Crop Pest Detection and Classification by K-Means and EM Clustering

Madhuri Devi Chodey[1], Humera Tamkeen[2]

[1]Madhuri Devi Chodey, Professor, Dept. of Electronics and Communication Engineering, Navodaya Institute of Technology, Raichur- 584103.
[2]Humera Tamkeen, Dept. of Electronics and Communication Engineering, Navodaya Institute of Technology, Raichur- 584103.

---***---

Abstract - ***India is agriculture based country which has majority of the people depend on the agriculture. It has direct effect on Indian economy. With the use of digital image processing techniques in the agriculture and addition of modern technologies will help to increase the total yield. This project focuses on crop pest detection and classification, providing information about the type of the pest and reducing the use of fertilizers in the field. The user/farmer can upload the image; from the uploaded image features are extracted. Color based image segmentation using L*a*b color space models is applied and then k-means and EM clustering is performed; the cluster providing maximum information about the pest is selected. Then percentage of affected region, various parameters such as mean, entropy, energy, RMS etc. and classification of the pest using multi SVM is performed.***

Key Words*: *Texture feature, affected region percentage, colour based segmentation, K-means, EM clustering, SVM classification.

1. INTRODUCTION

Plant disease has turned up a difficult circumstances or problem to the farmer reducing the growth in the field, plant pest and disease affecting the plant frightening the food security and significant losses. Much system has been proposed so far to control the pest in the field, most commonly and accurate method is to use digital image processing technique. Automatic pest and disease detection provide benefits in monitoring field, controlling growth of the pest at early stage, increase productivity. The Liu Liu et.al [1] has stated the risk of the agriculture due to the pest in the field. It will harm the farmer economic growth and also will cause plant growth. The author has state the region based end to end technique by using the pestnet also used the neural networks and the feature extraction for the classifying the various types of the pest species by image [10].

Sushma R Huddar et.al [2] has stated the segmentation of pest image also the counting the whiteflies in field. Pesticides have been used in agriculture which affects the plant and also the environment. With the help of the digital image processing technologies available in the system today the farmers will be able to reduce the loss. Mean filtering and adaptive thresholding has been used on the natural images which are initially processed by the image enhancement methods to identify pest has been proposed by the Yogesh Kumar [3].

K. Thenmozhi [4] has worked on the pest types which will infect sugarcane production. Images are initially preprocessed, segmented and the image features are extraction. The edge detection method of Sobel algorithm has been applied by using the 9 geometric features of the pest. Z Iqbal et.al [5] has proposed the survey on the citrus plant diseases which are caused by the different type of the pest. Mainly buy sing the shape and color feature of the image has been used for the image segmentation, image feature extraction for extracting the image components [7]. Image enhancement image de noising methods has been used for the detection of the pest images. Hanyurwimfura et.al [6] proposes system that will control the falloff the armyworm in the region of Rwanda. Internet of thing (IoT) enabled technology has been adopted by the author in this work by using sensors to detect fall armyworm in the field. The location of the pest falling will be informed to the user.

2. SYSTEM ANALYSIS

2.1 Existing system

In the past, many methods has been used for pest detection such as , Otsu's thresholding method, RGB color based, Histogram, adaptive thresholding, novel algorithm, edge detection and other methods. Using a single method for result analysis based on the system generated results will make the system complicated, as no pests will be of same shape, size or position. Hence need of user interface to detect the pest accurately.

2.2 Proposed System

In the proposed work we have used the K means clustering and EM clustering algorithms to compare and analyze the best possible result. The two algorithms will help us to understand the result based on the input type and the feature set generated. Based on selected the image cluster, the region of the pest area detected and the feature set will be generated. This will allow the system to work on the selected pixel instead of complete image to give the best result.

3. METHODOLOGY

The methodologies used in this work are as follows:

3.1 Image Preprocessing

The given input image is processed before passing it to the main system. The output of the preprocessing will be the l*a*b image which will be suitable for next processing. The preprocessing is a sequence of operation that performs on scanned input images. It primarily enhances the image illustration for higher segmentation. The task of preprocessing is to phase the required pattern from the image and perform normalization, noise filtering and smoothing. The preprocessing also defines a solid illustration of the segmented model. After segmentation, binarization procedure is used where it convert a grey scale to a binary image.

1. ***Image Resizing***: If the input image has the resolution of [1024*1600] then the processing of the system may be slow due to the higher resolution. Hence we need to convert it to the required size. For example 255*255 by using inbuilt mat lab functions.

2. ***Image Restoration***: It is a simple process of taking the corrupted pixels of image and cleaning them. The restoration process allows the user to adjust image contrast, brightness and the other features of the image. It is not a de-noising method, but allows the user to smoothening image by using the simple method like point spread function, venar filter and de convolution method etc.

3.2 Image Enhancement

The image enhancement technique is different from one field to another field according to its objective. Enhancement of the image includes the color transformation (if needed), image contrast enhancement using **imadjust().**

3.3 Color Based Image Segmentation

In this the input image is segmented based on the color bands of the lab a*b*. The segmentation will be achieved by using the k means bad the EM clustering at t multiple level. The cluster is nothing but the group of the pixel belonging to the same color bands. The equations for RGB are as follows which use in the color based segmentation is;

$$R= \sum_{x=1,y=1}^{x=w,y=h} \frac{R(I(x,y))}{w \times h} \quad (1)$$

$$G= \sum_{x=1,y=1}^{x=w,y=h} \frac{G(I(x,y))}{w \times h} \quad (2)$$

$$B= \sum_{x=1,y=1}^{x=w,y=h} \frac{B(I(x,y))}{w \times h} \quad (3)$$

3.4 K-Means and EM Clustering Models

K-means and EM algorithm are generally same and are used in common to find the natural clusters in the range of the given data by varying the input type parameters. The K-means algorithm starts with acquiring the digital image uploaded by the user. K-means is applied for color space transformation result l*a*b (luminosity and chromaticity layer) and the k-means clustering used to segment the plant pest images. Clusters are formed for pest images based on intensity or the color r the texture or the location of the input pixel.

For the proposed project we used color and intensity combination. For K-means algorithm use of the starting location of the partitions such as Euclidean distance is important to find the cluster. Based on selected number of the pass, the data value can switch the partitions with each successive pass. The partitions are represented red, green, blue and yellow axis of the color cube. Every pixel in the input image will be saved and compared with the initial partitions obtained and the nearest partitions are determined recorded. Then, RGB color of every pixel within selected partition will be used to determined and served as new value for the given partition. If the obtained partition has no pixel value associated with it, simply ignored in this implementation. The algorithm continues until the pixels are not changing their partitions associated with it or the partitions values changing by set of small amount are acceptable.

The image pixels are grouped into the similar grouped of ‘K’ as follows:

I(x,y)={image(p1,p2..pn=K(1,2,3..n)} (4)

Each of these pixels will exhibit the property based on the individual color band. Every Pixel in the image is compared with each other and side pixel for grouping. The K means clustering will use the mean value instead of the average value. The clustering of the pixel with the data value of the pixel in the same color band in the image will help the system to group the pixels as follows.

$$\phi(cluster, data) = \Sigma \left\{ \Sigma (xi - ci)^{T} \right\} (Xi\text{-}cj) \quad (5)$$

But due to range of the color value (data) in the RGB color band is 0 to 255. The data values include neighbor values, but the neighbor values are of same mean value, hence it make K-means algorithm less efficient in the exact grouping of the pixels.

The EM algorithm which is derived further on the aspects of the K-Means algorithm, First step is to choose partition and performs the processing of the pixel values on the input image in the color band of the RGB. It starts the EM cycle, first expectation is performed. For Expectation step equation is defined as

$$E[z_{ij}] = \frac{p(x = x_i \mid \mu = \mu_j)}{\sum_{n=1}^{k} p(x = x_i \mid \mu = \mu_n)} \quad (6)$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_j)^2}}{\sum_{n=1}^{k} e^{-\frac{1}{2\sigma^2}(x_i - \mu_n)^2}} \quad (7)$$

This is used to serve the weights for the lower expression. The sigma squared used in the equation of the expectation gives the covariance value of the pixel. E step will compute the weight or expectation of the pixel for every partition than next step is to perform the maximization or M step. The equation for the maximization is given by

$$\mu_j \leftarrow \frac{1}{m} \sum_{i=1}^{m} E[z_{ij}]\, x_i \quad (8)$$

Partition value of the pixel j is changed to an average value of weight pixel; the user is allowed to choose the best possible result for the feature extraction based on the clustering of the pixel.

3.5 Feature Extraction

Feature Extraction of the pest images are performed by using function regionprop(). Feature extraction is of two types.

- Extraction of Feature in pattern
- Extraction of Feature in Texture

With the help of GLCM, pixels of pairs information are collected, occurrence of the pixel brightness in an image exhibits by the GLCM. The valued matrix is created at distance d=1 and angles which are represented in the degree ranges of (0, 45, 90,135). It provides the stats like entropy, energy, contrast and correlation. For texture character profile such as smooth, silky, and rough GLCM is used. GLCM is prepared from the gray scale values and picks up the relationship between two neighboring pixel at a time. GLCM implementation for input image with 8 tones is shown below

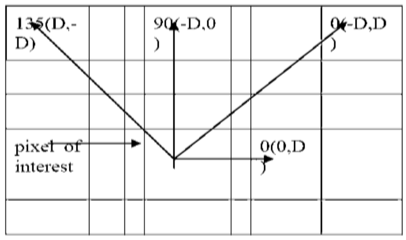


Figure.1. GLCM implementation for 8 tone image

Features are extracted using GLCM matrix such as correlation, contrast, energy, entropy, inverse different moment, homogeneity, smoothness, mean, standard deviation, variance, skewness, RMS.

3.6 SVM Based Classification

SVM (Support Vector Machine) is a type of the supervised machine learning method which will examine data and identify the similar type of the patterns, which are used for the later classification. SVM model has been able to model the complex structure of the non-linear decision boundaries with high accuracy and SVM is efficiently used for the binary classification. The SVM is to classify the data set with boundaries and extent it to nonlinear boundaries. SVM becomes prominent when pixel map is used as the dataset values as input. It gives high accuracy equivalent to neural network with elaborated features. By designing the kernel function, SVM can be applied to the complex data and this model is efficient in both linear and nonlinear data handling. Support vector machine consists of two approaches:

- linearly separable
- nonlinearly separable

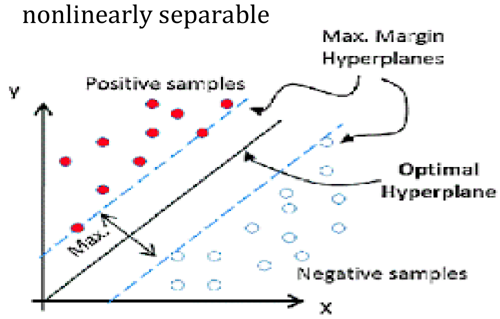


Figure.2. General SVM classsifier

As shown in the above diagram, the SVM classifier drawn the patterns from the input sample (in our case it is an image). SVM is composed of the followings:

- Hyper-planes: among them the optimal and Maximum- minimum margin hyper planes are used in the classification of the patterns.

- Negative samples, which will drop below the required threshold values
- Positive samples, which will drop above the threshold values

4. PROPOSED SYSTEM ARCHITECTURE

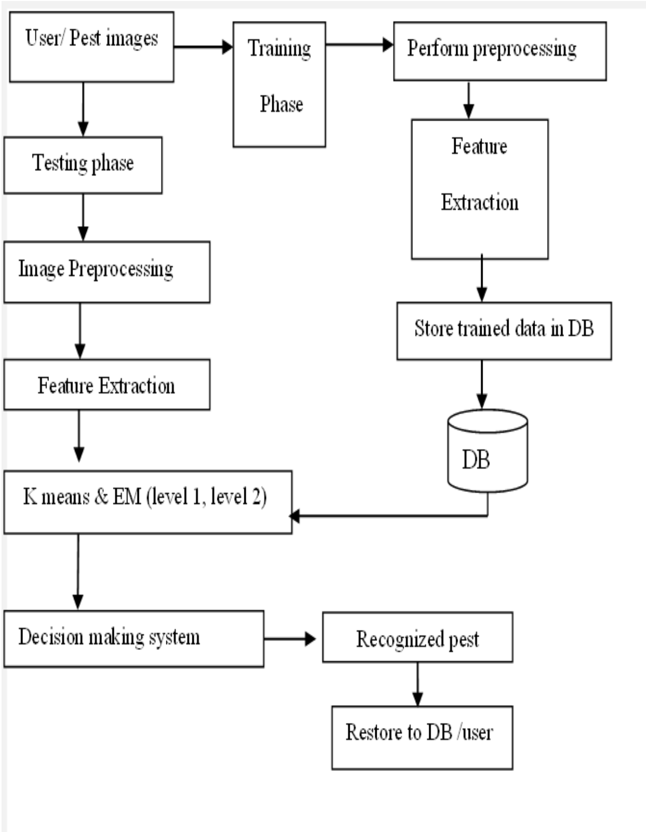


Figure.3. Proposed System Architecture.

Basic system architecture is shown in the figure. First the user pest images are collected and saved in the folder, while running the program user is allowed to detect the image if interest which will be subjected to pest detection and classification. During training phase the entropy value of all the images will be store in DB. While in testing phase, load the image of interest. Selected image then applied for image preprocessing; during the image preprocessing the image quality is improved by image enhancement feature and de-noising of the image by using median filter.

The output of image preprocessing is then applied to the feature extraction; where the feature such shape, texture and color are stored in DB which will be helpful for classification of the images. Color based segmentation is performed and detection is performed using K-Means clustering algorithm and EM (expectation and maximization).L*a*b method is used for color based segmentation. The clustering output is then used to measure the quantity such as entropy, smoothness, kurtosis, RMS value, energy, correlation, entropy and the other parameters.

The decision making system perform more the 500 iteration to detect the pest with the help of SVM classifier. The detected pest is restoring to the user. Hence using the proposed system pest can be easily detected and classification of pest is possible at the early stage with minimizing the effect of the pest on the agriculture field.

5. EXPERIMENTAL RESULTS

Initial GUI of the proposed work is shown in below figure, first step is the uploading the pest image.

Step 1: Click on the "LOAD IMAGE" to select the pest image from the image folder whose detection and classification is too performed

Figure.4.GUI of the proposed work

Step 2: Click on the "ENHANCE IMAGE" button on the GUI of the proposed work. EnhanceD image will be inserted in the "ENHANCE IMAGE" table shown in the figure.

Step 3: Click on the "CLUSTERING AND SEGMENTATION" on the GUI of the proposed work for performing clustering and segmentation, the clustering result is displayed as shown in the figure.

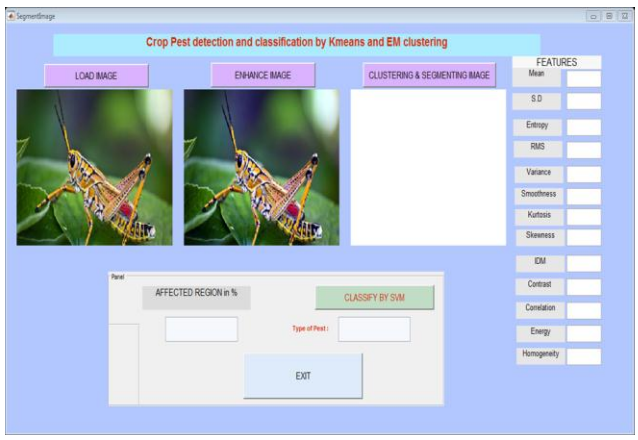


Figure.5. Input image and image enhancement

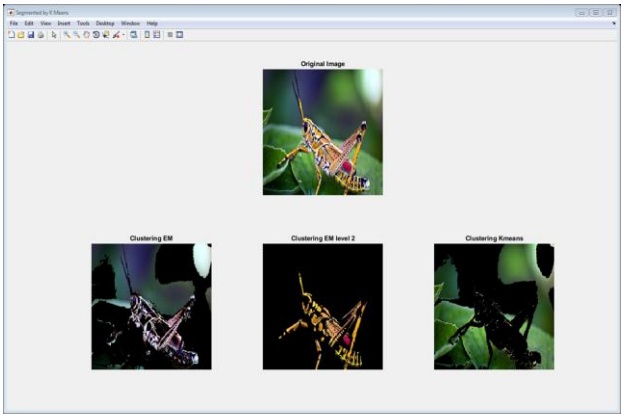


Figure.6. Image clustering by K-means and EM Clustering

Figure.7. Selected cluster image with feature set and affected region percentage

Step 4: Perform classification by using SVM; click on the "CLASSIFY BY THE SVM" button on GUI of the proposed work. Based on 500 iteration pest will be detected and classification will be displayed as shown in the below figure. Dialog box will be display with message "pest found" and another dialog box with name of the pest detected is display.

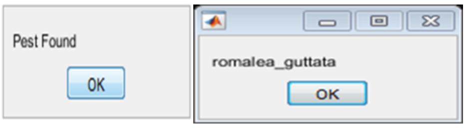
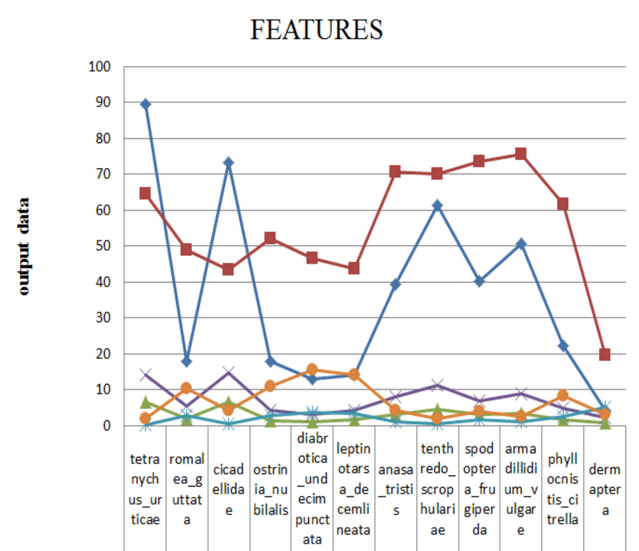


Figure.8. Classification of the image

The plot for the features listed in the table is shown below



	tetranychus_urticae	romalea_guttata	cicadellidae	ostrinia_nubilalis	diabrotica_undecimpunctata	leptinotarsa_decemlineata	anasa_tristis	tenthredo_scrophulariae	spodoptera_frugiperda	armadillidium_vulgare	phyllocnistis_citrella	dermaptera
MEAN	89.61	18.02	73.21	18.05	13.175	14.21	39.37	61.53	40.13	50.8	22.37	4.37
SD	64.63	48.88	43.39	52.23	46.61	43.66	70.8	69.94	73.64	75.5	61.68	19.58
ENTROPY	6.8	1.93	6.69	1.55	1.15	1.66	3.23	4.71	3.16	3.46	1.61	0.87
RMS	14.32	5.62	14.85	4.27	3.09	4.44	8.16	11.32	6.97	8.85	5.04	2.28
SKEWNESS	0.2	2.86	0.48	2.98	3.67	3.39	1.16	0.64	1.611	1.1	2.63	5.32
KURTOSIS	1.99	10.49	4.47	10.91	15.55	14.1	4.3	2.07	4.09	2.72	8.42	3.17

Figure.9. Plot for features obtained for twelve pest species images.

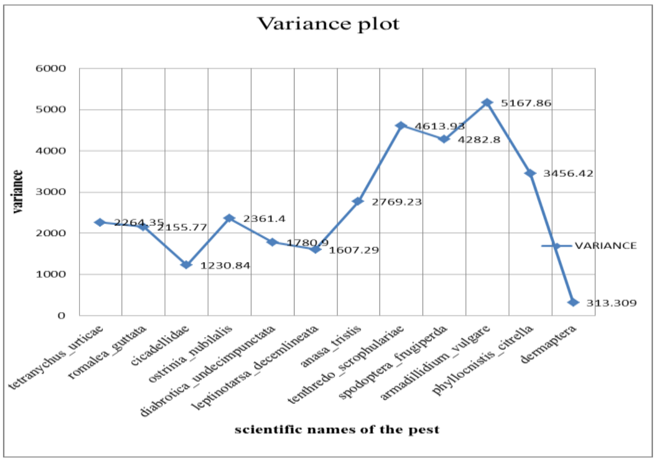
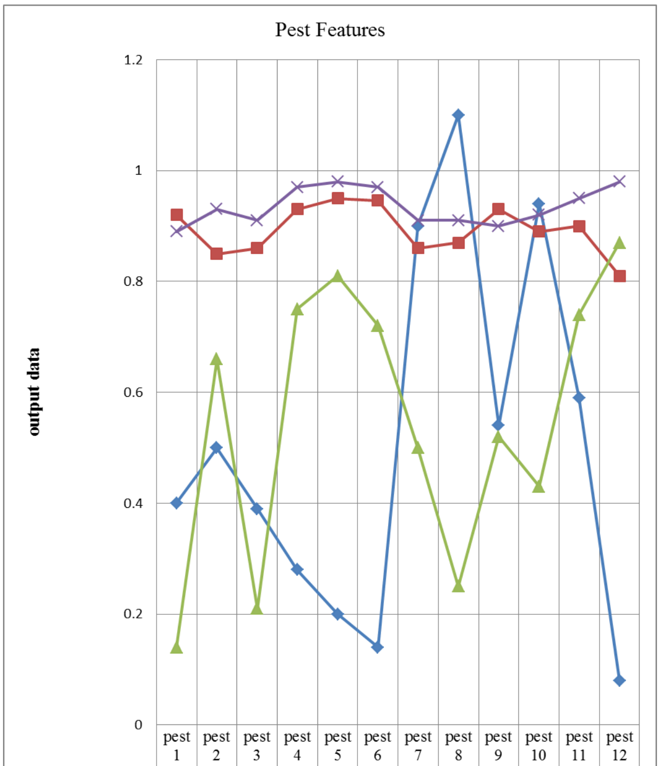


Figure.10. Plot for variance feature obtained for Twelve Species of pest images



	pest 1	pest 2	pest 3	pest 4	pest 5	pest 6	pest 7	pest 8	pest 9	pest 10	pest 11	pest 12
contrast	0.4	0.5	0.39	0.28	0.2	0.14	0.9	1.1	0.54	0.94	0.59	0.08
correlation	0.92	0.85	0.86	0.93	0.95	0.946	0.86	0.87	0.93	0.89	0.9	0.81
energy	0.14	0.66	0.21	0.75	0.81	0.72	0.5	0.25	0.52	0.43	0.74	0.87
homogeneity	0.89	0.93	0.91	0.97	0.98	0.97	0.91	0.91	0.9	0.92	0.95	0.98

Figure.11. Plot of contrast, correlation, energy and homogeneity for twelve species of the pest images

6. CONCLUSION

In this work we have proposed pest type detection and identification by using DIP methods as explained in the above segments. Proposed system of detection and identification of pest by using feature set is performed on the database as the training, in which the feature values are recovered and store in the system database. User's input sample image is passed to the system to extract input pest image features; the image is than processed in multiple stages by using region based segmentation of the image by using the K-means clustering and EM based level clustering to generate the dataset of features to give final result. With the help of SVM to classify the pest and the future work of the proposed method includes recognizing the disease in multiple pests, providing cure and precaution messages to user through mobile applications.

REFERENCES

[1] Liu Liu ,Rujing Wang , ChengjunXie, Po Yang, "PestNet: An End-to-End Deep Learning Approach for Large-Scale Multi-Class Pest Detection and Classification," IEEE Access, vol.7,pp. 2169-3536, 2019.

[2] Sushma R Huddar, Swarna Gauri, "Novel algorithm for segmentation and automatic identification of pests on plant using image processing," IEEE 2018.

[3] Yogesh Kumar, Ashwini Kumar Dubey,"Pest detection using adaptive thresholding," IEEE 2017.

[4] K.Thenmohiz,U.Srinivasaulu Reddy,"Image processing techniques for insect shape detection in field crops," IEEE 2017,pp. 978-1-5386-4031-9.

[5] Z. Iqbal et al.,"An automated detection and classification of citrus plant diseases usingimage processing techniques: A review," ELSEVIER 2018, Computers and Electronics in Agriculture 153 (2018) 12–32.

[6] Hanyurwimfura Damien, Nizeyimana Eric, Ndikumana Faustine,"Monitoring System to Strive Against Fall Armyworm in Crops Case Study: Maize in Rwanda," 2018 IEEE SmartWorld.

[7] Ji'An Xia, YuWang Yang, HongXin Cao,"Performance analysis of clustering method based on crop pest spectrum,"2018 Asian Agricultural and Biological Engineering Association. Published by Elsevier B.V.

[8] Junfeng Gao, David Nuyttens,"Recognizing weeds in a maize crop using a randomforest machine-learning algorithm and near-infrared snapshot mosaic hyper spectral imagery,"2018 IAgrE. Published by Elsevier Ltd.

[9] Preetha Rajan, Radhakrishnan. B."Detection And Classification Of Pests From Crop Images Using Support Vector Machine",2016 International Conference on Emerging Technological Trends [ICETT],IEEE 2016

[10] Milan Maric, Irena Orovic," Compressive Sensing based image processing in TrapView pest monitoring system",MIPRO 2016.

[11] Al-Hiary, H., S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. AlRahamneh. 2011. "Fastand accurate detection and classification of plant diseases". International Journal of Computer Applications, 17(1): 31-38

[12] Argenti, F., L. Alparone, and G. Benelli. 1990. "Fast algorithms for texture analysis using co-occurrence matrices." IEEE proceedings, 137, (6): 443-448

[13] Chih-Wei, H., and C. Lin. 2002." A comparison of methods for multiclass support vector machines." IEEE Transactions on Neural Networks, 13(2): 415-425