

Detection of Early Blight Tomato Leaf Using k-Means Clustering

Tesfay Lijalem¹ and Semere A Asefa^{2*}

¹College of Natural and Computational Sciences, University of Ethiopia

²Department of Biomedical Engineering University, Republic of Korea

*Corresponding Author

Semere A Asefa, Department of Biomedical Engineering Yonsei University, Republic of Korea.

Submitted: 2023, May 28; Accepted: 2023, Jun 30; Published: 2023, July 03

Citation: Lijalem, T., Asefa, S. A. (2023). Detection of Early Blight Tomato Leaf Using k-Means Clustering. *Int J Diabetes Metab Disord*, 8(3), 382-389.

Abstract

Early blight is one of the major diseases of tomatoes that affects the leaves and fruit quality. Detection and estimation of the disease severity are performed using the visual observation method. Visual detection requires significant time for visual inspection of a large cultivated area. Thus, image processing techniques have proven to be an effective method as compared to visual analysis. In this study, digital image processing methods and techniques were used to detect early blight of tomato (EBT), estimate the disease severity, and classify tomato leaves. Totally, 198 infected plants were randomly taken from the Haramaya University research site "Rare" at four different times. Diseased potato leaf images were captured, resized, and stored for experimentation. The stored images were processed using median filtering to remove noise while preserving useful features in an image and image enhancement. The RGB images were transformed to gray scale and CIELAB color space, and the k-means clustering was applied to estimate the disease severity of the potato leaves, and Otsu's thresholding algorithm was applied to estimate the disease severity of both the detached and live leaves. MATLAB algorithms will be developed to determine the total area and infected lesion area of the leaf samples.

Keywords: EBT, K-means Clustering, Segmentation.

1. Introduction

Tomato is a popular and widely grown vegetable crop in Ethiopia, and it is largely produced in the northern and central rift valley area, specifically along the wash River. However, there are a number of factors that limit tomato production and productivity. Diseases in the crop are among the major factors which affect production and quantity of crop yield. The most common diseases are early blight, late blight, and septoria leaf spot effect during the growing season, reducing quality and tomato crop yield [1].

Early blight (EB) is one of the most common destructive foliage tomato diseases caused by the fungus *Alteria solani* and major yield losses in most potato-growing areas worldwide. On the leaves of a tomato, early blight appears as Rings with yellow holes and it can also cause symptoms such as collar rot (basal stem lesions at the seedling stage), stem lesions on the adult plant, and fruit rot over a wide range of temperatures, i.e., 4–36. It occurring early in the growth cycle, spreading quickly and causing severe damage. Timely disease detection and for effective control the occurrence and development of EBT is important to identify the infectious on tomato leaves easily and improve the quality of crop. Traditionally, direct observation methods were carried out to detect diseases on crop leaves. According to Hitimana and Gwun (2014), the detection of severity of infected leaves has been done by the farmers

using naked eye and it is also largely dependent on the level of experience of agronomists and farmers. However, this method has been shown to be time-consuming, costly, inefficient for detection and difficult to monitor the large farm that are subject to multiple human errors. Additionally, in practice, farmers require continuous monitoring methods to efficiently identify tomato disease [2].

The development of an appropriate computer routine algorithm using MATLAB software image processing techniques for detection and segmentation of early blight of tomato leave are fast, less expensive, require less effort and more accurate. Thus, image processing techniques are present in this work. The image processing technique has many important and effective applications in the field of agriculture for the detection of foliar diseases. For example, in identifying the types of disease, finding the shape of the affected area, detecting the edges of the diseased leaf, calculating the diseased ratio, separating the layers of target images, and determining the color of affected areas [2]. Image processing techniques have been found increasingly useful in the fruit industry, especially for applications in quality inspection and defect sorting applications.

In this paper, an accurate disease detection method for the diagnosis of the early blight on tomato leaves is developed. Automatic

detection frame work for early blight of tomato leaf disease consists of the following steps: like image acquisition, preprocessing, and segmentation. First, the tomato leaf image was acquired using a digital camera and preprocessed to remove noise and enhance the contrast of the tomato leaf image. The image in the RGB color model was converted into the CIELAB color model that the color transformation allows us to measure the visual difference present in the RGB image. The CIELAB system is device independent, which is defined by the CIE to classify color according to human vision. The converted $L^*a^*b^*$ color values are returned as a numeric array of the same size as the input. L^* , a^* , and b^* stand for lightness, red/green value, and blue/yellow value, respectively. Secondly, the preprocessed image is k-means clustering was applied to segment the region of interest into two clusters, and Otsu's thresholding was used for segmentation.

2. Materials and Method

2.1 Description of the Study Site

The experiment was conducted at Haramaya University's "Rare" research station. Haramaya University (HU) is located in the East Hararghe zone of Oromia Regional State, Ethiopia. Haramaya University is located at 9o26'N latitude and 42o3' longitude at 9o 26'N latitude. The altitude on the campus varies from 1980 to 2000 meters above sea level (m.a.s.l.). The mean annual precipitation is 780 mm and the mean annual maximum and minimum temperatures are and, respectively [3].

2.2 Experimental Material and Procedure

A tomato variety named "money maker" was used as a test crop. This variety was released by the tomato improvement Programmed of Haramaya University in 2010. The variety was released by the University for mid to high altitude areas of eastern Ethiopia. Thirty-three equal sized pots were utilized to conduct the experiment. To maintain the uniformity of extraneous factors, the tomato was grown in greenhouse "Rare" in the pots filled with soil of similar proportions. The pathogen was inoculated on thirty days old tomato planted on the pot. On the eleventh day after the inoculation, TLB disease began to appear. This means the diagnosis and sample collection was began on the forty-one days old tomato plant after the inculcation of the pathogens and application of fungicides of different volume. The images were captured using Digital camera and stored for later processing using Personal computer and Matlab software is used for implementation.

2.3 Sampling Technique

The first images were taken on the forty-first day following the occurrence of early blight of tomato leaf disease. This was repeated every eight days for two consecutive weeks. In this work, thirty-three infected tomato plants images were randomly taken from the pool of pots on which the plants were grown. Within a week, two images of each tomato plant (one from the bottom and one from the middle) and 66 images per week were taken to determine the progression of the disease and classifying. In total, 198 infected plants of tomato samples were randomly were taken from 11 groups.

2.4 Image Analysis

The basic steps for detection, estimation of severity and classification of tomato plant disease is shown in the following figure 2.

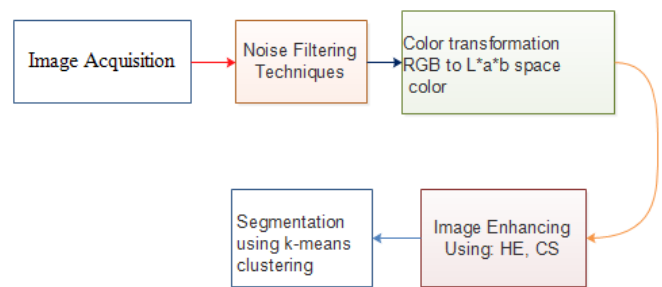


Figure 1: Frame Diagram To Detect And Classify Early Blight of Tomato Disease.

2.5 Image Acquisition

The tomato leaf samples were acquired by placing the samples on a white background (paper) using digital camera. These images were stored in JPEG (Joint Photographic Expert Group) format with 256x256 size.

2.6 Pre-Processing

Following acquisition, the next step is pre-processing image to remove the noise, poor resolutions of images, unwanted background of the image and enhancement to improve the quality of the image using median filtering and contrast limited adaptive histogram equalization. Then, the captured RGB color image is transformed into CIELA or Lab space. CIELAB are color independent space model and this are also derived from human perception, but RGB is color dependent space model. The LAB color is designed to approximate human vision and it aspires to perceptual uniformity and its L-component closely matches human perception of lightness. The transforming Formula from RGB to LAB color model is given by the equations [4,5].

$$L = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \quad (1a)$$

$$A = 1.4749 * 0.2213 * R - 0.3390 * G + 0.1177B + 128 \quad (2b)$$

$$B = 0.6245 \times 0.1949 \times R + 0.6057 \times G - 0.8006 \times B + 128 \quad (3c)$$

The detection of the damaged leaf area was done by avoiding defoliation using A-channel of the CIELAB color model by developing the appropriate MatLab algorithm. A-channel is selected because it is easy for human perception. The other advantage for using CIELAB color model is that the leaf veins are not mistaken as or with the damage.

$$G(i,j) = \begin{cases} 1, & \text{if } f(i,j) \geq T \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Where, T is the threshold value, $G(i,j) = 1$, for image elements of objects and $G(i,j) = 0$, for image elements of the background.

3. Color-Based Segmentation Using K-Means Clustering

K-Means clustering algorithm is an unsupervised algorithm and it is used to segment images based on the $L^*a^*b^*$ color space. K-Means clustering algorithm is an unsupervised algorithm and it was used to segment the region interest from the background and it clusters or partitions the given image into K-clusters or parts based on the K-centroids. K is the value which is used to create number of segments for clustering. Objects can be clustered to one of k groups primarily chosen and the cluster membership is determined by calculating the centroid for each group and assigning each object to the group with the closest centroid. This approach minimizes the overall within-cluster dispersion by iterative reallocation of cluster members [8]. This algorithm aims at minimizing an objective function known as squared error function expressed in equation (3). In this study, $k = 2$, two different centers of the clusters are created based on its distance of each color pixels from the center of the cluster is computed. Let $X = \{x_1, x_2, x_3, \dots, x_{k_i}\}$ be the set of data points and $V = \{v_1, v_2, v_3, \dots, v_k\}$ be the set of cluster centers. Then the objective function $J(V)$ is evaluated as

$$J(V) = \sum_{i=1}^k \sum_{j=1}^{k_i} |x_i - v_j|^2 \quad (3)$$

Where $\|x_i - v_j\|$ is the Euclidean distance between x_i and v_j 'ki' is the number of data points in ith cluster. 'k' is the number of cluster centers. The new cluster center is calculated based on the color intensity value of the i^{th} data point (pixel) and the j^{th} centroid using Euclidean distance equation (4).

$$v_j = \left(\frac{1}{k_i}\right) \sum_{i=1}^{k_i} x_i \quad (4)$$

2.7 Image Segmentation

Image segmentation means partitioning of an image into various parts of the similar features or having some similarity measure. The segmentation can be done using various methods like the otsu's method, K-means clustering unsupervised learning [6,7]. The A-channel is thresholded using Otsu's thresholding method (Otsu, 1979) to segment the damaged and normal parts of the leaves. Thresholding is the transformation of an input image f to an output (segmented) binary image G as follows:

The Euclidean distance (d) between the i^{th} data point $p(x_i, y_i)$, and the j^{th} the centroid $v(x_j, y_j)$ is calculated by (5).

$$d = \left[(x_i - x_j)^2 + (y_i - y_j)^2 \right]^{\frac{1}{2}} \quad (5)$$

• The algorithm of K-means is:

- Read the input image of tomato leaf.
- Transform image from RGB to LAB color space;
- Classify colors using k-means clustering in 'AB' space using Euclidean distance.
- Label each pixel in the image from the results of k-means.
- Generate the images that segment the image by color.
- Select disease containing segment.

4. Result and Discussion

In this study, early blight tomato leaf was acquired and MATLAB Software have been used to implement image enhancement. In the original RGB early blight tomato leaf images were resized to a standard format and converted into gray scale image using the MATLAB built-in function `rgb2gray`.

The acquired tomato leaf images are in jpeg format i.e. in true color format (24-bit color format). The color images are now decomposed into R-, G- and B- color component for Disease identification based on color histogram. This is shown in below screen shot of the mat lab

The captured image can be represented in Red, Green and Blue (RGB). The original 'leaf.jpg' converted into a gray scale image is shown in Figure 2.

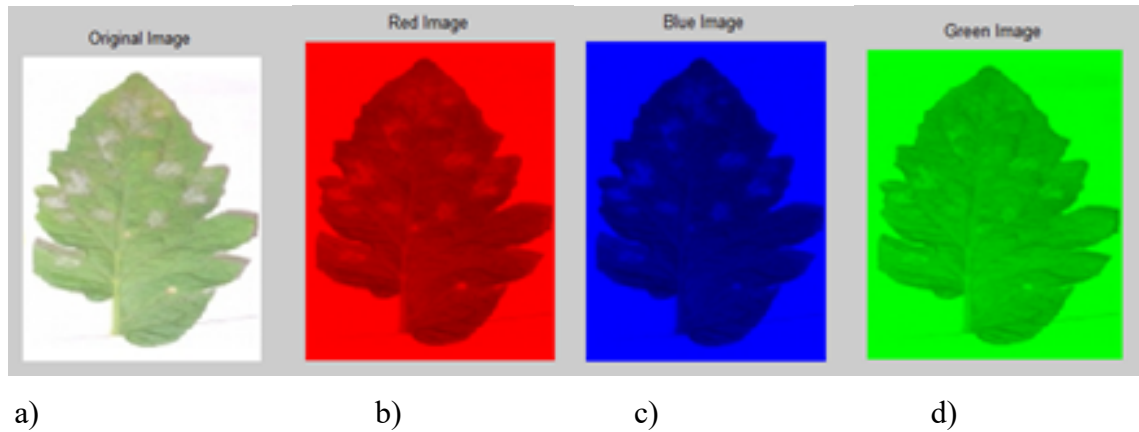


Figure 2: a) RGB image, b) Red image, c) Blue image, d) Green image

4.1 Pre-Processing

The acquired images are enhanced using contrast enhancement method with a contrast limit (range 0 to 1) to increase the local contrast of the tomato leaf image for easily identifying. Figure 6 shows the enhanced images using CLAHE.

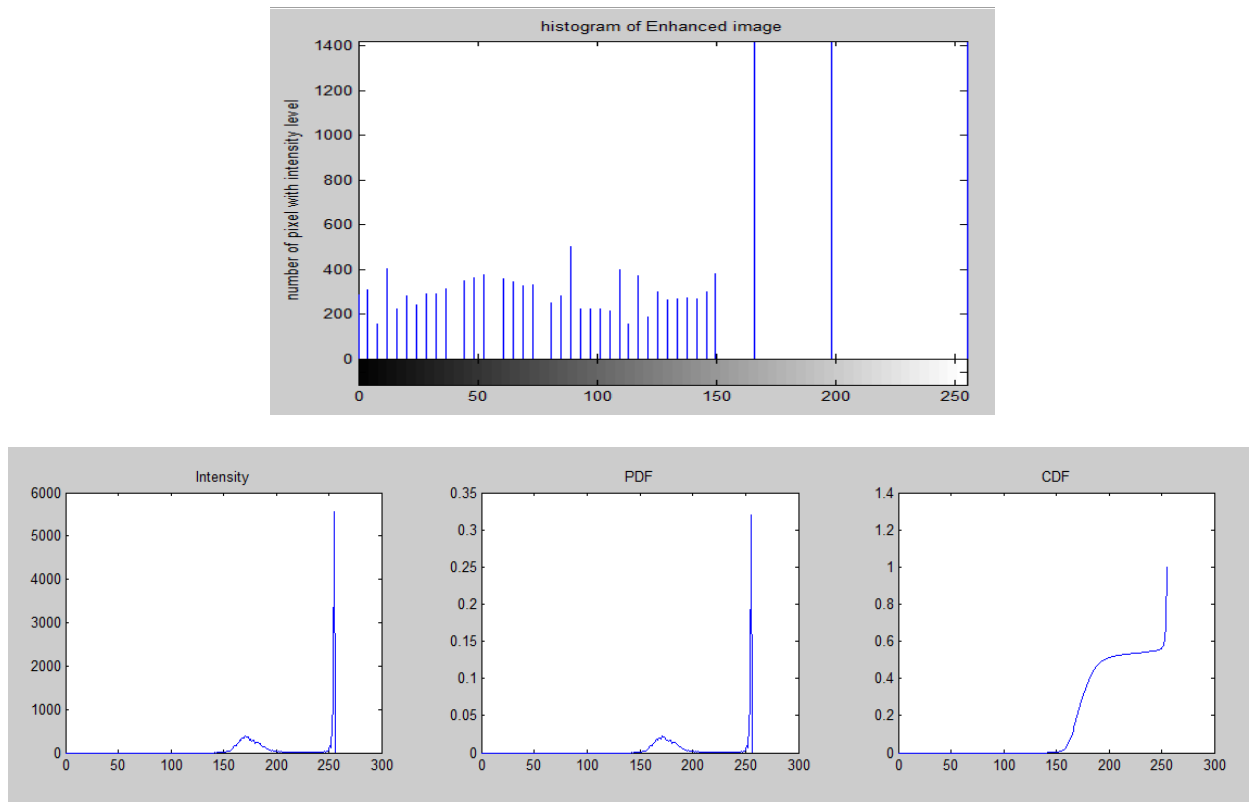


Figure 6: Indicates that the Contrast of Enhanced Image is Increased and the Visuality of The Image is Increased.

Figure 4 represents (a) the input image, b) the enhanced image using Histogram equalization, c) the gray scale image histogram d) the enhanced image's histogram. The horizontal axis of a histogram represents the intensity level or value of color (0-255), while vertical axis shows the number of pixels of each color.

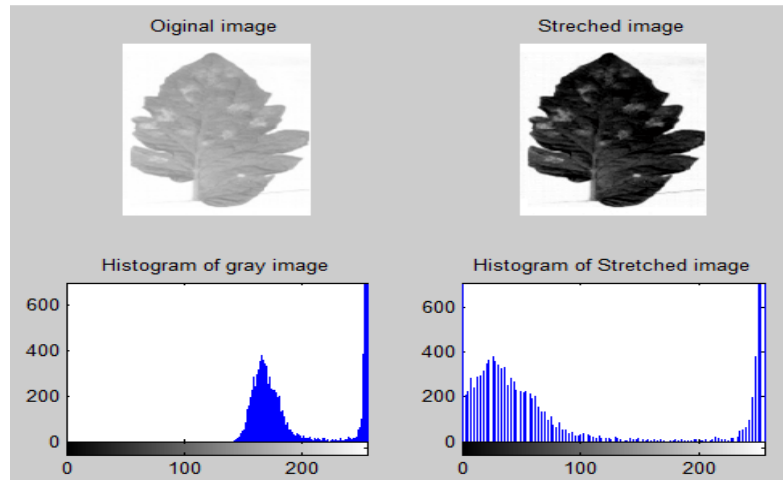


Figure 7: Image Colour Transformation

The enhanced image is transformed into the LAB color space and it is stored in only two channels (A and B components) and it causes reduced processing time for the detect segmentation. The LAB color helps to approximate human vision and it aspires to perceptual uniformity. The L-component closely matches human

perception of lightness and it was used to detect the lesion region of tomato leaf. The transformed CIE LAB color model was used to identify lesion part of the tomato leaf is clearly separated L-component and A-component from the normal as shown below in Figure 4.

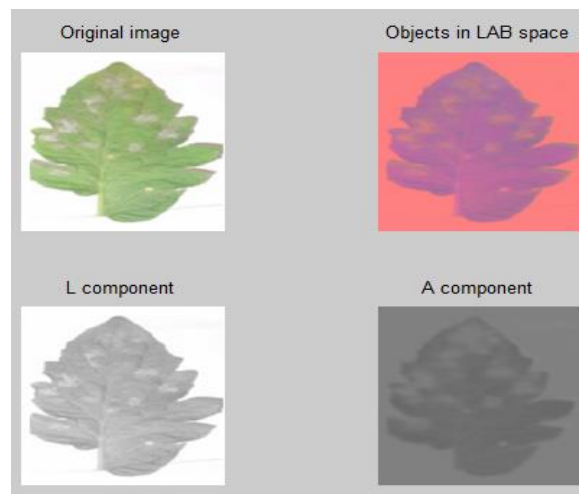


Figure 8: The Image Colour Transformation Result

The Lab* color space allows us to quantify the color utilizing an independent color space.

4.2 Result of Lesion Detection by K-Means Clustering Algorithm

After color Lab space transforming k-means clustering was applied to separate groups of tomato leaf and it treats each image as

having a location in space. The diseased area of the leaf is cluster separately using k-means clustering and it enabled us to obtain the area of the diseased and normal region by avoiding the defoliation parts. Tomato leaf image was clustered into two objects in cluster 1 and objects using k-means clustering with values. The clustered tomato leaf image using k-means clustering is shown in figure 2.

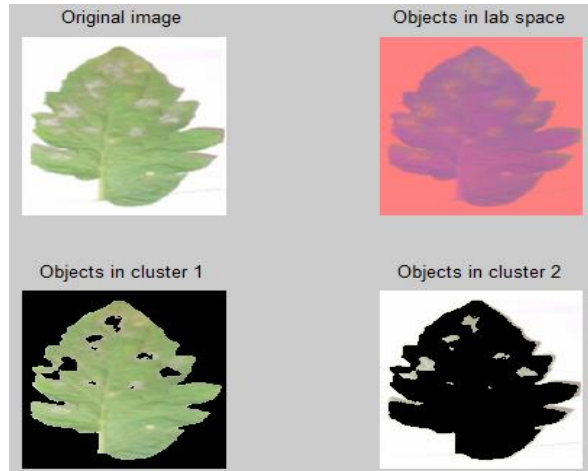


Figure 9: Infected Lesion Segmentation Using K-Means Clustering Algorithm (DI).

Then, A thresholding technique based on Otsu algorithm was applied to get the binary image with black as back ground and white as four grounds and the total area of the early blight of tomato leaf. Following is the results of segmented images using thresholding techniques used to determine the total area of tomato leaf image.



Figure 10: Extraction of Total Area Extraction of Normal Tomato Leaf Area.

Figure 10: Shows Determining the Area Helps to Estimate the Disease Severity of Tomato Leaf

4.3 Entropy

Entropy is another evaluation measure used to compute the randomness or information content in an image. It assists to calculate the uniformity measure in an image. This method of evaluation has derived its idea or concept from information theory and minimum description length principle where the data have discrete random distribution. It was calculated using

$$E = -\sum_{i=1}^a e_i \log e_i \quad (11)$$

Where ‘e’ represents the pixels frequency and ‘i’ represents the intensity value of pixel. Lower value of entropy assures less randomness in image information and vice versa for higher value of

entropy which shows more randomness in image. Therefore for segmentation method with better performance entropy value will be lesser and for poor performance segmentation method entropy value will be higher.

4.4 Diseased Area Calculation

The diseased area is computed by using the original image and the segmented image. This gives the degree of infection in tomato. The following table (table2) shows the result for all diseases.

Total tomato area = 27418q. Pixel units

Disease portion area = 3103 sq. pixel units

Infection % = (total tomato area/ disease portion) *100
 = (27418/3103) *100
 = 11.3173%

No.	Tomato Leaf	Total Area	Diseased Area	Difference	% Difference
1.	ML	27418	3103	0.113173827	11.31738274
2.	ML	20508	3083	0.150331578	15.03315779
3.	LL	18198	2189	0.120287944	12.02879437
4.	LL	15654	4483	0.286380478	28.63804778
5.	ML	18610	1333	0.071628157	7.16281569
6.	LL	19679	2325	0.118146247	11.81462473
7.	ML	18974	2295	0.120954991	12.0954991
8.	ML	18602	1577	0.084775831	8.477583056
9.	LL	11316	1855	0.163927183	16.39271828
10.	ML	18726	2759	0.147335256	14.73352558
11.	LL	20139	375	0.018620587	1.862058692
12.	ML	25004	3868	0.154695249	15.46952488
13.	LL	15897	4735	0.297854941	29.78549412
14.	ML	16166	286	0.017691451	1.769145119
15.	LL	14457	3937	0.272324825	27.23248253
16.	ML	21027	1550	0.073714748	7.371474771
17.	LL	20130	1195	0.059364133	5.936413313
18.	ML	16415	5565	0.33901919	33.90191898
19.	LL	10189	2544	0.249681029	24.96810286
20.	ML	21844	6359	0.291109687	29.11096869
21.	LL	18687	3918	0.209664473	20.96644726
22.	LL	18687	3918	0.209664473	20.96644726
23.	LL	20275	1164	0.057410604	5.741060419
24.	LL	18439	2034	0.11030967	11.03096697
25.	ML	17476	2175	0.124456397	12.44563973
26.	ML	20686	773	0.037368268	3.736826839
27.	ML	21314	1320	0.061931125	6.193112508
28.	ML	18523	2553	0.137828645	13.78286455
29.	ML	19277	1869	0.09695492	9.695492037
30.	ML	14478	2762	0.190772206	19.07722061
31.	ML	15875	1299	0.081826772	8.182677165
32.	ML	16881	1657	0.098157692	9.815769208

ML = Middle tomato leaves and LL = Lower tomato leaves.

Entropy Value of Different Segmentation Methods

Table 1: Results of Total and Diseased Area the Tomato Leaf

5. Conclusion

In this work, a method has been developed to detect the ETB images by means of automatic image segmentation techniques. The samples were collected from HrU research area "Raree". First, the early blight of tomato leaf images was acquired and pre-processed to enhance the contrast of the images. The RGB image was transformed into L*a*b space and then, K-means Clustering algorithm is used to segment the disease part of the leaves with two different cluster processes. After all iteration the final image is displayed in

the second cluster. The centroid weight is calculated to identify accurate portion and also to calculate infection percentage. The intensity value of each centroid is plot based on red, green and blue pixels for infected portion. The mathematical formula is used to calculate the infection percentage. With regards to future work different segmentation algorithm and different techniques to find infected area is used to obtain more accurate result. In this work the proposed work is limited to K-means clustering segmentation and mathematical formulations [9-13].

References

1. Patil, J. K., & Kumar, R. (2011). Color feature extraction of tomato leaf diseases. *International Journal of Engineering Trends and Technology*, 2(2), 72-74.
2. Huiyu Z., Wu, J. & Zhang, J. 2010. Digital image processing part I. Ventus publishing aps, ISBN 978-87-7681-541-4.
3. Asrat, P., & TEFERA, T. (2005). PROCEEDINGS OF THE 22ND ANNUAL RESEARCH AND EXTENSION REVIEW MEETING.
4. Liu, L., & Zhou, G. (2009, December). Extraction of the rice leaf disease image based on BP neural network. In 2009 International Conference on Computational Intelligence and Software Engineering (pp. 1-3). IEEE.
5. Vibhute, A., & Bodhe, S. K. (2012). Applications of image processing in agriculture: a survey. *International Journal of Computer Applications*, 52(2), 34-40.
6. Mollazade, K., Omid, M., Tab, F. A., Kalaj, Y. R., Mohtasebi, S. S., & Zude, M. (2013). Analysis of texture-based features for predicting mechanical properties of horticultural products by laser light backscattering imaging. *Computers and electronics in agriculture*, 98, 34-45.
7. Yasikka, M., & Santhi, E. (2015). Foliage Measurement Using Image Processing Techniques. *International Journal For Trends in Engineering and Technology*, 5(2).
8. Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k Means Clustering algorithm for prediction of Students Academic Performance. arXiv preprint arXiv:1002.2425.
9. Lemma, D. (2002). Tomatoes, Research Experiences and Production Prospects: Research Report, No. 43. Agricultural Research Organization, Ethiopia.
10. Hitimana, E., & Gwun, O. (2014). Automatic estimation of live coffee leaf infection based on image processing techniques. arXiv preprint arXiv:1402.5805.
11. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.
12. Ghodsi, R., Yani, R. M., Jalali, R., & Ruzbahman, M. (2012). Predicting wheat production in Iran using an artificial neural networks approach. *International Journal of Academic Research in Business and Social Sciences*, 2(2), 34.
13. Gonzalez, R. C. (2009). *Digital image processing*. Pearson education india.

Copyright: ©2023 Semere A Asefa, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.