Support Vector Machine Based Red Palm Weevil
(*Rynchophorus Ferrugineous, Olivier*) Recognition System

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

**Abstract:** Problem statement: Red palm weevil (*Rynchophorus Ferrugineous, Olivier*) is an insect which threatens the existence of palm trees. The proposed research is to develop a RPW identification system using Support Vector Machine method. The problem is to extract image features from an image and using SVM to find out the existence of RPW in an image. Approach: Images are snapped and image processing techniques of Regional Properties and Zernike Moments are used to extract different features of an image. The obtained features are fed into the SVM based system individually as well as in combination. The database used to train and test the system includes 326 RPW and 93 other insect images. The input data from database is selected randomly and fed into the system in three steps i.e., 25, 50 and 75% while remaining database is used for testing purpose. In SVM, polynomial kernel function and Radial Basis Function are used for training. Each experiment is repeated 10 times and the average results are used for analysis. Results: The optimal results are obtained by using Radial Basis Function in SVM at lower values of sigma ‘σ’ while Polynomial kernel function is not successful in returning adequate results. Further detailed analysis of results for ‘σ’ value of 10 and 15 revealed that proposed system works well with large training data and with inputs obtained by Regional Properties. The optimal value of ‘σ’ for proposed system is found to be 10 when training data ratio is 50%. The training time for proposed system depends on size of database and is found to be 0.025 sec per image while time consumed by proposed system for identification of RPW in an image is found to be 15 milli sec. The proposed system’s success in identification of RPW and other insect is found to be 97 and 93% respectively. Conclusion: It is concluded that SVM based system using Radial Basis Function having ‘σ’ value of 10 is optimal in identifying RPW from an image. The optimal input data for the proposed system needs to be obtained by Regional Properties only.

**Key words:** Red palm weevil, automated recognition system, machine learning algorithms, support vector machine, radial basis function, insect recognition

**INTRODUCTION**

Red Palm Weevil (RPW) [*Rynchophorus Ferrugineous (Olivier)*], is considered as one of the most destructive insect for palm trees. It was first identified in South and South-east of Asia in early 20th century. The existence of insect is reported in many countries around the world (Lefroy, 1907; Buxton, 1920; Abraham *et al*., 1998; Al-Ayedh, 2008; Li *et al*., 2009; Faleiro, 2006).

RPW spends its life cycle inside the trunk of palm tree where it feeds on tissues of the tree. Being inside the palm tree, RPW is protected and undetected from outside. (Faleiro, 2006; Esteban-Duran *et al*., 1998; Murphy and Briscoe, 1999). RPW usually remains inside the infested palm tree for generations as long as food is available and emerges when the tree is hollow, to find another host. Currently, infested palm trees are terminated to prevent the spread of RPW and save neighboring palm trees.

Many approaches have been proposed to control RPW, however, Abraham *et al.* (1989) has suggested that Integrated Pest Management (IPM) is the most successful method to control and manage RPW. Faleiro (2006) stated that early detection and trapping are among the main elements of the IPM for RPW. Therefore, major emphases were put for the development and improvement of this technique. In trapping, traps containing bait, pheromone and pesticide would be spread in the entire field and surveyed regularly. These traps would primarily be used to observe the existence of RPW and to know the
scale of its spread in the farm. This leads to take appropriate decisions accordingly. These traps may also be used to detect RPW at early stage. The recommended trap density is 1-2 traps ha\(^{-1}\) (Faleiro, 2006; Soroker et al., 2005). The inspection and maintenance process of traps requires low skilled field staff but is considered as labor-intensive and time-consuming.

In order to automate the inspection process, wireless image sensor network can be incorporated in the traps. The idea is to take an image of a trapped insect and process it for identification of RPW. In such a system, all motes (nodes) of image sensor network would communicate with each other and send the gathered information to the main server. The wireless sensor network in general, has been adopted in the field of Agriculture (Burrell et al., 2004), poultry (Murad et al., 2009), industry (Jan et al., 2010). The image sensor network has been utilized in different applications such as Fruit Flies surveillance (Liu et al., 2009), environment observation and surveillance (Feng et al., 2005), object detection and recognition (Kulkarni et al., 2005).

The success of any recognition system depends on its processing time and reliability. Some automated systems for identification and recognition of different insects have been proposed, such as Automated Bee Identification System (ABIS) proposed by Arbuckle et al. (2001) for identification of Bees; Digital Automated Identification System (DAISY) proposed by Watson et al. (2004) for identification of Ophioninae Automated Insect Identification through Concatenated Histograms of Local Appearance System (AIICHLA) proposed by Larios et al. (2007) for identification of Stonefly larvae; Species Identification Automated and Web Accessible System (SPIWA) proposed by Do et al. (1999) for identification of Spiders software system developed and proposed by Al-Saqer et al. (2010) for identification of Pecan Weevil.

Another promising method for pattern recognition applications, based on machine learning, was introduced by Cortes and Vapnik (1995) to solve two group classification problems. This method is known as Support Vector Machine (SVM) method. It is used in many pattern recognition applications, such as: Qin and He (2005) proposed face recognition method based on SVM; Ganapathiraju et al. (2004) proposed speech recognition method using SVM; Yuxia and Hongtao (2008) proposed Simulated Annealing Algorithm based on SVM for recognition of stored grain pests.

For the development of wireless image sensor network based RPW identification system, the initial phase is to develop an efficient recognition system of RPW via image processing techniques. The recognition system needs to be efficient in terms of processing time, computational resources and reliability.

Two different image processing techniques were implemented for the identification of RPW (Al-Saqer and Hassan, 2011a). The algorithm used for that system utilized regional properties of the insect’s image and the values of moment invariant (Zernike Moments). The processing time was found to be 0.47 sec with 97 and 88% recognition rate for RPW and other insects respectively. Another approach was tested to resolve the same issue by using ANN where pixel information was submitted to the ANN in binary form (Al-Saqer and Hassan, 2011b). The training of the proposed network was reported to take 183.4 sec but took decision swiftly. The best recognition rate for RPW and other insects was reported to be 99 and 93% respectively. The proposed ANN needed to have 4 layers of network and a total of 24,771 neurons. This approach proved to have better results on the expense of higher requirements of computing.

The aim of this research is to evaluate an alternative approach of SVM that would outperform the previous systems in terms of efficiency, time consumption and less computational requirements which are imperative for wireless image sensor network. Different sample size of data is used for training and testing and their results are compared. The system is expected to distinguish RPW from other insects which are normally found in habitat of palm tree.

**MATERIALS AND METHODS**

During the recent years SVM is reported to be performing well as compared to other techniques (Cho et al., 2006). In this method, the non linear pattern of data is mapped into multidimensional feature space via a kernel function. On the mapped feature space, a hyperplane is used to distribute the two classes for classification purpose. The boundary items are the only items which are considered while optimizing the boundary of the classes. The items of classes close of the boundary make a vector known as Support Vector. Using the support vector, it is easy to optimize boundary of classes and make hyperplane for classification. If the support vector changes or moves, then position of hyperplane will also change and consequently, the classification boundary. However, hyperplane will be unaffected by the movement of items other than support vector. This process also usually solves the regular problem of local minima of ANN (Cortes and Vapnik, 1995).

The main task in SVM method is to find an optimal kernel function which is used in mapping of data to
multidimensional feature space. There are different kernel function proposed over time but Radial Basis Function (RBF) and polynomial function are most commonly used for pattern recognition problems (Chin, 1998). Gaussian function is usually used as RBF. It is real valued function and is represented as Eq. 1:

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

(1)

Where:
- \(x\) and \(y\) = Support vector and data point to be tested respectively
- \(\sigma\) = The width of the Gaussian curve

As the value of \(\sigma\) increases, the decision boundary becomes more regular and decision surface becomes smoother. It is also inversely proportional to the number of support vectors (Buhmann, 2003).

Polynomial function is a directional function and its output is dependant on the direction of two vectors in low dimensional space and is represented as Eq. 2:

$$K(x, y) = (x.y + 1)^d$$

(2)

where, \(x\) and \(y\) are support vector and testing data point respectively while \(d\) is the degree of the polynomial. The magnitude of the output is dependent on the testing data point. For experiments, kernel function of RBF and polynomial were selected. Different values of degree \(d\) and sigma \(\sigma\) were tried for polynomial and RBF kernel functions respectively.

**Image acquisition:** It is recommended in pattern recognition problems that size of training database should be large and variant. To acquire this, large numbers of insects were collected and their images were taken after preparation by imaging system. The imaging system includes Sony Cyber-shot DSC-HX1 camera which can shoot at 10 frames per second and equipped with 9.1 megapixel resolution and 20x optical zoom. Images taken were of the size of 3456×2592 pixels. Original images were processed to convert into binary format and resized to 501×519 pixels.

For simulations, a computer ‘Dell Optiples 780’ having Core 2 Duo E8400 3.0 GHz processor of Intel was used having RAM of 4 GB. Simulations and image processing were conducted using MATLAB® Version 7.9.0.529 (R2006a).

**Data processing method:** The inputs for ANN and SVM are derived from two image processing techniques i.e., Zernike Moments and Regional Properties, that have been evaluated in earlier study (Al-Saquer and Hassan, 2011a).

Zernike Moments method is a competent technique because of invariance in rotation, efficiency in expression, robustness in noise and short processing time. In this method, a set of complex polynomials are introduced which form an orthogonal set over interior of a circle. The center of the image is taken as origin and coordinates of pixel are mapped to the unit circle’s range for purpose of computation of Zernike Moments. The pixels outside the unit circle are discarded. The orthogonal properties ensure that there is no redundancy or overlapping of information between moments with different orders and repetition. Thus, each moment will be unique and independent representation of a specific image (Kim and Kim, 2000). Zernike Moments of order 3 was applied to all images and the resultant six unique values for each image were used as inputs.

Regional Properties method is a technique that uses regional descriptors of the object. It deals with the region of the image instead of the boundary of the object in an image. This method extracts important properties of regions of image e.g., area, orientation, centroid.

For unique representation of RPW, area of the region and lengths of major and minor axes of RPW are used. The region is obtained by calculating the number of connected pixels in the image. The length of major axis and minor axis are calculated as length (in pixels) and width (in pixel) of the elliptical considered region in the image respectively (Gonzalez and Woods, 2002). The three values obtained were also used as inputs. The results of major axis, minor axes and normalized area were used as inputs.

A database of 419 images of RPW and other insects was used. All images were processed by Zernike Moments and Regional Properties methods. The database comprised of 326 images of RPW and 93 images of other insects as mentioned in Table 1. Three sets of tests were conducted. Values obtained by Zernike Moments, Regional Properties and combination of both were applied as inputs in each set of test. Furthermore, each set of input used randomly selected training sizes of 25, 50 and 75% of entire database. The remaining data was used for testing purpose. For consistency of experiments, the selected training data remained unchanged for entire set of test and each set of test was repeated 10 times. The training data was selected randomly and average results are considered are analysis.
Table 1: Other insects for experiments

<table>
<thead>
<tr>
<th>Scientific name (Family/order)</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calosoma Chlorostictum (Carabidae/Coleoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Gryllus Bimaculatus (Gryllidae/Orthoptera)</td>
<td>10</td>
</tr>
<tr>
<td>Conocephalus Conocephalus (Tettigoniidae/Orthoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Gryllotalpa Gryllotalpa (Gryllotalpidae/Orthoptera)</td>
<td>3</td>
</tr>
<tr>
<td>Gryllotalpa Africana (Gryllotalpidae/Orthoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Orates Nasiconiris (Scarabaeidae/Coleoptera)</td>
<td>13</td>
</tr>
<tr>
<td>Cybister Triungulatus Africana (Dytiscidae/Coleoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Scarites Eurytus (Carabidae/Coleoptera)</td>
<td>3</td>
</tr>
<tr>
<td>Lanelater Motodenta (Elateroidea/Coleoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Mlabri Tenebrosa (Meloidae/Coleoptera)</td>
<td>3</td>
</tr>
<tr>
<td>Hyles Lineata Livornica (Sphingidae/Lepidoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Coccotrypes advena (Curculionidae/Coleoptera)</td>
<td>3</td>
</tr>
<tr>
<td>Gopneloleon Sp. (Myrmeleontidae/Neuroptera)</td>
<td>3</td>
</tr>
<tr>
<td>Blepharopsis Mindica (Mantidae/Mantodea)</td>
<td>2</td>
</tr>
<tr>
<td>Anax Sp. (Aeshnidae/Odonata)</td>
<td>3</td>
</tr>
<tr>
<td>Xylocopa Hottentota (Anthophoridae/Hymenoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Poikiloderma (Pamphiliidae/Orthoptera)</td>
<td>4</td>
</tr>
<tr>
<td>Lophyra Sp. (Carabidae/Coleoptera)</td>
<td>5</td>
</tr>
<tr>
<td>Scarabaens Sp. (Scarabaeidae/Coleoptera)</td>
<td>3</td>
</tr>
<tr>
<td>Cerceris Rybyensis (Sphecidae/Hymenoptera)</td>
<td>3</td>
</tr>
</tbody>
</table>

**Error and its criteria:** In this research, the Error is defined as incorrect identification i.e. either RPW is not identified as RPW and is classified as other insect or vice versa. This error of classification can be categorized into two classes: Type-I and Type-II errors. If any other insect is classified as RPW then Type-I error occurs while if RPW is not correctly classified then it is marked as Type-II error.

As the focus of research is identification of RPW in an image, so Type-II error plays more critical role as compared to Type-I error. The system’s inefficiency may be described as Type-II error while oversensitivity of system (false alarm) may be described as Type-I error. To place a unified criteria for selection and comparison of results, a criteria is proposed where Type-II error is given double weight as compared to Type-I error. The system giving better results is marked by Eq. 3 low value of criteria of selection i.e.:

\[
\text{Criteria}_{\text{selection}} = 2(\text{Error}_{\text{Type-II}}) + \text{Error}_{\text{Type-I}} 
\]

**RESULTS**

SVM method was used in this research where the input data was obtained by different regional descriptor techniques i.e. Zernike Moments, Regional Properties and combination of both. In SVM, the polynomial kernel function was unable to indentify RPW and provide appropriate results for all values of degree ‘d’. On the other hand, RBF kernel function was successful in identifying RPW and providing the results.

For training data size of 25%, the experimental results obtained for the cases when input was obtained by Regional Properties, Zernike Moment and combination of both are plotted in Fig. 1. Similarly for training data size of 50 and 75%, the results obtained are plotted in Fig. 2 and 3 respectively. The recognition rates for RPW and other insects for selected sigma values ‘c’ are presented in Table 2.

Training time for SVM is found to be dependent on size of training data as well as imaging techniques i.e., 0.205, 0.2 and 0.225 sec image\(^{-1}\) when input data is obtained by Regional Properties, Zernike Moments and combination of both while processing time of an image was approximately equal to 50 micro sec when network is trained by values obtained by Regional Properties and 0.2 sec for the other two case studies.
Table 2: Results for different ‘σ’ values for all case studies

<table>
<thead>
<tr>
<th>Training data size</th>
<th>Regional properties</th>
<th>Zernike moments</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RPW†</td>
<td>Other Insects†</td>
<td>RPW†</td>
</tr>
<tr>
<td>‘σ’ = 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>97.75</td>
<td>63.43</td>
<td>91.31</td>
</tr>
<tr>
<td>50%</td>
<td>97.12</td>
<td>93.26</td>
<td>96.20</td>
</tr>
<tr>
<td>75%</td>
<td>97.28</td>
<td>89.57</td>
<td>97.90</td>
</tr>
<tr>
<td>‘σ’ = 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>97.58</td>
<td>87.14</td>
<td>91.27</td>
</tr>
<tr>
<td>50%</td>
<td>96.75</td>
<td>90.22</td>
<td>97.30</td>
</tr>
<tr>
<td>75%</td>
<td>96.79</td>
<td>86.52</td>
<td>98.27</td>
</tr>
</tbody>
</table>

†: Recognition rates are mentioned in percentage

**DISCUSSION**

The results presented in Fig. 1 mentions that input data obtained by Regional Properties provide better results as compared to the input data obtained by Zernike Moments. While comparing results of case studies when input data was obtained by Regional Properties and combination of both Regional Properties and Zernike Moments, it is observed that the pattern of errors with respect to change in value of ‘σ’ is similar. It is also observed from Fig. 1 that Type-I error is higher than Type-II error for all case studies which is optimal for the solution of problem. Furthermore, both types of errors are higher in case study when input data was obtained by Zernike Moments while for other case studies, both type of errors are closer in values.

The results presented in Fig. 2 refers that Type-I error for the case study when input data was obtained by Zernike Moments, is highest while Type-II error for the same is lowest. For the other two case studies, the Type-II errors are slightly higher while Type-I errors are lower comparative to the case study when input data was obtained by Zernike Moments. Computing the criteria mentioned in equation 4, it is found that case studies when input data was obtained by Regional Properties and by combination of both Regional Properties and Zernike Moments, outperform the results when input data was obtained by Zernike Moments. Type-II errors for all three case studies is always lower than 5% while variation in Type-I errors is higher with respect to variation of ‘σ’ for all case studies. The results of Type-I and Type-II errors are close to each other for case studies when input data was obtained by Regional Properties and by combination of both Regional Properties and Zernike Moments. Figure 3 mentions the results when training data is 75% and it suggests that the behavior of the system is consistent with the findings when the training data is 50%.

In additional, when system was trained using 25% data, as mentioned in Fig. 1, it is observed that best results are obtained when the value of ‘σ’ is 15, whereas, from Fig. 2 and 3, the best results are obtained when value of ‘σ’ is 10. Table 2 summarizes the results of all case studies when ‘σ’ value is taken as 10 and 15.

Overall, the recognition rate for RPW and other insects are found to be the lowest for the case study when input data was obtained by Zernike Moments. Furthermore, recognition rates when input data was obtained by Regional Properties, do not change significantly when input data is obtained by combination of both Regional Properties and Zernike Moments. It is also observed that recognition rate of RPW is not varying significantly for all case studies and the performance of the system mainly depends on variation of recognition rate of other insects.

Moreover, it is found that recognition rates improve when training data is increased from 25% to 50% while similar behavior is not observed when training data is increased from 50% to 75%. This may be due to the reason that data left for testing is reduced with increase of training data and each wrong result contributes more in error percentages. The best result is obtained for the case when training data is 50% at value of 10 for ‘σ’ and input data was obtained using Regional Properties only. The training time required by proposed system is found to be 0.025 sec image⁻¹ which is 95% less than the ANN based RPW identification system (Al-Saquer and Hassan, 2011b).
CONCLUSION

This research is focused on developing a software system that can identify Red Palm Weevil by using Support Vector Machine method. The input to the proposed system is prepared by using two image processing techniques i.e., Regional Properties and Zernike Moments. The outputs of these techniques are fed into the proposed system individually as well as in combination. The Polynomial kernel function and Radial Basis function are used in SVM by varying degree 'd' and sigma 'σ' respectively. However, the Polynomial kernel function did not produce adequate results. The database of images consists of 326 RPW and 93 other insects. The training of system is conducted by randomly selecting three percentages of data i.e., 25, 50 and 75%, while remaining data is used for testing. It is observed from the results that the inputs to the proposed system obtained by Regional Properties produce better results as compared to inputs obtained by using Zernike Moments while results produced by combination of both techniques were close to the former. Radial Basis function performs efficiently in SVM at lower 'σ' values of 10 and 15. For all experiments, the critical Type-II error is lower than the Type-I error. The best recognition rate obtained are 97 and 93% for RPW and other insects respectively. The optimum solution is found when Radial Basis function with 'σ' value of 10 is used for SVM with 50% training data. The time obtained for training the proposed system is 0.025 sec image\(^{-1}\) while testing time of an image is 15 milliseconds.

ACKNOWLEDGEMENT

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

REFERENCES


Lefroy, H.M., 1907. The more important insects injurious to Indian Agriculture Memoirs of the Department of Agriculture in India, vol i, No 2, Ento Se, Pp 139 + v; 80 figures. Nature, 76: 588-588. DOI:10.1038/076588a0


Murphy, S.T. and B.R. Briscoe, 1999. The red palm weevil as an alien invasive: Biology and the prospects for Biological Control as a component of IPM. Biocontrol News Inform., 20: 35N-45N.


