SVM-based Pest Classification in Agriculture Field

B. Divya, M. Santhi

Abstract--- Integrated Pest Management is currently used to reduce the use of harmful pesticides and chemicals in the agriculture environment. However, the early detection of pest and controlling of the pest population is the critical task and time consuming process and the judgment is mostly based on the manual process which is highly prone for error. In this paper, an image based classification is used to for detection and classification of the pest species which is commonly available in the felid. Digital images were obtained. Detection of pest in the images, segmentation, feature extraction was performed by the algorithms for the detected pest. Finally, SVM was used for classification and results were compared with K-Neural Network. Compared to the KNN, SVM achieved accurate results with combined features with accuracy as its metrics.

Keywords--- IPM, Pest, GLCM, LTP, CCV, SVM.

I. INTRODUCTION

Integrated pest management (IPM) is an efficient and accurate for pest control. Rather than using regular spray, it uses spray on the presence of pest. Traditional methods of trapping pest using light trap and counting pest are commonly used [1]. Regular checking of traps and identifying species requires biological knowledge and is time consuming. It also has poor accuracy and low recognition rate. IPM requires the monitoring of pest insect species with best pesticide option that is suitable in both ecological and social concern. Therefore, IPM main consideration is for improving recognition and classification rate. Computer vision and machine learning has emerged as automated technology for the issue [2]. In agriculture and ecology, image based insect recognition has wide applications with minimal operator training [3, 4]. There are several challenges with reference to the image based insect recognition. These include pose variance problem that provides to imperfect feature extraction and mis-identification; species similarity which leads to inefficient classification; selection of architecture when large numbers of insects are detected.

Early insect recognition was done based on threshold, which is opted as the best approach for detection and classification. Resulting of false alarm due to shadows and physical consideration, researchers used other methods to solve this issue [5]. Low-level image representation such as Gray-Level Co-occurrence Matrices (GLCM) [6], colour histogram, geometric shape [7,8]. Hu moment invariants, Eigen-images, wavelet coding or other simple features[9-14] based on colour, shape, texture are used. Many automated web development programs are also developed such as the automated bee identification system (ABIS)[13], digital automated identification system (DAISY)[14], identification, automated and web accessible (SPIDA)[15] which are highly effective for controlled environment such as controlled lighting, green house environment etc., and small database projects. However, the features extracted were not sufficient and detailed which will lead to the false alarm and most of these systems required manual manipulation. So, further image feature extraction algorithms are needed for precise recognition.

Local feature based representation methods were used that allow finding local image structure in repeatable fashion and encode them in a representation that is invariant to a range of image transformations, such as translation, rotation, scaling, and affine deformation and minimal learning of pest image[16-27]. Bag of words is commonly used method which involves partitioning by local operators, encoding using dictionary of visual words and pre-process of histogram with the minimum encoding length [28]. The resulting features then form the basis of approaches for recognizing specific parts like legs, antennae, tails, wing pads, etc. However, patch descriptors have to be chosen carefully for feature extraction. Due to background cluster, location changes and partial occlusion [29], the pre-processing procedures can represent compact information. So the classifier can be dominated by irrelevant information and refining of useful information is incapable by pre-process procedure. The major challenges the most applications fail to meet are variation in intra-species and pose variance.

Improvement in the feature selection is done by advancement in algorithms and combining of features for identification like Multiple-task sparse representation and multiple-kernel learning[30], Deep residual learning[31], Bio-inspired methods[32]. But still the efforts remain lacking due to unrefined recognition of image. All the previous algorithms lack in some factors either in visualizing the important features of pest and applying the learning model which will discriminative the information of features without human intervention. Therefore, SVM will be an alternative method for classification. In this work we developed an SVM based identification algorithm to avoid the false positive results. First the possible features like colour, shape is obtained and trained with the SVM classifier to improve the accuracy of prediction. We used the same set of images, feature sets and feature selection methods for both ANN and SVM based models. Therefore, a fair comparison and the performances evaluation of the two well optimized machine learning methods (ANN and SVM) in identifying species are obtained.

Manuscript received February 01, 2019

B. Divya, Electronics and Communication Engineering, Saranathan College of Engineering, Trichy, India. (e-mail: divyabalasubramaniam05@gmail.com)

M. Santhi, Electronics and Communication Engineering, Saranathan College of Engineering, Trichy, India. (e-mail: santhiphd@gmail.com)

The rest of this paper is organised as follows. In section 2, we summarise material and methodologies overview that are used. Section 3 summarises feature extraction methods that are used for the images. Section 4 provides classification. In section 5, the empirical evaluation with the experimental setup, a brief description of the dataset, and the testing results are discussed. Finally, section 6 concludes the paper with a discussion of the findings towards future extensions.

II. MATERIALS AND METHODS

The computational steps and methodologies used in the experiment are schematically demonstrated in fig.1. The proposed method consists of two major sections:

i. segmentation
ii. classification

A. Implementation environment and materials

The experiments were carried out in Matlab 2017b. Briefly, 6 different species of insects commonly found in storage and open field are collected. The images are obtained from online resources which contained different set of images varying in size, angle and lighting. For better identification, the feature parts and background were extracted. This resulted in a final set of images. Subsequently the images feature like colour, patterns, size, and texture are obtained through taxonomical identification.

B. Image segmentation

Portioning an image having same features or similarity into various parts is segmentation. Some of the methods for segmentation process are like otsu' method, k-means clustering, converting RGB image into HIS model and so on, In this paper, k-means clustering is used and they are given as input image for further processing.

C. K -means clustering

K means clustering algorithm is commonly used to classify the image into k number of classes based on the features. Classification of image is done by reducing the sum of square of distance between the object and corresponding cluster. The algorithm as follows:

Step 1. Read input image.

Step 2. Transform image from RGB to L*a*b* colour space.

Step 3. Classify colours using K-Means clustering in 'a*b*' space.

Step 4. Label each pixel in the image from the results of K-Means.

Step 5. Generate images that segment the image by colour.

In this algorithm, Euclidean distance metric is used for k-means clustering. L*a*b is used because colour space is stored only in a* and b* component which reduces processing time for segmentation.

III. IMAGE FEATURE EXTRACTION

Identifying species through the feature and pattern s that are minute is the key point for recognition. The process of converting the image into digital descriptors from the micro structural patterns is a feature extraction. Global feature approach describes the overall characteristics of the pest image. Several metrics like statically and spectra methods, colour distribution, density of specific objects and oriented edge response are computed to quantify the global appearance. To do this effectively Colour Coherence Vector (CCV), GLCM, LTP features are computed. Standard pre-processing and filtering required for the feature are done.

A. Colour Coherence Vector (CCV)

CCV incorporates spatial information by histogram based method by comparing the images. In this technique, each pixel has to be classified as coherent or incoherent pixel.

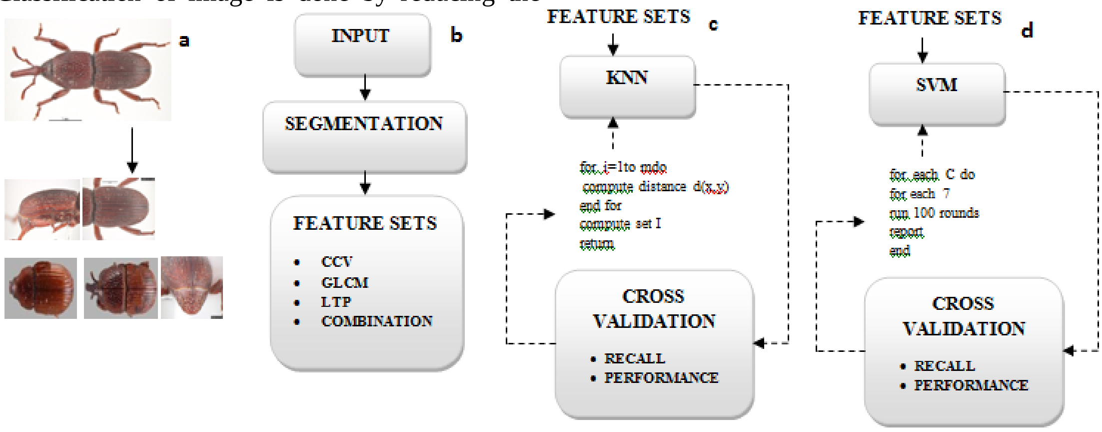


Fig. 1. The experimental procedure (a) obtaining inpt images; (b) features set collection; (c,d) showing the training and cross validation using ANN and SVM models.

Based on the, similarity- coloured region classification of each pixel is done. Coherent pixel is sizable contiguous region whereas incoherent pixels are not belonging to some sizable region. For computing CCV, first image is blurred. Then, image is discretized based on the n distinct colours to eliminate the variation in the neighbouring pixels. Then, finds the connected components based on τ which is constant value.

Based on the value, the coherent and incoherent pixels are classified from connected component. After classification, CCV computes the coherent and incoherent pixels' histogram. In single histogram, the two histograms are stored.

B. Local Ternary pattern (LTP)

Local ternary patter is extension of local binary pattern (LBP) which is developed to overcome the centre pixel small gray changes that may lead to different codes and LBP is Sensitive to noise [34]. LTP examines the similarity of the pixels. Pixel with small variations are accounted as alike and degree of similarity is with the threshold value 'τ'.

$$LTP_{P,R,\tau} = \sum_{i=0}^{P-1} s(p_i - p_c) * 3^i, s(x) = \begin{cases} 1, x \geq \tau \\ 0, |x| < \tau \\ -1, x \leq -\tau \end{cases} \quad (1)$$

Where, τ is a threshold specified by the user. The ternary pattern is divided into two parts to reduce the feature dimension: the positive part and negative part.

C. Gray Level Co-occurrence Matrix (GLCM)

Image is analysed by first order and second order statistics. First order includes colour at every pixel, standard deviation and entropy. Second order feature includes gray level co-occurrence matrix (GLCM). Contrast, correlation, energy and inertia are derived from GLCM.

First order statistics are used to find the region of interest with an image. In second order statics, GLCM is the probability that a pixel of particular grey level occurs at a specified direction and distance from its neighbouring pixel.

$$Cij = \frac{Pij}{\sum_{i,j}^{N} Pij} \quad (2)$$

Where Cij is Co-occurrence probability, Pij is number of occurrence of grey level i and j.

Energy is the sum of squared elements in GLCM (ranging from 0 to 1) and Inertia is the inverse difference moment within an image.

$$Energy = \sum_{i,j=0}^{N-1} Cij^2 \quad (3)$$

$$Inertia = \sum_{i,j=0}^{N-1} \frac{Cij}{1+(i-j)^2} \quad (4)$$

Entropy is used to characterize the texture of the input image by statistical measure of randomness.

$$Entropy = \sum_{i,j=0}^{N-1} Cij \log Cij \quad (5)$$

Correlation is a whole field deformation measurement technique that extracts whole field displacement data by comparing a pair of digital images of a specimen before and after deformation (ranging from -1 to +1).

$$Correlation = \sum_{i,j,x,y=0}^{N-1} \frac{(i-\mu_x)(j-\mu_y)Cij}{\sigma_x \sigma_y} \quad (6)$$

Where μ_x, μ_y and σ_x, σ_y are mean and standard deviation of sums of rows and columns in the matrix respectively. N is the dimension of square matrix of GLCM.

IV. SVM CLASSIFICATION AND PARAMETER OPTIMIZATION

A. SVM Classification

SVMs are best among the classifier in handling the task ranging from genomic data to text. By designing the kernel function, SVM can be applied to the complex data beyond feature vectors. These models are efficient in both linear and nonlinear data handling. The classifier aims to draw decision boundaries between data points. LIBSVM open source is used for the feature set obtained [33]. There are several kernels available. Some of the general kernels are given at below:

Pth Degree polynomial:

$$K(x_i, x_j) = (x_i x_j + 1)^p \quad (7)$$

Radial bases (RBF):

$$K(x_i, x_j) = e^{\left\| \frac{x_i - x_j^2}{2\sigma^2} \right\|} \quad (8)$$

Multi-layers Perceptron (MLP):

$$K(x_i, x_j) = \tanh(\beta x_i x_j + \delta) \quad (9)$$

To separate the classifiers with maximum margin, RBF is adopted and performance best among the other kernel functions in nonlinear classification. The experiment was initiated by using all features [34]. Besides the three feature selection methods, combination of all the feature sets is used for better representation of features.

B. Quality Prediction

The classification can be evaluated by the confusion matrix that was calculated for each round of validation. Due to the classification, they were calculated for each species of insects from an $M \times M$ confusion matrix as follows.

Let $CM{\times}M$ be the confusion matrix for M species where rows represent the actual class label and columns stand for the predicted labels. Then, for a given class label i,

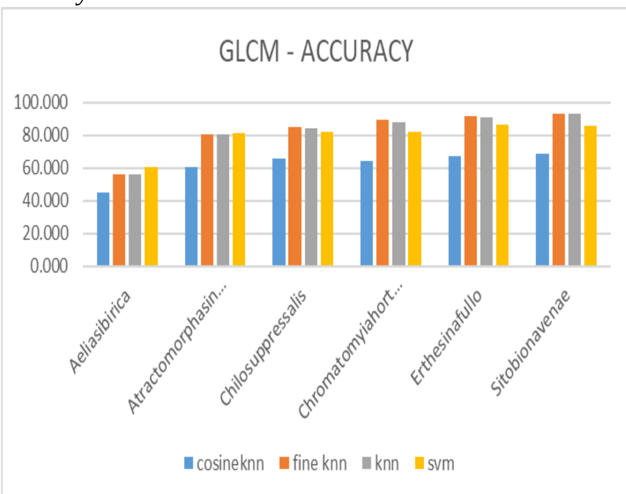
$$FalsePositive(FP) = \sum_{j=1}^{M} C_{ji} - C_{ii} \quad (10)$$

$$FalseNegative(FN) = C_{ii} \quad (11)$$

$$TruePositive(TP) = \sum_{j=1}^{M} C_{ij} - C_{ii} \quad (12)$$

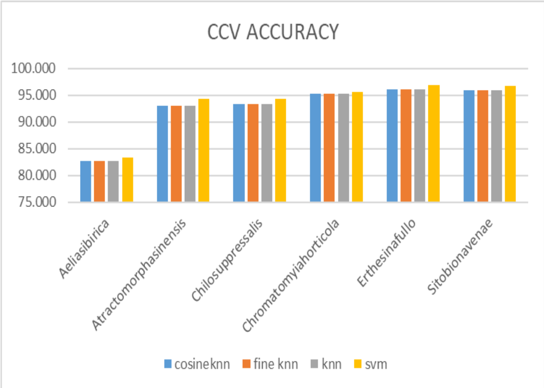
$$TrueNegative(TN) = \sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij} - (FP + FN + TP) \quad (13)$$

The evaluation of the performance of each approach is by *Accuracy*.

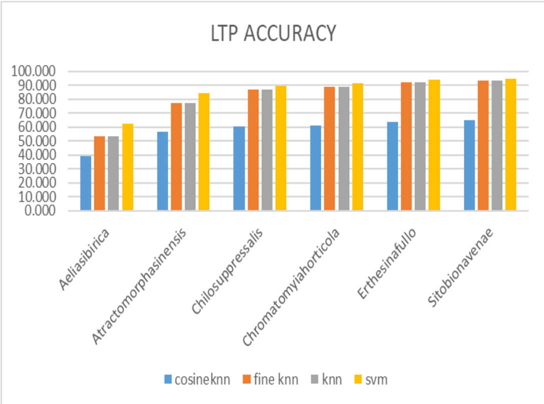


(a)

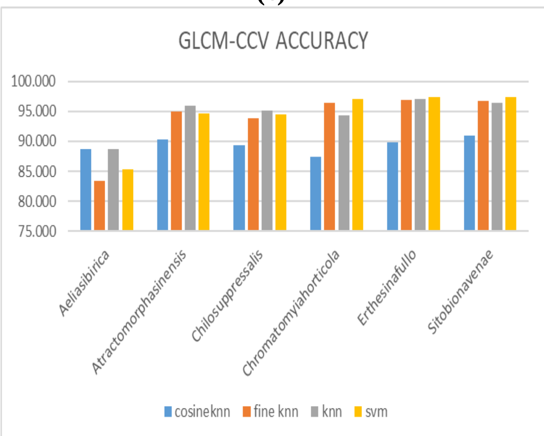
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



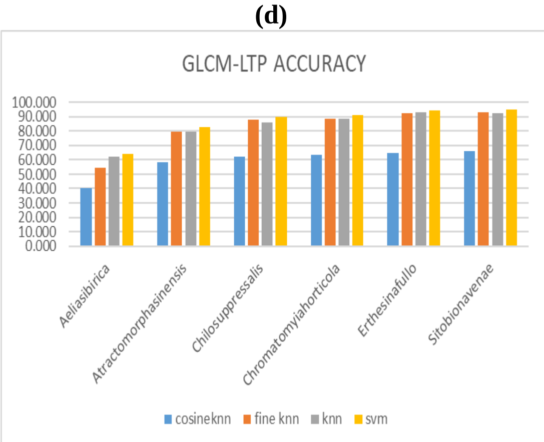
(b)



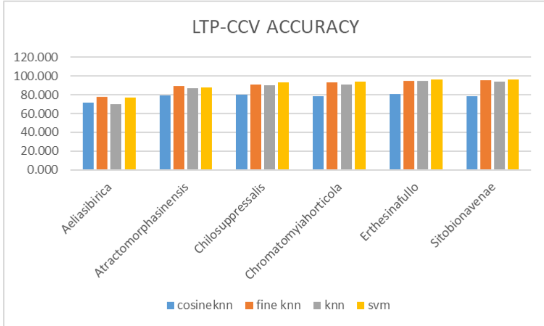
(c)



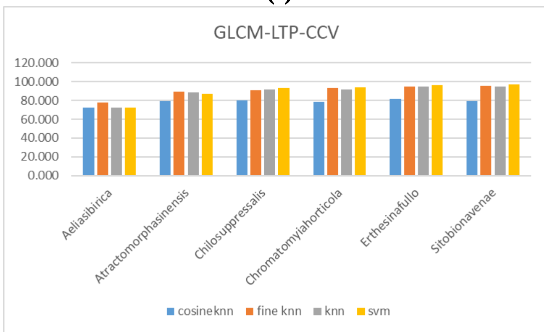
(d)



(e)



(f)



(g)

Fig.2 The results (a) GLCM accuracy (b)CCV accuracy (c) LTP accuracy (d) GLCM-CCV accuracy (e) GLCM-LTP accuracy (f) LTP-CCV accuracy (g) GLCM-LTP-CCV accuracy.

$$Accuracy = \frac{\sum_{i=1}^{M} TP_i}{\sum_{i=1}^{M} TP_i + FN_i} X\ 100\% \qquad (14)$$

The overall accuracy is computed by Eq (14). and the global statistics such as mean, standard deviation variance are given. For species evaluation, the remaining metrics are used. Accuracy and sensitivity are same as they measure the true prediction of class. The validation metrics can also be calculated based on the accuracy data.

V. RESULTS AND DISCUSSION

A total of 7 features were generated from the 6 species of images. Features Selection methods were applied to extract the features in each image. The union of three features sets were yield 7 features in common and the majority of them are unique to the particular set. Statistical methods may not highlight the common features that bear relevance to their identities. For KNN and SVM, each feature was sought for each species of test and train images

The KNN was carried out for the species identification using the distance function. Totally three functions were used to the analysis of image. The performance assessment was done based on the accuracy for the six different species. Among the three function, the Weighted KNN shows the better performance. But, the performance assessment was not satisfied for all the species. The work was carried out to through next class identifier.

The model was carried forward for the identification of species using the multiclass SVM method. The prediction performance was assessed through accuracy for all the 6 different species. almost near perfect value was obtained for all the species. over all the multiclass SVM method suggest that some species are difficult to identify and distinguish between the others.

Compared to the KNN method, the accuracies for the multiclass SVN were found to be slightly higher. For individual feature sets and combined feature sets the results are better for SVM. The Accuracy for the combined features are high than the individual features as shown in Fig.2.

The broader objective of this work is to develop a machine learning method that would eventually identify the pest in field even to the species level and provide the necessary actions. To achieve this the training machines to recognize features that are associated with a particular species were designed. KNN and SVM classifiers are used to as machine learning algorithms for these feature sets obtained from each pest image. Both have their own advantages and disadvantages. First KNN was used to identify the insects and explore the feasibility of machine learning algorithm. KNN provided with problems like nature complexity and insight of extent. So, different methods are needed to overcome such problems. SVM based species identification was adopted and also the accuracy of prediction has been improved. This work overcomes the previous works challenges in identification of pests. When large dataset was used the improvement is needed which is currently underway. The limited number of images per species not only restricted us in sample sizes, but also made it harder to distinguish those difficult pairs. Besides the quantity of images, quality has appeared to be another issue that might have introduced artifacts in the feature extraction stage. Even though prediction models were compared under the same conditions, having actual pest fragments rather than sub-images could be closer to a real-world scenario. Some other pest images are needed to be added to the work. Hence, collection, storage, and analysis of high quality images remain as some of the future works.

VI. CONCLUSION

In summary, both KNN and SVM could be used to identify the species of pest images. The multi-class SVM method was found to be a good strategy for the species identification. It had average overall accuracy of 85%, when features selection methods were consolidated with the consensus approach. For individual species, it showed excellent genus level accuracy but could not distinguish between the pest with very similar appearance. At an individual level, the SVM method worked better for most of the species. This anomaly could be due to the vivid similarity between their patterns and only highlights that one method may not be sufficient to completely address this problem. Hence, the comparative study between these two machine learning methods builds an excellent platform for the future studies in the convolutional neural network.

REFERENCES

1. Ziyi Liu, Junfeng Gao, Guoguo Yang, Huan Zhang, Yong He, “Localization and Classification of Paddy Field Pests using a Saliency Map and Deep Convolutional Neural Network”, Scientific Reports ,vol. 6, February 2016.
2. Chenglu Wen, Daoxi Wu, Huosheng Hu, Wei Pan, “Pose Estimation-dependent Identification Method for Field Moth Images using Deep Learning Architecture”, Journal of Biosystems Engineering, Vol. 136, August 2015.
3. SitiN.A.Hassan, NadiahS.A.Rahman, ZawZawHtike and Shoon Lei Win , “Advances in Automatic Insect Classification”, Electrical and Electronics Engineering: An International Journal (ELELIJ) Vol 3, No 2, May 2011.
4. Ying Yang, Bo Peng, and Jianqin Wang, “System for Detection and Recognition of Pests in Stored-Grain Based on Video Analysis”, pp. 119–124, International Conference on Computer and Computing Technologies in Agriculture ,2011.
5. Katarina Mele, “Insect Soup Challenge: Segmentation, Counting, and Simple Classification”, IEEE International Conference on Computer Vision Workshops ,2013.
6. Zhu, L. & Zhang, Z., “Auto-classification of insect images based on color histogram and GLCM” International Conference on Fuzzy Systems and Knowledge Discovery, 2010.
7. Solis-Sánchez, L. O., García-Escalante, J., Castañeda-Miranda, R., Torres-Pacheco, I. & Guevara-González, R. “Machine vision algorithm for whiteflies (BemisiatabaciGenn.) scouting under greenhouse environment”, Journal of Applied Entomology, vol. 133,pp. 546–552,2009.
8. Zhang, H. & Mao, H, “Feature selection for the stored-grain insects based on PSO and SVM”, International Workshop on Knowledge Discovery and Data Mining ,2009.
9. Yang, H. et al., “Research on insect identification based on pattern recognition technology.”, International Conference on Natural Computation, 2010.
10. Do, M., Harp, J. & Norris, K, “A test of a pattern recognition system for identification of spiders.”, B. Entomol. Res 89, pp.217–224,1999.
11. Yu, Z. & Shen, X.,” Application of Several Segmentation Algorithms to the Digital Image of Helicoverpaarmigera”, Journal of China Agricultural University 5, 2001.
12. Arbuckle, T., Schroder, S., Steinhage, V. &Wittmann, D. “Biodiversity informatics in action: identification and monitoring of bee species using ABIS.”, International Symposium Informatics for Environmental Protection, Zurich, Switzerland. Marburg: Metropolis-Verlag., 2001.
13. Yao, Q. et al. An insect imaging system to automate rice light-trap pest identification. J INTEGR AGR 11, pp.978–985,2012.
14. Kumar, R., Martin, V. &Moisan, S., “Robust insect classification applied to real time greenhouse infestation monitoring”, International Conference on Pattern Recognition: Visual Observation and Analysis of Animal and Insect Behavior, IEEE. 2010.

15. Solis-Sánchez, L. O. et al., “Scale invariant feature approach for insect monitoring”,Comput. Electron. Agr75, pp.92–99,2011.
16. Cheng, L. &Guyer, D.,” Image-based orchard insect automated identification and classification method”,Comput. Electron. Agr, pp.110–115, 2012.
17. Xia, C., Lee, J., Li, Y., Chung, B. & Chon, T.,” In situ detection of small-size insect pests sampled on traps using multifractal analysis.”, Opt. Eng,2012.
18. Venugoban, K. &Ramanan,, “A. Image classification of paddy field insect pests using gradient-based features.”, International Journal of Machine Learning and Computing, 2014.
19. Zhang, J., Wang, R., Xie, C. & Li, R.,” Crop pests image recognition based on multifeatures fusion.”, Journal of Computational Information Systems, pp.5121–5129, 2014.
20. Yao, Q. et al.,” Automated counting of rice planthoppers in paddy fields based on image processing.”, J INTEGR AGR, pp.1736–1745, 2014.
21. Sivic, J. & Zisserman,” A. Video Google: a text retrieval approach to object matching in videos”, International Conference on Computer Vision (ICCV 2003), IEEE, 2003.
22. Ojala, T., Pietikäinen, M. &Mäenpää, T.,” Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.”, IEEE T PATTERN ANAL,pp. 971–987, 2002.
23. Lowe, D. G.,” Distinctive image features from scale-invariant keypoints”, Int. J. Comput. Vision, pp. 91–110, 2004.
24. Dalal, N. &Triggs, B.,” Histograms of oriented gradients for human detection.”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), IEEE, 2005.
25. Zhang, W., Deng, H., Dietterich, T. G. & Mortensen, E. N.,” A hierarchical object recognition system based on multi-scale principal curvature regions.”, International Conference on Pattern Recognition (ICPR 2006), IEEE. 2006.
26. Larios, N. *et al.* Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. Mach. Vision. Appl19,pp. 105–123 2008.
27. XiChengYouhua, ZhangYiqiong, ChenYunzhi, WuYiYue, “Pest identification via deep residual learning in complex background” Computers and Electronics in Agriculture, Vol. 141, pp. 351-356,2017,
28. LimiaoDeng, YanjiangWang, ZhongzhiHan, RenshiYu,” Research on insect pest image detection and recognition based on bio-inspired methods” Biosystems Engineering Volume 169, Pages 139-148,May 2018.
29. ChengjunXie, JieZhang,RuiLi, JinyanLicPeilinHong, JunfengXiaaPengChena” Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning Computers and Electronics in Agriculture, pp.123-132,2015.
30. A. Ebrahimi a, M.H. Khoshtaghazaa, S. Minaei a, B. Jamshidi b “, Vision-based pest detection based on SVM classification method”, Computers and Electronics in Agriculture, pp. 52–58,2017.
31. He-Ping Yang, Chun-Sen Ma, Hui Wen, Qing-Bin Zhan &Xin-Li Wang,” A tool for developing an automatic insect identification system based on wing outlines”, Scientific Reports volume 5, 2015.
32. Manisha Bhangea, H.A.Hingoliwalab, “Smart Farming: Pomegranate Disease Detection Using Image Processing”, Procedia Computer Science, pp.280 – 288,2015.
33. Zalak R. Barot, NarendrasinhLimbad, “An Approach for Detection and Classification of Fruit Disease: A Survey”, International Journal of Science and Research (IJSR), pp.44-50,2017.
34. HalilBisgin, TanmayBera, Hongjian Ding, “Comparing SVM and ANN based Machine Learning Methods for Species Identification of Food Contaminating Beetles”, Scientific Reports,2018.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication