

# A Well-Organized Approach for the Detection of Plant Leaf Disease using K-Means Clustering Algorithm

Sofia Mathew

Research Scholar

Department of Computer Science  
 Bharathiar University, Coimbatore

**Abstract**— The study of Plant Diseases refers to the studies of visually observable patterns of a particular plant. Now a day’s crops face many diseases. It also affected the agriculture sector of our country. This paper provides a method to detect disease by calculating leaf area through pixel number statistics. The method studied is for increasing throughput and reducing subjectiveness arising from human experts in detecting the leaf disease. Leaf spots can be indicative of crop diseases, where leaf batches are examined manually and subjected to expert opinion. In this paper an Image Processing system is developed to automate the inspection of these leaf spots.

**Key words:** Leaf Disease, Noise Filter, Clustering, Image Segmentation, Image Processing

## I. INTRODUCTION

In computer science, image processing is one outward appearance of signal processing for which the input is an image, such as a scanned photograph or video frame. The output of image processing may be either an image, a set of characteristics, values or parameters related to that input image. Nearly all image processing techniques treated the image as a two-dimensional signal and using some standard signal processing techniques to it. In any image processing application the important thing is an input that is an ‘image’.

Disease detection of plants is a major challenge in horticulture. Traditionally, finding is done by physically and this is very time consuming and extremely complex process. Outcome may not be precise most of the period as human gets exhaustion. Thus, advanced machine vision is the best way to get accurate results. Image segmentation is one of the enviable process in image analysis which is required for plant disease detection. The purpose of image segmentation is to decompose the image into different small partitions that are meaningful with respect to a particular function.

In this research work focused on disease detection in various types of plants. First, this implies to commonly monitor the disease affected plant leaves. Disease images are acquired using cameras, scanners or any other digital devices. Then the acquired image has to be processed to interpret the image contents by different image processing techniques. The primary objective of this work is to detect the disease affected in the plants using color segmentation operation. In my proposed work collect images of disease affected leaves from various plants. These images will be preprocessed for enhancement. Then segmentation will be carried out for extraction of disease part from foreground of image. After segmentation, various features of disease including color and shape will be extracted. Disease affected area is locate by using different type of edge detection methods.

## II. METHODOLOGY

The architecture of this work is represented below figure.

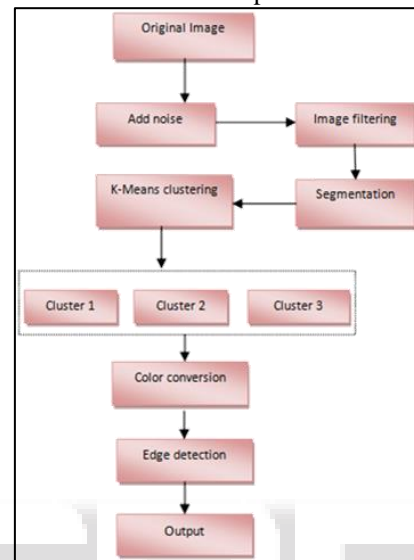


Fig. 1: Architecture of Plant Pest Identification

### A. Salt & Pepper Noise

It represents erratically occurring white and black pixels. An effectual noise reduction method for this type of noise involves the usage of a median filter. Salt and pepper noise creep into images in situations where rapid transients, such as faulty switch, take position. In image following distortion from salt and pepper noise looks like the image attached. This type of noise contains arbitrary occurrences of both black & white passion values, and often caused by threshold of noise image. Salt & Pepper distribution noise can be expressed by

$$P(x) = \begin{cases} p1, & x = A \\ p2, & x = B \\ 0, & \text{otherwise} \end{cases}$$

Where P1, P2 are the Probabilities Density Function (PDF) p(x) is allocation salt and pepper noise in image and A, B are the array size image. In this paper salt & pepper noise in picture is at random occurred in white and black pixels of an image [6]. The challenge in removing salt & pepper noise from image is due to the fact that image data as well as the noise, allocate the similar tiny set of value, which obscure the practice of discover and eliminate the noise.

### B. Linear Filtering

Linear filter can be achieved through complication in the spatial domain. In the complication method the significance of output pixels is calculate as the weighted sum of nearest pixels from the input image.

- The medium of weighting feature is referred to as the intricacy kernel and represents the filter performance. For our purposes these matrices will be of odd dimension (e.g. 3x3, 5x5, 7x7, etc.).
- The weighting process may be expressed as:

$$R = w_1 I_1 + w_2 I_2 + \dots + w_{mn} I_{1mn} = \sum_{i=1}^{mn} w_i I_i$$

Where m,n are the indices of the kernel (often a square matrix) and I represents the input image pixel intensities. The weighting factors are given by w.

### C. Clustering

Clustering is a division of data into groups of similar objects. Representing the data by less clusters inevitably misplace assured fine fact, but accomplish simplification. It representation of information by its clusters. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and arithmetic investigation. From a apparatus knowledge perception cluster converse to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a data concept. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a records notion. From a realistic perception cluster acting an dazzling task in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database appliance, Web analysis, CRM, advertising, medical diagnostics, computational science, and many others.

### D. Fuzzy C-Means Clustering Algorithm

Fuzzy logic deals with the ambiguity and elusiveness present in the predicament. The image in grey color has an uncertainty in brightness and darkness of a pixel value. Segmentation is carried out by firstly converting the RGB image to HIS image which has perplexity in decisive whether pixel is normal or infected. This sort of pensiveness is known as spatial ambiguity. To resolve this, image is considered as a fuzzy set. Fuzzy clustering obtained more evenhanded results for vague cluster boundaries. In a fuzzy set there is a degree of association for every member. Value of membership deceit between 0 and 1 and the sum of membership of each object is 1. Larger the membership values indicates higher assurance to the cluster. Fuzzy C-Means objective function is

$$\sum_{j=1}^m \sum_{i=1}^n u_{ij}^m |x_i - c_j|^2$$

Where m is any real number greater than 1, c is the number of clusters,  $u_{ij}$  is the membership degree of  $x_i$  in the cluster j,  $|x_i - c_j|^2$  is euclidean distance between any data object present in the cluster and the centroid of jth cluster.

Fuzzy C- means clustering algorithm is

- 1) Select a number of clusters.
- 2) Assign erratically to each point in the coefficients for being in the clusters.
- 3) Do again until the algorithm has convergence.
- 4) Compute the centroid for each and every cluster.
- 5) For every point calculate its co-efficient in the cluster.

### III. K-MEANS CLUSTERING ALGORITHM

K-means algorithm aspires in the partition of surveillance into k-clusters. This algorithm is the simplest method in the

segmentation technique. In which the segmentation is used to breakdown the image into smaller components. K-means clustering [10] treats each object as having a location in space. It finds partitions such that objects within each cluster are as secure to each further as feasible, and as distant from objects in new cluster as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other. The K-means clustering algorithm is a simple method for estimating the mean (vectors) of a set of K-groups. The cluster mean of  $K_i = \{t_{i1}, t_{i2}, \dots, t_{im}\}$  is defined as

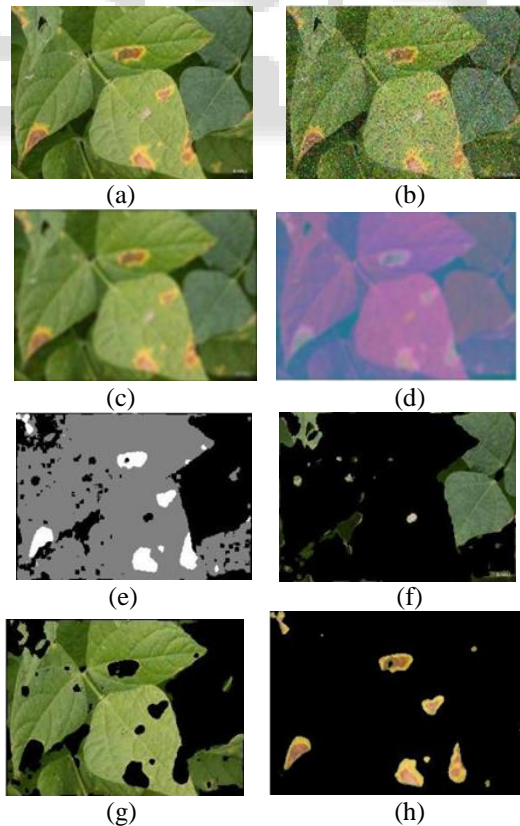
$$m_i = \frac{1}{m} \sum_{j=1}^m t_{ij}$$

### A. Algorithm

- Step 1: Select K points as the initial centroid.
- Step 2: Repeat
- Step 3: From K clusters by assigning all points to the closest centroid.
- Step 4: Recomputed the centroid of each cluster.
- Step 5: Until the centroid don't change

Initial centroids are often chosen randomly where the centroid is (typically) the mean of the points in the cluster. The closeness is measured by Euclidean Distance, cosine similarity and correlation. K-Means will converge for common similarity measures which are mentioned. Where the Euclidean Distance metrics are used to cluster the objects into three clusters in K-Means.

## IV. RESULTS



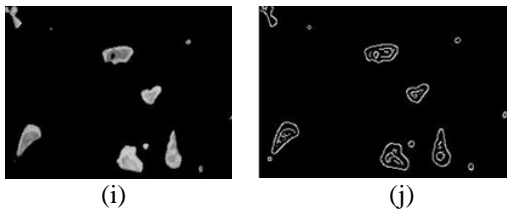


Fig. 2: (a) Typical input image present in the database. (b) Salt and pepper noise. (c) Linear filter is used to remove noise. (d) Color conversion. (e) Cluster index. (f) Objects in cluster 1. (g) Objects in cluster 2. (h) Objects in cluster 3. (i). Color conversion. (j) Canny edge detection.

## V. CONCLUSION

In this research work the area of plant leaf disease detection is introduced. The system developed here is for plant disease detection, the development of good segmentation methods and precise features is very important in order to run the system in real time. One of the key concepts in image processing is segmentation in which particular homogeneous regions are grouped together based on the defined number of clusters. The performance of the novel unsupervised clustering algorithm K-Means Clustering Algorithm is used to segment the plant disease. The main technique used was segmentation, which is done using a method based on K-Means algorithm and edge detection operation. Many methods exist for image segmentation that attempts to segment an image into homogenous regions.

This technique gives efficient results when compared to the previous work. Experiments are applied on various images and results show a clear segmentation. Our proposed work is easy to execute thus can be managed easily and giving effectual results in segmentation of the diseases in plant leaves.

## REFERENCES

- [1] Chu XiaoLi "Method of image segmentation based on Fuzzy C Means Clustering algorithm" (ICISS), International Conference on Digital Object Identifier, Page(s):469-472.2012.
- [2] Jingpan Liu; "An image mosaic block detection method based on fuzzy c-means clustering". International Conference on Computational Problem-solving (ICCRD), Page(s):469-472.2012.
- [3] Muthukannan, K" Color image segmentation using k-means clustering and Optimal Fuzzy C-Means clustering" International Conference on Communication and Computational Intelligence (INCOCCI), Page(s): 229-234,2010.
- [4] Feng Zhao; "Spatial improved fuzzy c-means clustering for image segmentation" International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT), Page(s):4791-4794, 2011.
- [5] Chaira, T. "A novel intuitionistic fuzzy c means color clustering on human cell images World Congress on Nature & Biologically Inspired Computing, 2009.NaBIC, Page(s):736-741, 2009.
- [6] Xin-bo Zhang" An Image Segmentation Algorithm Based on Fuzzy C-Means Clustering" International Conference on Digital Image Processing, Page(s):22-26, 2009.

- [7] Ming Li "Fuzzy-C-Means Clustering Based on the Gray And Spatial Feature For Image Segmentation" International Conference on Computational Intelligence and Security Page(s):1641-1646,2006.
- [8] Schowengerdt, Robert A.(2007).Remote sensing: models and methods for image processing (3rd ed.). Academic Press.p.2.ISBN 978-0-12-369407-2.
- [9] Schott, John Robert (2007).Remote sensing: models and methods for image processing (3rd ed.). Academic Press.p.2.ISBN 978-0-19-517817-3.
- [10] Liu, JianGuo & Mason, Philippa J.(2009).Essential Image Processing for GIS and Remote Sensing. Wiley-Blackwell.p.4.ISBN 978-0-470-51032-2
- [11] BegniG.Escadafal R. Fontannaz D. and Hong-Nga Nguyen A.-T.(2005). Remote sensing: a tool to monitor and assess desertification. Les dossiers the matiques du CSFD. Issue 2.44 pp.