A Robust Recognition System for Pecan Weevil using Artificial Neural Networks

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: Pecan Weevil is a widely found pest among pecan trees and these pests are known to cause significant damage to the pecan trees resulting in enormous annual losses to pecan growers. Traditional identification techniques for pecan weevil include traps with pheromones to detect the infestation of these pests. However, these traditional methods require expensive labor hours to set-up the traps and their monitoring. These techniques are also unreliable for early detection of pecan weevil infestation. Early detection of these pests is essential in minimizing the potential losses to the pecan trees. **Approach:** In this study, we develop a neural network-based identification system for pecan weevils. The neural networks require 3-9 image descriptors as input for successful recognition of pecan weevil. The nine image descriptors originate from standard image processing techniques such as Regional Properties (RP) and Zernike Moments (ZM). For training purposes, a comprehensive database was assembled comprising of 205 images of pecan weevil and 75 other insects commonly found in the same habitat. The networks were trained by two algorithms and several training ratios were studied to investigate the efficacy and robustness of the developed neural networks. **Results:** The neural networks developed in this study are capable of 100% recognition of pecan weevil as well as 100% recognition of other insects in the database. These recognition rates were achieved by using 75% of the data for training and using the Scaled Conjugate Gradient (SCG) algorithm and nine image descriptors as input. The average training times for these networks with the SCG algorithm was only 2-4 sec. and the testing time for a single image was only 0.16 sec. **Conclusion:** The neural network-based pecan weevil identification system developed in this study provides a reliable and robust method to identify pecan weevils and the proposed system should prove useful in designing an automated, wireless sensor network for detecting pecan weevil in the field.

Key words: Pecan weevil, automated recognition system, artificial neural networks, image processing techniques, insect recognition, image descriptors, regional properties, zernike moments, conjugate gradient algorithm

INTRODUCTION

Pecan weevil is recognized as one of the most destructive pests infesting pecans. It is also believed to be the most serious late-season pest because it attacks the nuts (Harris, 1979). Pecan weevil spends most of its life underground in the soil and its life cycle lasts from two to three years. The infestation of this pest starts when the adult pecan weevil emerges from the soil and attacks the nuts. The weevil punctures the nuts, feeds on the nut and the female lays eggs in the nuts. It takes about 30 days for the larvae to be developed, which feed inside the nut.

Current detection methods of pecan weevil involve identifying its emergence and then applying pesticides to control the infestation. Multiple pesticide applications may be required for effective control of these pests (Mulder, 2004). The time at which pesticides are applied are recognized by inspecting the dropped nuts for the appearance of pecan weevils. These pests are also detected by using traps with pheromones. Different types of traps are available- wire cone trap, pyramid trap, circling trap. Typically, 1-2 traps are used for each tree and in each orchard block, traps are set-up on about 3-5 trees (Mizell, 2003). Traps are required to be monitored over several weeks.

As evident from the above, detecting pecan weevil by using traps is a labor-intensive and expensive process. The automation of this process would result in reliable and efficient control of pecan weevil infestation. The first step in automating this process is to develop a reliable identification/recognition system for the insects. Similar efforts have been made in the literature to develop recognition systems for other

insects. For example, Liu *et al.* (2009) developed an identification system for fruit flies, Arbuckle *et al.* (2001) developed a system for identifying bees and Do *et al.* (1999) developed an identification system for spiders. Recently, a recognition system has also been developed for red palm weevil (Al-Saqer and Hassan, 2011a; 2011b).

A particularly attractive method to develop a recognition system is to use an image processing-based identification system that can then be integrated in a wireless sensor network for practical field applications. An identification system for pecan weevils was recently proposed by Al-Saqer *et al.* (2010) that utilized image processing techniques based on the template matching method (Ashaghathra, 2008). In that study, it was shown that regional properties and Zernike moments were sufficient to successfully recognize the pecan weevil. 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.*, 2010). In another study (Al-Saqer and Hassan, 2011c), a system based on the Support Vector Machine (SVM) method was used for recognition of pecan weevil and promising results were obtained. However, there is clearly a need to develop more reliable, predictive methods to identify pecan weevil. Towards this end, in this study, an identification system is developed that can successfully recognize pecan weevils among a host of other insects found in the same habitat. In particular, artificial neural networks are developed that utilize very few, easily available descriptors as inputs. The descriptors are calculated based on two standard image processing techniques. The artificial neural networks developed in this study are capable of high recognition rates for pecan weevil based on only limited, easily available information from the images of pecan weevils.

MATERIALS AND METHODS

Image acquisition: For the purposes of collecting images of insects, a wide range of samples of pecan weevil and other insects found in the same habitat were collected. The details of other insects' images used in this study are listed in Table 1. The imaging system included a CCD black and white camera from Allied Vision Technologies with model number F-145B, which is equipped with 1.45 megapixel 2/3 inch progressive CCD sensor. The raw images from the camera were processed to convert them into binary format and resized to 114×134 pixels. The processing was conducted using a computer Dell Optiplex 780 equipped with an Intel Core 2 Duo E8400 3.0 GHz processor and 4 GB of RAM. MATLAB® Version 7.9.0.529 (R2006) software was utilized for the simulations.

Table 1: Other insects used in the database for developing ANN for pecan weevil recognition system

Scientific name (Family/Order)	No. 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

Data processing method: Image processing techniques were used to obtain the descriptors that will serve as inputs for developing the ANN. The two standard techniques used were the Regional Properties and Zernike Moments. In the Regional Properties method, three descriptors derived from the binary image of the insects were obtained. The descriptors were the lengths of major and minor axes of the region and area in an image. The value of the area is determined by counting the number of pixels connected with each other in the image. Similarly, the lengths of major and minor axes are obtained by counting the pixels in the major and minor axes of the elliptical region in the image (Gonzalez and Woods, 2002). The lengths of major and minor axes and the normalized value of area were used as inputs to the ANN.

Zernike Moments was the second image processing technique used to obtain descriptors for the insects' images. The method requires the introduction of a set of complex polynomials which formulate an orthogonal set over the interior of an object's circle. The center of circle is considered as the origin and pixel coordinates are mapped to a unit circle. Any pixel found outside the circle is not considered in computation of Zernike Moments. Due to orthogonal properties, overlapping and repetition of information among moments with different orders is avoided. Hence, the representation of an image is unique and independent for each moment (Kim and Kim, 2000). Zernike Moments of third order were used in this study and the resultant six values of the moments were used as inputs to the ANN. These two image processing techniques were utilized in image recognition applications due to their rotational invariance, expression efficiency and noise robustness.

Artificial neural networks: The development of ANN or Neural Network (NN) is inspired by the working of human brain. This data modeling technique models complex relationships between input and output and has been used for several pattern recognition applications (Bishop, 1995) and is capable of solving highly non-linear problems. A single neuron in ANN can be represented mathematically as shown in Eq. 1:

$$y = f(\sum_{i=j}^{n} w_i x_i + b) \tag{1}$$

where, x is input, w is the weight and b is bias of a neuron. The output y is dependent on inputs, their weights, bias and its transfer function, f, which is generally a sigmoidal function.

Neural networks represent an attractive method for pattern recognition problems since the multilayer networks can be used to describe arbitrarily complex decision surfaces by using the complex network architecture contained in them. The network architecture is stated in terms of the number of hidden layers and neurons for a typical feed-forward multilayer perceptron based on the back-propagation algorithm (Hagan *et al.*, 2002). The network architecture contains information about the complexity of the neural network and the number of parameters that have been used for its design. Although finding the network architecture is a trial and error procedure and is problem dependent, recently certain simpler guidelines have been presented to decide the number of hidden neurons in a network layer. Xu and Chen (2008) reported that the optimal number of neurons 'n' in a hidden layer is dependent on the dimension of the input 'd' and number of training pairs 'N' for small or medium size dataset i.e., n = N/d

if its value is below or close to 30; else $n = \left(\dfrac{N}{d \log N}\right)^{1/2}$.

They categorized the training dataset to be medium or small if training pairs are less than 5000. It is also reported that their method does not consider the problem of local minima (Xu and Chen, 2008). In this study, we have tested the method developed by Xu and Chen (2008) for the number of hidden neurons in single hidden-layer neural networks. To avoid local minima, the neural networks developed in this study were trained 20 times by using advanced, gradient-based search algorithms in combination with the back-propagation updates of the network weights and biases. In particular, the Scaled Conjugate Gradient Algorithm (SCG) and Conjugate Gradient with Powell/Beale Restart Algorithm (CGB) were selected since they have been found useful in pattern recognition problems (Al-Saqer and Hassan, 2011b; Johansson *et al.*, 1991;

Moller, 1993; Powell, 1977). At the end of simulations, the best 50 and 100 trials out of the 200 networks generated were selected for additional analysis. In all cases, an early stopping criterion with validation check was used to prevent over-fitting of the data by the neural network (Hagan *et al.*, 2002). This method stops the training and the subsequent over-fitting of the network when the error on a validation set increases continuously. The method ensures that the network parameters are small in magnitude and thus, avoids over-fitting. All the transfer functions used in the network were sigmoidal functions and a single hidden layer was used in all cases.

The number of images used for training and testing were 280, which included 205 images of pecan weevil and 75 images of other insects that are normally found in the habitat of pecan trees. Three different training ratios of 25, 50 and 75% were used for training the neural networks while the remaining data were used for testing the network. Further, the training set was randomly divided for 10 times in each of the 20 trials. In this manner, a total of 200 networks were generated. Three different scenarios were considered for the network inputs. In particular, the inputs to ANN were provided from the descriptors obtained by only the Zernike Moments (ZM), descriptors obtained by only the Regional Properties (RP) and a combination of both RP and ZM (RPZM). Thus, the number of inputs in the networks were 3, 6 or a total of 9 image descriptors depending on the methods included in the image analysis.

RESULTS

Artificial neural networks were developed in this study to identify pecan weevil among other insects. The feed-forward neural networks developed used up to nine standard image descriptors as inputs. These nine descriptors included three descriptors from region properties and the remaining six descriptors were derived from the Zernike moments of order 3, as described earlier. Several different networks were designed that used variable number of inputs. In particular, three types/levels of inputs were considered for the network: (1) Three descriptors from regional properties were used as input, (2) Six descriptors from Zernike moments were used as network inputs and (3) Nine descriptors from both regional properties and Zernike moments were used as inputs.

In each of the above cases, three different training ratios were used to test the network's performance and robustness. In particular, 25, 50 and 75% of the data were used for training purposes and in each case two different training algorithms (SCG and CGB) were used to test the sensitivity of neural networks to the training algorithms used.

Table 2: Summary results for recognition of pecan weevil with different techniques

| Technique | Recognition rate* | | Time (sec)** | | Source |
	Pecan weevil	Other insects	Training	Testing	
Template matching	99	97	-	0.44	Al-Saqer *et al*. (2010)
SVM	99	97	31	0.15	Al-Saqer and Hassan (2011c)
ANN	100	100	38.5	0.16	This study

*; Recognition rates are in percentages; **; Times include time for image processing + neural network

Table 3: Recognition results for Pecan Weevil Using CGB algorithm: average of the best 50 networks of the 200 runs

| Training data (%) | Inputs to the ANN | Avg. training time (sec) | Recognition rates | |
			Pecan Weevil	Other Insects
25%	RP	0.48	96	94
	ZM	0.50	99	94
	RPZM	0.54	100	97
50%	RP	0.66	96	96
	ZM	0.58	99	95
	RPZM	0.76	100	99
75%	RP	0.84	99	96
	ZM	0.57	99	97
	RPZM	0.87	100	99

Table 4: Recognition results for pecan weevil using CGB algorithm: average of the best 100 networks of the 200 runs

| Training data (%) | Inputs to the ANN | Avg. training time (sec) | Recognition rates | |
			Pecan Weevil	Other Insects
25%	RP	0.43	94	96
	ZM	0.44	97	94
	RPZM	0.49	98	95
50%	RP	0.58	94	96
	ZM	0.51	97	94
	RPZM	0.65	98	97
75%	RP	0.70	97	96
	ZM	0.57	99	94
	RPZM	0.76	99	96

Table 5: Recognition results for pecan weevil using SCG algorithm: average of the best 50 networks of the 200 runs

| Training data (%) | Inputs to the ANN | Avg. training time (sec) | Recognition rates | |
			Pecan Weevil	Other Insects
25%	RP	2.40	99	93
	ZM	2.42	100	96
	RPZM	2.19	100	98
50%	RP	3.92	99	97
	ZM	3.65	100	97
	RPZM	3.16	100	100
75%	RP	3.83	99	100
	ZM	4.01	100	100
	RPZM	3.58	100	100

Table 2 lists the summary results obtained with the neural networks in this study. The table also compares the results obtained in this study with our earlier work where template matching (Al-Saqer *et al*., 2010) and support vector machine method (Al-Saqer and Hassan, 2011c) were used to identify pecan weevil. As evident from Table 2, the current study provides the best results in terms of recognition of pecan weevil. Note that the networks developed in this study require only three-nine image descriptors for reliable identification of pecan weevil. Thus, the networks developed in the current study are ideally suited as part of an image recognition system for field applications. Note that the support vector machine-based identification system developed earlier (Al-Saqer and Hassan, 2011c) exhibits similar performance than the neural networks developed in this study, as shown in Table 2. However, the recognition rates are slightly higher for the neural network than the support vector machine method. Table 2 also lists the execution times for each of the methods. Note that these times include both the image processing and neural network execution times.

In each case, the neural networks were trained 20 times and the training data was randomly selected 10 times. Thus, the best network performance of the 200 trials as well as the average of the best 50 and 100 trials (in terms of recognition rates) were analyzed to investigate the sensitivity of the results to the training algorithm used as well as the effect of training ratios and the number of descriptors used.

Table 3 and 4 list the recognition results obtained with the CGB algorithm when the average of best 50 and 100 trials, respectively, is considered. As shown in these tables, the recognition rates of pecan weevil is always greater than 95%. Remarkably, training of the neural network with only 25% of the data and using descriptors from both RP and ZM methods gave 100% correct recognition results, as shown in Table 3. Thus, the results of the best 50 networks when the training data varies from 25-75% appear to be quite similar. However, when the average of best 100 networks is considered, a larger training data has about 5% higher recognition rates as shown in Table 4. Thus, the convergence of the networks to the best solution is more likely when 75% of the training data is used. The tables also show the average training times for these trials and they all are below 1 sec. for the CGB algorithm.

Table 5 and 6 contain the corresponding results obtained with the SCG algorithm. The results with the SCG algorithm were slightly superior to those from the CGB algorithm. For example, the recognition rates for pecan weevil were always 98%

or higher when the SCG algorithm was used. This was true regardless of the amount of data used for training or the number of trials (best 50 or 100) over which the average is taken. Thus, the results from the SCG algorithm were more robust, although this increased robustness was attained at the cost of larger training times. In particular, the average training times varied from 2 to 4 sec. for the SCG algorithm compared to the less than 1 sec. for the CGB algorithm. Figure 1 and 2 present the recognition rates for the best 50 networks for CGB and SCG algorithms, respectively. These figures and Table 3-6 show that the recognition rates for ZM were slightly higher than RP method. Furthermore, the best results are obtained when

network used inputs from both RP and ZM method. Table 5 and 6 show that 100% recognition rates for pecan weevil as well as other insects are possible when 75% of the training data is used with the SCG algorithm and descriptors from both RP and ZM methods are included for training. These results also show that the training times for a given algorithm are very similar when only RP and ZM descriptors are used or when both are used as network inputs. In summary, 100% successful recognition of pecan weevil as well as 100% recognition of other insects was obtained by using 75% of the training data with the SCG algorithm and utilizing a total of nine descriptors from the RP and ZM methods.
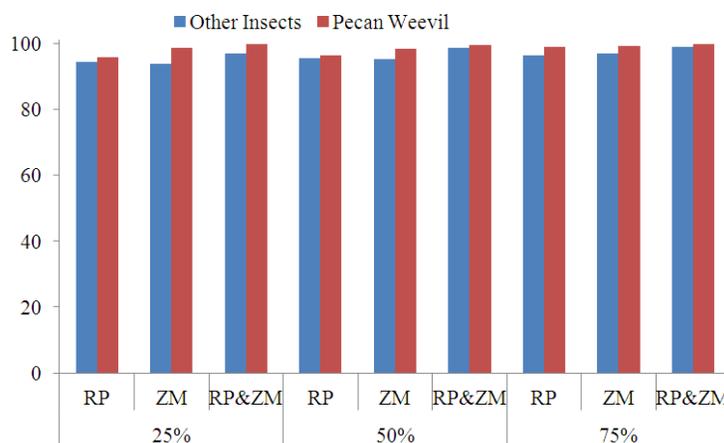
Fig. 1: Recognition Results for Pecan Weevil Using the CGB Algorithm: Average of the 50 Best Networks out of 200 Runs

Fig. 2: Recognition Results for Pecan Weevil Using the SCG Algorithm: Average of the 50 Best Networks out of 200 Runs

Table 6: Recognition results for pecan weevil using SCG Algorithm: average of the best 100 networks of the 200 runs

Training data (%)	Inputs to the ANN	Avg. training time (sec)	Recognition rates	
			Pecan Weevil	Other Insects
25%	RP	2.28	99	91
	ZM	2.19	99	94
	RPZM	2.12	99	97
50%	RP	4.38	98	96
	ZM	3.40	100	95
	RPZM	3.09	100	98
75%	RP	3.61	99	98
	ZM	4.16	100	99
	RPZM	3.47	100	100

DISCUSSION

The neural networks developed in this study are computationally more efficient and require the least amount of information as inputs. In fact, the set of input information can be derived by the three region properties only, namely, the length of major and minor axes and the area of the image. Using these three image descriptors, the neural network-based identification system was capable of 99% recognition of pecan weevils and 100% recognition of other insects in the database, as shown in Table 5. These statistics were obtained when 75% of the data was used for training purposes and the SCG algorithm was used for training the network. The training time was less than 4 sec. in this case indicating that the method is efficient.

Since the identification system works with image descriptors rather than images directly, the memory requirements are lower than a network that would include processing of image data directly. Thus, once the image descriptors are supplied to the network, a reliable identification system is obtained based on the trained networks presented in this study. In an earlier study, we presented a support vector machine-based method for identifying pecan weevil. As shown in Table 2, the method presented in this study based on neural networks is comparable to that approach; albeit our analysis indicated that the current method is superior in terms of its robustness since multiple trials provided recognition results that were similar. This is evident from Table 3-6 which show the averages of top 25 and 50% of the networks. In other words, in more than half of the trials, high recognition rates were achieved suggesting reliability and robustness of this method. Thus, the results reported in this study are promising new developments in recognizing pecan weevil and would be helpful in developing reliable wireless sensor networks that utilize the recognition capabilities developed here.

CONCLUSION

This research focused on developing robust identification system for pecan weevil. In this study, nine image descriptors derived from standard image processing techniques were used as inputs to artificial neural networks. The descriptors were obtained from the Regional Properties (RP) and Zernike Moments (ZM) methods. The neural networks were trained with two different training algorithms and three different training ratios (25-75%) were studied. In each case, the network was trained 20 times to avoid convergence to local minima and the training data was randomly selected 10 times.

The neural networks developed in this study are capable of 100% recognition rates for pecan weevil and 100% recognition of other insects' images, when only nine descriptors originating from the RP and ZM methods are used and the networks are trained with SCG algorithm and 75% of the data. The SCG algorithm outperforms the CGB algorithm in terms of recognition rates; however, the training times for SCG algorithm are slightly larger when compared to the CGB algorithm. The average training times were less than 1 sec. for CGB algorithm and ranged from 2 to 4 sec. for the SCG algorithm. Further, the testing time for each image was only 0.16 sec. Thus, the networks developed in this study appear to be reliable for developing wireless sensor networks for field applications.

ACKNOWLEDGMENT

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. and G.M. Hassan, 2011c. Pecan weevil recognition using support vector machine method. Am. J. Agric. Biol. Sci., 6: 521-526. DOI: 10.3844/ajabssp.2011.521.526

Al-Saqer, S.M., P. Weckler, J. Solie, M. Stone and A. Wayadande, 2010. Identification of Pecan weevil through image processing. Am. J. Agric. Biol. Sci., 6: 69-79. DOI: 10.3844/ajabssp.20 11.69.79

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

Ashaghathra, S.M., 2008. Identification of Pecan Weevils Through Image Processing. Ist Edn., ProQuest, ISBN-10: 0549560947, pp: 148.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. 1st Edn., Oxford University Press, USA., ISBN-10: 0198538642, pp: 482.

Do, M.R., 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

Gonzalez, R.C. and R.E. Woods, 2002. Digital Image Processing. 2nd Edn., Printice Hall, NJ., ISBN-10: 0201180758, pp: 793.

Hagan, M.T., H.B. Demuth and M. Beale, 2002. Neural Network Design. 1st Edn., Beijing, ISBN-10: 7111108418.

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

Johansson, E.M., F.U. Dowla and D.M. Goodman, 1991. Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method. Int. J. Neural Syst., 2: 291-301. DOI: 10.1142/S0129065791000261

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 Communication, Networking and Mobile Computing, Sep. 24-26, IEEE Xplore, Beijing, pp: 1-5. DOI: 10.1109/WICOM.2009.5303034

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

Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. J. Neural Netw., 6: 525-533. DOI: 10.1016/S0893-6080(05)80056-5

Mulder, P., 2004. Pecan Weevil season in full swing. Plant Dis. Insect adviso.

Powell, M.J.D., 1977. Restart procedures for the conjugate gradient method. J. Math. Program., 12: 241-254. DOI: 10.1007/BF01593790

Xu, S. and L. Chen, 2008. A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. Proceeding of the 5th International Conference on Information Technology and Applications, (ICITA, 08), Queensland, Australia, June 23-26.