

# Tomato Disease Segmentation using K-Means Clustering

Binne M. James  
Research Scholar  
PSGR Krishnammal  
College for Women  
Coimbatore

Punitha S. C.  
Head, Assistant Professor  
PSGR Krishnammal  
College for Women  
Coimbatore

## ABSTRACT

Tomato production is daunting task because of various attacks from different bacteria. The symptoms of infected tomato are usually changed in color, infected spots, special kind of specks, and hollow areas with concentric rings having different colors on the outer surface area. The main aim of this research work is to detect tomato diseases using segmentation algorithm. Algorithm development includes two main steps; preprocessing step like contrast enhancement and segmentation using k-means clustering. Tomato samples are collected from local market and data acquisition has been carried out for database preparation. There are six different types of tomato disease are focused in this work such as Anthracnose, Bacterial canker, Bacterial speck, Bacterial spot, Early blight and Late blight. The segmentation techniques based in pixel locality are applied to get the segmented tomato and infected part of the tomato. The experimental result shows that the segmentation of tomato disease using K-means clustering performs well in finding out different diseases accurately.

## Keywords







Bruises, LDA, Bacterial speck, FFB, Canker, Centroids, Epidemics.

## 1. INTRODUCTION

Fruits and vegetables production face significant losses in India due to bacterial infection. Vegetable quality is frequently referred to size, shape, color and bruises from which it can be classified and sorted. The accurate and timely detection of diseases is the basis for built-in management of a farm. In general, the features of any fruits or vegetables are individually extracted by manual inspection [1]. To make this task more accurate and efficient various studies are carried out to detect diseases automatically by applying image processing techniques. For production, it is used to develop automatic sort system for examining the quality of fruits to increase customer satisfaction level [2].

Bacterial diseases of tomatoes can be some of the most serious and destructive diseases affecting both greenhouse-grown and field crops. Under most field conditions they can cause localized epidemics affecting young rising fruit; the greenhouse total crop losses can occur. The accurate detection of tomato disease is challenging task to researchers. The following table (table1) explains that the different diseases of tomato fruits with the description about the disease.

Table1. Tomato diseases

Disease Name	Image	Description
Anthracnose		Small, circular, slightly sunken spots...
Bacterial Canker		Wilted, browning plants and fruit spots are like bird's-eye
Bacterial Spot		Small, darken, slightly increased dots on fruit.
Early Blight		Stems appear as wound with dark border and gray center.
Late Blight		Brown, black stem lesions.
Bacterial Speck		Tiny, dark brown to black spots with contiguous yellow hollow.

This work presents tomato fruit disease identification using image processing techniques to monitor tomato diseases based on color space segmentation. The datasets used for this experiment was collected based on real sample images for tomato at different disease, which were collected from a local market. The proposed approach consists of three different phases; like pre-processing, segmentation, and calculating disease area. In the pre-processing phase the input images are resized to 250x250 pixels for reduce their color index. Contrast enhancement technique is used to enhance the color edges. In the segmentation phase clustering algorithm is used to segment affected portion of the tomato image from the original image. Finally the infected area is calculated to know about the disease severity.

## 2. RELATED WORK

This section reviews the existing approaches used to tackling the difficulty of disease categorization of various leafs, plants, fruits and vegetables.

Polder et al [3], proposed an approach to measure the ripeness of tomatoes. In this the standard RGB images and hyper spectral images are used to classify the ripeness stages using Linear discriminant analysis (LDA) technique.

Jaffar et al [4], depict the relationship between the color of palm oil fruits and their ripeness using K-means clustering for segmented the fruit fresh bunches (FFB) colors and also classify ripeness stages as under ripe, ripe and overripe depending on different color intensity using fuzzy classification. The system achieves 88.74% of accuracy.

Paulraj et al [5], proposed an approach for identifying ripeness of banana fruit using neural network model with back propagation method and achieves 96% of accuracy. Rizam et al [6], approach to measure ripeness of watermelon fruits depends on YCbCr Color Space and the neural network classification achieves 86.51% of accuracy.

R.Kalaivani et al [7], presented a tomato crop grading system based on color based segmentation of k-means clustering algorithm. Mahdi M et al [8], proposed a hand-held system to measure nutrient level and the health status based on leaf colors.

Chawaraj Jaisin et al [9], proposed a sorting system to find individual longans from bunches of longan fruits using various image processing techniques. Hadi Izadi et al [10], determined the relationships between fruits and vegetables based on size and projected area of single and multiple dimensions. Meenu Dadwal et al [11], present a histogram based classification of ripen and unripen tomatoes using ripeness levels.

## 3. TOMATO DISEASE IDENTIFICATION SYSTEM

The proposed work includes seven different phases to detect tomato disease. In first phase the image is resized as 250x250 pixels to reduce color intensity values. Second phase, image is preprocessed by using contrast enhancement technique to enhance the contrast of disease portion. Third phase, k-means clustering algorithm is used to segment the disease portion from original image.

In fourth phase, the boundary is extracted for segmented image to get better accuracy. Fifth phase weighted centroid is calculated for output image and the sixth phase image is displayed based on standard deviation value<50. Finally the infected area is calculated using mathematical formula by

using the original image area and disease area. The following figure (Fig1) shows the proposed framework for finding tomato diseases.

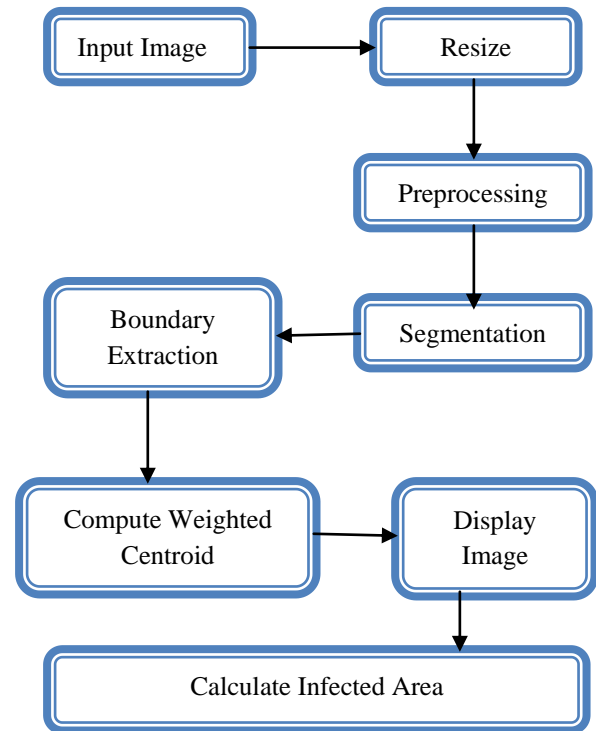


Fig1. Proposed Framework

### 3.1 Data Collection

The data was collected from local market place. The dataset contains six different diseases of tomatoes like anthracnose, speck, spot, canker, early bight and late blight.

### 3.2 Preprocessing

The proposed system is implemented in Matlab 2014a software. The input image is resized using imresize() function and the contrast is enhanced using imadjust() function with various parameters.

### 3.3 Methodology

#### K-Means Clustering

K-means clustering provides partitioning of the image based on color space. K is the value which is used to create number of segments for clustering. In this application k=5, five different centers of the clusters are created based on its distance of each color pixels from the center of the cluster is computed. The mostly used distance measure is Euclidean distance, formula to calculate Euclidean distance is given below,

$$(E. d.)^2 = \min\{\mu_1 \dots \mu_k\} \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} |X(i, j) - \mu^2|$$

$$\mu_k = \frac{\sum_{i=1}^{i=n} P(i) * i}{\sum_{i=1}^{i=n} P(i)}$$

Based on the Euclidean distance of each pixel color from the center, the pixel is marked to the cluster where the Euclidean distance is value minimum. For each iteration in segmented process the center of mass is updated based on average of each color pixels. The process is repeated until the mass comes same, at that moment the process is stopped.

Basically  $k$ -mean clustering algorithm classifies the objects into  $k$  number of classes based on a set of features. The classification is performed by minimizing the sum of squares of distances among the data objects and their equivalent clusters. In the cluster extraction block first transforms the segmented image from  $R*G*B$  to  $L*a*b$  color space. Then system classifies different colors using  $k$ -mean clustering in " $L*a*b$ " space. Based on the result of previous step system performs labeling process in the segmented image. Typically  $k$ -mean clustering algorithm selects random mean pixel in every cluster image and finally mined the number of clusters accessible in image based on the value of  $k$ .

#### 4. EXPERIMENTS AND RESULTS

The experiment is carried out by using different tomato disease by implementing algorithms in Matlab platform. First the input image is resized by using matlab function `imresize()` with 250x250 pixel rates. The following figure (Fig2) shows the resized original image.



Fig2. Resized Image

##### 4.1 Contrast Enhancement

In this preprocessing phase, the image is enhanced based on contrast enhancement technique by using `imadjust()` function with the increase and decrease parameters. The following figure (Fig3) shows the contrast enhanced image.



Fig3. Contrast Enhanced Image

##### 4.2 Segmentation

The segmentation is done by implementing K-Means Clustering algorithm with the  $k$  value of 5. The below figure (Fig4) demonstrate the segment infected portion with fifth cluster.



Fig4. Segmented Image

##### 4.3 Boundary Extraction

The segmented image contains some unwanted pixels so the boundary extraction is used to extract the boundary to find accurate infected boundary. The following figure (Fig5) illustrates the boundary extracted image.

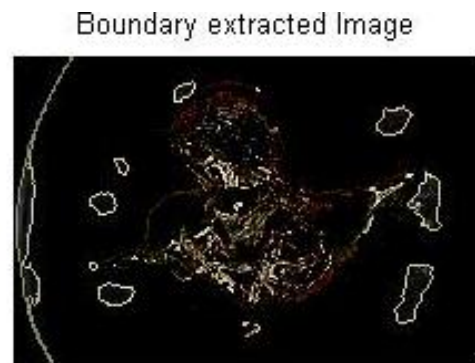


Fig5. Boundary Extracted Image

##### 4.4 Weighted Centroids

The weighted centroid is calculated for each resultant image to find the infected area. The centroids are calculated by computing standard deviation for each centroid. The following figure (Fig6) gives the weighted centroid calculated image.

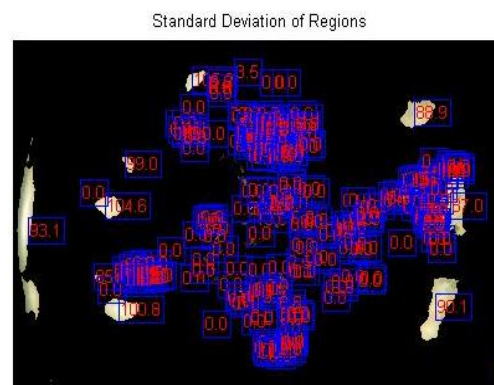


Fig7. Weighted Centroid

##### 4.5 Centroids and Intensity Values

The centroid for particular region is traced to calculate the intensity level for each red, green and blue pixel of the image. The following figures (Fig7 & Fig 8) display the traced centroids with blue plots and the intensity values for the each centroid.

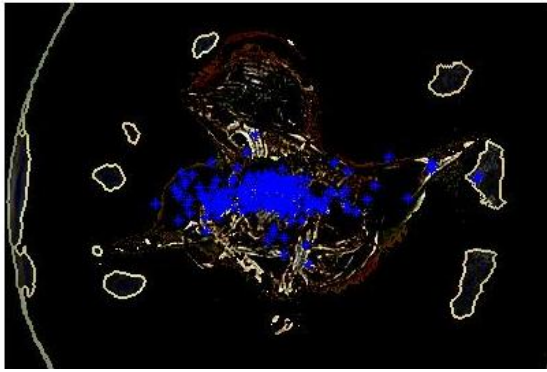


Fig7. Centroids

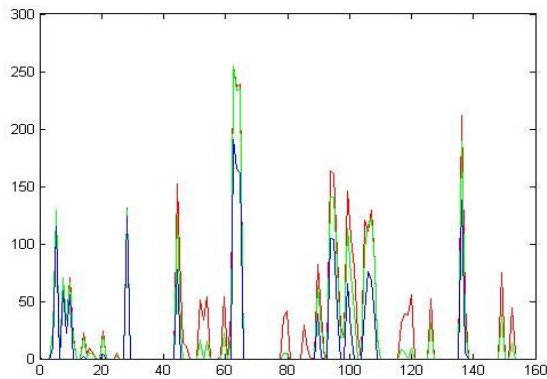


Fig8. Intensity values

#### 4.6 Display Image

Finally the resultant image is displayed based on the standard deviation. The region with <50 of standard deviation is

displayed to showed as infected area. Sometimes the unwanted portion of the image may have high standard deviation so the threshold value is set to be 50. In the Fig7 the weight centroids hide the standard deviation of <50 regions since it will display high standard deviation values. The following figure (Fig9) exemplifies the result image.

Objects Having Standard Deviation < 50

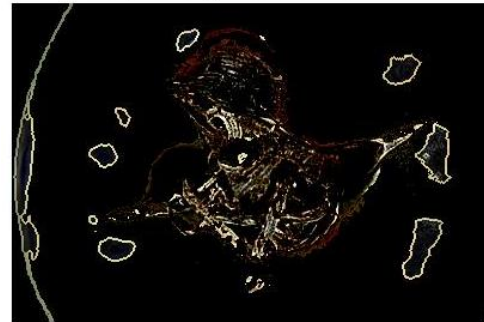


Fig9. Result Image

#### 4.7 Infected Area Calculation


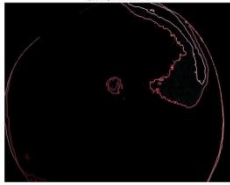

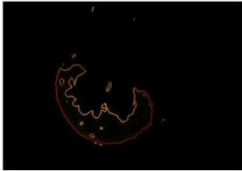

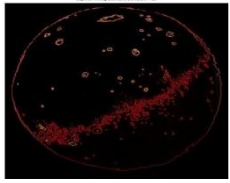
The infected area is calculated by using the original image and the segmented image. This gives the degree of infection in tomato. The following table (table2) shows the result for all diseases.

Total tomato area=480000 sq. pixel units




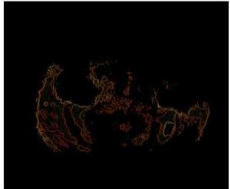

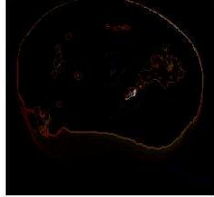
Disease portion area=4352 sq. pixel units

$$\begin{aligned} \text{Infection \%} &= (\text{total tomato area} / \text{disease portion}) * 100 \\ &= (480000 / 4325) * 100 \\ &= 0.9\% \end{aligned}$$

Table2. Infection Percentage for different tomato diseases

Disease Name	Image	Infected Area	Total Area (Sq. pixel units)	Infected Area (Sq. pixel units)	Infection Percentage
Anthracnose			480000	4325	0.9%
Bacterial Canker			117932	9353	7.9%
Bacterial Spot			380400	74352	19.5%



Early Blight		 Objects Having Standard Deviation < 50	106795	2021	1.8%
Late Blight		 Objects Having Standard Deviation < 50	86141	3324	3.8%
Bacterial Speck		 Objects Having Standard Deviation < 50	50576	4857	9.6%

## 5. CONCLUSION

The proposed work demonstrates the tomato disease identification using segmentation. In this work the real samples of tomato images are collected from local market and used for further processing. The K-means Clustering algorithm is used to segment the disease portion with 5 different cluster processes. After all iteration the final image is displayed in the fifth 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.

## 6. REFERENCES

- [1] Shruti, Nidhi Seth, Fungus/Disease Analysis in Tomato Crop using Image Processing Techniques, International Journal of Computer Trends and Technology (IJCTT) – volume 13 number 2 – Jul 2014, pp:61-63.
- [2] Brezmes, J., Llobet, E., Vilanova, X., Saiz, G., Correig, X.: Fruit ripeness monitoring using an electronic nose. Sensors and Actuators B-Chem. Journal 69(3), 223–229 (2000).
- [3] Polder, G., van der Heijden, G.W.A.M., Young, I.T.: Spectral Image Analysis for Measuring Ripeness of Tomatoes. Transactions-American Society of Agricultural Engineers International Journal 45(4), 1155–1162 (2002)
- [4] Jaffar, A., Jaafar, R., Jamil, N., Low, C.Y., Abdullah, B.: Photogrammetric Grad-ing of Oil Palm Fresh Fruit Bunches. International Journal of Mechanical & Mecha-tronics Engineering (IJMME) 9(10), 18–24 (2009).
- [5] Paulraj, M.P., Hema, C.R., Krishnan, R.P., Radzi, S.S.M.: Color Recognition Algo-rithm using a Neural Network Model in Determining the Ripeness of a Banana. In: Proc. the International Conference on Man-Machine Systems (ICoMMS), Penang, Malaysia, pp. 2B7-1–2B7-4 (2009).
- [6] Rizam, S., YAsmin, A.R.F., Ihsan, M.Y.A., Shazana, K.: Non-destructive Water-melon Ripeness Determination Using Image Processing and Artificial Neural Net-work (ANN). International Journal of Intelligent Technology 4(2), 130–134 (2009).
- [7] R.Kalaivani, Dr.S. Murugan and, Dr.Azha.Periasamy, “IDENTIFYING THE QUALITY OF TOMATOES IN IMAGE PROCESSING USING MATLAB”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 8, August 2013.
- [8] Mahdi M. Ali, Ahmed Al-Ani, Derek Eamus and Daniel K.Y. Tan, “A New Image Processing Based Technique to Determine Chlorophyll in Plants”, American-Eurasian J. Agric. & Environ. Sci., 12 (10): 1323-1328, 2012.
- [9] Chawaraj Jaisin , Siwalak Pathaveerat , and Anupun Terdwongworakul ,” Determining the size and location of longans in bunches by image processing technique “, Maejo Int. J. Sci. Technol. 2013, 7(03), 444-455.
- [10] Hadi Izadi, Saadat Kamgar, Mohamad Hossien Raoufat,” Mass Modeeling of Tomato Based on Physical Characteristics”, International Journal of Agronomy and Plant Production. Vol., 4 (10), 2631-2636, 2013.
- [11] Meenu Dadwal, V.K.Banga,” Estimate Ripeness Level of fruits Using RGB Color Space and Fuzzy Logic Technique”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-1, October 2012.