OpenCV Based Disease Identification of Mango Leaves

Jayaprakash Sethupathy [#1], Veni S [*2]

[#] Department of Mechanical Engineering [*] Department of Electronics & Communication Engineering,
Amrita School of Engineering, Ettimadai, Coimbatore
Amrita Vishwa Vidyapeetham, Amrita University, India – 641112
[#] Robert Bosch Engineering and Business Solutions Private Ltd
Keeranatham Village, Coimbatore, India – 641035
[1] jayaprakash.settupathy@in.bosch.com
[2] s_veni@cb.amrita.edu

Abstract—This paper aims in classifying and identifying the diseases of mango leaves for Indian agriculture. K-means algorithm is chosen for the disease segmentation, and the disease classification and identification is carried out using the SVM classifier. Disease identification based on analysis of patches or discoloring of leaf will hold good for some of the plant diseases, but some other diseases which will deform the leaf shape cannot be identified based on the same method. In this case leaf shape based disease identification has to be performed. Based on this analysis two topics are addressed in this research paper. (1) Disease identification using the OpenCV libraries (2) Leaf shape based disease identification.

Keyword- k-means, Principal Component Analysis (PCA), feature extraction, shape detection, disease identification, Elliptic fourier analysis, Support Vector Machine(SVM), Artificial Neural Network (ANN)

I. INTRODUCTION

Mango (Mangifera indica) is one of the delicious and most important fruit crops cultivated in Indian agriculture. It is exported to many countries in the form of raw or ripe fruits and also in the form of processed consumables like ripe mango slices or juice, raw mango Pickle, etc. Mango is rich in vitamin A and C, it also has rich medicinal values in Indian traditional Ayurvedic medicine, and Mango leaves are mostly used during rituals since these are having antibacterial activity against gram positive bacteria. In recent times the export value of Indian mango is declined due to the uncontrolled use of pesticides, hence it is a right time for the researchers to come up with ideas for early identification of diseases and control the use of dangerous pesticides which causes threat to human health.

Common diseases of mango include gall midge infestation, black mildew hopper attack, mango malformation disease, pulp weevil, stem miner, anthracnose, alternaria leaf spots etc. These diseases are occurring due to the insects, bacterial, fungal and viral infection and these diseases affect the crop yield by infecting the leaves, flowers, fruits and stem. Infection in leaf causes the photosynthesis to be blocked and in due course of time it causes the plant to die.

Identification of diseases or deficiency is usually carried out by farmers by frequent monitoring of the plant leaves, flowers, fruits or stem. For small scale farmers, early identification of disease is very much possible and able to control the insects by organic pesticides or by the use of minimal amount of chemical pesticides. For large scale farmers frequent monitoring and early identification of disease is not possible and it results in a severe outbreak of the disease and pest growth which cannot be controlled by organic means. In this situation farmers are forced to use the poisonous chemicals to eradicate the disease in order to retain the crop yield. This problem can be solved by automating the monitoring process by use of advanced image processing techniques.

The proposed work aims in making the automated system easily available for the farmer's using the low cost devices or boards such as android devices and raspberry pi. The steps involved in disease detection are Digital image acquisition, Image pre-processing (noise removal, Color transformation, and histogram equalization), k-means Segmentation, Feature extraction, and classification using the support vector machine algorithm which is a supervised learning algorithm.

In case of disease identification based on discoloring of leaf, steps mentioned above holds good, but in case of disease identification based on leaf deformation or leaf shape, a separate learning algorithm for the leaf deformation analysis is needed; In this case the image pre-processing step is slightly different, where the pre-processed image is converted to a binary image for the leaf shape identification. Further to the morphological feature extraction, principal component analysis (PCA) is done to extract the relevant features of the plant. The classification of the plant is done using the SVM classification algorithm.

The disease name and the corresponding feature vectors are added to the Matlab or OpenCV database using the learning algorithm. Different segmentation and classification algorithms are studied for the disease clustering and identification. K-means is one of the simple and robust segmentation algorithms to implement for low cost

development and it uses unsupervised learning method to solve known clustering issues. Disease classification is achieved by SVM classifier, because of the implementation simplicity in both the Matlab and OpenCV libraries compared back propagation neural networks.

Leaf shape based disease identification are done in spatial and in frequency domain image processing. In case of the spatial domain image processing, basic geometric features of leaf such as diameter, area, perimeter, physiological length, width, etc. are identified. Based on these geometric features, digital morphological features such as rectangularity, circularity, etc. are calculated for identifying the diseased leaf.
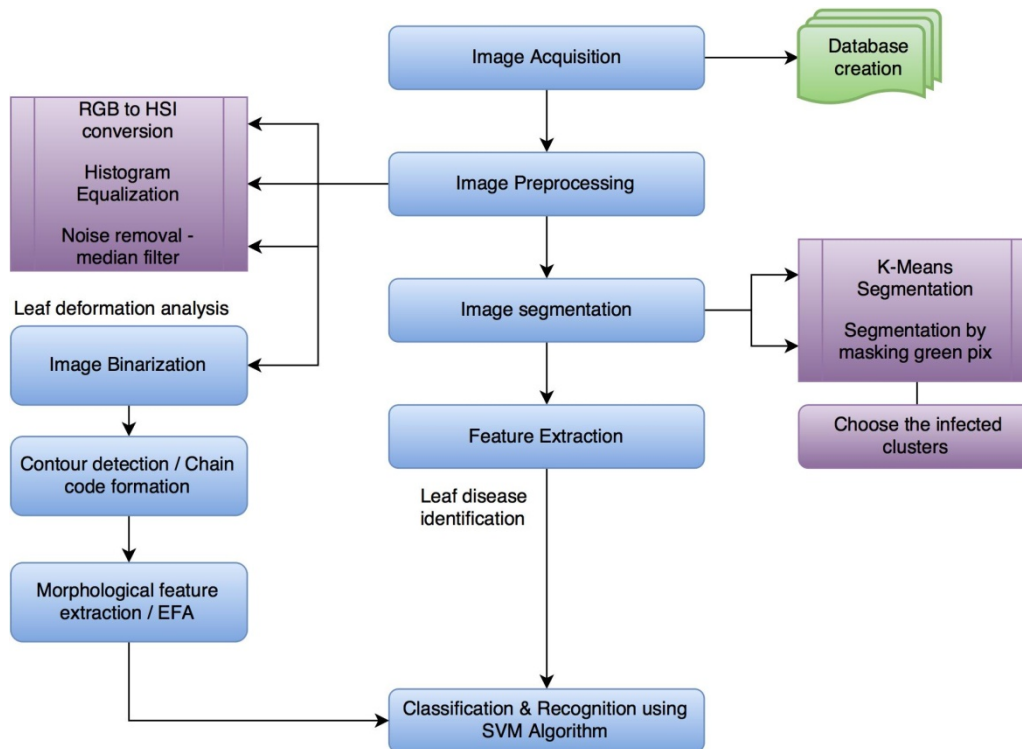

Fig. 1. Implementation block diagram

The frequency domain analysis of the diseased leaf identification is done using the Elliptic Fourier descriptor (EFD) or elliptic Fourier analysis (EFA) [6]. The chain code of the leaf shape contour is extracted and the Fourier transform is applied for the extracted chain code. The number of harmonics required for the analysis is user definable, and each harmonic has 4 coefficients known as Fourier descriptors. Precise descriptions can be obtained with more harmonics, but usually ten harmonics are required to identify the leaf shape or disease accurately. This method is used only for the Matlab based implementation. For android based application morphological feature extraction technique is used because of the non availability of Fourier transform libraries in OpenCV.

Implementation of the algorithm is done using both Matlab and OpenCV libraries and the results are compared. Since OpenCV libraries are open source, and implementation is possible in android devices and other low cost open hardware development boards such as Raspberry Pi and Beagle bone, The OpenCV implementation is fine tuned to get the accurate results compared to Matlab, so that it can be made available to farmers easily.

II. STEPS IN DISEASE IDENTIFICATION

A. Image acquisition

The diseased leaf image is acquired using the digital camera interfaced to the Raspberry Pi hardware or the smart phone camera, the image is acquired from a certain uniform distance with sufficient lighting for learning and classification. The image background should provide a proper contrast to the leaf color. Mango leaf disease dataset is prepared with both black and white background, based on the comparative study black background image provides better results and hence it is used for the disease identification of mango leaf.

Jayaprakash Sethupathy et al. / International Journal of Engineering and Technology (IJET)

Fig. 2. Disease affected leaf image

B.   Image pre-processing

Image acquired using the digital camera is pre-processed using the noise removal with averaging filter, color transformation and histogram equalization. The color transformation step converts the RGB image to HSI (Hue, Saturation and intensity) representation as this color space is based on human perception. Hue refers to the dominant color attribute in the same way as perceived by a human observer. Saturation refers to the amount of brightness or white light added to the hue. Intensity refers to the amplitude of light. After the RGB to HSI conversion, Hue part of the image is considered for the analysis as this provides only the required information. S and I component are ignored as it does not give any significant information [2]. Fig 3 shows the H, S and I components of the mango leaf infected with Gall Midge infestation.
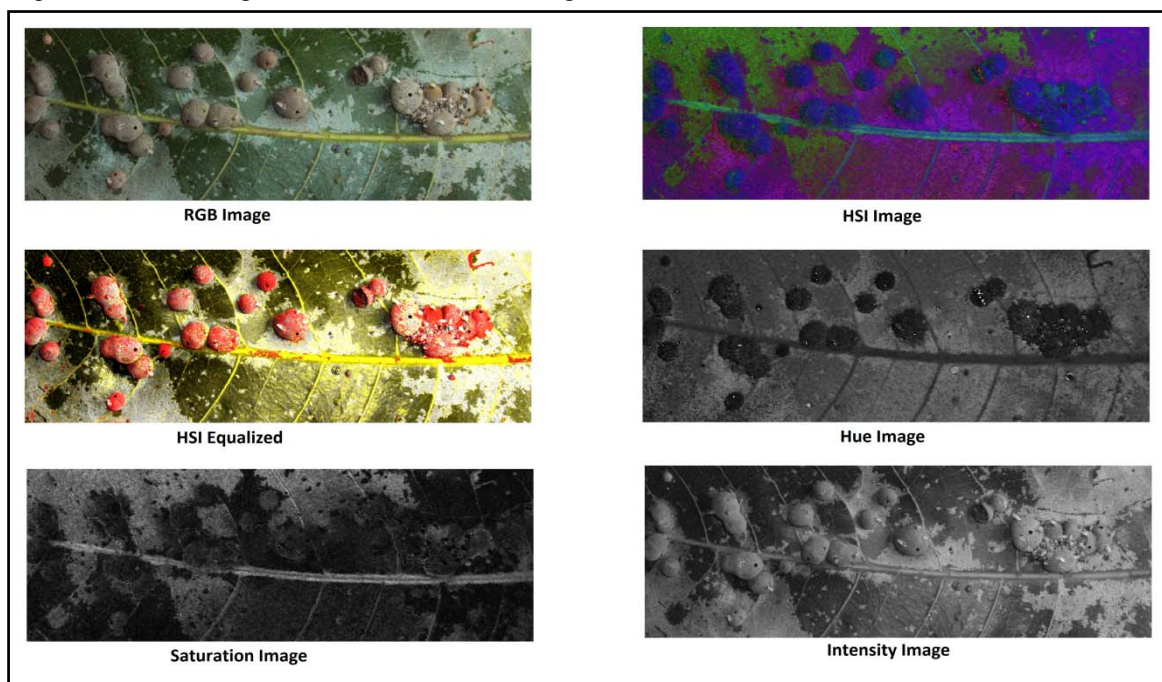

Fig. 3. HSI Image of Gall midge infected mango leaf

Masking green pixels: Since most of the green colored pixels refer to the healthy leaf and it does not add any value to the disease identification techniques, the green pixels of the leaf are removed by a certain masking technique, this method significantly reduces processing time [2]. The masking of green pixels is achieved by computing the intensity value of the green pixels, if the intensity is less than a predefined threshold value, RGB component of that particular pixel is assigned with a value of zero. The green pixel masking is an optional step in our disease identification technique as the diseased part of the leaf is able to be completely isolated in the segmentation process.

C.   Segmentation

There are different image segmentation techniques like threshold based, edge based, cluster based and neural network based. One of the most efficient methods is the clustering method which again has multiple subtypes, k-means clustering, Fuzzy C-means clustering, subtractive clustering method etc. One of most used clustering algorithm is k-means clustering.
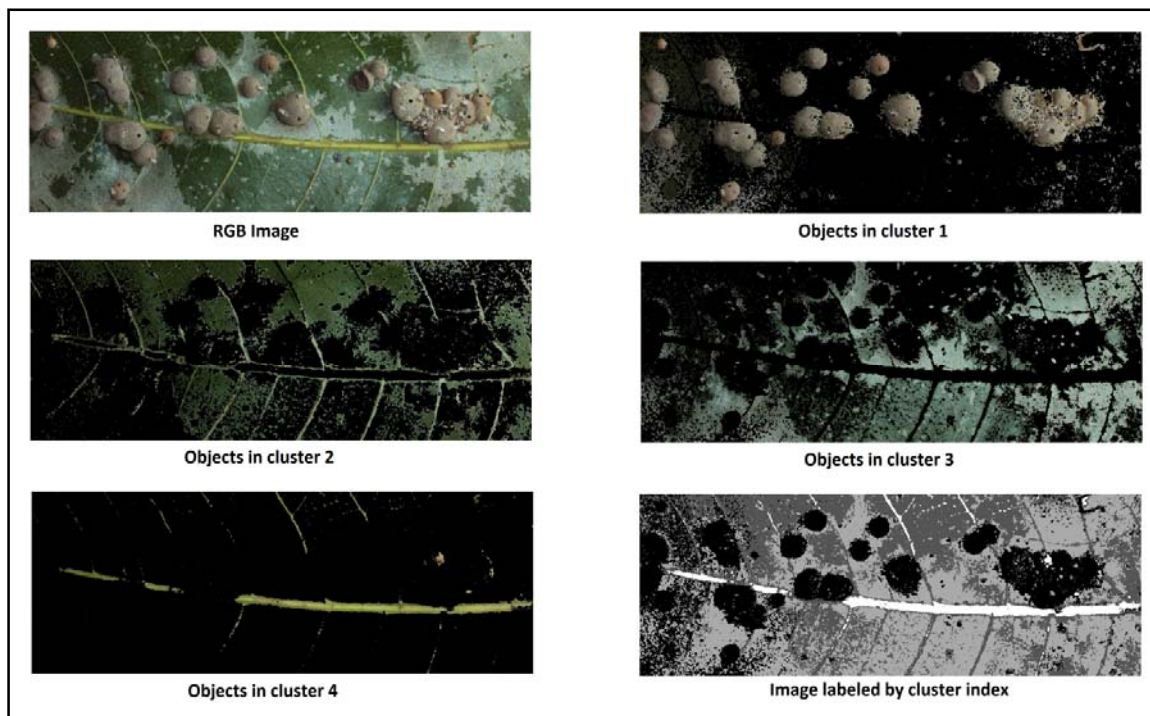
Fig. 4. K-Means segmentation of RBG image

K-means clustering is simple and computationally faster than other clustering techniques and it also works for large number of variables. But it produces different cluster result for different number of number of cluster and different initial centroid values. So it is required to initialize the proper number of number of cluster k and proper initial centroid. [10]
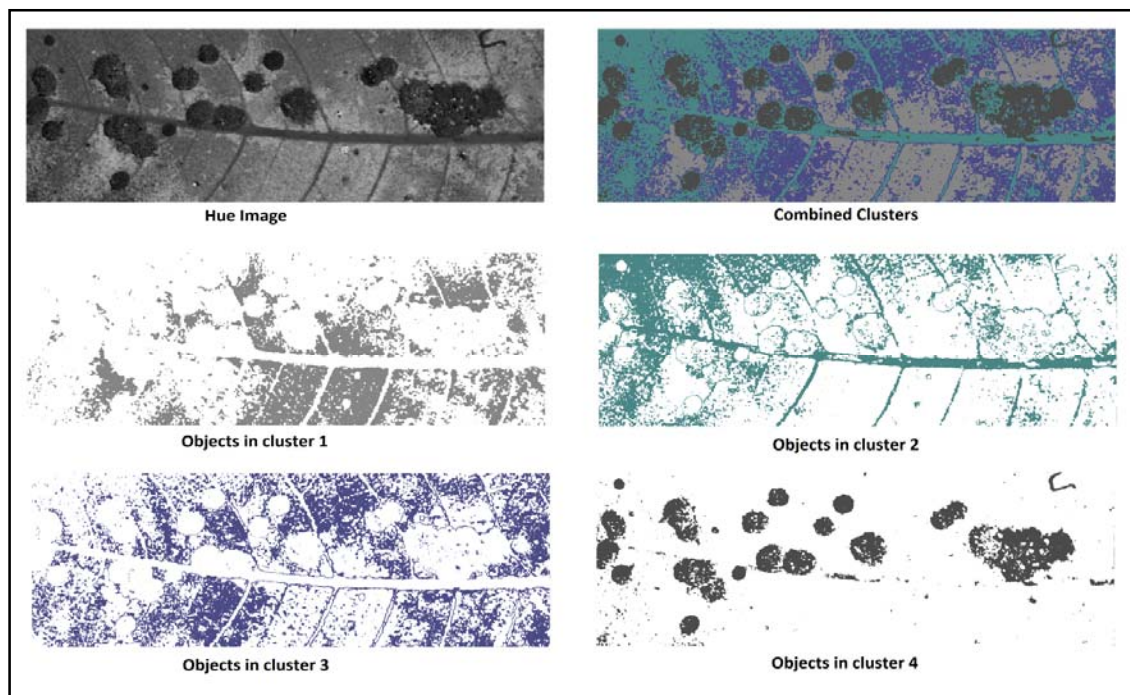

Fig. 5. K-Means segmentation of hue image

Many different segmentation options are experimented with different inputs to the k-means clustering algorithm [11], each of the different inputs results in different results. Fig 4 shows the clustering applied to the original leaf image and Fig 5 shows the clustering applied to the hue part of the HSI image. Classification result for these two clustering input types is compared in Table1.

K-means clustering is an unsupervised learning algorithm to classify the image patches based on attributes/features into K number of group. K is positive integer number. The grouping is done by minimizing

the sum of squares of distances between data and the corresponding cluster centroid. Observations in the same cluster are similar in some sense. K in our problem denotes the number of colored leaf patches which results in a perfect separation of infected leaf part leaving out the non infected image pixels in other clusters. Value of K can be chosen in runtime based on the diseased leaf image.
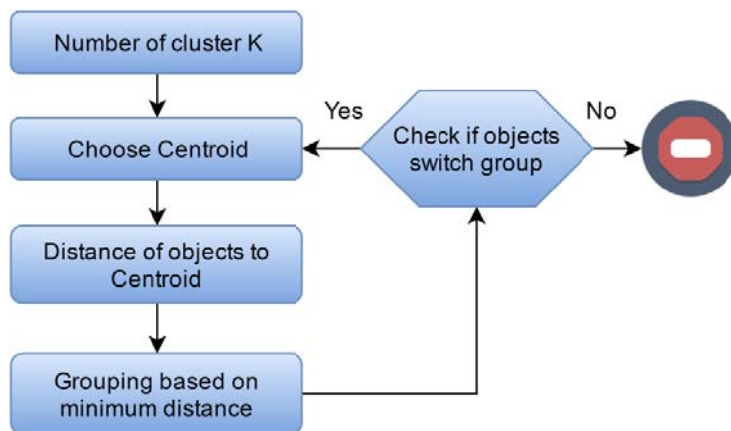

Fig. 6. K-Means segmentation steps

D. Feature extraction

There are many feature extraction techniques implemented. [7][8] and [9] implements the texture features based on gray-level co-occurrence matrix (GLCM) and spatial gray-level dependence matrices (SGDM), whereas in [2] texture features such as Contrast, energy, local homogeneity, cluster shade and prominence are used as features. In our implementation statistical features such as Standard Deviation, Variance, Kurtosis, Skewness, and Entropy are used as features for the disease identification of the mango leaf.

E. Support vector machine classifier

K-means algorithm will itself serve as very simple classification algorithm, in which each object can be assigned to the nearest cluster centroid and can be labeled accordingly. This is nothing but the 1NN classifier (k-NN – k Nearest Neighbor) trained based on the cluster centroid only. But this method is not robust enough to properly classify the input images. Artificial neural network (ANN)is a fast learning classification algorithm which works based on the back propagation of the multilayered neural networks[1][7]. Support vector machine (SVM) is a binary classifier and supervised learning algorithm in which the training examples are used for building a separating hyper plane equation. This hyper plane serves in categorizing the new test samples. [2]. Based on many parameters such as accuracy of classification, implementation in both OpenCV and Matlab SVM classifier is chosen for the disease classification problem.

The Table1 provides the classification accuracy with different number of training samples in OpenCV and Matlab and the different types of clustering inputs, clustering of the original diseased leaf RGB image and clustering of the Hue part of the RGB to HSI color space converted image. Result shows that the usage of HSI color space converted image provides a better result than the RGB image.

TABLE I Classification Percentage with Matlab and OpenCV

Number of training data	Success percentage in image classification			
	Matlab with RGB image	OpenCV with RGB image	Matlab with HSI image	OpenCV with HSI image
15	70	70	80	78
25	75	75	85	83

In Fig 7 data in grey shows the classification percentage of the RGB image and data in orange shows the classification percentage of the HSI image.
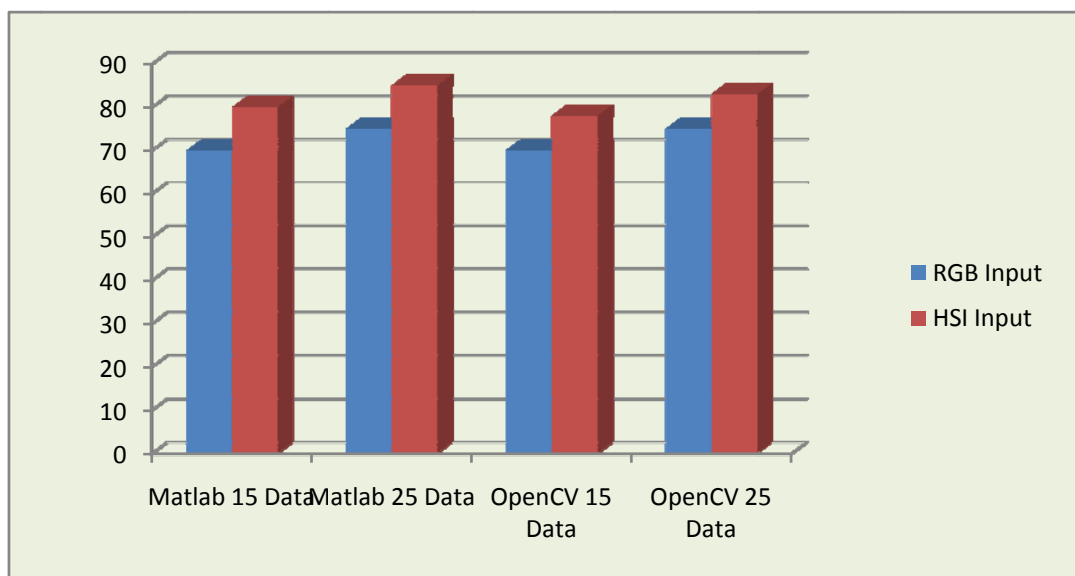
Fig. 7. Classification percentage with Matlab and OpenCV

III. LEAF DEFORMATION ANALYSIS

Leaf deformation is caused mainly because of the plant diseases and insects. Shape detection for the leaves is mainly executed for the species identification of the plants [3] [5] [6]. The same algorithm is reused in this implementation for the leaf deformation analysis.
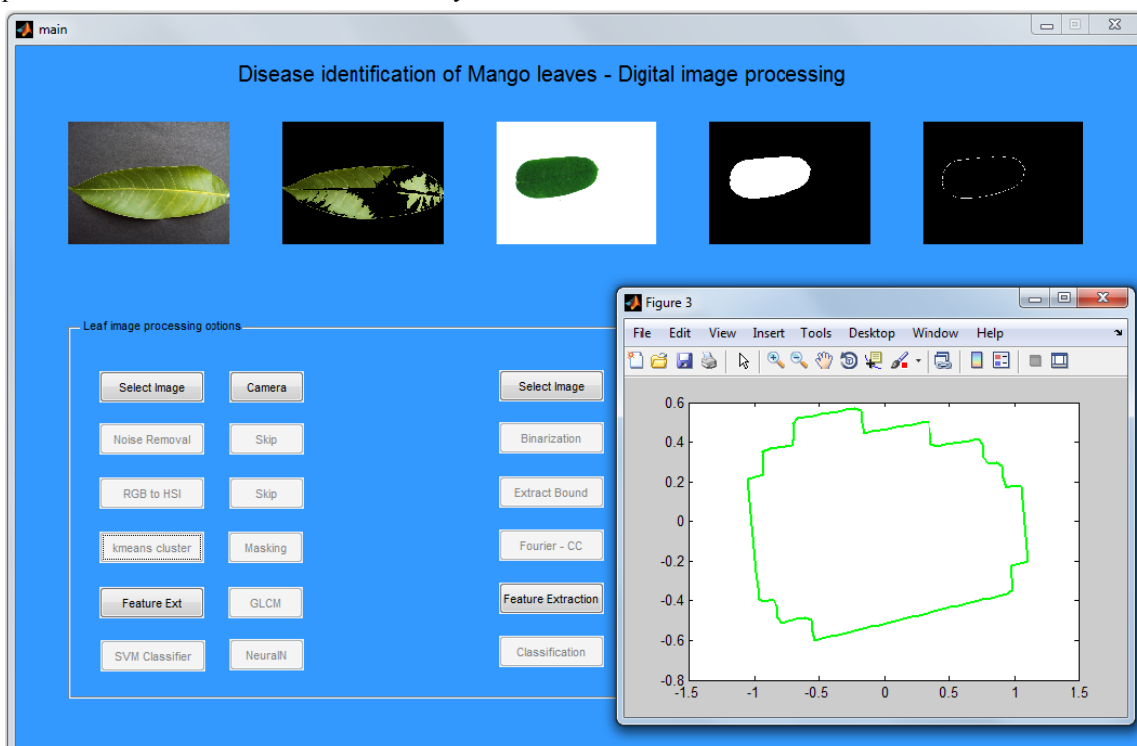

Fig. 8. Leaf boundary and chain code extraction

A. Image pre-processing and boundary extraction

The acquired digital image is first applied with a 3x3 averaging filter to remove the unwanted noise in the acquired image, the preprocessed image is then converted to grey scale image and then further converted to a binary image using the thresholding technique. The contour is then extracted from the binary image by subtracting the actual binary image with the eroded image as shown in Fig 8.

B. Feature extraction

The morphological feature extraction has the following steps, which extracts the features of the plant leaf based on the leaf shape or contour. Digital morphological features include geometric features and invariable moment

features. Geometric features that are considered for disease identification based on leaf shape are rectangularity, sphericity, circularity, Aspect ratio, convex area ratio, convex perimeter ratio, form factor. The invariable moments are regional moments of inertia and angle code histogram which are not used in our project [3][5]. These digital morphological features are derived from a basic set of geometric features like diameter, physiological length and width, leaf area and perimeter [5].

With a large number of variables (morphological features of leaf) mentioned above, the dispersion matrix will become too large to study and interpret the data properly. It is very much possible that the variables under study may be correlated or provides the same meaning, which implies that the data are redundant. Having such variables for analysis increases the dataset size with no much importance in the data. There will be a total of 210 three-dimensional scatter plot combinations with 12 variables, compared to 35 combinations with 7 variables. We can see a drastic reduction of the process overhead by reducing the redundant data by applying the principal component analysis. So PCA will result in a more meaningful feature sets which are necessary to classify the disease under consideration.

Principal component analysis (PCA) uses the simple statistical calculations like calculation of mean subtracted data and calculation of correlation matrix of the variables under consideration and the calculation of Eigen values and vectors for the correlation matrix and a max calculation of the data based on Eigen vector. This procedure converts the correlated variables in to a set of uncorrelated variables called principal components. The resulting principal components will be less than the number of original variables.

1) Aspect Ratio (AR):The aspect ratio is the proportionality relation between the width and height, it is the ratio of maximum length L_{max} to the minimum length L_{min} of the minimum bounding rectangle (MBR).

$$AR = \frac{\mathbf{L}_{max}}{\mathbf{L}_{min}}$$

2) Rectangularity (R):Rectangularity is the measure of similarity of the leaf to a Rectangle; it is the ratio of region-of-interest (ROI) area to the MBR area.

$$R = \frac{\mathbf{A}_{ROI}}{\mathbf{L}_{max}/\mathbf{L}_{min}}$$

3) Circularity (C):It is based on the bounding points of the ROI and is the ratio of the mean distance between the center of the ROI and all of the bounding points (μR)and the quadratic mean deviation of the mean distance (sR).

$$C = \frac{\mu R}{sR}$$

4) Convex Area Ratio *(CAR)*:The convex area ratio is the ratio of the ROI area and the convex hull area (AC).

$$CAR = \frac{\mathbf{A}_{ROI}}{AC}$$

5) Convex Perimeter Ratio *(CPR)*:The ratio of the ROI perimeter (PROI) and the convex hull perimeter (PC).

$$CPR = \frac{\mathbf{P}_{ROI}}{PC}$$

6) Sphericity (S):The ratio of the radius of the incircle of the ROI (r_i) to radius of the excircle of the ROI (r_c).

$$S = \frac{\mathbf{r}_i}{\mathbf{r}_c}$$

7) Form Factor (FF):Form factor is a well-known shape description characteristic given by

$$FF = \frac{4\pi \mathbf{A}_{ROI}}{\mathbf{P^2}_{ROI}}$$

Elliptic Fourier analysis (EFA) is another popular method used for leaf shape identification in which the frequency domain analysis of the leaf shape is done in contrast to the spatial domain in morphological feature extraction. In EFA the chain code of the leaf contour is first extracted [6]. Using the chain code extracted as shown in Fig8, a Fourier series expansion of the chain-code extracted from the closed-contour of the binarized leaf image in x and y axis is obtained as below.

$$x_N(t) = A_0 + \sum_{n=1}^{N} \left(a_n \cos\frac{2n\pi t}{T} + b_n \sin\frac{2n\pi t}{T} \right)$$

$$y_N(t) = C_0 + \sum_{n=1}^{N} \left(c_n \cos\frac{2n\pi t}{T} + d_n \sin\frac{2n\pi t}{T} \right)$$

The harmonic coefficients a_n, b_n, c_n and d_n for each of the harmonics are obtained using the formula defined in [6]. Principal component analysis is then applied to choose the uncorrelated harmonic coefficients. Feature vectors are then used for the classification and identification of the leaf deformation. The EFA implementation is possible only with the Matlab as there are no library functions available in OpenCV.

IV. FUTURE WORK

The future research topics are aimed at resolving the drawbacks in the current work and to add some new inventions to the project. The main drawback in the current work is the background dependant image processing. I.e. during the image acquisition, leaf background should be arranged in such a way that it provides sufficient contrast to the leaf and the diseased part of the leaf so that a proper segmentation can be achieved. The future work aims at a real time image acquisition directly in the agricultural field and the implementation should be able to distinguish the background and the leaf area.

Implementation of the robotic arm is planned as future advancement to this research topic. This implementation provides an autonomous robot which would be able to survey the agricultural field and identify the disease of the plants. The identified diseases will be automatically updated to the disease database which would be communicated to the farmer.

The vehicle carrying the robotic arm could be a tractor in case of the fields with spacious crops like mango. In case of the other crops where the tractor cannot maneuver inside the field, a line following robot is implemented [mini project of the course] to carry the robotic arm. There is no special arrangement needed to create the lines inside the field for robot movement, drip irrigation pipes used inside the field will provide the direction for the robot movement inside the field.

V. CONCLUSION

A comparison study of the disease classification between Matlab and OpenCV implementation is carried out and OpenCV implementation is fine tuned to obtain the similar results of the Matlab implementation. OpenCV implementation provides easy availability for the usage by farmers in the form of Android App or in the form of a low cost application specific board. Leaf shape based deformation analysis is implemented in Matlab and it provides better results to classify the deformed leaf against the normal leaf.

REFERENCES

[1] Pawan P. Warne1 and Dr. S. R. Ganorkar (2015) "Detection of Diseases on Cotton Leaves Using K-Mean Clustering", International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 04

[2] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini (2013) "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features" Agric Eng Int: CIGR Journal, 15(1): 211−217.

[3] Shailesh Sangle, Kavita Shirsat, Varsha Bhosle. (2013) "Shape Based Plant Leaf Classification System Using Android", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 8.

[4] Pooja Kamavisdar, Sonam Saluja, Sonu Agrawal (2013) "A Survey on Image Classification Approaches and Techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1

[5] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Qiao-Liang Xiang (2007) "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network" arXiv:0707.4289v1

[6] Jõao Camargo Neto, George E. Meyer, David D. Jones, Ashok K. Samal (2006) "Plant species identification using Elliptic Fourier leaf shape analysis" Computers and Electronics in Agriculture 50 121–134

[7] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh (2011) "Fast and Accurate Detection and Classification of Plant Diseases" International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011

[8] PradnyaRavindra Narvekar, Mahesh Manik Kumbhar, S. N. Patil (2014) "Grape Leaf Diseases Detection & Analysis using SGDM Matrix Method" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2014

[9] PriyaP, Dony A Dsouza (2015) "Study of Feature Extraction Techniques for the Detection of Diseases of Agricultural Products", International Journal of innovative research in electrical, electronics, instrumentation and control engineering, Vol. 3, Special Issue 1, April 2015

[10] Nameirakpam Dhanachandra, Khumanthem Manglem and Yambem Jina Chanu, (2015) "Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm", ScienceDirect, Procedia Computer Science 54 (2015) 764 – 771

[11] Anand R, Veni S, Aravinth J, "An Application of image processing techniques for Detection of Diseases on Brinjal Leaves Using K-Means Clustering Method", Fifth International Conference on Recent Trends in Information Technology 2016 (ICRTIT 2016), April 8 - 9 2016 at Anna University, Chennai campus sponsored by IEEE Madras Section

[12] Thi-Lan Le, Duc-Tuan Tran, Van-Nam Hoang, (2014) "Fully automatic leaf-based plant identification, application for Vietnamese medicinal plant search", SoICT

[13] Thi-Lan Le, Duc-Tuan Tran, Ngoc-Hai Pham, (2014) "Kernel descriptor based plant leaf identification", IEEE

[14] James S. Cope, David Corney, Jonathan Y. Clark, Paolo Remagnino, Paul Wilkin, (2012), "Plant species identification using digital morphometrics: A review", Expert Systems with Applications 39 (2012) 7562–7573

[15] Vijay satti, Anshul satya, Shanu sharma, (2013) "An automatic leaf recognition system for plant identification using machine vision technology", International Journal of Engineering Science and Technology (IJEST)

[16] Savita N. Ghaiwat, Parul Arora, (2014) "Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review", ISSN: 2347 - 2812, Volume-2, Issue - 3, 2014

[17] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, Paulus Insap Santosa, (2011), "Leaf Classification Using Shape, Color, and Texture Features", International Journal of Computer Trends and Technology- July to Aug Issue 2011

[18] Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang, (2006), "Leaf shape based plant species recognition", Applied Mathematics and Computation 185 (2007) 883–893

AUTHOR PROFILE

Jayaprakash Sethupathy Completed BE in Electronics and Communication Engineering in AC College of Engineering and Technology in the year 2006. Started the career in HCL Technologies, Chennai, in the year 2006 as a member technical staff under the embedded system domain. Joined in Robert Bosch Engineering and Business solutions Private Ltd, Coimbatore, in the year 2008 as an embedded system Engineer in the power train domain of the diesel automotive systems and continuing as a project manager. Pursuing Masters in Automotive systems and embedded systems in Amrita University Coimbatore on a part time basis.

Dr. S. Veni joined Amrita Vishwa Vidyapeetham as a faculty in the year 2001 under the Department of Electronics and Communication Engineering with 8 years of experience in the field of teaching. She is currently working as Associate Professor. She received her AMIE degree from Institution of Engineers, Calcutta in the year 1994 and M.E degree (Applied Electronics) from Bharathiar University, Coimbatore in the year 1998. She obtained her Ph.D degree in the area of Image Processing from Amrita Vishwa Vidya Peetham in January 2012. Her areas of interest include Signal and Image Processing, hardware implementation of signal and image processing algorithms. She has published nearly 30 papers in International Journals and conferences.