Feature Optimization of Whitefly Detection Algorithm using Image Segmentation and Feature Analysis

Honse-Al-Walid
Associate Professor,
Department of Computer
Science & Engineering
Ahsanullah University of
Science and Technology
Dhaka, Bangladesh

Md. Tanjim-Al-Akib
CSE Graduate,
Department of Computer
Science & Engineering
Ahsanullah University of
Science and Technology
Dhaka, Bangladesh

Nusratul Islam
CSE Graduate,
Department of Computer
Science & Engineering
Ahsanullah University of
Science and Technology
Dhaka, Bangladesh

ABSTRACT

With the increase of agricultural production worldwide, the economic importance of early detection of crop pests is more evident than ever before. The purpose of this study is to improve the performance of crop pest detection algorithm using image processing techniques. This study combines the object extraction, feature extraction and classification of a particular crop pest species using the SVM classifier. The end point of this study is to achieve optimal performance and accuracy.

Keywords

Image Processing, Image Segmentation, Feature Extraction, Pest Detection, Whitefly, Blob Extraction, SVM Classifier, Performance Analysis

1. INTRODUCTION

About 75 percent of the world's poor people live in rural areas, and most of them are involved in farming. [1] But in order to attain the best production, people need to protect plants from various kinds of pests. In recent times, many researches have focused on detecting pests in agricultural crops. Crops infected by pests must be detected at an early stage in order to facilitate direct action for prevention of further infection. There are many existing manual methods to detect pests in crops but they all tend to be time consuming. With the help of latest technologies and rising advancement of image processing research field, a more efficient system for early detection of pests in crops can be developed.

Pesticides is one of the preventive measures taken against pest occurrence, but its long-term usage is not feasible in many ways because they can be very harmful for both the environment and the animals that come in contact to the plants. Usage of pesticides reduces nitrogen fixation and thus it decreases the production of crops. Therefore, early detection of crop pests can play a huge role to achieve better productivity in the field of agriculture. In this study, the primary focus is on studying the existing pest detection algorithms, optimizing them by applying different techniques and improving performance through analysis.

2. IMPORTANCE OF EARLY PEST DETECTION

Huge amount of crops are destroyed yearly around the world because of pests and diseases. The number is quite shocking and it ranges from 10%-16% of the total worldwide agricultural production [2]. Farmers tend to use pesticides to prevent pests and diseases to attack the crops but as a matter of fact those pesticides sometimes become more harmful than the pests or the disease itself.

Whiteflies generally suck phloem sap and a large number of them can turn leaves into yellow, appear dry and sometimes to fall off the plants. When these whiteflies lay eggs or the time when nymphs grow bigger to an adult whitefly, they suck out a lot of phloem sap causing plant distortion, silvering of leaves and possibly serious losses in some vegetable crops. Some whiteflies even transmits viruses to some plants. Sometimes small group of them doesn't affect much but it's very important to notice issues that may cause infestation later on [3].

Heavy infestation of whiteflies can lead to the cause of death of plant. Even late detection of these pests can cause heavy loss for the farmers. After a certain delay in taking actions sometimes farmers panic and use a lot of pesticides which is also very harmful for the crops. The goal of this work is to develop a method which will be robust, easy to adopt and automate the identification process based on computer vision. Nowadays smartphone cameras are very sophisticated and they provide high resolution photos. Using image processing techniques it is possible to detect pests in their early stages of infestation. But in order to apply pest detection algorithms to portable devices like smartphones, the detection algorithm needs to have optimized performance and accuracy.

3. LETARATURE OVERVIEW

Insect detection using static images method was first given by Paul Boissard, Sabine Moisan [4]. The actual algorithm computation took hours and the slow scanning process included risks of pests flying away from the crops, resulting in a faulty image database.

Another study conducted by Sandip P. Bhamare and Samadhan C. Kulkarni [5] used shape and texture extraction methods for detecting Black Sigatoka insect in banana trees. Their algorithm showed promise, but lacked a viable image database and a detection accuracy rate. In another study conducted by Goutham and Tejaswini, detects pests in tomato plants by extracting Region of Interest (ROI) from tomato plant images.

Usage of pan tilt camera to acquire images in a greenhouse was introduced by Rupesh G. Mundada and Dr. V. V. Gohokar [6]. Their algorithm is automated and shows a 100% accurate detection rate, but is not optimized.This study

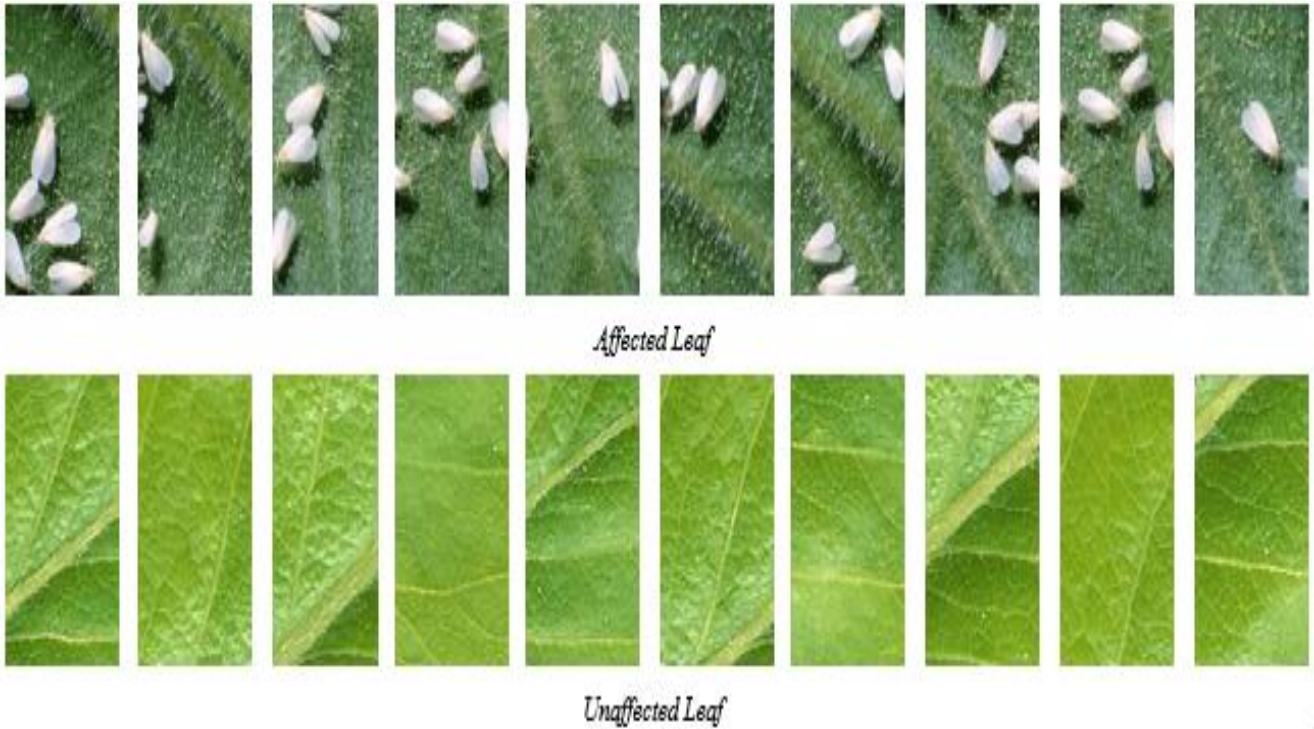primarily focuses on [6] and seeks to optimize its performance and accuracy.

Fig 1: Image Database

4. OVERVIEW OF EXISTING ALGORITHM

The existing algorithm for detection and classification of pests in greenhouse is [6]. Their study focuses on whiteflies as early detection of this pest which is crucial for effective intervention.

Image database for their study was acquired through pan tilt camera with zoom within the boundaries of a greenhouse. For this study the same image database was chosen as the image quality is optimal. A portion of the acquired image database for is shown in Fig 1.

The algorithm [6] use several stages of image processing techniques. Their methodology, consists of three main working stages, namely: pre-processing, detection and classification. The elaborate steps of this algorithm are displayed in Fig. 2.

The images used in [6] were saved in common image formats (JPEG, PNG) and the image capture system was directly interfaced with the detection system. Pre-processing stage includes the following steps

- RGB to Grayscale Conversion
- Image Resizing
- Image Filtering.

A set of seven image properties are chosen as features. They are:

- Entropy [7]
- Mean
- Standard Deviation [8]
- Contrast [9]
- Correlation [10]
- Eccentricity [6]
- Euler Number

Based on the feature set defined in the previous section, [6] uses the binary SVM classifier for detection of whiteflies. Using this classifier, [6] demonstrates a 100% success rate in detecting leaves affected with whitefly.
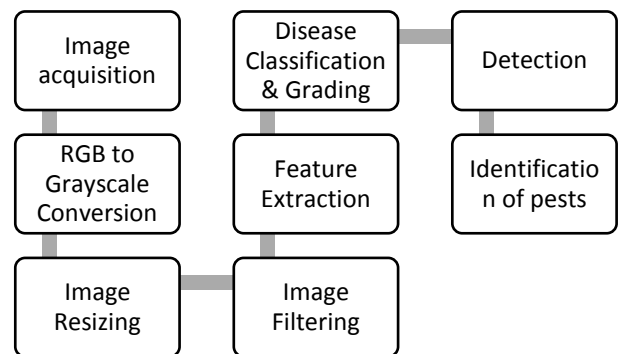

Fig 2: Steps of the Mundada-Gohokar Detection Algorithm

Table 1. Feature Set for Affected Leafs

	Entropy	Standard Deviation	Contrast	Mean	Correlation	Euler Number	Eccentricity
	7.3803	51.942	51.942	110.17	0.99668	78	0.5918
	7.3504	46.056	46.056	126.58	0.99556	121	0.6452
	6.963	33.666	33.666	108.48	0.99075	199	0.69273
	7.1334	43.296	43.296	118.91	0.99535	49	0.54657
	6.9486	37.231	37.231	106.56	0.99417	101	0.66372
	7.1385	40.904	40.904	121.46	0.99488	107	0.56102
	7.1221	43.344	43.344	121.01	0.99483	84	0.6187
	6.5884	28.999	28.999	122.46	0.98892	213	0.6486
	7.0988	40.127	40.127	127.06	0.99418	105	0.63682
	6.6421	28.735	28.735	122.99	0.99032	208	0.66292
Mean	**7.03656**	**39.43**	**39.43**	**118.568**	**0.993564**	**126.5**	**0.626808**

Table 2. Feature Set for Empty Leafs

	Entropy	Standard Deviation	Contrast	Mean	Correlation	Euler Number	Eccentricity
	5.4556	11.94	11.94	147.76	0.95978	92	0.61338
	5.5588	12.674	12.674	143.25	0.961	1	0.86603
	6.0448	17.31	17.31	147.71	0.97499	133	0.68753
	6.4417	21.716	21.716	149.81	0.98329	1	0.86603
	6.6703	24.807	24.807	145.36	0.98702	55	0.68813
	6.4218	20.884	20.884	142.16	0.98215	81	0.7262
	6.618	25.721	25.721	154.9	0.987	10	0.78578
	6.903	31.958	31.958	142.13	0.9919	43	0.77838
	6.7352	27.689	27.689	136.81	0.98968	79	0.67525
	6.4034	20.657	20.657	150.06	0.98463	91	0.67251
Mean	**6.32526**	**21.5356**	**21.5356**	**145.995**	**0.980144**	**58.6**	**0.735922**

5. ANALYSIS OF EXISTING ALGORITHMS

This study takes a more in-depth look at [6] determine the accuracy and aims make [6] more efficient by reducing the number of features used in the study.

5.1 Importance of Feature Reduction

According to a study made by Andreas G. K. Janecek and Wilfried N. Gansterer [11] in application fields where the search for the optimal feature set is still a topic of active research. It can be observed that the accuracy of classification achieved with a reduced feature set is often significantly better than using a full feature set. In an active field such as image processing and computer vision, the search for the perfect feature set is continually progressing. Reducing the size of the feature set can make classifier algorithms more accurate and reduce time needed for the full execution of the classification algorithm

5.2 Percentage of Varience

For understanding detection algorithm [6], at first, the image database that was used in [6] is obtained. Then the images are converted to grayscale, resized and filtered according to the specifications mentioned in the previous study. Then the

seven features are extracted from the image and prepared into a simple feature vector. By repeating this process, a feature vector for each of the sample images which are shown in Fig.1 is obtained. A portion of the sample feature vector set obtained through this method is shown in Table 1. Note that since the current concern is with training images, it is already known which data are from images affected with whitefly and which are empty leaf data. The feature set for the whitefly affected leaves is given in Table 1 and the feature set for empty leaves is given in Table 2.

Now for each of the columns in Table 1 and Table 2, the mean of the features is calculated. After the mean is calculated their variation with respect to one another is obtained.

That is, the percentage of variation is defined by the ratio of the difference between two means and the highest value of mean among the two data. This percentage provides a basic idea about which feature varies more in an image for a leaf affected with whitefly versus an empty leaf. The higher the percentage of variation, the more viable that feature is in terms of significance to whitefly detection.

5.3 Result Analysis

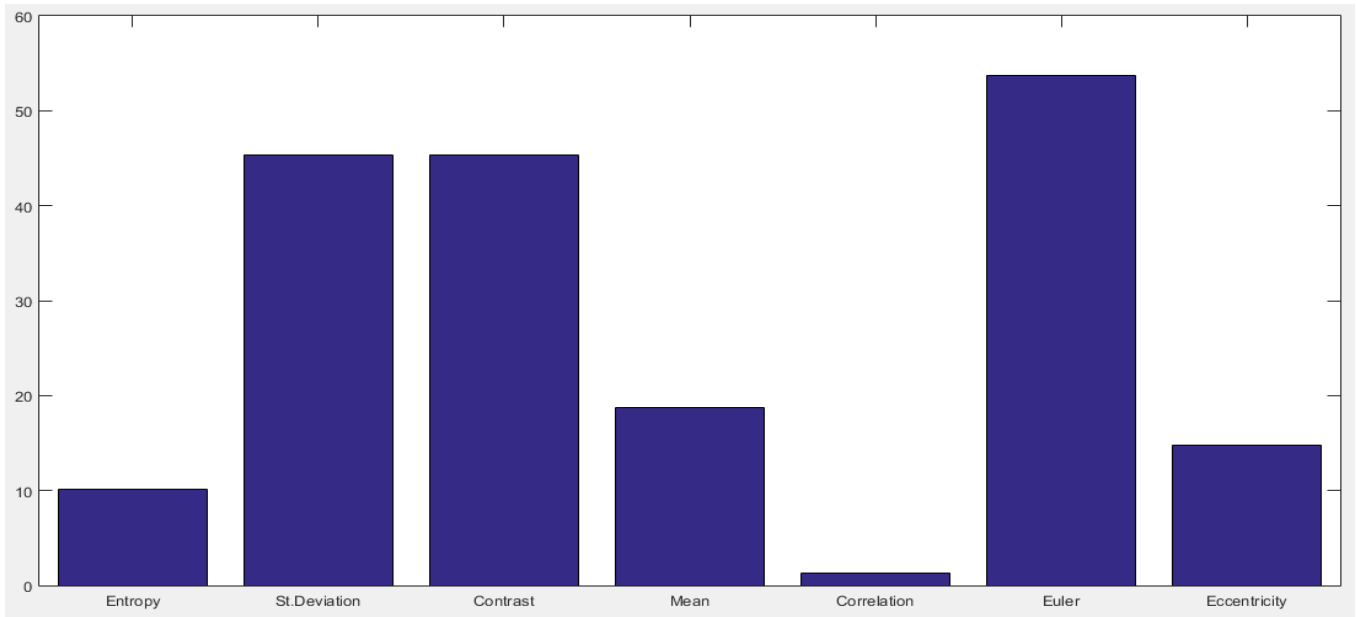The results obtained through the calculation of percentage of variation are shown in Fig 3.

Fig 3: Percentage of Varience

Note that the bar for standard deviation and contrast have the same exact percentage of variance. It is also noted that the percentage of variance is surprisingly low for the features correlation and entropy. With the percentage of variance being under 10% these two features can be eliminated from the feature set for their insignificance to pest detection. Fig 1.

So in conclusion to this analysis, it is noted that the features contrast, entropy and correlation do not vary much between affected and unaffected images, and as such, can be eliminated from the optimized feature vector. The goal of this study will now be to suggest new significant feature extraction methods and optimize the existing algorithm.

6. PROPOSED ALGORITHM

In image processing, there are two principal techniques for object detection in images. One of them is the aforementioned feature extraction method, the other method is known as image segmentation, or object extraction. While feature detection deals with extracting data points from an image based on some of its property, object extraction takes a more physical approach to the image with a view to focus on detecting actual edges and other visible characteristics in the image. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze [12]. Unlike other greenhouse pests, whiteflies are visually significant. For this reason image segmentation and object extraction is a feasible technique for detecting whiteflies in leaf images. Therefore, with the aim of improving the existing [6], this study aims to combine the two principal image processing techniques of image segmentation and feature extraction. The proposed algorithm, as depicted in Fig 4, uses the blob extraction method of image segmentation, obtaining physical characteristics data from the image.

6.1 Image Acquisitions

This study uses the same image database as [6]. This image database, as depicted in Fig 1 is of .jpeg format and consists of two classes of images: a) Affected with whitefly, and b) Healthy, unaffected leaf.

6.2 Image Pre-processing

The images are converted to grayscale, resized and then smoothened to eliminate background noise. The pre-processing steps are discussed in detail here:

6.2.1 Resize

Different sample of different sizes needed to be brought uniformity to the images so that computation and image processing can be same for all images. For this purpose all the images are resized to 512x720px at the very beginning of processing.

6.2.2 RGB to Grayscale Conversion

Digital images are classified in four main types according to its format of pixel: binary images, grayscale images, indexed images, and true color images. [13]. In grayscale images the value of a pixel is a scalar who represents an intensity or gray level within some range. The value of a pixel in indexed images is a scalar who represents an index. It indicates the color of the pixel in a color map. In true color images the value of a pixel is a vector. This vector indicates the color of the pixel according to some color space (e.g. RGB, CMYK).

A true color image can be converted to a grayscale image by eliminating the hue and saturation information while retaining the luminance. The RGB image is 3 dimensional, while a grayscale image is 2 dimensional.

There is a formula to convert RGB to grayscale image used in this study is as follows:

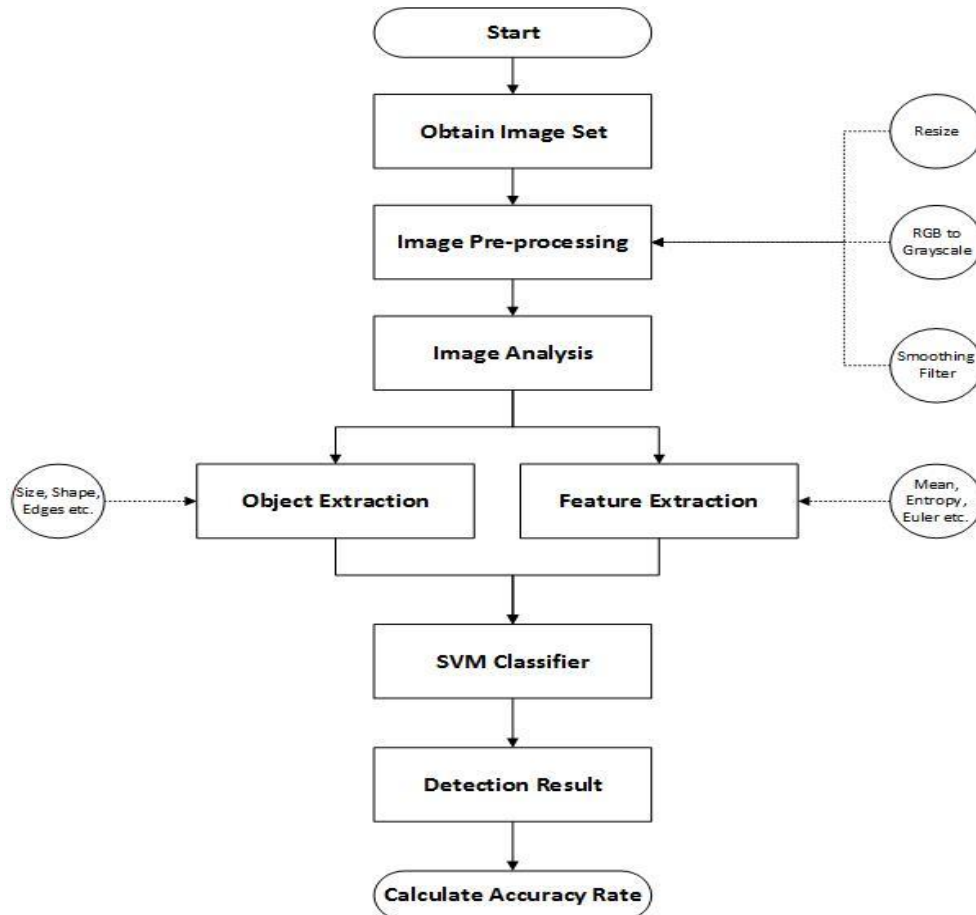**I(x,y)=(0.2989×Red)+(0.5870×Green)+(0.1140×Blue)[14]**

Fig 1: Proposed Algorithm

6.2.3 Filtering

The image is filtered to get rid of the environmental noises, which helps to detect the objects. Low pass filters, high pass filters, median filters etc are some methods of noise reduction. The low pass filters are smoothening filters whereas the high pass filters are sharpening filters. The traditional median filters are used to reduce salt-paper noise. In this study smoothening filter is used to reduce noise and to improve the visual quality of the image. The simplest smoothening filter is average filter which consists of a 3x3 matrix of 1 and is divided by 9sample of different sizes needed to be brought uniformity to the images so that computation and image processing can be same for all images. For this purpose all the images are resized to 512x720px at the very beginning of processing.

6.3 Image Analysis

In this stage of the algorithm two different image processing techniques are followed to obtain a significant and optimized feature vector. For image segmentation, the blob detection method is applied for detecting the region containing whitefly in images.

6.3.1 Blob Detection

Informally, a blob is a region of an image in which some properties are constant or approximately constant; all the points in a blob can be considered in some sense to be similar to each other. [15]. After the sample images have been put through the aforementioned pre- processing procedures, the resultant images are fixed-size, grayscale and have minimum noise. The images are now analyzed by drawing a histogram using the intensities of the grayscale image. Histogram is an estimate of the probability distribution of a continuous variable and was first introduced by Karl Pearson [16]. The histogram of an affected image vs. an empty leaf image, as demonstrated in   Fig 5, provides an interesting insight to the distribution of the pixels in the image. In case of the whitefly affected leaf image, the histogram demonstrates a clear disparity between intensity of the background of the leaf and foreground of the pest. This forms two peaks in the histogram. The greater peak culminating at intensity level 100 and the smaller peak occurring near 220- 250 intensity level. The boundary of this disparity is used as a threshold value, which in the case of Fig 5 is marked   near 200. The threshold value in turn is used to convert the grayscale images to binary images. The result of this conversion is shown in Fig 6. It is observed that with the proper use of thresholding, the whitefly affected regions have been almost accurately segmented out of the image. The segmented region of interests, or more informally 'blobs' can now be studied for further physical features.
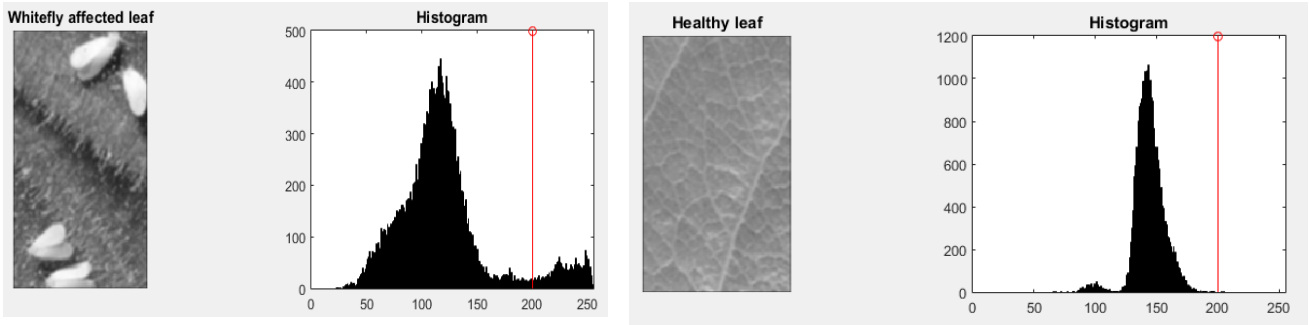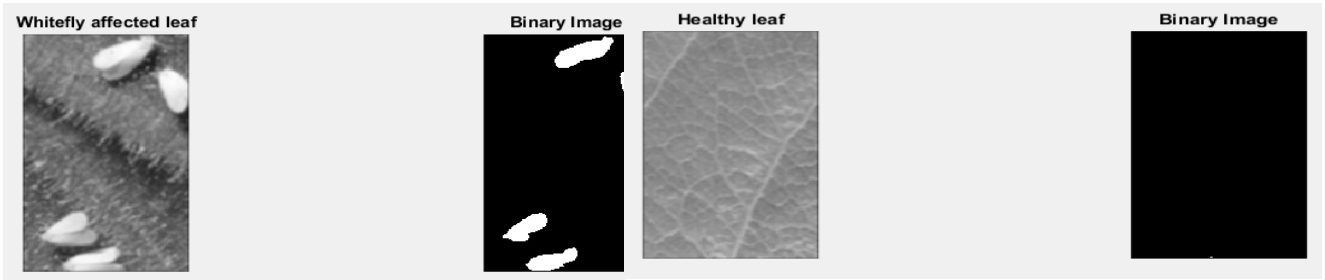
Fig 5: Histogram of Sample Image

Fig 2: Conversion to Binary Image

6.3.2 Blob Property Extraction

Now that the outlines of the whiteflies have been detected, mathematical analysis to measure their diameter, centroid and area can be performed. These data provide insight to the measurement of the pests, as well as the total affected area. These data can be significant in application of a pest-detection program because it is important to know how much a crop leaf has been affected by a pest for consideration of possible treatment methods.

The data extracted from the sample images using image segmentation is displayed in Table 3 and Fig 7 demonstrates a demo application that can count the number of whiteflies affecting the crop and the total affected area using image segmentation. This application gives a measure of how much area is affected by the pest. If the pixel density of the captured image, the focal length and the camera distance is known then the real world affected area in mm^2 can be calculated.
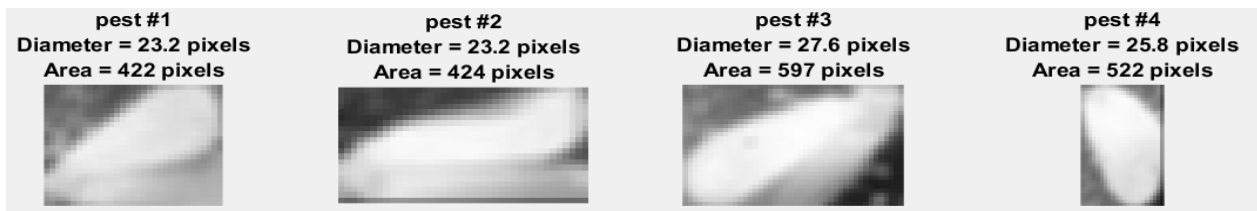
Table 3. Features Obtained from Image Segmentation

Probable Pest Number	Mean Intensity	Affected Pixel Area	Pest Diameter
14	230.37	2437	40.518
12	231.27	2492	44.883
7	230.2	470	19.214
7	233.68	2296	44.796
3	236.12	1254	24.482
6	231.35	1390	32.495
7	234.63	2165	27.194
3	234.99	870	24.597
5	234.36	2026	36.984
1	234.03	785	31.615

a) For Affected Leaf

Probable Pest Number	Mean Intensity	Affected Pixel Area	Pest Diameter
1	228.5	2	1.5958
0	0	2	0
1	226	1	1.1284
1	227	1	1.1284
3	226.67	3	1.5595
2	230	2	1.3621
4	228.19	10	2.778
5	229.15	14	3.0668
3	229.92	12	3.119
1	228	4	2.2568

b) For Empty Leaf

pest #1 Diameter = 23.2 pixels Area = 422 pixels
pest #2 Diameter = 23.2 pixels Area = 424 pixels
pest #3 Diameter = 27.6 pixels Area = 597 pixels
pest #4 Diameter = 25.8 pixels Area = 522 pixels

Total number of pest detected = 4
Total Affected Area = 1965.0

Fig 3: Application of Blob Detection

6.4 Feature Optimization & Classification

After elimination of features who had little to no effect on the classifier's performance, the features that have been selected for the optimized feature set are: Mean, Standard Deviation, Euler number and Eccentricity. These are the features that are provided to a SVM classifier for classification and identification of whitefly in images. These features are interfaced with learning algorithms that are able to analyze data and detect known patterns [17]. If given a set of training examples, where each entry is labeled to belong to a particular class, an SVM classifier builds a model that assigns new test data to one category or the other, making it a non-probabilistic binary linear classifier. New test data samples are then mapped into that same space and predicted to belong to a category based on which side of the decision boundary they fall on. In addition to linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick.

This trick consists of implicitly mapping training and test data into higher-dimensional feature spaces, when data is not linearly separable. This study makes use of a binary SVM classifier which takes as training input the reduced feature sets.

7. RESULT ANALYSIS

The training set for SVM classifier is shown in Fig.8 and Table 4 and Table 5 shows the test result and output of the SVM classifier using the reduced feature set. Note that the feature 'Blob Diameter' has been added to the feature set. It is found that the SVM has been able to classify all test datasets successfully using these features. So the accuracy rate is 100%. When both [6] and the new revised algorithm were run in 30 attempts, their run times varied significantly, as shown in Fig 9. The average difference between the runtime of the two algorithms is approximately 2.32 seconds, which is a 36.14% performance improvement.
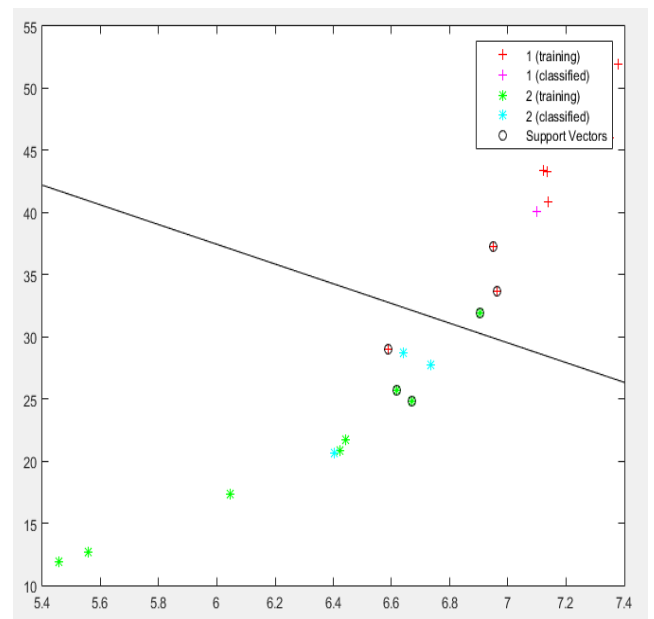
Table 4. SVM Train Set

Standard Deviation	Mean	Euler Number	Eccentricity	Diameter	Labels
51.942	110.17	14	0.68268	40.518	Whitefly
46.056	126.58	12	0.81433	44.883	Whitefly
33.666	108.48	7	0.50458	19.214	Whitefly
43.296	118.91	7	0.72846	44.796	Whitefly
37.231	106.56	3	0.94435	24.482	Whitefly
40.904	121.46	6	0.69777	32.495	Whitefly
43.344	121.01	7	0.89161	27.194	Whitefly
28.999	122.46	3	0.97309	24.597	Whitefly
11.94	147.76	1	0.86603	1.5958	Empty
12.674	143.25	0	0	0	Empty
17.31	147.71	1	0	1.1284	Empty
21.716	149.81	1	0	1.1284	Empty
24.807	145.36	3	0	1.5595	Empty
20.884	142.16	2	0	1.3621	Empty
25.721	154.9	4	0.18298	2.778	Empty
31.958	142.13	5	0.32047	3.0668	Empty

Table 5. SVM Test Set & Result

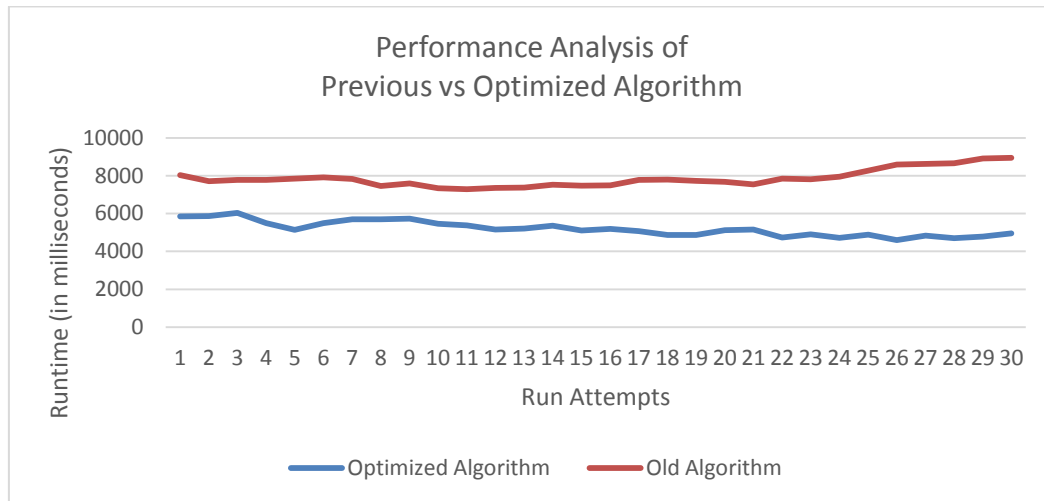| Standard Deviation	Mean	Euler Number	Eccentricity	Diameter	Result
40.127	127.06	5	0.89505	36.984	Whitefly
28.735	122.99	1	0.89132	31.615	Whitefly
27.689	136.81	3	0.88947	3.119	Empty
20.657	150.06	1	0	2.2568	Empty

Fig 4: SVM Output

Fig 5: Performance Analysis Chart

8. FUTURE SCOPE

The importance of a fast and reliable solution that can effectively detect crop pests is undeniable. With that in mind, the optimization of detection algorithms is crucial. In order to implement detection algorithms in feasible user applications, the algorithm needs to be runtime efficient. The goal of this study is to achieve such optimization that allows portable devices like cell phones with working camera implement these pest detection algorithms. So the future scope of this study will include further optimization using other non-linear classifiers, and to apply these algorithms into feasible mobile applications, so that the general user can benefit from modern day image processing researches.

9. REFERENCES

[1] Kwadwo, Asenso-Okyere, K. Davis, and D. Aredo, "Advancing agriculture in developing countries through knowledge and innovation," Synopsis of an international conference. International Food Policy Research Institute (with the German Agency for Technical Cooperation (GTZ) among others) in Addis Abeba. 2008.

[2] Chakraborty, Sukumar, and Adrian C. Newton, "Climate change, plant diseases and food security: an overview," Plant Pathology 60.1 (2011): 2-14.

[3] Rasdi, Mohd Z., et al., "Population ecology of whitefly, Bemisia tabaci, (Homoptera: Aleyrodidae) on brinjal," Journal of Agricultural Science 1.1 (2009): 27.

[4] Boissard, Paul, Vincent Martin, and Sabine Moisan, "A cognitive vision approach to early pest detection in greenhouse crops," computers and electronics in agriculture 62.2 (2008): 81-93.

[5] S. P. Bhamare and S. C. Kulkarni, "Detection of Black Sigatoka on Banana Tree using Image Processing Techniques," IOSR Journal of Electronics and Communication Engineering, pp. 60-65.

[6] Mundada, Rupesh G., and V. V. Gohokar, "Detection and classification of pests in greenhouse using image processing." IOSR Journal of Electronics and Communication Engineering 5.6 (2013): 57-63.

[7] Shannon, Claude Elwood, "A mathematical theory of communication," ACM SIGMOBILE Mobile Computing and Communications Review 5.1 (2001): 3-55

[8] Bland, J. Martin, and Douglas G. Altman, " Statistics notes: measurement error, " Bmj 313.7059 (1996): 744.

[9] Peli, Eli, "Contrast in complex images," JOSA A 7.10 (1990): 2032- 2040.

[10] Pearson, Karl, "Note on regression and inheritance in the case of two parents," Proceedings of the Royal Society of London 58 (1895): 240- 242.

[11] Janecek, Andreas, et al., "On the Relationship Between Feature Selection and Classification Accuracy," FSDM. 2008.

[12] Peli, Eli, "Contrast in complex images," JOSA A 7.10 (1990): 2032- 2040.

[13] Pearson, Karl, "Note on regression and inheritance in the case of two parents," Proceedings of the Royal Society of London 58 (1895): 240- 242.

[14] Pham, Dzung L, Chenyang Xu, and Jerry L. Prince. "Current methods in medical image segmentation 1," Annual review of biomedical engineering 2.1 (2000): 315-337.

[15] N. Ray, "Digital Image Basics: File Formats," in CMPUT, 2010.

[16] Kumar, Tarun, and Karun Verma, " A Theory Based on Conversion of RGB image to Gray image, " International Journal of Computer Applications 7.2 (2010): 7-10.

[17] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," European conference on computer vision. Springer Berlin Heidelberg, 2006

[18] Pearson, Karl, "Contributions to the mathematical theory of evolution," Philosophical Transactions of the Royal Society of London. A 185 (1894): 71-110.

[19] Burges, Christopher JC, "A tutorial on support vector machines for pattern recognition," Data mining and knowledge discovery 2.2 (1998): 121-167