RESEARCH ARTICLE OPEN ACCESS

K-Means Segmentation Method for Automatic Leaf Disease Detection

Jimita Baghel[1], Prashant Jain[2]
[1,2] Department of Electronics and Communication Engineering Jabalpur Engineering College, Jabalpur (M.P.), India

ABSTRACT
Automatic detection of leaf disease is an essential research topic in agricultural research. It may prove benefits in monitoring fields and early detection of leaf diseases by the symptoms that appear on the leaves. Defect segmentation is carried out in two steps. This paper illustrates K-means clustering method for segmentation of diseased portion of leaf. At first, the pixels are clustered based on their color and spatial features, where the clustering process is accomplished. Then the clustered blocks are merged to a specific number of regions. This approach provides a feasible robust solution for defect segmentation of leaves.
Keywords—Leaf disease, image processing, k-means clustering segmentation.

I. INTRODUCTION

The application of machine vision is very important in agricultural industry. Seed analysis and classification can provide additional knowledge in their production, seeds quality control and in impurities identification. Generally these activities are performed by specialists by visually inspecting each sample, which is a very tedious and time consuming task. So, automation is required in this field. Now a days, computer vision technology is applied in a large variety of fields to increase the efficiency of the work. The naked eye observation of experts is the main approach adopted in practice for detection and identification of plant diseases. However, this requires continuous monitoring of experts which might be prohibitively expensive in large farms. Further, in some developing countries, farmers may have to go long distances to contact experts, this makes consulting experts too expensive and time consuming. Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant[2].

Therefore looking for fast, automatic, less expensive and accurate method to detect plant disease cases is of great realistic significance[2]. Digital image processing techniques has found a number of applications in various fields such as medical imaging, remote sensing, industrial inspection and agricultural processing etc. In the field of agriculture digital image processing techniques have been established as an effective means for analyzing purposes in various agricultural applications like plant recognition, crop yield estimation, soil quality estimation etc. The image processing can be used in agricultural applications for following purposes [3]:

1. To detect diseased leaf, stem, fruit
2. To quantify affected area by disease.
3. To find shape of affected area.
4. To determine color of affected area
5. To determine size & shape of fruits etc.

The leaf image processing system can been classified into two phases:- (1) Training Phase, which includes Image Acquisition, Image Pre-Processing, Feature Extraction and Artificial Neural Network based training and (2) Testing Phase, which includes Test Image Acquisition, Test Image Pre-processing, Feature Extraction, K- means based Segmentation and Classification, Percentage Infection Calculation and Disease Grading using Fuzzy Logic Toolbox[1].

II. IMAGE SEGMENTATION

Segmentation divides an image into its constituent regions or objects. It is the key step in image analysis. The level to which subdivision is carried depends on the problem being solved. That is, segmentation should stop when the objects of interest in an application have been isolated. Image segmentation algorithms are based on two basic properties of intensity values: discontinuity and similarity. In the first category, the approach is to divide an image based on abrupt changes in intensity, such as edges in an image. The second approach is based on dividing an image into regions that are similar according to a fixed criterion[5]. In the current work, the very purpose of segmentation is to identify regions in the image that are likely to qualify as diseased regions. There are various techniques for image segmentation such as clustering methods, compression-based methods, histogram-based methods, region growing methods etc. *K-means clustering* method has been used in the present work to carry out segmentation. Various image segmentation techniques are shown in Figure 1.

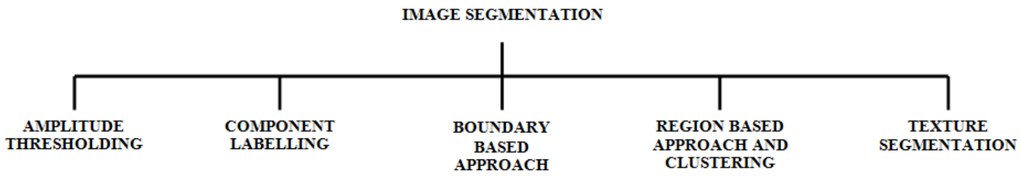

Figure 1: Image Segmentation Techniques

2.1 Region based approach and clustering

The main idea in region based segmentation technique is to identify various regions in an image that have similar features.

A cluster is a set of points in the feature space for which their local density is large compared to the density of feature points in the surrounding region. Clustering techniques are suitable for image segmentation and for classification of raw data to establish classes and prototypes. The success of clustering technique rests on partitioning of the feature space into cluster subsets. General clustering algorithms is based on split and merge technique (Figure 2).

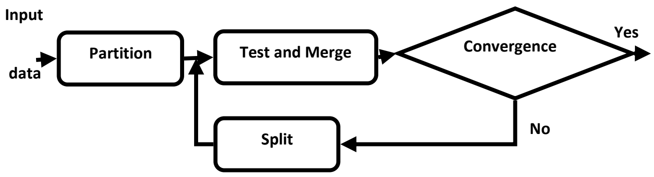

Figure 2: A Clustering Approach

Using a similarity measure, the input vectors are portioned into subsets. Each partition is tested to check whether or not the subsets are sufficiently distinct. Subsets that are not sufficiently distinct are merged. The procedure is repeated on each of the subsets until no future subdivisions result or some other convergence criterion is satisfied. Thus a similarity measure, a distinctiveness test and a stopping rule are required to define a clustering algorithm. For any two feature vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ some of commonly similarity measures are [6]:

1) Dot Product:

$$<\mathbf{x}_i, \mathbf{x}_j> \triangleq \mathbf{x}_i^T \mathbf{x}_j = ||\mathbf{x}_i|| \, ||\mathbf{x}_j|| \, \cos(\mathbf{x}_i, \mathbf{x}_j)$$

2) Similarity rule:

$$S(\mathbf{x}_i, \mathbf{x}_j) \triangleq \frac{<\mathbf{x}_i, \mathbf{x}_j>}{<\mathbf{x}_i, \mathbf{x}_i> + <\mathbf{x}_j, \mathbf{x}_j> - <\mathbf{x}_i, \mathbf{x}_j>}$$

3) Weighted Euclidean distance:

$$d(\mathbf{x}_i, \mathbf{x}_j) \triangleq \sum_{\mathrm{k}} [x_i(k) - x_j(k)]^2 \; w_k$$

4) Normalized Correlation:

$$\rho(\mathbf{x}_i, \mathbf{x}_j) \triangleq \frac{<\mathbf{x}_i, \mathbf{x}_j>}{\sqrt{<\mathbf{x}_i, \mathbf{x}_i><\mathbf{x}_j, \mathbf{x}_j>}}$$

2.2 K- means clustering method

Assume the number of clusters, K, is known. The partitioning of data is done such that the average spread or variance of partition is minimized. Let $\boldsymbol{\mu}_k(n)$ denote the k^{th} cluster center at the n^{th} iteration and R_k denote the region of k^{th} cluster at a given iteration. Initially, we assign arbitrary values to $\mu_k(0)$. At the n^{th} iteration take one of the data points $\mathbf{x}_i$ and assign it to the cluster whose center is closest to it, that is,

$$\boldsymbol{x}_i \in R_k \Leftrightarrow d\left(x_i, \mu_k(n)\right) = \min_{j=1,\ldots,K} [d(x_i, \mu_j(n)] \qquad (1)$$

where $d(\mathbf{x},\mathbf{y})$ is the distance measure used. Recompute the cluster centers by finding the point that minimizes the distance for elements within each cluster. Thus,

$$\mu_k(n+1): \sum_{x_i \in R_k} d(x_i, \mu_k(n+1) = \min_y \sum_{x_i} d(x_i, y), \qquad k = 1, \ldots, K \qquad (2)$$

The procedure is repeated for each $\mathbf{x}_i$, one at a time, until the cluster and their centers remain unchanged. If $d(\mathbf{x},\mathbf{y})$ is the Euclidean distance, then a cluster center is simply the mean location of its elements. If K is not known, we start with a large value of K and then merge to K-1, K-2 ... clusters by suitable cluster distance measure.

III. PROPOSED METHODOLOGY

The steps involved in image processing of leaf are shown in Figure 3. The basic procedure for leaf disease detection is as follows[1]:

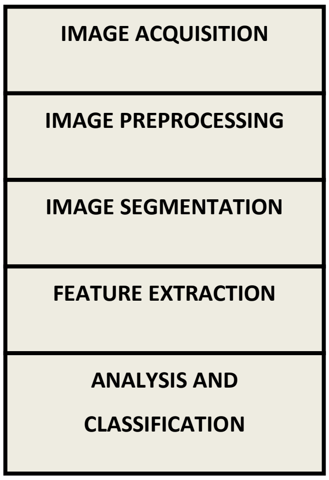

Figure 3: Image Processing Steps

A. Image Acquisition

Image Acquisition refers to acquiring an image by means of camera of a leaf. Image of a leaf can be taken with the help of a digital camera. Other methods can also be used.

B. Image Pre-Processing

Image Pre -processing refers to working on images in order to convert it in a suitable form on which the algorithm can be worked. The captured images will be cropped and be resized so that it can be effectively tested. In Digital image processing, computer algorithms are applied to execute image processing on digital images. Pre-processing consists of

• Resize Image
• Filter Image
• Segment Image
• Crop Image
• Binarization

C. K-Means based Segmentation

Following are the steps in k-means clustering:

1. Read input image of leaf.
2. Convert RGB color space to L* a* b* color space: This conversion enables to quantify the visual difference present in the RGB image.
3. Classify the colors in a* b* space using k-means clustering. K-means clustering treats each object as having a location in space. K-means finds partition such that objects within each cluster are as close to each other as possible and as far from objects in other clusters as possible. Since the color information exists in the a* b* space. The objects of interest are the pixels with 'a*' and 'b*' values.
4. Label every pixel in the image using the results from k-means. For every object in the input image, k-means returns an index corresponding to a cluster and label every pixel in the image with its cluster index.
5. Create image that segment the original image by color. This step will result in k number of images of each of which is a segment of the original image that is partitioned by color.
When segmentation is completed, one of the clusters contains the diseased spots being extracted [8].

D. Feature Extraction

Extraction of features of an image is a property in image processing where the major attributes which have to be analyzed are extracted.

E. Percentage Infection Calculation(P)

After calculation of the total leaf area (A_T) as well as the diseased area (A_D) of the leaf, the percentage infection (P) is calculated by using the following equation:

$$P = \frac{A_D}{A_T} \times 100 \quad (3)$$

IV. CONCLUSION

In the present scenario it is very important to have an established approach for grading the defects on the plant leaves automatically. For automatically detecting the leaf plant as well as for leaf disease detection, Machine Vision Technology is of great use. These systems are going to be very helpful for agriculturist since it is efficient than the manual method. The proposed system uses Euclidean distance technique and K- means clustering technique for segmentation of image to segment the leaf area, disease area and background area of the input leaf image in order to calculate the percentage infection of the disease in the leaf and to grade them into various classes.

These systems can be used to replace the manual leaf recognition technique and can be used by agricultural experts in identifying correct pesticide and its quantity to overcome the problem in an efficient and effective manner.

REFERENCES

[1]. Aakanksha Rastogi, Ritika Arora , Shanu Sharma, "Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic",IEEE 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN).

[2]. Al-Hiary H., S. Bani-Ahmad, M. Reyalat, M., Braik and Z. Al Rahamneh, "Fast and accurate detection and classification of plant disease", International Journal of computer Application, ISSN 0975-8887,vol.17, No1, pp. 31-38, 2011.

[3]. Jayamala K. Patil1 and Raj Kumar, "Advances in image processing for detection of plant diseases", Journal of Advanced Bioinformatics Applications and Research, ISSN 0976-2604Vol 2, Issue 2, pp 135-141, June-2011.

[4]. Tapas Kanungo,David M.Mount, "An Efficient K-Means Clustering Algorithm: Analysis and Implementation.",IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol.24,No.7,July 2002.

[5]. Rafael C. Gonzalez, Richard E.Woods and Steven L.Eddins, "Digital Image Processing", Pearson Education, 2nd Edition

[6]. Anil K. Jain, "Fundamentals of Digital Image Processing",PHI.

[7]. Darshana. A, Dr.Jharna Majumdar,Shilpa Ankalak, "Segmentation Method for AutomaticLeafDisease Detection." ,IJIRCCE ,Vol.3,Issue 7,July 2015.

[8]. Shiv Ram Dubey,Pushkar Dikshit,Nishant Singh, "Infected Fruit Part Detection Using k-means Clustering Segmentation Technique." ,IJAMAI,Vol.2,2013.