Maize Leaf Images Segmentation Using Color Threshold and K-means Clustering Methods to Identify the Percentage of the Affected Areas

Amel H Abbas¹, Nabeel M Mirza², Samar A Qassir³, Layla H Abbas⁴

¹,³,⁴ Department of Computer Science, College of Science, University of Mustansiriyah, Baghdad, Iraq.
² Department of Physics, College of Education, University of Mustansiriyah, Baghdad, Iraq.

Abstract. Control of diseases affecting plants is very important as it relates to the issue of food security, which is a very serious threat on human life. According to the International Maize and Wheat Improvement Center (CIMMYT), most of the maize diseases are caused by mildew. There are more than fifty different mildew diseases affecting maize. The diseases that infect plants go through different stages of the degree of infection. Therefore, determining the degree of injury helps decision-makers spend money on controlling the disease properly. The common method used to determine the degree of injury is visual examination. It's known that visual examination leaves a large area for error. It is unable to determine the degree of injury with high accuracy as a percentage for example. In this study, we used a new segmentation method to determine the percentage of affected areas for maize leaves. This methodology consist of three main stages: (a) classifying the image by K-Means clustering, (b) segmentation using Color Threshold, and (c) Estimate the affected area. Different color spaces (RGB, HSV, YCba, and L*a*b*) have been used in the segmentation and classification process. The results obtained indicate that the color threshold segmentation is more efficient than K-means clustering in terms of determining the injury ratio, especially with HSV color space. While it contrary to the RGB color model because it gives the worst result.

1. Introduction

Agriculture is the source of the economy of most countries. About (60-70) percentage of the earth's population depends on agriculture. Research in agriculture aims to increase production and quality while reducing spending and thus increasing profits. Plant diseases influence the growth of plants and lead to significant losses in production, which threaten the issue of food security. Therefore, the search for quick and accurate ways to detect diseases automatically through the indicators that appear on the plant leaf is significant. Computer vision is a solution that relies on automatic inspection. The image processing techniques are very effective in determining the degree of injury of plant leaves. Often times, to separate objects and identify the regions of interest with high accuracy, an image segmentation technique is used. The common segmentation algorithms that used for crop leaf diseases involve edge detection or threshold methods [1]. Some researchers [2], [3], [4], and [5] had investigated C-means and K-means clustering methods for automatic recognition. Other researchers such as [6] and [7] they used Artificial Neural Network (ANN) method and Support Vector Machine (SVM) method for identifying disease. Segmentation of color image provides more information than gray image, thus it is an essential technique for identifying diseases [8], and [9].
There are many types of maize diseases, but the focus will be on three common types (Hyalothyridium leaf spot, eyespot, and Macrospora leaf streak). Our study focuses on the separation of diseased regions from the healthy regions of maize leaves, which includes content of information and then comparing between them by color threshold segmentation and K-means clustering. The four-color models that will contribute to the detection of the affected area and calculate the percentage of an affected area are RGB (Red, Green, Blue), HSV (Hue, Saturation, Value), L*a*b (Luminance, Color components) and YCbCr (Luminance, Chrominance).

2. Problem Statement
Just as plants go through different stages of their life cycle, the diseases that they infect also go through different stages of infection [10]. Plant diseases can also appear in any season because of environmental and climatic changes [11]. Consequently, specialists in plant disease control and decision-makers resort to spend large amounts in order to obtain pesticides and equipment for the control of these pests, in addition to the costs of raw materials before injury such as seeds, watering, workers, continuous monitoring and effort, etc. Expenditure and loss rates depend on the degree and stage of injury. Sometimes the degree of infection is very high and spending more money to buy pesticides will not be beneficial because the pesticides at the wrong time will not be able to cure the disease and there will be two losses: crops loss and money. While, if the disease in the plant is simple and detected at a primitive stage of plant life before the outbreak and development of the epidemic to an advanced stage, there is no concern about the future of the agricultural product.

The common method used to determine the degree of injury in a plant is visual vision; this method may not be ideal because visual vision is not able to diagnose infection at the plant as an accurate percentage. This research offers a suggestion to be able to determine the degree of infection with high accuracy in the plant with the appearance of symptoms on the leaves of the plant.

In order to expand methods of diagnosis, and identifying with reliable and quick for maize eyespot rust, leaf streak rust, and Hyalothyridium mildew in farm. The images of these three types of corn disease are segmented using the based on K-means clustering method and color threshold segmentation according to the conditions of the disease images without the need to enhance the images or extract features as in the references [12], [13], and [14].

3. Methodology
The maize leaf diseases caused by fungi, bacteria and viruses are a major concern due to the impact of maize yield and quality [15]. Therefore, it is necessary to prevent these diseases and identify the types of these diseases. According to the (CIMMYT), most of the maize diseases are caused by mildew. There are more than fifty different mildew diseases affecting maize [16], including Eyespot leaf, Macrospora leaf, and Hyalothyridium leaf spot. Their images are shown in figure 1. These images obtained from the international maize and wheat organization website. The symptom of Eyespot sores characterized by concentric rings surrounded by yellowish halos [17], Macrospora lesions are an elongated stripe up to several centimetres, while the Hyalothyridium lesions characterized by elliptical brown spots [18].

| Type of disease | Description of the disease |
|-----------------|-----------------------------|
| Healthy Leaf    | ![Healthy Leaf Image]        |
| Eyespot         | ![Eyespot Image]            |

2
3.1 K-means Clustering

Clustering is the capacity of grouping a set of targets so that all targets in the same group are more similar to each other than to those in other groups. K-means is common methods of clustering; they are used when the data is without specific categories or groups [19]. Figure (2) illustrates the process of K-means clustering; the method works frequently to place each data point to a single cluster according to the feature that provided. Data points are clustered based on feature similarity. The K-means measurement consists of two separate parts [20]: First part is assigning each data point to its nearest centroid, and the second part requires computing the centroids using the mean of all data points assigned to that centroids cluster. In addition, Euclidean distance is one of the most used methods to assign the distance of the closest centroid.

Assumed that \((x, y)\) represents the original coordinates of an image that have to be cluster into \(k\) number of cluster, \(I(x, y)\) represents input pixels to be cluster the color image, and \(C_i\) be the centroid of color image. The distance between centroid and each pixel of the image can be determined as Euclidean distance (\(d\)) [20], [21]:

\[
d = \|I(x, y) - C_i\|
\]  

(1)
After calculating the distance, each data points are distributing into K clusters, and then all elements are appointed to the nearest centroid. Next is updating the position of centroids based on newly created clusters, using the following equation:

\[
C_k = \frac{1}{k} \sum_{y \in C_k} \sum_{x \in C_k} I(x, y)
\]  

(2)

Repeat the process until it meets the allowable value, and then reshape the cluster pixels into image.

3.2 Color Threshold Application

Color threshold application used to extract parts of the image that show within a definite color range. This application can be used to detect objects with consistent color values based segmentation, this is done by other color spaces such as HSV, YCbCr, and L*a*b.

There are many color spaces that are related to color model RGB (Red, Green, Blue). For example, HSV is transformation of a Cartesian RGB color model; their components are relative to the RGB color model from which derived. YCbCr is a scaled version of YUV and uses widely in video and image compression schemes such as MPEG and JPEG [22]. Where Y indicates to intensity = 0.299 x R + 0.587 x G + 0.114 x R. While, each of Cr (B - Y), and Cb (R - Y) indicate to color difference signals for blue and red respectively [23]. L*a*b* is also known as CIE: used to produce a perceptual color space. Where: L* is luminosity, its value equal to (0) for darkest and (100) for brightest. While, both of (a* and b*) represent the color components for gray [24], and [25].

Conversion of Hue Saturation Value (HSV) color space from RGB color model done by appointing the following values [23]:

\[
\begin{align*}
R' &= \frac{R}{255} \\
G' &= \frac{G}{255} \\
B' &= \frac{B}{255}
\end{align*}
\]  

(3)

\[
C_{\text{high}} = \max (R', G', B') \quad , C_{\text{low}} = \min (R', G', B')
\]  

(4)

\[
\Delta = C_{\text{high}} - C_{\text{low}}
\]  

(5)

\[
H = \begin{cases} 
0 & \Delta = 0 \\
60^\circ \times \left( \frac{G' - B'}{\Delta} \mod 6 \right) \quad , C_{\text{high}} = R' \\
60^\circ \times \left( \frac{B' - R'}{\Delta} + 2 \right) \quad , C_{\text{high}} = G' \\
60^\circ \times \left( \frac{R' - G'}{\Delta} + 4 \right) \quad , C_{\text{high}} = B'
\end{cases}
\]  

(6)

\[
S = \begin{cases} 
\frac{\Delta}{C_{\text{high}}} & C_{\text{high}} > 0 \\
0 & \text{otherwise}
\end{cases}
\]  

(7)

\[
V = \frac{C_{\text{max}}}{C_{\text{high}}} \quad , \text{where } C_{\text{max}} = 255
\]  

(8)
Conversion of Red Green Blue (RGB) to YC_rC_b color space as [22]:

\[
Y = 0.299 \times R + 0.587 \times G + 0.114 \times B
\]

\[
C_b = B - Y
\]

\[
C_r = R - Y
\]  \hspace{1cm} (9)

Conversion of Red Green Blue (RGB) to \( L^*a^*b^* \) color space as [26]:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} =
\begin{bmatrix}
X_r & X_g & X_b \\
Y_r & Y_g & Y_b \\
Z_r & Z_g & Z_b
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\hspace{1cm} (10)

### 3.3 Estimation of Affected Area

The color threshold application consists of three main steps: (a) choosing the color threshold command (b) uploading the image to be segmented and (c) choosing the color space. This application able to get the best color threshold based on many color spaces available such as (RGB, HSV, YC_rC_b, and \( L^*a^*b^* \)). The CTA removes each pixel whose color components value lie outside the threshold range and allows only pixels within the threshold to appear in the image with display their histograms of each channel of HSV, YC_rC_b, and \( L^*a^*b^* \) as illustrated in figures (3) - (5). The following is a summary of the most important steps to evaluate the affected area:

a) Convert the input image to a binary (black and white).

b) Fill holes and image regions.

c) Determine the affected areas, which are representing the white pixels \( W_p \).

d) Calculate the number of white pixels.

e) Divide white pixels per the total number of pixels in the image to get a percentage of the affected area.

\[
PAA = \frac{W_p}{T_p} \times 100\% \]  \hspace{1cm} (11)

where \( PAA \) denotes the percentage of the affected area, \( W_p \) represents the total number of white pixels of the affected area within the paper disease image, and \( T_p \) is the total elements of image (affected and healthy).

**Figure 3.** Color threshold of HSV color space for Macrostomia leaf disease.
Figure 4. Color threshold of YC\textsubscript{b}C\textsubscript{r} color space for Hyalothyridium leaf spot disease.

Figure 5. Color threshold of L*a*b* color space for eyespot.

4. Experimental Results
For the detection of injury ratio in maize leaves, MATLAB tool was used. K-means clustering and color threshold segmentation were applied. Figure (6) shows the original and resulted images after apply K-means clustering, and Figure (7) shows the original and resulted images after applied color threshold segmentation. Tables 1 and 2 illustrate the estimated values of the disease-affected area images for color spaces for color spaces (RGB, HSV, YC\textsubscript{b}C\textsubscript{r}, and L*a*b*) after apply color threshold segmentation and K-mean segmentation respectively.

| Type of Disease | Original Image | RGB | HSV | YC\textsubscript{b}C\textsubscript{r} | L*a*b* |
|-----------------|----------------|-----|-----|-----------------|--------|
| Hyalothyridium  | ![Image 1](image1.png) | ![Image 2](image2.png) | ![Image 3](image3.png) | ![Image 4](image4.png) | ![Image 5](image5.png) |

Table 1

Table 2
Figure 6. Resulted images for different color spaces after apply k-mean clustering.

| Type of Disease | Original Image | RGB | HSV | YCxCr | L*a*b* |
|-----------------|----------------|-----|-----|--------|--------|
| Hyalothyridium  | ![Image](image1.jpg) | ![Image](image2.jpg) | ![Image](image3.jpg) | ![Image](image4.jpg) | ![Image](image5.jpg) |
| Macrospora      | ![Image](image1.jpg) | ![Image](image2.jpg) | ![Image](image3.jpg) | ![Image](image4.jpg) | ![Image](image5.jpg) |
| Eyespot         | ![Image](image1.jpg) | ![Image](image2.jpg) | ![Image](image3.jpg) | ![Image](image4.jpg) | ![Image](image5.jpg) |

Figure 7. Resulted images for different color spaces after apply color threshold segmentation.

From the results, obviously that the K-Mean clustering method yielded good segmentation in most color spaces in terms of determining the affected area. According to the results obtained which shown in Tables (1) and (2); the comparison among color spaces appear that the HSV color space given the best results among other color spaces for both segmentation methods K-means clustering and color threshold method.
Table 1. The evaluation values of affected area for color spaces (RGB, HSV, YCbCr, and L*a*b*) after apply K-mean clustering.

| Image Name     | RGB Area | HSV Area | YCbCr Area | L*a*b* Area |
|----------------|----------|----------|------------|--------------|
| Hyalothyridium Leaf Spot |          |          |            |              |
| Leaf 0         | 7030     | 24000    | 30100      | 29200        |
| Leaf 1         | 2340     | 26400    | 7650       | 9630         |
| Macropora Leaf |          |          |            |              |
| Leaf 2         | 27200    | 91300    | 95900      | 69400        |
| Leaf 3         | 16200    | 39500    | 32200      | 31400        |
| Eyespot Leaf   |          |          |            |              |
| Leaf 4         | 22300    | 46500    | 1270       | 5920         |
| Leaf 5         | 9350     | 85400    | 24700      | 93400        |

Table 2. The evaluation values of affected area for color spaces (RGB, HSV, YCbCr, and L*a*b*) after apply color threshold segmentation.

| Image Name     | RGB Area | HSV Area | YCbCr Area | L*a*b* Area |
|----------------|----------|----------|------------|--------------|
| Hyalothyridium Leaf Spot |          |          |            |              |
| Leaf 0         | 14638    | 36000    | 36000      | 33900        |
| Leaf 1         | 1060     | 9955     | 9880       | 9750         |
| Macropora Leaf |          |          |            |              |
| Leaf 2         | 25600    | 99000    | 97600      | 76800        |
| Leaf 3         | 13700    | 40500    | 36200      | 31068        |
| Eyespot Leaf   |          |          |            |              |
| Leaf 4         | 450      | 1420     | 1050       | 411.25       |
| Leaf 5         | 7270     | 27900    | 26800      | 11600        |

5. Discussions of Results

Our methodology yielded excellent results in detection of the affected areas using segmentation of the color threshold, specifically in HSV color space, as well as the YCbCr and L*a*b* color spaces have given acceptable results, while RGB was the worst among them. As for the K-Mean clustering method, the color space L*a*b* yielded excellent results, as well as the HSV and YCbCr color spaces have given acceptable results, while RGB was also the worst among them.

There is a lot of previous works related to the percentage of the affected areas for plant leaves, but each research is quite different in terms of the study of disease type and plant. So, the results of this system cannot compare with the results of previous researches because of the different rates of damage from one image to another. For example, the researcher [12] detected the affected areas of the plant leaf diseases (Alternaria Alternata, Bacterial Blight, Fungal Leaf Spot and Fungus Anthracnose) and reached the following percentage results: 15.01, 19.60, 15.61, and 15.00 respectively.
Table 3. Percentage of the affected area.

| Disease       | Hyalothyridium (a) | Hyalothyridium (b) | Hyalothyridium (c) |
|---------------|--------------------|--------------------|--------------------|
| Percentage (%)| 60.40              | 26.28              | 41.90              |

| Disease       | Macrospora leaf (a) | Macrospora leaf (b) | Macrospora leaf (c) |
|---------------|---------------------|---------------------|---------------------|
| Percentage (%)| 25.20               | 21.40               | 08.99               |

| Disease       | Eyespot leaf (a)    | Eyespot leaf (b)    | Eyespot leaf (c)    |
|---------------|---------------------|---------------------|---------------------|
| Percentage (%)| 14.59               | 20.78               | 03.90               |

6. Conclusions

In this paper, the images of three types for maize diseases have been converted from RGB color model to HSV, YCrCb, and L*a*b* color spaces in order to detect and calculate the ratio of affected areas. To obtain the best result, we recommend using color threshold segmentation with HSV color space, and avoid using RGB color space as it produces the worst results. The calculation of the ratio of the affected area helps the specialists in the control of plant diseases to make the right decision. In future, other quality measurement methods such as Entropy, C-Means, etc., can be used as a similar effective to replace both of K-means and color threshold segmentation method, or experimenting with the same system more broadly to include a variety of plant diseases such as wheat, olives, potatoes and others.

7. References

[1] Mori F Mori T 2012 Region Segmentation and Object Extraction Based On Virtual Edge and Global Features In: Computer Vision ACCV 2012 Workshops (Springer Berlin Heidelberg), pp 182–193

[2] Jing H Daoliang L Guifen Ch Qingling D Yeiqi H 2012 Image Segmentation Method for Crop Nutrient Deficiency Based on Fuzzy C-Means Clustering Algorithm Intelligent Automation & Soft Computing, vol 18, no 8:35 pp 1145-1155

[3] Ji Z Xia Y Chen Q Sun Q Xia D Feng D 2012 Fuzzy C-Means Clustering with Weighted Image Patch for Image Segmentation Applied Soft Computing vol 12 no 6 pp 1659-1667

[4] He Y Hussain M Y Ma J Shafei B Steidl G 2012 A New Fuzzy C-Means Method with Total Variation Regularization for Segmentation of Images with Noisy and Incomplete Data Pattern Recognition vol 45 no 9 pp 3463–3471
[5] Max M 2011 A De-Texturing and Spatially Constrained K-means Approach for Image Segmentation Pattern Recognition Letters vol 32 no 2 pp 359-367

[6] Dheeb A Malik B Sulieman B 2011 Detection and Classification of Leaf Diseases using K-means-based Segmentation and Neural-networks-based Classification Information Technology Journal vol 10 no 2 pp 267-275

[7] Changqi O Daoliang L Jianlun W Shuting W Yu H 2013 The Research of the Strawberry Disease Identification Based on Image Processing and Pattern Recognition Computer and Computing Technologies in Agriculture vol 392 pp 69-77

[8] Xiaohong H Bingjun L Fang L 2013 Image Segmentation Based on Graph Theory in Multi-Color Space for Maize Leaf Disease Transactions of the Chinese Society of Agricultural Engineering vol 44 no 2 pp 177-181

[9] Li G Ma Z Huang C 2010 Segmentation of Color Images of Grape Diseases Using K-means Clustering Algorithm Transactions of the Chinese Society of Agricultural Engineering vol 26 no 14 pp 32-37

[10] Dilani D Pedro W Peter K Kevin D Paul W 2017 Life Styles of Colletotrichum Species and Implications for Plant Biosecurity Fungal Biology Reviews (Elsevier) British Mycological Society vol 31 pp 155–168

[11] Raquel G Emilia H Wagner B 2008 Climate Change and Plant Diseases Scientia Agricola vol 65 pp 98–107

[12] Iqbaldeep K Gifty A Amit V 2016 Detection and Classification of Disease Affected Region of Plant Leaves using Image Processing Technique Indian Journal of Science and Technology vol 9 no 48 pp 1–13

[13] Dange D Sayyad M 2015 Computer Vision Image Enhancement and Plant Leaves Disease Detection International Journal of Modern Trends in Engineering and Research (IJMTER) vol 2 no 6 pp 106–110

[14] Kiran R Ujwalla G 2014 An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques IOSR Journal of Computer Engineering (IOSR-JCE) vol 16 no 1 pp 10–16

[15] Jagadeesh D Rajesh Y Abdulmunaf S 2014 Automatic Fungal Disease Detection based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops International Journal of Image, Graphics and Signal processing, vol 6 no1 pp24-31

[16] The CIMMYT Maize Program 2004 Maize Diseases: A Guide for Field Identification 4th edition Mexico, D. F.: CIMMYT

[17] Martha P 2016 Stenocarpella Maydis: Identification, Management, and Population Diversity Ph.D. thesis Botany and Plant Pathology department Purdue University West Lafayette, Indiana.

[18] Munkvold G White D 2016 Compendium of Corn Diseases the American Phytopathological Society 4th Edition (St. Paul Minnesota) (U.S.A)

[19] Archana S Avantika Y Ajay R 2013 K-means with Three Different Distance Metrics International Journal of Computer Applications vol 67 no 10 pp 13-17

[20] Nameirakpam D Yambem J 2015 Image Segmentation Method using K-means Clustering Algorithm for Color Image Advanced Research in Electrical and Electronic Engineering vol 2 no 11 pp 68-72

[21] Nabeel M Ali Abid D Hazim G 2013 Underwater Images Quality Assessment Based on the Minimum Distance Classifier (MDC) and Quality Factor Algorithms College of Education Journal, Vol. 4, P.785-803

[22] Kolkur S Kalbande D Shimpi C Jatakia J 2017 Human Skin Detection Using RGB, HSV and YCbCr Color Models Advances in Intelligent Systems Research, (Atlantis Press) vol 137 pp 324-332

[23] Rashmi H Shashidhar M Prashanth K 2014 Automatic Tracking of Traffic Signs Based on HSV International Journal of Engineering Research & Technology (IJERT) vol 3 no 5 pp 914-917
[24] Granato D Masson M 2010 *Instrumental Color and Sensory Acceptance of Soy-based Emulsions: A Response Surface Approach* Food Science and Technology vol 30 no 4 pp 1090–1096

[25] Vivek S Messala S Ashwini V 2012 *Colour Based Image Segmentation Using L*A*B* Colour Space Based on Genetic Algorithm* International Journal of Emerging Technology and Advanced Engineering vol 2 no 6 pp 156–162

[26] Ford A Roberts A 1998 *Color space conversions* Westminster University London United Kingdom