Proc. of the 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)
28-30 November 2017, Putrajaya, Malaysia

Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM Classifier

Ahmad Nor Ikhwan Masazhar#, Mahanijah Md Kamal*

#,* *Faculty of Electrical Engineering,*
Universiti Teknologi MARA
40450 Shah Alam
Selangor Malaysia
[1]iwwaaannnn@gmail.com
[2]mahani724@salam.uitm.edu.my

***Abstract*— Disease in oil palm sector is one of the major concerns cause it effects the production and economy losses to Malaysia. The problem of disease that arises in oil palm plantation are . Nowadays plant diseases detection has received a lot of attention in monitoring the symptoms at earlier stage of plant growth. This work presents the use of digital image processing technique for detection and classification of oil palm leaf disease symptoms. Here, the disease detection used k-means clustering and multiclass SVM classifier to determine two palm oil diseases based on the symptoms of the disease through its leaf. By using k-means clustering technique, thirteen types of features are extracted from the leaf images. The classification of the disease is carried out by using multiclass SVM classifier. The detection shows that SVM achieves accuracy of 97% for Chimaera and 95% for Anthracnose.**

Keywords; oil palm disease, image processing, k-means clustering, SVM

I. INTRODUCTION

Oil palm or its scientific name Eleais guineensis, is an important and useful economic crop which used as for food industry as well as non-food industry [1]. It was first introduced to Malaysia as a decorative plant in 1870. From 1960, the planted area in Malaysia is measured to have more than 5 million hectares of the palm oil tree planted. Large plantation areas in Malaysia make it the main exporter [2]. Palm oil industry contributes significantly to Malaysia's economic growth. The sustainability of palm oil industry depends on how well the Malaysian economy adapts to ever-changing world demand [3]. The production of oil palm can be maximize by taking care of the oil palm trees health. This is because palm oil trees are exposed to several diseases. Generally, the disease microorganism are fungi, bacteria and virus [4]. Disease infection can disrupt the production of palm oil and that is why to obtain maximum production, the health of the palm oil tree is crucial. All growing stages of oil palm plants can be infected by a disease [5]. However, in advanced disease infections, it has no satisfactory methods that can be used in reducing the effect of the infections on the yield. The disease infections on palm oil tree usually has specific symptoms and the most basic method to detect the infected palm oil tree is by visual inspection which is known to use intensive labour and consume more time [6-7]. Besides, by using labour force to classify the disease can also can cause human error.

Accurate detection of the palm oil tree disease is crucial so that accurate preventive action can be taken [8]. By using image processing, the pattern of the disease infection symptoms on the leaf is learned by Support Vector Machine (SVM) classifier so that the palm oil disease can be accurately detect and classify in a short period [9]. The best platform in which many image processing algorithms and SVM classifier algorithm that have been developed is Matlab. Major advantage of using Matlab as an image processing algorithm development environment is it's built in image processing functions [10]. Diseases on plant leaves cause major production and economic losses to the country. These diseases cannot be detected by human eyes and the symptoms is not visible. There are several ways to detect the diseases on plant leaves such as image processing techniques for earlier detection. Fruit disease identification using images based on k-means clustering technique were studied [11]. Here, images of an apple are classified using a multiclass SVM for detection and identification with apple scab disease. Various types of leaf spot diseases in agriculture such as bacterial, fungal and viral were studied for automatic leaf diseases detection using digital image processing [12]. Classification of fungal disease of cotton uses mobile captured symptoms of cotton leaf spot images and categorize the disease using SVM. Feature extraction techniques are used to extract features such as boundary, shape, colour and texture for the disease spots to recognize the cotton leaf disease [13]. SVM approach is used to distinguish crops and weeds in digital images to perform classification. Based on the classification rate, it shows that SVM achieves 97% accuracy with no misclassification of crops as weeds [14].

Hence, the objectives of this paper is to design a vision system that can help determine the palm oil leaf disease infection using digital image processing and SVM classifier. By this approach, the time taken to determine the leaf disease infections on oil palm tree can be reduced and easier. In this paper, two common types of palm oil leaf diseases will be discussed. Anthracnose and chimaera attacked the palm oil leaf during the nursery stage. The images of leaf diseases were taken from Tapak Semaian Benih Felcra Sendayan, Negeri Sembilan. These two disease can be classify based on the symptoms spotted on the oil palm leaf.

A. Anthracnose

Anthracnose disease is caused by the pathogenic fungi which is Botryodiplodia spp., Melanconium elaeidis and Glomerella cingulate. This disease mainly attacks seeds in age of 2 months. The symptom usually appears as small brown spots and will expand wider with pale yellow border.

Figure 1: Anthracnose disease leaf symptom

B. Chimaera

Figure 2: Chimaera disease leaf symptom

Chimaera disease is caused by the genetic problem of the oil palm tree seed. The symptom of the disease are the leaves have white stripe or yellowish-white, and the lack of chlorophyll. The leaves will not recover as normal tree leaves.

II. Methodology

In this section, the methodology will be explain the process in completing this project. It will be explain in three subsection which is block diagram, flow chart and also theoretical background.

A. Block Diagram

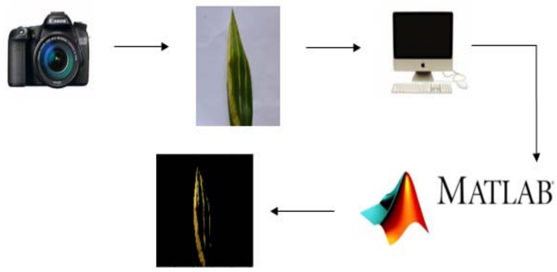


Figure 3: Block diagram of the project

Figure 3 shows the block diagram of the whole project. The palm oil disease can be spot on the palm oil tree leaves by observing the leaf shape and colour. The certain disease can be classifies based on every spot on the leaf. The proposed system to tackle this problem is based on image processing and SVM classifier technique. Basically, in image processing algorithm, the process of the project start with taking the picture of the palm oil tree leaf that was infected by a disease. The camera is used as an image acquisition device or the input. Camera is the most important hardware in this project. The camera used can be any type of camera such as an inbuilt camera of laptop or a mobile phone as well. Then the picture will be transfer into the computer to be process. Computer used as the receiver and also used to run the Matlab software. Matlab software was chosen as the platform. After getting an image, the next part is to do the calibration for detecting the colour of the object. The detection was perform by using k-means clustering. After that the SVM classifier take place. The feature extracted from the sample leaf image is observe by the classifier and the classification and accuracy for the palm oil disease is generated by using this method.

B. Flow Chart

Figure 4 shows flow chart of the project. The project start with the image acquisition phase. In this phase, various images of the selected oil palm disease that are to be classifies were taken using camera. The image then will be used as the input for the image processing phase. The first step of the image processing is to do the image enhancement that is contrast enhancement. The contrast enhancement was done to reduce the noises in the image which make the region of interest of the image not clear that can lead to the quality of the image dropped. Then the input image will be convert from RGB colour space to L*a*b colour space. The L*a*b* colour space enables the users to quantify these visual differences of the image. After that the image will be clustered by using K-means clustering. This is a way to separate groups of things or objects in which in this project the image will be separated based on colour. Based on the selected cluster, then the feature will be extract using gyrocomatrix. By using this technique, the gray-level co-occurrence matrix (GLCM) was created from image. Lastly the classification of the disease will be done by using multiclass SVM classification. The feature extracted from the image will be classifies and compare by each row of the dataset.

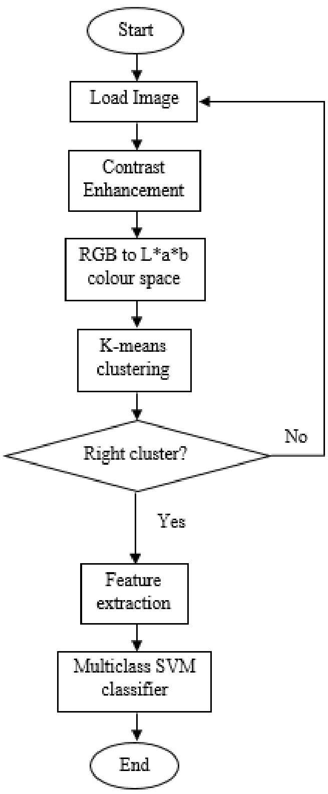


Figure 4: Flow chart of the project

C. Theoretical Background

1) Image processing

The symptom of the disease can be detected from the oil palm leaf image based on the colour and the shape. In this project, image processing technique that will be used is k-means clustering. The first step in the image processing phase is image acquisition. The image that was taken by camera will be load to the Matlab to be process.

Secondly, the image enhancement phase will take place. This technique will adjust the image intensity values or colour map. In this project, imadjust(I) function was use. This function is to maps the intensity values in grayscale image i to new values in j such that 1% of data is saturated at low and high intensities of i. This increases the contrast of the output image j. The function use to do the contrast enhancement is imadjust (I,stretchlim(I)). The stretchlim(I) function is to returns Low High, a two-element vector of pixel values that specify lower and upper limits that can be used for contrast stretching image i. By default, values in Low High specify the bottom 1% and the top 1% of all pixel values. The gray values returned can be used by the imadjust function to increase the contrast of an image.

Next, the image segmentation based on colour-based segmentation using k-means clustering was conducted. The first step in this phase is to convert the image from RGB to L*a*b colour space. Basically there are three colours for the input image which is white, green, and yellowish brown. The L*a*b* colour space enables the users to quantify these visual differences. The L*a*b* colour space is derived from the CIE XYZ tristimulus values. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' indicating where colour falls along the red-green axis, and chromaticity-layer 'b*' indicating where the colour falls along the blue-yellow axis. All of the colour information is in the 'a*' and 'b*' layers. You can measure the difference between two colours using the Euclidean distance metric. The conversion is in makecform and applycform format before converting the image to L*a*b* colour space.

By using k-means clustering the colours in a*b* space can be classify. Clustering is a method to differentiate groups of objects. K-means clustering manage each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires specific the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other. Since the colour information exists in the 'a*b*' space, the objects are pixels with 'a*' and 'b*' values. In this project, k-means used to cluster the objects into three clusters using the Euclidean Distance Metric. The results from k-means clustering then will be label for every pixel in the image. This was done to separate objects in the image by colour.

The dataset of the image pixel is clustered according to nearest respective cluster centres is the first step of fitness computation such that each pixel x_i of colour image is put into the respective cluster with cluster centre z_j for $j = 1, 2, \ldots, K$ by the following equations:

$$\text{If } \|x_i - z_j\| < \|x_i - z_l\|, \quad (1)$$

$$i = 1,2,\ldots,m \times n \quad l = 1,2,\ldots,K, \text{ and } p \neq j$$

By computing the mean of each pixel of the assigned clusters, new cluster centers can obtained. The new center of cluster Z_i is given by for the cluster C_i as:

$$Z_i(r,g,b) = \frac{1}{ni}\sum_{xj\epsilon\, ci}(xj(r,g,b)) \quad i = 1,2,\ldots k \quad (2)$$

By calculating Euclidean distance between the pixels and their respective cluster, the fitness function is calculated by the following equations:

$$M = \sum Mi \quad (3)$$

$$M_i = \sum_{xj\,\epsilon\, ci} |(xj(r,g,b))| \quad (4)$$

For feature extraction, the method used is gyrocomatrix. By using this technique, the gray-level co-occurrence matrix (GLCM) from image was created from image. It is a technique in which the image texture and colour are considered to come out with the features that represents the image. By using this technique, GLCM was created from image by using glcms = graycomatrix(I) function. Features called as texture features, which include local homogeneity, contrast, cluster shade, Energy, and cluster prominence are computed for the image as given in equations:

$$\text{Contrast} = \sum_{i,j=0}^{N-1}(i,j)^2\, C\,(i,j) \quad (5)$$

$$\text{Energy} = \sum_{i,j=0}^{N-1}(i,j)^2 \quad (6)$$

`

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1}(i,j)^2/(1+(i-j)^2) \quad (7)$$

$$\text{Entropy} = \sum_{i,j=0}^{N-1}(i,j)\,\log C(i,j) \quad (8)$$

2) *Multiclass SVM Classifier*

SVM is a binary classifier to analyze the data and recognize the pattern for classification. The classification of the feature extracted from the input image was made by classifying each row of data in the dataset which is in a matrix of data form. The data in the matrix was presented in a row and column where row corresponds to the number of sample used for training while column correspond to the feature extracted from the image. Thus, the training data used should consist the similar number of columns as the input data. Each row of data in a matrix of data of the dataset is classifies using the information in a support vector machine classifier structure SVMStruct, created using the svmtrain function. The svmclassify function uses results from svmtrain to classify vectors x according to the following equation:

$$\text{Entropy} = \sum_i a_i k(s_i,x) + b \quad (9)$$

where s_i are the support vectors, α_i are the weights, b is the bias, and k is a kernel function. In the case of a linear kernel, k is the dot product. If $c \geq 0$, then x is classified as a member of the first group, otherwise it is classified as a member of the second group.

III. Result and Discussion

To evaluate the proposed method, the samples were divided into training and test dataset. The training set is used to train the SVM classifier while the test set is used to predict the accuracy of the classifier. The steps involved are as follow:

A. *Image Acquisition*

(a) (b)

Figure 5: Sample images of palm oil leaf disease

Figure 5 shows two samples of oil palm diseases that attacked during the nursery stage. This image is uploaded into Matlab platform and analysed. Figure 5(a) represents Chimaera disease symptom while Figure 5(b) represents Anthracnose disease symptom.

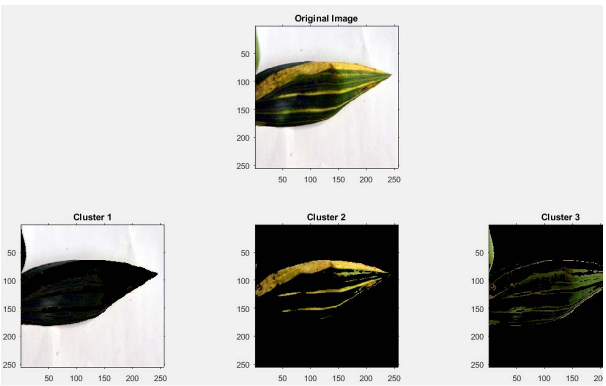
B. Image Enhancement

(a) (b)

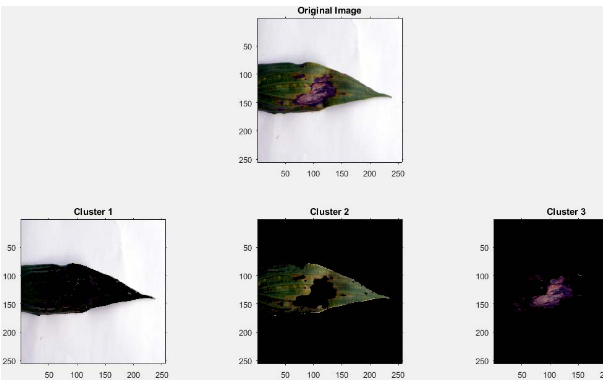
Figure 6: Image enhancement for the sample disease image.

Figure 6 referring to the sample image after going through the image enhancement process. This technique adjusts the colour intensity values or colour map of the leaf. The function in image enhancement used to maps the intensity values in grayscale image to new values such that 1% of data is saturated at low and high intensities of the grayscale image. Figure 6(a) shows the Chimaera disease while Figure 6(b) shows the Anthracnose disease after the contrast has been adjust.

C. Image Segmentation Using K-means Clustering



(a)



(b)

Figure 7: Image segmentation using K-means clustering for the sample disease image

Figure 7 shows the sample image undergoing image segmentation using K-means clustering. This is a method to differentiate things or object in the image. In this process, the K-means clustering differentiate the sample image into 3 cluster with respect to the colour pixel of every image by using Euclidean Distance Metric. From figure 7(a), the sample image of Chimaera disease was clustered into 3 cluster. The selected cluster for this disease with respect to the disease symptom which is shown in cluster 2. Meanwhile, for Figure 7(b) it is shown the Anthracnose disease also was separated into three cluster. The selected cluster for this disease with respect to the disease symptom was shown in cluster 3.

D. Feature Extraction

Table 1 illustrates the extracted feature from both Anthracnose and Chimaera disease for the sample image. The feature extracted from the selected cluster in the k-means clustering with respect to the disease symptom represent the selected cluster image. This can be done by using GLCM. By using this technique, the image texture and colour are considered to come out with the features that represents the image. The feature extracted from the image was represented in thirteen parameters which is tabulate as in Table 1.

Table 1: Feature extracted from both anthracnose and chimaera disease sample image.

Disease	**Anthracnose**	**Chimaera**
Mean	2.05693	10.0361
Standard Deviation	12.6362	37.678
Entropy	0.451482	1.08534
RMS	1.67194	4.37123
Variance	144.61	1376.69
Smoothness	0.999998	0.999999
Kurtosis	58.4101	19.1248
Skewness	7.12043	4.04505
IDM	255	255
Contrast	0.0126838	0.109559
Correlation	0.920698	0.956385
Energy	0.95093	0.824291
Homogeneity	0.994795	0.983154

E. Multiclass SVM Classifier

Table 2 shows the Multiclass SVM classifier accuracy for both Chimaera and Anthracnose disease for the sample image used. The accuracy was compute with 500 iterations for each sample image. The tested sample image undergoing classification can produce an average accuracy up to more than 90% and the accuracy for both are good.

Table 2: Multiclass SVM classifier accuracy for both anthracnose and chimaera disease sample image

Disease	Chimaera	Anthracnose
Accuracy	96.7742	95.1613

IV. Conclusion

Oil palm disease symptom are usually can be seen and spotted on its leaves. Basically, this paper focus on two oil palm diseases which is Anthracnose and Chimaera. Therefore, the main objective is develop a system that can detect the diseases and infections on the oil palm leaf at nursery stage and isolate the affected plant earlier. The detection and classification of these diseases was done by using digital image processing technique based on k-means clustering approach and multiclass SVM classifier. By using the proposed technique, the result shows that the SVM classifier achieves accuracy of 97% for Chimaera and 95% for Anthracnose. The performance of the system demonstrate a reliable system that can be used to detect and classify the disease symptoms and manage the effected plant before transferring it to the plantation.

Acknowledgment

I would like to express my special thanks and gratitude to Universiti Teknologi MARA which has granted me a fund with code project: 600-IRMI/DANA 5/3/LESTARI (35/2015).

References

[1] B.A.M. Suhaila, "The Palm Oil Industry fkrom The Perspective of Sustainable Development: A Case Study of Malaysian Palm Oil Industry", 2012.

[2] S.A. Sulaiman, S. Balamohan, M.N.Z. Moni, S.M. Atnaw and A.O. Mohamed, "Feasibility study of gasification of oil palm fronds". Journal of Mechanical Engineering and Sciences. Vol. 9, pp. 1744-1757, 2015.

[3] M. A. S. Zaidi, K Zurina, I. M. Adib and K. Z. Abdul,"Response of Malaysia's Palm Oil Export to Domestic and Foreign Macroeconomic Shocks". Prosiding Perkem ke-9, 2014.

[4] N. I. Nadhrah, N. Rosimah, R.Nurrashyeda and I. A Seman, "Effect of Formulated Bioorganic Containing Burkholderia GanoEB2 in Suppressing Ganoderma Disease in Oil Palm Seedlings. vol. 51, no. 2, pp. 80–87, 2014.

[5] I. Ishaq, M. S. Alias, J. Kadir, and Kasawani, "Detection of Basal Stem Rot Disease at Oil Palm Plantations Using Sonic Tomography". Journal of Sustainability Science and Management. Vol 9, No. 2, p. 52. 2014.

[6] S. Liaghat, S.B. Mansor, H.Z.M. Shafri, S. Meon, R. Ehsani, S.H.M.N. Azam and N.M. Noh, "Early Detection Of Oil Palm Fungal Disease Infestation Using A Mid-Infrared Spectroscopy Technique", 2011.

[7] N. Laila, S. Shafiquzzaman, Y. U. Kalsom and M. M. A. Mondal., "Issues of Ganoderma spp. And Basal Stem Rot Disease Management in Oil Palm", American Journal of Agricultural Science. vol 2, no. 3, pp. 103-107, 2015.

[8] S. Heri, G. Totok, J. R. Heru, D. Witjaksana and M. Budiman, "Mapping and identifying basal stem rot disease in oil palms in North Sumatra with QuickBird imager", Precision Agric. no 12, pp. 233–248, 2010

[9] A. F. Alham, M. Qorib, P. P.Ardhi, H. Kalamullah, Aprinaldi, J. Setiyawan, and A. M. Aniati, "Detection of Palm Oil Leaf Disease with Image Processing and Neural Network Classification on Mobile Device", International Journal of Computer Theory and Engineering, vol. 5, no. 3, 2013.

[10] R. Rojas, "Neural Networks: A Systematic Introduction",1996

[11] J.G.A. Barbedo, "Digital image processing techniques for detecting, quantifying and classification plant disease", SpringerPlus, vol. 2, no. 1, p. 60, 2013

[12] A.N. Rathod, B. Tanawal, V. Shah, "Image Processing Techniques for Detection of Leaf Disease", vol. 3, no. 11, pp. 397-399, 2013.

[13] S.P. Patil, R.S. Zambre, "Classification of Cotton Leaf Spot Disease using Support Vector Machine", Int. Journal of Engineering Research and Applications, vol. 4, pp. 92-97, 2014.

[14] F. Ahmed, H. A. Al-Mamun, A.S.M.H. Bari, E. Hossain, P. Kwan, "Classification of crops and weeds from digital image: A support vector machine approach", Crop detection, vol. 40, pp. 98-104, 2012.