Disease Detection in Soya Bean using K-Means Clustering Segmentation Technique

Jimita Baghel
Department of Electronics
and Communication Engineering
Jabalpur Engineering College, Jabalpur (M.P.),
India

Prashant Jain
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.In this paper K-means clustering method is used for segmentation of diseased portion of soya bean leaf. Percentage calculation and accurate grading of leaf disease is necessary for determination of pest control measures. This approach provides a feasible robust solution for defect segmentation of leaves.

Keywords

Leaf disease, image processing, k-means clustering segmentation**,** soya bean leaf

1. INTRODUCTION

India is the fifth largest producer of soya bean in the world today. Three major soya bean producing states in India are Madhya Pradesh, Maharashtra and Rajasthan[1]. Soya bean (*Glycine max*) is a species of legume widely grown for its edible bean which has numerous uses. Early detection of disease is essential for good production and quality control. Generally the disease detection activities are performed by specialists by visually inspecting each sample, which is a very tedious and time consuming task. Hence automatic machine vision technology has acquired a significant role in this field[3].

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[4]. The image processing can be used in soya bean production for following purposes: 1). To detect diseased leaf, stem, seed. 2) To find diseased area. 3) To compare shape of seeds,etc. The entire image processing system has been divided into two parts: Training and Testing[2].

2. IMAGE SEGMENTATION

Segmentation divides an image into its constituent regions or objects. It is the key step in image analysis. In the current work, the very purpose of segmentation is to identify diseased regions in the image[5]. There are various techniques for image segmentation such as clustering methods, compression-based methods, histogram-based methods, region growing methods etc[6]. *K-means clustering* method has been used in the present work.

2.1. K-Means Clustering Segmentation Technique

In this technique colors are segmented in an automated fashion using L*a*b*color space. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible[8].

2.2. The K-Means Algorithm Process

- The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points.

- For each data point calculate the Euclidean distance from the data point to each cluster.

- If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster. Repeat the above step until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends.

- The choice of initial partition can greatly affect the final clusters that result, in terms of inter-cluster and intra-cluster distances and cohesion.[10]

2.3. Euclidean Distance

The Euclidean distance is the straight-line distance between two pixels,

$$ED = \sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)} \ (1)$$

where (x_1,y_1) & $(x_2,y_2))$ are two pixel points or two data points.

So,

$$\text{Euclidean distance} = (P - Q) * (P - Q)'$$

Here P is a data point and Q is the center of a cluster.

S is a vector containing all the data points the 3-d color space.

3. PROPOSED METHODOLOGY

The steps involved in image processing of leaf are:

3.1.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.

3.2.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.

3.3.Image Segmentation

Here we have used K-Means technique to perform segmentation. Following are the steps in k-means clustering in the course of disease detection[8]:

1. Read image.

2. Convert the image from RGB Color Space to L*a*b* Color Space.

3. Classify the Colors in a* b* Space using K-Means Clustering.

4. Label every pixel in the image using the results from k-means.

5. Create images that segment the image by color.

6. Separate the infected part and uninfected part.

3.4.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

$$P = \frac{A_D}{A_T} \times 100 \qquad (2)$$

4. EXPERIMENT

The step by step procedure performed in this experiment is given below. The entire procedure has been performed using MATLAB version 7.10.0.499(R2010a).

4.1 Original Image

Two images have been taken into consideration: infected soya bean leaf and soya bean pod (Fig.1).

(a)   (b)

Fig.1: Original Image (a) Leaf (b) Pod

1. Cluster Generation
After loading the image, clusters are generated using K-Means segmentation technique.

2. Segmented area cluster
Used to determine area of leaf without infection(Fig.2)

Fig.2: Segmented area cluster

3. Diseased cluster
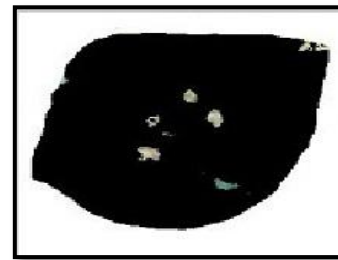Used to find the diseased area (Fig.3)

Fig.3: Diseased part cluster

4. Binary Image of Diseased Part
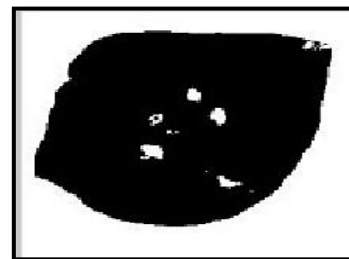This will give the area of the infected part(Fig.5)

Fig.4: Binary image of diseased part

5. Binary Image of Background
This will give area of the whole leaf(Fig.6).
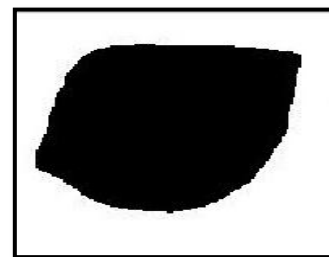

Fig.5: Binary image of background

5. RESULT AND ANALYSIS

The parameters considered for evaluation are as follows:

5.1. Percentage Area Affected

This shows area of leaf is infected by pest & how much percentage of leaf is affected by pest.

Percentage infection calculation is achieved by,
$A_D/A_T *100$ (3)

Where, AD is the area of diseased part and AT is the area of whole leaf. The area calculation of both the parts is shown below (Fig.7 (a) and (b)).

```
Q=im2bw(P);
imshow(Q);
>> bwarea(Q)

ans =

    1.9745e+003
```

(a)

```
>> B=im2bw(I);
>> imshow(B)
>> bwarea(B)


ans =

    3.2348e+005
```

(b)

Fig.7: Calculation of area (a) Diseased part, A_D (b) Whole leaf, A_T

Hence, the infected area can be calculated as shown below (Fig.8):

```
>> U= 1.97e+003/(3.2348e+005)

U =

     0.0061
```

Fig.8: The ratio of infected area to the whole plant area

The percentage of infected area for the given soya bean leaf is given by,

$$0.0061*100=0.61 \% \qquad (4)$$

5.2. Energy

The gray level energy is calculated using the equation,

$$E(x) = \sum_{i=1}^{x} p(x) \qquad (5)$$

where E(x) represents the gray level energy with 256 bins and p(i) refers to the probability distribution functions, which contains the histogram counts. The larger energy value corresponds to the lower number of gray levels, which means simple. The smaller energy corresponds to the higher number of gray levels, which means complex. It specifies the sensitiveness of clustering techniques to noise data.

5.3. Entropy

Entropy is used to characterize the texture of the input image based on the statistical measure of randomness. This measure quantifies the expected value of the information contained in a image. Considering discrete probability distribution of the image p the entropy is formulated as,

$$H(x) = \sum_{i=1}^{k} p(i)log_2 p(i) \qquad (6)$$

5.4. Mutual Information

MI is the measure used between two random variables that are mutually dependent. The symmetric function is formulated as,

$$I(x) = \sum_{XY} p_{XY} (X,Y)log_2 \frac{p(X,Y)}{p_x(X)p_y(Y)} \qquad (7)$$

5.5. Evaluation Time

Evaluation Time (ET) is defined as the time taken by a system to process the operation. The measurement of any event depends on the memory-size, the input data size, and the memory access time. The execution time taken should be low for online and real-time image processing applications. Hence

lower ET is better than higher ET value when all other performance-measures are identical.

Table 2: Result and comparative analysis of above parameters

S.No.	Features	K-means	Edge Detection
1	Energy	0.4467	0.8582
2	Entropy	0.1352	0.1034
3	Mutual Information	0.67	0.52
4	Evaluation time(sec)	1.907967	2.106

It is evident that time consumed in K-means method is less than other algorithm.

Table 3: Area of leaf and percentage infection

Affected area	1.9745e+003
Total area	3.2348e+005
Percentage infection	0.61

6. CONCLUSION

From the above segmentation technique it is concluded that K-Means clustering works well if the boundaries are well defined. Effective grading of leaf disease can be very helpful in reducing the damage of the crop plants. Computer vision technology and digital image processing is of great use for automatically detecting leaf as well as pod disease. This system is very useful in the field of agriculture since it is efficient than manual procedure.

In the proposed system, K-Means clustering segmentation technique is used to separate the leaf disease area and the background area and percentage infection is calculated. Further fuzzy logic can also be used to grade the leaf disease. This system can be used to determine the amount of pesticide needed to treat the infected area.

7. REFERENCES

[1] B.B Singh, "Success of Soybean in India: The Early Challenges and Pioneer Promoters."International Institute of Tropical Agriculture (IITA), Kano Station, Sabo Bakin Zuwo Road, PMB 3112, Kano,Nigeria.

[2] 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).

[3] 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.

[4] Jayamala K. Patil 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.

[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.JharnaMajumdar,ShilpaAnkalak, "Segmentation Method for Automatic Leaf Disease Detection.",IJIRCCE,Vol.3,Issue 7,July 2015.

[8] www.mathsworks.com, k-means clustering method.

[9] Sachin.B. Jadhav, Prof.Dr.Sanjay B Patil,"Grading of Soybean Leaf Disease Based on Segmented Image Using K-means Clustering.",IJARECE,Volume 4, Issue 6, January 2015.

[10] Manjit Chintalapalli,"Color Image segmentation by Clustering".

[11] R. Pydipati, T.F. Burks and W.S. Lee, "Identification of citrus disease using color texture features and discriminant analysis", Computer and Electronics in Agriculture, Elsevier, Vol 52, Issue 2, pp. 49-59, 2009.

[12] Santanu Phadikar and Jaya Sil, "Rice disease identification using pattern recognition techniques", Proceedings Of 11th International Conference On Computer And Information Technology,25-27, 2008.

[13] Shen Weizheng, Wu,Yachun, Chen Zhanliang and Wei Hongda, "Grading method of leaf spot disease based on image processing". Proceedings Of 2008 International Conference On Computer Science And Software Engineering, Volume 06,2008.

[14] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu – Xuan Wang and Yi – Fan Chang, "A leaf recognition algorithm for plantclassification using probabilistic neural network", IEEE 7th International Symposium on Signal Processing and Information Technology,2007.

[15] Xu Pengyun and Li Jigang, "Computer assistance image processing spores counting",International Asia Conference on Informatics in Control, Automation and Robotics, IEEE computer society, pp-203-206, 2009.