



Image Processing Technique on Identification of Leaf Types and Detection of Diseases Portion on Tomato Leaves

Dr. Santosh K C¹, Bharathi N²

¹Associate Professor, Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India.

²Assitant Professor, Computer Science and Engineering, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India.

Emails: kcsantoo@gmail.com¹, bharathinghorphade@gmail.com²

Corresponding Author Orcid ID: 0000-0001-5012-8449

Abstract

This study presents novel techniques for leaf type identification, employing color histogram and edge histogram approaches. A color histogram serves as a model to represent color through intensity values, while each image is associated with a descriptive caption. Signatures, encompassing shape, color, and texture, provide a basis for comparing images. The edge histogram, delineating the distribution of five edge types in localized sub-images, further enhances the identification process. To address disease detection on tomato leaves, a color-based segmentation method utilizing the k-means clustering technique is proposed. This iterative approach partitions images into k clusters, facilitating the identification of diseases affecting tomato plants. Beyond environmental factors like rain and temperature, crop diseases emerge as primary influencers on production quality and crop yield. Early detection of diseases is crucial for effective control and mitigation. Leveraging technological advancements, the paper emphasizes the potential of using images of diseased leaves for accurate disease identification. This involves feature extraction from images, which can be subsequently employed in classification algorithms or content-based image retrieval systems.

Keywords: Color Histogram, K-Means Clustering, Segmentation, Crop Disease, Laboratory Color Space Model

1. Introduction

India, predominantly an agrarian nation, sustains a significant portion of its population through agriculture, a pivotal sector that profoundly influences the country's economy. The success of agricultural endeavors is intricately tied to environmental variables like rainfall, temperature, and other weather parameters, which lie beyond human manipulation. Apart from these factors, diseases pose a considerable threat to crop productivity, albeit one that can be mitigated through human intervention. Effective diseases management is a complex undertaking, but image processing emerges as a potent tool to simplify this challenge [1]. By leveraging image processing

techniques, Detecting diseases that impact different plant components including leaves, stems, roots, and fruits becomes achievable. This technology enables the detection of the affected area, the specific type of disease, and an assessment of its severity. Typically manifesting on leaves or stems, diseases pose a significant risk to crops, with tomatoes, as the world's most popular and widely cultivated vegetable, becoming a focal point for such investigations. The farming of tomato plants holds immense socioeconomic significance for diverse stakeholders, including households, horticulturists, agriculturists, laborer's, retailers, chefs, and other participants in the food production



industry. Recognized as the tenth highly valuable agricultural asset globally, tomatoes face various disorders throughout the growing season. Hence, this experiment specifically targets tomato leaves as a representative subject for comprehensive analysis and disease detection. Apart from environmental variables like rainfall and temperature, the primary factor influencing agricultural productivity is the occurrence of leaf diseases. Consequently, disease management emerges as a critical concern in agriculture. Effectively addressing this issue requires early identification of diseases, enabling prompt and appropriate treatment to control the spread [2]. In the contemporary technological landscape, leveraging the advancements available, it is now feasible to utilize images of diseased leaves for the precise identification of the type of disease. This objective is attainable through the features extraction from digital images, which can subsequently be employed in conjunction with classification algorithms or content-based image retrieval systems. Plants play a crucial role in sustaining life on our planet, offering crucial benefits such as oxygen production, a source of food, fuel, and medicine. Additionally, plants contribute to climate regulation, serve as habitats and sustenance for various insects and animals, and act as a natural mechanism for flood control. A good understanding of plants is required for enhancing agricultural productivity and sustainability, for discovering new medicines, planning and mitigating the worst. Every leaf possesses unique characteristics that contain valuable information, enabling individuals to identify and categorize plants through visual observation. Among these characteristics, leaf shape stands out as a distinctive feature commonly employed by people for plant recognition and classification [3]. Basic geometric information such as diameter, physiological length and width, the leaf area, and circumference further contribute to the classification process. Additionally, factors like leaf color, texture, and veining are regarded as significant indicators. The amalgamation of these features proves instrumental

in the Identification and categorization through image analysis [4].

1.1 Late Blight of Potato and Tomato

Late blight is the most destructive potato disease in British Columbia. It occurs in all areas of the province, but depends more on wet weather, especially rainfall, high humidity and cool to mild temperature regime. Original Image of Tomato image is shown in Figure 3.

1.2 Late Blight of Management

- Separate early and late crop fields as much as possible to slow the spread of diseases from early to late crops. Early crops should not be planted upwind of late crops. Make sure the potatoes are well grated
- Monitor your field regularly for early detection of early blight symptoms. Pay close attention to the weather forecast to predict the onset of mold. Adopt an effective fungicide spray program before or soon after disease outbreak and before row closure.
- If the blight gets out of control, kill the crop and then apply a fungicide when the tops are about half dead.
- Do not fertilize or overwater. Thick fresh leaves stay moist longer and are more likely to become infected. Avoid overhead irrigation if possible.

2. System Design

The envisioned methodology for disease identification involves the acquisition of images of various leaves through a digital camera [5]. After image capture, advanced digital image processing terminology are deployed to extract pertinent features essential for subsequent analysis as shown in Figure 1. Feature Extraction Using Color Histogram is shown in Table 1.

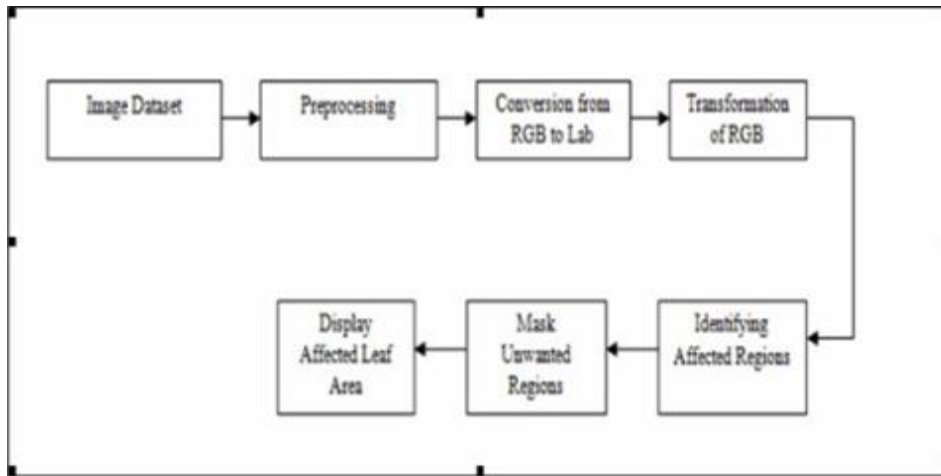


Figure 1 Methodology for Disease Identification

2.1 Pseudocode

- **Step 1:** Initiate the process by reading the target image.
- **Step 2:**
 - Convert the image from the R G B color space to the Lab* color space.
 - Utilize the 'srgb2lab' color transformation for accurate representation.
- **Step 3:**
 - Perform color classification in the 'ab' color space using the K-means clustering algorithm.
 - Identify the most suitable number of clusters (K) by considering the features exhibited in the image.
- **Step 4:** Assign every pixel in the digital image a label obtained on the results fetched from the K-means clustering.
- **Step 5:** Utilize the obtained pixel labels to create segmented images, highlighting dominant colors.

Step 1: Begin the process by reading the image using the imread function.

- `on = imread('hestain.png');`
- `imshow(on), title('H&E image');`

```

matlab
Copy code

on = imread('hestain.png');
imshow(on), title('H&E image');
  
```

Step 2: Convert the digital image from R G B color space to Lab* color space

The Lab* color space originates from the CIE XYZ tristimulus values, offering a comprehensive representation of color. Comprising a luminosity layer denoted as 'L*', a chromaticity layer 'a*' representing color along the red-green axis, and another chromaticity layer 'b*' indicating color along the blue-yellow axis, Lab* encapsulates all color information within its 'a*' and 'b*' layers. The Euclidean distance metric serves as a reliable measure for gauging the dissimilarity between two colors. To seamlessly transition the image into the Lab* color space, the 'makecform' and 'applycform' functions are employed, ensuring a nuanced representation of color features.

- `cform = makecform('srgb2lab');`
- `lab_he = applycform(he, cform);`

```

matlab
Copy code

cform = makecform('srgb2lab');
lab_image = applycform(original_image, cform);
  
```

Step 3: Classify color in 'a*b*' space using K-stands for clustering

Clustering serves as a method to effectively group objects, aiming to distinguish distinct sets within a dataset [6]. Within the framework of K-means

clustering, every independent object is treated as a spatial location, and the algorithm seeks to establish partitions where the items within each cluster exhibit a high degree of similarity to one another and distinctly separated from objects in other clusters. Implementing K-means clustering necessitates defining the number of clusters for segmentation and selecting a suitable distance metric to measure the proximity between two objects. Given that color information is represented in the 'ab' space, the objects in focus are pixels characterized by values in 'a*' and 'b*'. To accomplish the clustering task, the K-means algorithm is applied, categorizing objects into three clusters [7]. The choice of the Euclidean distance metric enhances the precision of measuring proximity between objects, facilitating a meaningful and accurate grouping of pixels based on their color attributes.

Step 4: Label every pixel in the image using the results from k means for each object in your input, k means returns the index of the corresponding cluster. Label every pixel in the digital image with its cluster index.

Step 5: Create images that segment the dominant color.

You can use pixel labels to separate objects by color, resulting in three images.

- `imshow (segmented_images {1}), title ('objects in cluster 1');`
- `imshow (segmented_images {2}), title ('objects in cluster 2');`
- `imshow (segmented_images {3}), name ('objects in cluster 3');`

```
matlab Copy code
imshow(segmented_images(1)), title('Objects in Cluster 1');
imshow(segmented_images(2)), title('Objects in Cluster 2');
imshow(segmented_images(3)), title('Objects in Cluster 3');
```

2.2 K Means Clustering

The K-means clustering method guarantees a consistent presence of K clusters, ensuring that each cluster comprises a minimum of one item. These clusters are defined as non-hierarchical and non-overlapping entities. In this approach, each member within a cluster is categorized by greater proximity to its cluster than to any other, emphasizing that the concept of proximity extends beyond the traditional idea of a central cluster point. Essentially, K-means clustering is a technique within cluster analysis designed to divide n observations into K clusters, where each observation aligns with the cluster possessing the closest mean. K-means stands out as a fundamental unsupervised learning algorithm, efficiently tackling the clustering problem. The methodology provides a straightforward approach to classifying a given dataset into a predefined number of clusters, where the magnitude of k is predetermined. The core concept involves establishing k centroids, each corresponding to a distinct cluster. Strategic placement of these centroids is crucial, as their locations significantly impact the final clustering outcomes [8]. Ideally, centroids are positioned to maximize separation. The process unfolds through linking each point in the dataset with the nearest centroid, completing an initial grouping. Subsequently, the centroids are recalculated as the barycenters of the previously formed clusters. This iterative loop continues until the centroids no longer shift, indicating convergence as shown in fig. 2. K-means produces disjoint flat clusters through this iterative, numerical, and non-deterministic approach. In contrast, hierarchical clustering is another widely adopted technique for image segmentation. Despite its prevalence, the k-means method remains particularly popular in image segmentation due to its simplicity and effectiveness. Hierarchical clustering, like k-means, contributes to the realm of unsupervised learning, providing iterative and versatile solutions for diverse clustering challenges in image analysis [9]. The results are shown in Figures 4, 5, 6, & 7.

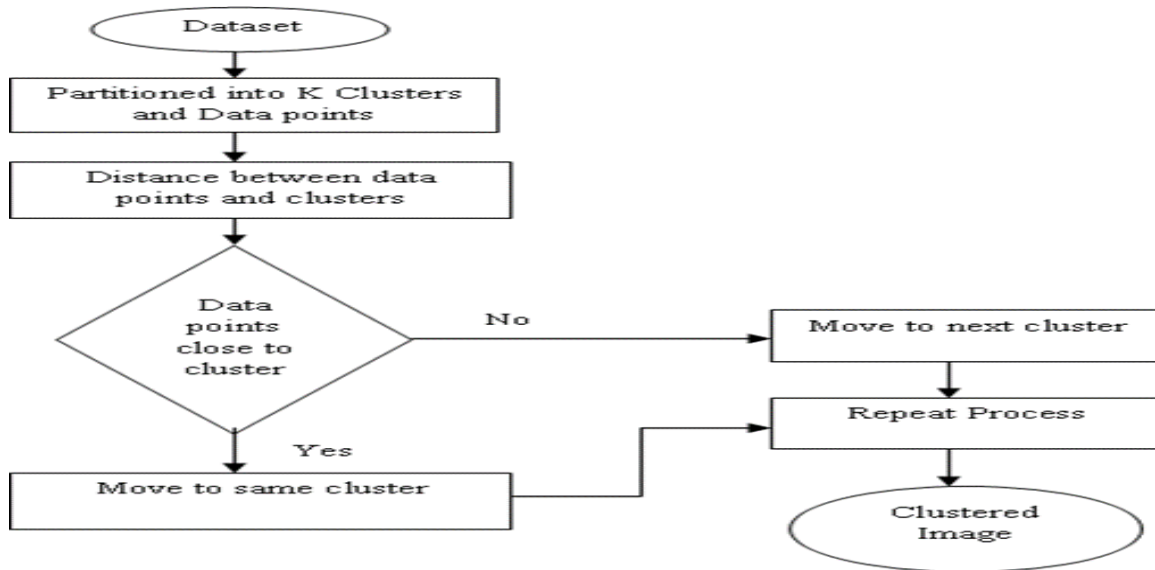


Figure 2 Flow Chart for K means Algorithm

2.3 Histogram Intersection Algorithm

Step 1: First, we insert various Images O_i objects and then create a Bm-Block Matrix. Compute the mean μ of the block matrices. All obtained block matrices are concatenated.

Step 2: Distances between pixels are subsequently computed for each image in the dataset and averages are found. Rows indicate the number of objects in the images. There is a threshold value for which a matching image has been detected. The matching criterion for matching images is set to 80 percent. This can be set to different desired levels of compliance.

Step 3: Now Calculate Euclidean Distance (D)

$$Deuclid(r, s) = \sqrt{\sum_{i=1}^N (r_i - s_i)^2} \quad (1)$$

Where r and s represent the mean values of the feature vectors respectively while $T_i = O_i$ where T_i is the test query image and O_i are the object images. Repeat the above procedure for n object images. We now have an "N" image of the object and its Euclidean distance matrix.

Step 4: The above features are combined to match the image. In this work, the above similarity is measured and then averages are selected for the final output

3. Feature Extraction

3.1 Color Histogram

A color space is known as a model for representing colors in terms of intensity values. A color space usually defines a one- to four-dimensional space. A color component or color channel is one of the dimensions. Color spaces are interconnected using mathematical formulas. Many histogram distances were used to define the similarity of two-color histograms. Euclidean distance and its variations are most commonly used [10].

3.1.1 Definition Of Colour Histogram

An image histogram refers to a probability mass function of image intensities. This is extended for color images to capture the joint probabilities of the intensities of the three-color channels.

$$A, B, C(a, b, c) = N \cdot \text{prob}(A=a, B=b, C=c) \quad (2)$$

Where A, B and C represent the three-color channels R.G.B or H.S.V. and N is the number of pixels in the image. Since a typical powerful image-mining computer represents color images with up to 224 colors.

3.1.2 Histogram Euclidean Distance

Let h and g represent two color histograms. The Euclidean distance between the color histograms h and g can be calculated as in this distance formula, there is only a comparison between identical bins in the respective histograms

$$d^2(h, g) = \sum_A \sum_B \sum_C (h(a, b, c) - g(a, b, c))^2 \quad (3)$$

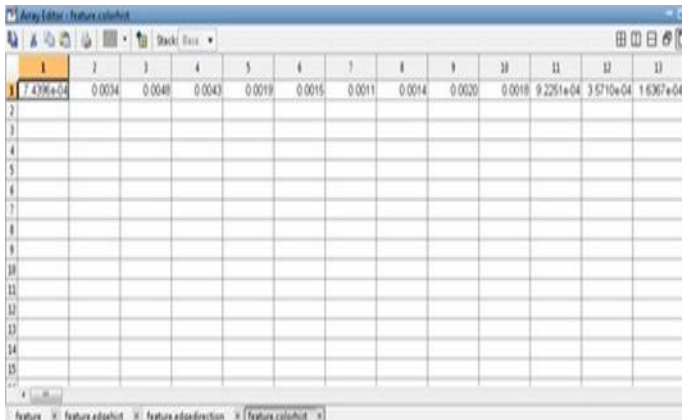
3.1.3 Histogram Intersection Distance

The intersection of the color histogram was designed to obtain a color image. The intersection of the histograms h and g is given by the relation

$$d(h, g) = \frac{\sum A \sum B \sum C \min(h(a, b, c), g(a, b, c))}{\text{Min}(|h|, |g|)}$$

Where $|h|$ and $|g|$ specify the size of each histogram, correlating directly with the number of samples. Colors that are not present in the user query image do not contribute to the intersection distance. This reduces the contribution of background colors. The sum is normalized by the histogram with the least number of samples.

4. Results and Discussion



1	2	3	4	5	6	7	8	9	10	11	12	13
7.4306e+04	0.0034	0.0040	0.0043	0.0019	0.0015	0.0011	0.0014	0.0020	0.0010	9.2251e+04	3.5710e+04	1.6307e+04

Table 1 Feature Extraction Using Color Histogram



Figure 4 Segmentation Using Cluster Index

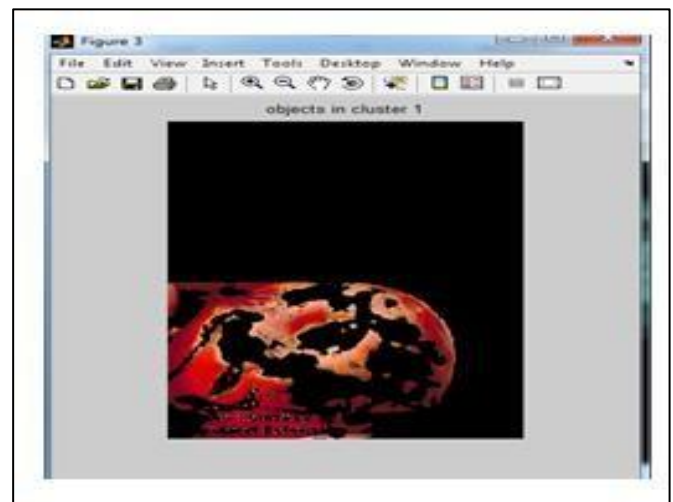


Figure 5 Object in Cluster 1



Figure 3 Original Image of Tomato

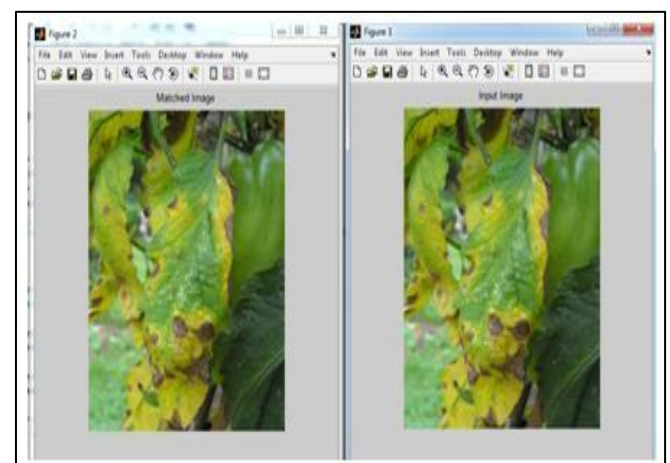


Figure 6 Matching with Query Image

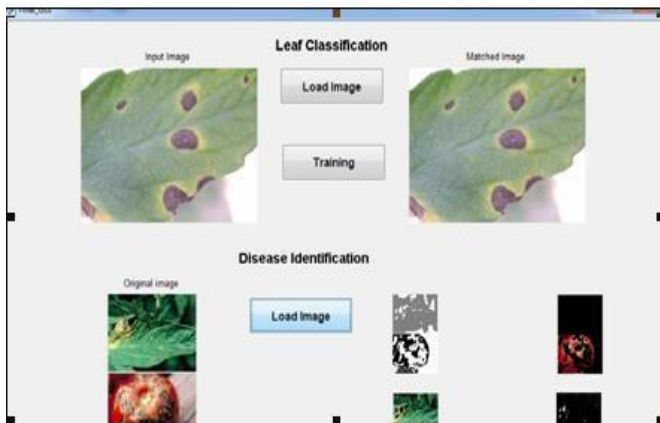


Figure 7 Identification of Leaf Disease

Conclusion

The proposed system is implemented considering the color and texture properties of the sheet, which can be extracted using the color histogram and edge histogram methods. Types of extraction and classification methods that can be used for leaf recognition and classification. Affected parts of tomato leaves were identified using a K-means clustering algorithm and a color transform structure where RGB is converted to the Lab color space. For smaller values of k, the algorithms give good results. For larger values of k, the segmentation is very coarse, with many clusters appearing at discrete locations in the images. Diverse initial partitions can result in varied final clusters. The K-means algorithm is valued for its simplicity and high effectiveness. It demonstrates effectiveness particularly in scenarios where the clusters exhibit limited separation from one another.

References

- [1]. Mohamad Faizal Ab Jabal, Suhardi Hamid, Salahuddin Shuib and Illiasaak Ahmad "Leaf Features Extraction and Recognition approaches to classify Plant" Faculty of Computer Sci. and Mathematics, University Technology MARA, Malaysia Journal of Computer Science 9 (10) 1295-1304, 2013 ISSN: 1549-3636 © 2013 Science Publications.
- [2]. Ehsanirad, A., 2010. "Plant classification based on leaf recognition". Int. J. Compute. Sci. Inform. Security, 8: 78- 81.
- [3]. Kadir, A., L.E. Nugroho, A. Susanto and P.I. Santosa, 2011b. "Leaf classification using shape, color and texture features". Int. J. Compute. Trends Technol. Seventh Sense Res. Group, 1: 225-230.
- [4]. Lee, C.L. and S.Y. Chen, 2003. "Classification for

leaf images". Proceedings of the 16th IPPR Conference on Computer Vision Graphics and Image Processing, (CVGIP '03), pp: 355-363.

- [5]. J. R. Smith and S.F. Chang. "Automated image retrieval using color and texture", Technical Report CU/CTR 408-95-14, Columbia University, July 1995.
- [6]. Santosh KC, Pradeep N, "Novel approach to determine age and gender from wrist x-ray images" Turkish Journal of Computer and Mathematics Education Vol.12 No.3(2021), 5215-5229
- [7]. Cope, J.S., D. Corney, J.Y. Clark, P. Remagnino and P. Wilkin, 2012. "Plant species identification using digital morphometrics" A review. Expert Syst. Applic., 39: 7562- 7573. DOI: 10.1016/j.eswa.2012.01.073.
- [8]. Du, J.X., X.F. Wang and G.J. Zhang, 2007. "Leaf shape-based plant species recognition". Applied Mathem. Comput., 185: 883-893. DOI: 10.1016/j.amc.2006.07.072
- [9]. Santosh K C and Pradeep N. "Machine Learning Techniques for Human Age and Gender Identification Based on Teeth X-Ray Image", Hindawi, Journal of Healthcare Engineering, Volume 2022, Article ID 8302674, 14 pages <https://doi.org/10.1155/2022/8302674>
- [10]. Santosh K C, Pradeep N, "Development of Human Age and Gender Identification System from Teeth, Wrist and Femur Images" Hosting by Elsevier SSRN, 2019 SUSCOM.