engineers.

[3] Stafford, J. V., Miller, P. C. H., 1993. Spatially selective application of herbicide to cereal crops. *Computer & Electronics in Agriculture* 9 (3), 217 – 229.

[4] Naeem, A. M., Ahmad, I., Islam, M., and Nawaz. S. 2007. Weed Classification using Angular Cross Sectional Intensities for Real-Time Selective Herbicide Applications. *Proceedings of the International Conference on Computing: Theory and Applications (ICCTA'07)*, 731-736.

[5] Tang, L., Tian, L. F., Steward, B. L., 1999. Machine vision-based highresolution weed mapping and patch-sprayer performance simulation. *Transactions of the ASAE* 108 (2), 317 – 318.

[6] Lee, W. S., Slaughter, D. C., Giles, D. K., 1999. The watershed transform: Definitions, algorithms and parallelization strategies. Precision Agriculture 1, 95 – 113.

[7] Søgaard, H. T., 2005. Weed Classification by Active Shape Models. *Biosyst Eng.* 91 (3), 271–281.

[8] Naeem, A. M., Ahmad, I., and Islam, M. 2007. Weed Classification using Two Dimensional Weed Coverage Rate (2D-WCR) for Real-Time Selective Herbicide Applications. *Proceedings of the 19th International Conference on Computer, Information and Systems Science and Engineering*, 335-339.

[9] Siddiqi, M. H., Ahmad, I. and Sulaiman, S. B. 2009. Edge Link Detector Based Weed Classifier. *Proceedings of the International Conference on Digital Image Processing (ICDIP'09)*, 255-259.

[10] Islam, M., Mustafa, M. M., Hussain, A., Hanif, M., and Mustafa, A. 2005. Real time machine vision for automatic weed-detection using weed coverage rate. *Malaysian Science and Technology Congress*, Malaysia.

[11] Charis, G. EEC206 Project Report, Friday 5PM, March 21st 2003. Color Segmentation of Plant and Soil. 3177 EUII.

[12] Shen, F. and Wang, H. 2001. A local edge detector used for finding corners. *Proceedings of the 3rd International Conference on Information, Communications and Signal Processing (ICICS'01)*, 65–71.

[13] Ghazali, K. H., Mustafa, M. M., and Hussain, A., 2008. Implementation of feature extraction technique from gray level co-occurrence matrix to classify narrow and broad weed in oil palm plantation. *Europ. J. Sci. Res.* 20 (1), 68-75.

[14] Ghazali, K.H., Mustafa, M.M., Hussain, A., 2008. Machine Vision System for Automatic Weeding Strategy using Image Processing Technique. *American-Eurasian Journal of Agriculture and Environmental Sciences* 3 (3), 451-458.

[15] Siddiqi, M. H., Sulaiman, S. B, Faye, I. and Ahmad, I. 2009. A Real Time Specific Weed Discrimination System Using Multi-Level Wavelet Decomposition. *J. Agric. Bio.* 11 (5), 559–565.

[16] Siddiqi, M. H., Ahmad, I., and Sulaiman, S. B. 2009. Weed recognition based on erosion and dilation segmentation algorithm. *Proceedings of the International Conference on Education Technology and Computers (ICETC'09)*, 224-229.

[17] Siddiqi, M. H., Ahmad, W., Ahmad, I., 2008. Weed classification using erosion and watershed segmentation algorithm. *Proceedings of the International Joint Conferences on Computer, Information, Systems Sciences, and Engineering. Springer Netherlands*, pp. 366 – 369.

[18] Ahmad, I., Naeem, A. M., Islam, M., and Nawaz. S. 2007. Weed Classification Using Histogram Maxima With Threshold For Selective Herbicide Applications. *Proceedings of the 19th International Conference on Computer, Information and Systems Science and Engineering* vol. 19, 331- 334.