American Journal of Agricultural and Biological Sciences 6 (4): 521-526, 2011
ISSN 1557-4989

Pecan Weevil Recognition Using Support Vector Machine Method

Ghulam Mubashar Hassan and Saleh Mufleh Al-Saqer
Department of Agricultural Engineering,
College of Food and Agriculture Sciences,
King Saud University, Riyadh, Kingdom of Saudi Arabia

Abstract: Problem statement: The Pecan weevil was considered as the most dangerous pest of Pecan fruits. The aim of this research is to evaluate Support Vector Machine method (SVM) for identifying Pecan Weevil among other insects. Eventually, this recognition system will serve in a wireless imaging network for monitoring Pecan Weevils. **Approach:** SVM has been evaluated using two different kernel functions i.e., Polynomial Function and Radial Basis Function. Database of 205 Pecan Weevils and 75 other insects which typically exist in pecan habitat has been used. Three sets of input data for SVM have been generated by two standard region-based recognition methods. These sets are comprised of output obtained by Zernike Moments, Regional Properties and combination of these two methods. For each kernel function, the system had been trained by 25, 50 and 75% of data and remaining ratio in each case has been used for testing. Each experiment is repeated ten times and average results are considered for comparisons and analysis. **Results:** The optimum recognition rate had been found when system is trained by 75% of data. The results are approximately similar when the input data is obtained by Regional Properties and combination of Regional Properties and Zernike Moments methods. The optimum results are obtained when input data has been obtained by Zernike Moments alone for lower values of sigma 'σ'. The proposed system is able to successfully recognize 99% of Pecan Weevil and 97% of the other insects using the radial basis function. The proposed system took approximately 31 sec for processing 75% of the data which include the time for training. The testing time is found to be 0.15 sec. **Conclusion:** Promising results can be obtained when input data is obtained by Zernike Moments and SVM is trained by RBF and 75% of data.

Key words: Support Vector Machine method (SVM), optimum recognition, Pecan Weevil, trapping, automated recognition system

INTRODUCTION

Pecan Weevil has been classified as one of the most destructive pests of pecans. It is also believed to be the most serious late-season pest because it attacks the nuts (Harris, 1979). This insect spends most of its life underground in soil and its life cycle lasts two to three years. The damage of the pecan nuts starts when the adult pecan weevil emerges from soil and attacks the nuts. It drills hole in the nut and feed itself. The female lays eggs in the nuts and it takes about 30 days for the larvae to be developed which feed inside the nut.

Currently, Pecan Weevil is controlled by detecting its emergence and subsequently applying insecticides. For efficient control of Pecan Weevil, one to four insecticide applications at precise time of emergence are required. The appropriate time of applying insecticide can be determined by inspection of dropped nuts and appearance of trapped Pecan Weevils. The most common method of detecting appearance of Pecan Weevil is by using traps. They are of different types such as wire cone trap, pyramid trap, circling trap. It is recommended to have 1-2 traps per tree while 3-5 trees in each orchard block need to have traps (Mizell, 2003). Traps need to be monitored after 2-3 days and should be positioned 1-2 weeks before nut reaches gel stage. This indicates that trap monitoring is a laborious and time consuming technique. The automation of this process would result in efficient and reliable control of Pecan Weevil.

An important component of any automated recognition system for monitoring would require recognition of the target insect. Few approaches have already been developed such as Surveillance of Fruit

Corresponding Author: Ghulam Mubashar Hassan, Department of Agricultural Engineering, College of Food and Agriculture Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia

Flies System for identification of Fruit Flies (Liu *et al*., 2009), Automated Bee Identification System for identification of Bees (Arbuckle *et al*., 2001), Species Identification Automated and Web Accessible System for identification of Spiders (Do *et al*., 1999), Red Palm Weevil Recognition System for identification of Red Palm Weevil (Al-Saqer and Hassan, 2011a; 2011b).

In 1995, Cortes and Vapnik introduced a machine learning algorithm based recognition method know as Support Vector Machine (SVM). It solves problems related to two groups classification. Several SVM based pattern recognition techniques are adopted for machine vision applications such as Face recognition problem (Qin and He, 2005), speech recognition method (Ganapathiraju *et al*., 2004), Simulated Annealing Algorithm for stored grain pest recognition (Yuxia and Hongtao, 2008).

For identification of Pecan Weevil, a recognition system was proposed by (Al-Saqer *et al*., 2011). which utilized several image processing techniques based on template matching (Ashaghathra, 2008). That study concluded that Regional Properties and Zernike Moments methods are sufficient to identify Pecan Weevil. The identification rates for Pecan Weevil and other insects were 90 and 93% by using Regional Properties method and 97 and 99% by using Zernike Moments methods respectively. The total processing time was found to be 0.44 sec. However, only 15% of the Pecan Weevil images were used for testing and the two recognition methods had to be used together (Al-Saqer *et al*., 2011).

The motivation for this research is to explore the abilities of SVM to classify Pecan Weevil among insects. Considering earlier studies of SVM, it is expected that the proposed research will yield in developing a robost and reliable recognition system. The proposed solution will be trained and tested by different sizes of randomly selected data.

MATERIALS AND METHODS

The Support Vector Machine (SVM) can be used as linear and non linear classifier. The fundamental idea of SVM is to classify a given data into multidimensional feature space. This method has been implemented in many machine vision applications recently and has comparable performance with other techniques (Cho *et al*., 2006). The hyperplane is used on the mapped feature space to classify the two possible distinguished classes. When optimizing the margin of the classified space, items on the margin are considered only. The items of classes close to the margin make a vector known as Support Vector (SV). The SV makes the optimization of the margin easier which results in making hyperplane for classification. The classification margin is determined by the position of hyperplane which will always correspond to any change or relocation of the SV. However, hyperplane will remain independent of any change occuring to any item other than SV.

Finding an optimal kernel function is a core task to use SVM. The kernel function is used to map the data to multidimensional feature space. Several kernel functions can be used for this purpose; however, Radial Basis Function (RBF) and Polynomial Function (PF) have been selected as they are reported to perform better for pattern recognition problems (Chin, 1998). Typically, Gaussian function is used as RBF. This function can be represented by Eq. 1 as:

$$K(x,y) = \exp\left(-\frac{\| x-y \|}{2\sigma^2}\right) \quad (1)$$

where, x and y are SV and targeted data point to be classified respectively whereas, ‘σ’ known as sigma represents the width of the Gaussian curve. With the increase in value of sigma ‘σ’, the decision surface becomes smoother and decision margin becomes more regular. The value of sigma ‘σ’ is also inversely proportional to the number of SVs (Buhmann, 2003).

The output of a PF, a directional function, is dependent on the direction of two vectors in low dimensional space and is mathematically presented in Eq. 2 as:

$$K(x,y) = (x, y+1)^d \quad (2)$$

where, x and y are SV and target data point respectively whereas d represents the degree of the polynomial. The scale of the output is reliant on the testing data point. For experiments, kernel functions of RBF and PF are selected. Different values of degree ‘d’ and sigma ‘σ’ are used for PF and RBF respectively. Initially, PF is used in which the degree ‘d’ is varied over a range from 2-6 with increment of 0.1 in each step. Afterwards, RBF was tested and the value of sigma ‘σ’ is varied from 1-150 with increment of 1.

Image acquisition: In this project, large and diverse numbers of insects were collected and their images were acquired for training purpose. The imaging system included Allied Vision Technologies (AVT) F-145B CCD black and white camera which is equipped with 1.45 megapixel 2/3 inch progressive CCD sensor. Original images were processed to convert into binary format and resized to 114×134 pixels. The processing was conducted using a computer 'Dell Optiples 780' having Core 2 Duo E8400 3.0 GHz processor of Intel with RAM of 4 GB. MATLAB® Version 7.9.0.529 (R2006) software was utilized for the simulations.

Data processing method: The inputs for SVM are the descriptors of each insect's image derived from two standard regional descriptor methods, i.e., Zernike Moments and Regional Properties. The two adopted methods are characterized with some advantages such as their invariance for rotation, high reliability, noise resilience and short processing time.

For Zernike Moments, an orthogonal set over interior of a circle would be formed by a set of complex polynomials. The origin is considered to be the center of the image and coordinates of pixel are mapped to the unit circle's range to calculate the values of Zernike Moments. Pixels outside the unit circle would not be included in the computation process. The orthogonality property guarantees that there is no redundancy or overlapping of information between moments with different orders and repetition. As a result, each moment will be distinctive descriptor for a given image (Kim and Kim, 2000). The output of processing the images with Zernike Moments at order 3, resulted in six unique values representing each image.

Furthermore, Regional Properties present each image by a set of values that have been derived from the regions of that image. In specific, the area of the region and lengths of major and minor axes of insects' image were measured and formed to characterize each individual image. In this process, the number of connected pixels in the region represent the first value (area). Whereas the second and third value (major axis and minor axis) are calculated as length (in pixels) and width (in pixel) of the elliptical considered region in the image respectively (Woods, 2002). These three values are used as inputs.

The database is comprised of 205 Pecan Weevils covering wide range of variations in terms of insect's size, age and gender. Furthermore, the database includes 75 other insects representing many types of insects normally present in the pecan habitat. The names of insects used in the experiment and their number of replicates are presented in Table 1. The images of these insects were acquired and then processed by Zernike Moments and Regional Properties methods.

The experiment of this study involved conducting three sets of tests in which the inputs were obtained by Zernike Moments, Regional Properties and combination of both methods. In addition, the database was divided into training and testing sets.

Table 1: Insects used for testing the algorithm

Insect	Number of replicates
Acrosterunum hilaris (Say)	5
Apis mellifera L	4
Brochymena guadripustulata (Fab)	5
Chortophaga viridifasciata (Deg)	4
Chrysobothris femorata (Oliv)	5
Coleoptera carabidae	1
Compsus auricephalus (Say)	3
Condoerus lividus (Deg)	5
Conotrachelus elegans (Say)	5
Cyrtepistomus castaneus (Roolofs)	2
Green June, Hemiptera Reduvlldae	1
Hyphantria Cunea (Drury)	4
Leptoglossus Opposites (Say)	2
Lepyronia Gibbosa (Ball)	5
Metealfa Pruinosa (Say)	4
Naupactus Leucoloma (Boh)	5
Pantomorus Pallidus (Horn)	5
Plathypena Scabra (Fab)	5
Tomostethus Multicinctus (Rohwer)	4

The selection of image for training purpose was done by selecting randomly a group of 25, 50 and 75% of entire database. The remaining portion of the data in each case was used for testing. For consistency and robustness of experiments, the selected data remained unchanged for entire set of test and each set of test was repeated 10 times. The average results were considered for analysis. The time consumed for training and processing an image is found to be dependent on the size of training data, imaging techniques, size of the image and SVs of training data.

Error: Error can be defined as the misclassification of either Pecan Weevils and other insects. Typically, this error of classification can be recognized as Type-I or Type-II errors. Type-I error occurs when any other insect is classified as Pecan Weevil, while Type-II error happens when Pecan Weevil is not correctly classified. Clearly, Type -II error is more crucial in this research.

RESULTS

The recognition system for identification of Pecan Weevil using SVM is evaluated by using two kernel functions i.e., PF and RBF. The input data to the SVM are derived by using two different image processing techniques i.e., Regional Properties and Zernike Moments. In the first experiment, PF is used and the results are presented in Fig. 1 and 2. Results presented in Fig. 1 refer to the case when input data is derived by Zernike Moments and training of the system is conducted by using 25 and 50% of the data.

Whereas, inputs derived by Regional Properties and combination of both image processing techniques did not provide adequate results for the same training ratios.

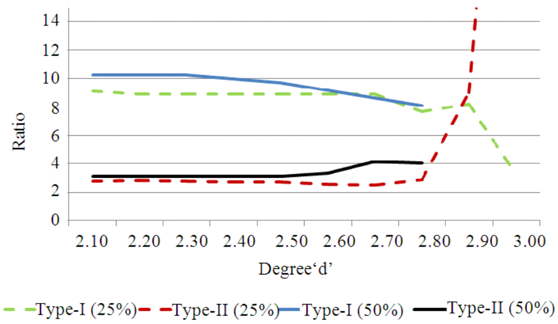

Fig. 1: Errors w.r.t. different values of degree 'd' for PF at 25 and 50% training data obtained by ZM

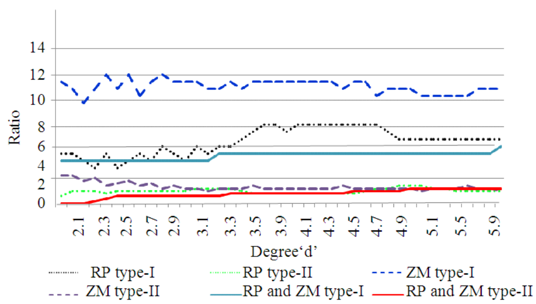

Fig. 2: Error rate w.r.t. different values of degree 'd' for PF at 75% training data

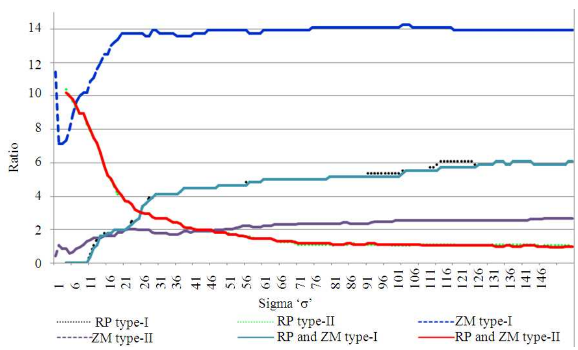

Fig. 3: Error rate w.r.t. different values of 'σ' for RBF at 25% training data

Table 2: Best recognition rates for PF with parameters

Training data ratio (%)	Input data method	Degree 'd'	Pecan weevil†	Other insects†
25	RP	-	-	-
50	RP	-	-	-
75	RP	2.4, 2.6	99.02	97.22
25	ZM	2.8	97.12	92.32
50	ZM	2.6	96.67	90.81
75	ZM	5.3	99.02	91.67
25	RP and ZM	-	-	-
50	RP and ZM	-	-	-
75	RP and ZM	2.1--2.3	100	96.67

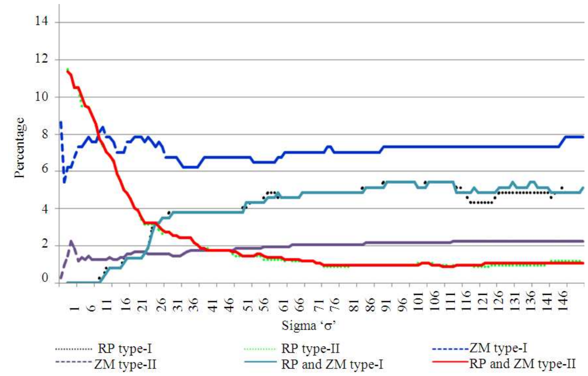

Fig. 4: Error rate w.r.t. different values of 'σ' for RBF at 50% training data

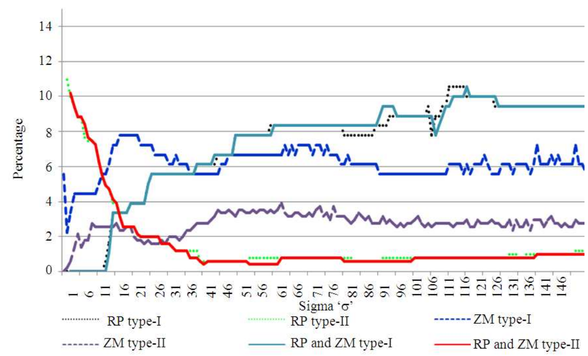

Fig. 5: Error rate w.r.t. different values of 'σ' for RBF at 75% training data

Table 3: Best recognition rates for RBF with parameters

Training data ratio (%)	Input data method	Sigma 'σ'	Pecan weevil†	Other insects†
25	RP	71--77	98.89	95.00
50	RP	119--123	93.91	99.12
75	RP	40--44	99.41	93.89
25	ZM	3	99.15	92.86
50	ZM	2	99.02	94.59
75	ZM	2	99.80	97.78
25	RP and ZM	71--79	98.82	95.00
50	RP and ZM	53	98.53	96.22
75	RP and ZM	41	99.61	93.89

Results shown in Fig. 2 are achieved when system is trained by 75% of data for all sets of inputs. The best recognition rates for all cases accompanied with their parameters' values are mentioned in Table 2.

The results of the recognition system, when RBF is used, are presented in Fig. 3-5 when system is trained by 25, 50 and 75% of data respectively. The best results for all cases are presented in Table 3 with their respective sigma 'σ' values. The time required for training and testing are mentioned in Table 4.

Table 4: Results for time for 75% training data

Kernel function	Input data method	Training time (sec)	Testing time (sec)
PF	RP	9.3312	0.0197
PF	ZM	32.3818	0.1469
PF	RP and ZM	35.1801	0.1665
RBF	RP	4.1578	0.0197
RBF	ZM	31.0252	0.1469
RBF	RP and ZM	35.1262	0.1665

DISCUSSION

The results presented in Fig. 1 mentions that input data obtained by Zernike Moments provide adequate results for small range of degree ‘d’ when PF is used in SVM. The range of degree ‘d’ providing adequate results is between 2 and 3 while training data used is 25 and 50%. It is observed that both types of errors are lower when training data used is 25% as compared to the errors when training data used is 50%.

The results in Fig. 2 presents that Type-I Error is always higher than Type-II Error for all cases when 75% data is used for training. The Type-II Error is mostly below 2% while Type-I Error is always above 2.5%. The Type-I and Type-II Errors are mostly highest when input data is obtained by Zernike Moments for different values of degree ‘d’. The Type-I Errors for the cases when input data is obtained using Regional Properties or combination of both Zernike Moments and Regional Properties is close to each other and always remain below 7%.

For the RBF, it is noticed in Fig. 3 that for smaller values of sigma ‘σ’, the error rates are low when input data is obtained by Zernike Moments while input data obtained by Regional Properties and combination of both Regional Properties and Zernike Moments have high error rates for low values of sigma ‘σ’. As the value of sigma ‘σ’ increases the error rates increases for the case when input data is obtained by Zernike Moments. Type-II Errors decreases with the increase in sigma ‘σ’ for cases when input data is obtained by Regional Properties and combination of both Regional Properties and Zernike Moments while opposite behavior is observed for Type-I Errors.

The results for 50% of training data mentions in Fig. 5 that both types of errors follow the same pattern as depicted in Fig. 4 for the cases when input data was obtained by Regional Properties and combination of both regional properties and Zernike Moments. The error rates remain consistent and their fluctuation remains below 2% for the case when input data was obtained by the Zernike Moments while the lowest errors are found at lower values of sigma ‘σ’. Similar observations are noticed when system is trained by 75% of the data as shown in Fig. 5.

Comparing the three graphs mentioned in Fig. 3-5, it can be concluded that performance of the system improves with the increase in training data. Both types of errors are low at small values of sigma and follow the same pattern for cases when inputs are obtained by Zernike Moments. Whereas for the other cases, when inputs are obtained by regional properties and combination of both regional properties and Zernike moments, both types of errors have opposite trends and intersect each other between sigma ‘σ’ values of 15 and 30.

The values of degree 'd' and recognition rates of best results for each case when system is trained using PF mentions in Table 2 that adequate results are not always obtained and few cases did not provide adequate results for any value of degree ‘d’. For some experiments, the best results occurred and repeated for different values of degree ‘d’. The adequate results are always obtained for the case study when system is trained by 75% of the data. The best recognition results are obtained for low values of degree ‘d’ of PF for all cases except when system is trained by 75% of data and inputs are obtained by Zernike Moments. Overall, the highest recognition rates using PF are obtained when system is trained with 75% of data and inputs are obtained by combination of both Regional Properties and Zernike Moments. These recognition rates are recorded for the range of degree ‘d’ for PF i.e., 2.1-2.3.

On the other hand, system trained using RBF provides adequate results for all the cases as mentioned in Table 3. Recognition system show the tendency of higher recognition rates for low values of sigma 'σ' when inputs are obtained by Zernike Moments. Moreover, the performance of system improves as the ratio of training data is increased. However, recognition rates has not improved with the increase of training ratio for the cases when the inputs are obtained by Regional Properties and combination of both Regional Properties and Zernike Moments. In general, the highest recognition rates are recorded when input data is obtained by Zernike Moments and system is trained with 75% of data with sigma ‘σ’ value of 2.

After analyzing all the results, the best results are obtained for the case when system is trained using RBF having sigma ‘σ’ value of 2 when system is trained by 75% of data and inputs are obtained using Zernike Moments only. These results are 99 and 97% for recognizing Pecan Weevil and other insects respectively. At these settings, the recorded time for processing and training is approximately 31 sec while testing time for an image is 0.15 sec. These promising results encourage the adoptation of proposed system as an alternative for the earlier proposed template matching based system (Ashaghathra, 2008).

CONCLUSION

This study concluded that SVM method is a reliable method for the recognition of Pecan Weevil. The descriptors derived by Zernike Moments at order 3 and Regional Properties were proven to be simple and unique representatives of a given insect image. PF and RBF kernel functions have been tested individually and have produced some significant recognition rates for both Zernike Moments and Regional Properties methods. Furthermore, recognition rates when using inputs from Zernike Moments provide better results as compared to the cases when using inputs from Regional Properties or combination of both Regional Properties and Zernike Moments. The higher recognition rates are obtained when system is trained using RBF. The proposed system is able to successfully recognize 99% Pecan Weevil and 97% of the other insects when sigma 'σ' is taken as 2 and inputs are derived by Zernike Moments. This system took about 31 sec for processing and training 75% of the data while the testing time for an image is found to be 0.15 sec.

ACKNOWLEDGEMENT

This project was supported by Research Center of College of Food and Agriculture Sciences, Deanship of Scientific Research, King Saud University.

REFERENCES

Al-Saqer, S.M. and G.M. Hassan, 2011a, Red palm weevil (*Rynchophorus ferrugineous*, *Olivier*) recognition by image processing techniques. Am. J. Agric. Biol. Sci., 6: 365-376. DOI: 10.3844/ajabssp.2011.365.376

Al-Saqer, S.M. and G.M. Hassan, 2011b. Artificial neural networks based red palm weevil (*Rynchophorus ferrugineous*, *olivier*) recognition system. Am. J. Agric. Biol. Sci., 6: 356-364. DOI: 10.3844/ajabssp.2011.356.364

Al-Saqer, S.M., P. Weckler, J. Solie, M. Stone and A. Wayadande, 2011. Identification of pecan weevils through image processing. Am. J. Agric. Biol. Sci., 6: 69-79. DOI: 10.3844/ajabssp.2011.69.79

Arbuckle, T., S. Schroder, V. Steinhage and D. Wittmann, 2001. Biodiversity informatics in action: Identification and monitoring of bee species using ABIS. Proceedings of the 15th International Symposium on Informatics for Environmental Protection, (ISIEP'01), Zurich, pp: 425-430.

Ashaghathra, S.M., 2008. Identification of Pecan Weevils through Image Processing. 1st Edn., ProQuest, ISBN: 0549560947, pp: 148.

Buhmann, M.D., 2003. Radial Basis Functions: Theory and Implementations. 1st Edn., Cambridge University Press, New York, ISBN: 0521633389, pp: 259.

Chin, K.K., 1998. Support Vector Machines applied to Speech Pattern Classification. University of Cambridge.

Cho, M.Y., T.F. Lee, S.W. Gau and C.N. Shih, 2006. Power Transformer Fault Diagnosis Using Support Vector Machines and Artificial Neural Networks with Clonal Selection Algorithms Optimization. Knowl.-Based Intell. Inform. Eng. Syst., 4251: 179-186. DOI: 10.1007/11892960_22

Do, M.T., J.M. Harp and K.C. Norris, 1999. A test of a pattern recognition system for identification of Spiders. Bull. Entomol. Res., 89: 217-224. DOI: 10.1017/S0007485399000334

Ganapathiraju, A., J.E. Hamaker and J. Picone, 2004. Applications of support vector machines to speech recognition. IEEE Trans. Signal Proc., 52: 2348-2355. DOI: 10.1109/TSP.2004.831018

Woods, R.E., 2002. Digital Image Processing. 2nd Edn., Printice Hall, Upper Saddle River, NJ., ISBN: 0201180758, pp: 793.

Harris, M.K., 1979. Pecan Weevil Distribution on Pecan Across the Pecan Belt. 1st Edn., Texas A and M University, USA., pp: 11.

Kim, W.Y. and Y.S. Kim, 2000. A region-based shape descriptor using zernike moments. Signal Process.: Image Commun., 16: 95-102. DOI: 10.1016/S0923-5965(00)00019-9

Liu, Y., J. Zhang, M. Richards, B. Pham and P. Roe *et al*., 2009. Towards continuous surveillance of fruit flies using sensor networks and machine vision. Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing, Sep. 24-26, IEEE Xplore Press, Beijing, China, Sep., pp: 1-5. DOI: 10.1109/WICOM.2009.5303034

Mizell, R.F., 2003. Traps for Monitoring plum curculio and pecan weevils.

Qin, J. and Z.S. He, 2005. A SVM face recognition method based on Gabor-featured key points. Proceedings of the International Conference on Machine Learning and Cybernetics, Aug. 18-21, IEEE Xplore Press, Guangzhou, China, pp: 5144-5149. DOI: 10.1109/ICMLC.2005.1527850

Yuxia, H. and Z. Hongtao, 2008. Recognition of the stored-grain pests based on simulated annealing algorithm and support vector machine. Trans. Chinese Soc. Agric. Mach.