Segmentation of Lesion Portion from Plant Leaves using Clustering Techniques

K. Muthukannan
Associate Professor,
Dept of ECE,
Einstein College of Engineering,Tirunelveli,India

P. Latha
Associate Professor,
Dept of ECE,
Government College of Engineering,Tirunelveli,India

P. Nisha
PG scholar,
Dept of ECE,
Einstein College of Engineering,Tirunelveli,India

R. Pon Selvi
PG scholar,
Dept of ECE,
Einstein College of Engineering,Tirunelveli,India

***Abstract*- Plant leaf diseases are the major problem that, threaten any plant crop cultivation, this leads to heavy loss in crop production and economic degradation. Digital plant leaf analysis resides a major task in medical as well as in agricultural field. Segmenting the disease affected portion from the leaf image is a difficult task. In this initially the leaf image is removed from its background using Thresholding technique. Image segmentation remains one of the major challenges in image analysis. Clustering is a method of creating groups of objects, or clusters, in such a way similar objects are grouped together while those that are different are segregated in their distinct clusters. In this to segment the lesion portion, clustering techniques such as K-Means, Improved K-Means, FCM** and **Improved FCM techniques are used. To measure the quality of the segmented image the performance metrics such as Rand Index (RI), Variation of Information (VOI) and Boundary Displacement Error (BDE) are measured.**

Index Terms— Thresholding, K-means, Improved K-means, FCM, Improved FCM and performance metrics.

I. INTRODUCTION

The important problems in plant cultivation are mainly due to Plant leaf Diseases, insects and pests, which cause heavy loss in the crop production. Image segmentation constitutes an important step and an essential process of image analysis. The image is subdivided into number of meaningful segments in image segmentation [14]. The segmented parts give some meaningful information in the form of texture, intensity or color. Image segmentation is a significant operation for study and interpretation of image acquired. This is a most challenging task in image processing; scholars have widely worked over this problem. Some of the segmentation methods are region growing [4], clustering techniques, fuzzy approaches, neural network, genetic algorithm, thresholding technique, and so on. In this paper clustering techniques such as K-Means, Improved K-Means, FCM [5] and Improved FCM techniques [23] are used to extract the disease affected portion from the leaf image. Different plant leaf images are collected on daily basis using digital camera. These images are given as input image and these images are used for pre-processing to remove the additional noise present in the input image samples. In this median filter is used to remove the salt and pepper or impulsive noise [5]. This will preserve the edge information compared to other filters. The clustering techniques [1] are used for the segmentation process. The performance is calculated for the segmented image, the performance parameters measured are Rand index, Variation of information and Boundary displacement error [20]. Based on these measured parameters the segmentation is compared and result is analyzed.

II. PROPOSED METHOD

The proposed work involves four modules: Image Acquisition and Image Preprocessing, Background removal, Image Segmentation, performance analysis. This section briefly explains the above modules. The block diagram for the proposed method is given below.

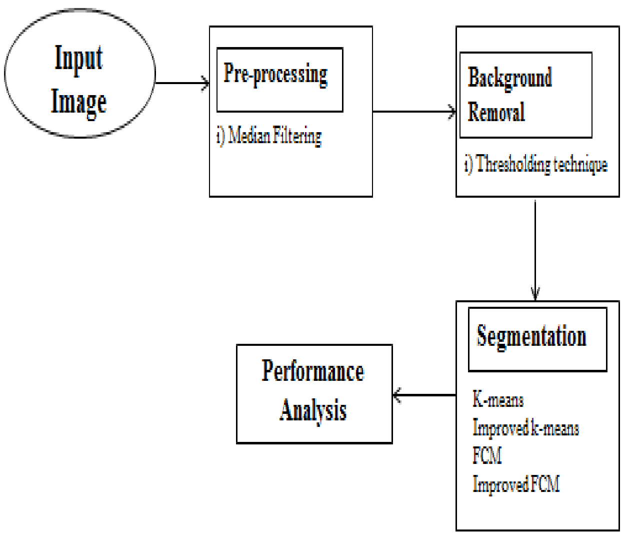


Fig 1. Workflow model for the proposed system.

The above Figure.1 explains the proposed work flow method for the diseased leaf segmentation. The step by step process is explained below.

1. Acquire the input image of diseased leaf

2. Pre-processing of image to convert it into proper format
- Resize the image
- Removal of noise using median filter

3. Segmentation of lesion portion using clustering techniques.

4. Performance analysis of the segmented image using rand index, variation of information and boundary displacement error is measured.

A. *Image acquisition and Image pre-processing*:

The various plant leaf images are collected directly from the field using digital camera. The white background is set to take the flash of each leaf images for better segmentation result. In this only five leaf images of different agricultural plant leaves are considered. The leaf samples are tomato leaf, bitter gourd leaf, lady finger leaf, chilly leaf and bean leaf. The input sample images are shown below

Fig 2. Input image samples: a)Tomato leaf b)Bitter gourd leaf c)Ladyfinger leaf d)chilly leaf e)Bean leaf

Preprocessing techniques are used to remove the unwanted noise from the input image. In this median filter is used to remove the impulse, salt and pepper noise from the leaf images. The preprocessed median filtered output image samples are given below

Fig 3. Median filtered output images: a) Tomato leaf b)Bitter gourd leaf c)Ladyfinger leaf d)chilly leaf e)Bean leaf

B. Background removal*:*

For the background removal thresholding technique is used[17]. In this only white background images are taken which provides good background removal result. In this technique the threshold value is set as fixed value. The pixel value above the threshold value is set as leaf image. The pixel value below the threshold value is considered as background. In this green and white pixels are taken. Thus the image is removed from the background. The background removed output leaf images are given below

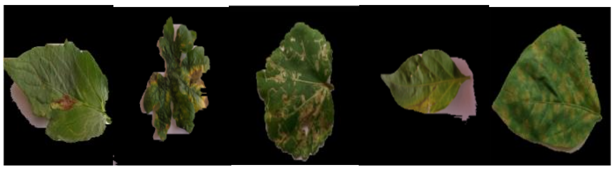


Fig 4. Background removed output images: a) Tomato leaf b)Bitter gourd leaf c)Ladyfinger leaf d)chilly leaf e)Bean leaf

The mean and standard deviation values are calculated for both the input leaf samples and background removed sample images to analyze the performance. The performance analysis for background removal tabulation is given below.

TABLE I. PERFORMANCE ANALYSIS FOR BACKGROUND REMOVAL

LEAF IMAGES	**ORIGINAL IMAGE**		**BACKGROUND REMOVED IMAGE**	
	Mean	***Standard deviation***	***Mean***	***Standard deviation***
Bean	93.72	9.5	70.4	10.16
Bitter gourd	101.91	9.46	87.78	12.81
Chilly	113.37	9.74	110.74	12.65
Lady finger	105.5	9.304	90.89	13.06
Tomato	109.23	9.56	56	12.3

The graph is plotted for the above values to show the difference between the original leaf samples and the background removed image samples. It is given below

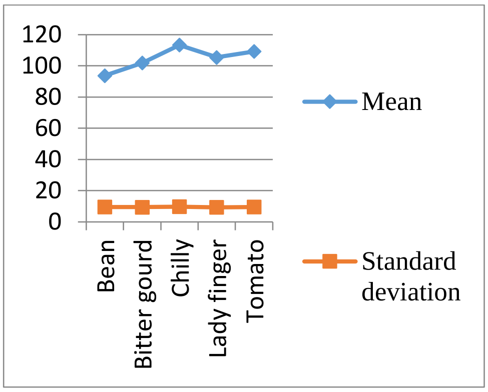


Fig 5. Performance evaluation graph for input leaf images

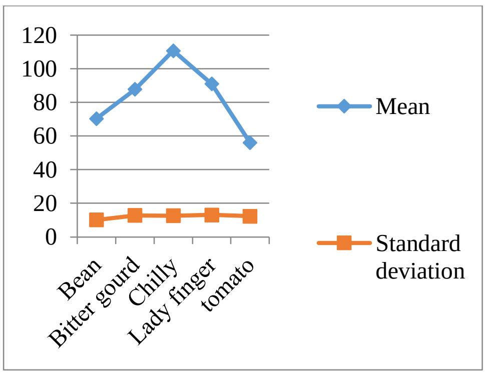


Fig 6. Performance evaluation graph for background removed image

From the above graph its understood that background removed image provide slightly poor result. This is due to the error produced due to the less brightness during the leaf image capture. To overcome this, the images should be captured under high brightness to avoid the error during the background removal.

C. Image segmentation

In this proposed method the clustering techniques are used to segment the disease affected portion from the input leaf image samples [4]. Clustering means grouping the similar pixels into clusters. The clustering techniques such as K-means clustering, Improved K-means clustering and Fuzzy c-means clustering and improved fuzzy c-means segmentation techniques are used.

1) K-means clustering

The similar pixels are grouped into clusters based on intensity values, RGB values, distance, connectivity and texture measurements. In this K-means segmentation cluster center is randomly selected [8]. Collection of points close to the centroid forms a cluster. The centroid gets updated according to the points and continues until the points stop changing their cluster [10], [12].

The k-means segmentation algorithm composed of following steps. Let $X = \{x_1,x_2,x_3,\ldots\ldots,x_n\}$ be the set of data points and $V = \{v_1,v_2,\ldots\ldots,v_c\}$ be the set of centers

1) Randomly select *'c'* cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

Recalculate the new cluster center.

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3).

2) Improved K-means clustering

This segmentation method is based on the distance calculation from the data point to the cluster center. This is same as above segmentation method but to improve the result here the objective function is minimized by reducing the number of iteration [1]. In the first iteration the objective function is calculated by

I1 =1/n (x-centroid) (1)

Where n= number of cluster and x= data point. In the second iteration the objective function is calculated by

I2= 1/(n+1) [x-centroid + previous iteration value] (2)

In the second iteration the previous iteration value is also calculated to provide better segmentation result. But in K-means the previous value is not added in objective function calculation.

3) Fuzzy C-means clustering

Fuzzy c-means clustering (FCM) is a method of clustering which allows a single data belongs to two or more clusters [1]. It allows the pixels belong to multiple classes with varying degrees of membership. It is also based on the minimization of the objective function:

$$J(U,c1,c2,\ldots.cc) = \sum_{i=1}^{c} Ji = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} \qquad (3)$$

Where, m is any real number greater than 1. Uij is the degree of membership of xi in the cluster j, Xi is the ith of d- dimensional measured data, and Cj is the d-dimension center of the cluster.

- In this each object belongs to each cluster to a certain degree.
- Object belongs to no clusters are considered as outliers.
- Not expected to overlap
- Child cluster belongs to the parent cluster.

4) Improved Fuzzy C-means clustering

An improved FCM is based on clustering centroids updates with the use of particle swarm optimization which is proposed in this paper [23]. This algorithm is designed to support multi dimension feature data and the accessible through parallel computation. The experimental results suggests that compared to the conventional FCM algorithm, the proposed algorithm leads to higher chances of global optimum clustering and its less computationally intensive when large clustering number is needed.

Particle swarm optimization is a population-based search algorithm and is initialized with a population of randomly selected solutions, called particles [22]. In PSO, each single solution is like a 'bird' in the search space, which is called 'particle'. All particles in PSO have their own fitness values which can be evaluated by the fitness function to be optimized, and also have velocities which direct the flying of the particles. These particles fly through the entire problem space by following the particles with the best solutions so far. PSO is initialized with a group of random particles and then searches for optima by updating each generation.

In this the member function is optimized. In this the conventional FCM the centroid value and clustering data are calculated. Randomly the values of centroids are generated. In this the best function is chosen and alpha value is optimized. In this thefitness function is basad on the minimum distance. The maximum iteration used in this technique is ten.

D. Performance analysis:

To measure the quality of the segmented leaf images [18], [20] the performance is analysed by using three parameters, which includes Rand Index (RI), Boundary Displacement Error (BDE) and Variation Of information (VOI).

1) Rand Index(RI)

The Rand index counts the fraction of pairs of pixels whose labeling are consistent between the computed segmentation [1]. The rand index or rand measure is a measure of the similarity between two data clusters.

$$R = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{n}{2}} \tag{4}$$

Where a+b as the number of agreements between X and Y, n is the number of elements and X and Y are cluster sets.

2) Boundary Displacement Error(BDE)

The Boundary Displacement Error (BDE) [1] measures the average displacement error of one boundary pixels and the closest boundary pixels in the other segmentation.

3) Variation of Information(VOI)

The Variation of Information (VOI) [3] metric defines the distance between two segmentations as the average conditional entropy of one segmentation given the other, and thus measures the amount of randomness in one segmentation which cannot be explained by the other. Suppose we have two clustering (a division of a set into several subsets) X and Y where X = {X1, X2... Xk}, pi = | Xi | / n, n = Σk | Xi |. Then the variation of information between two clustering is:

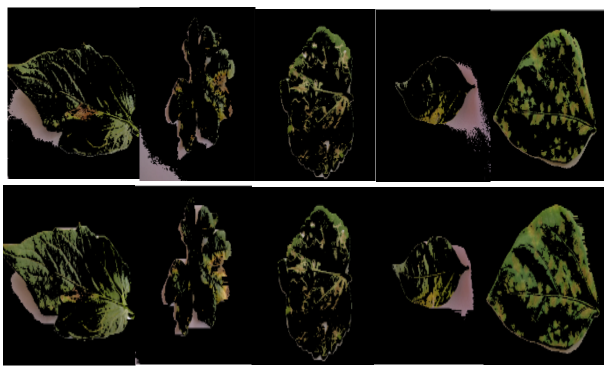
$$VI(X; Y) = H(X) + H(Y) - 2I(X, Y) \tag{5}$$

Where, H(X) is entropy of X and I(X, Y) is mutual information between X and Y. The mutual information of two clustering is the loss of uncertainty of one clustering if the other is given. Thus, mutual information is positive and bounded by

{H(X),H(Y)}_log2(n). (6)

III. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of various image segmentation techniques and the experimental results are analyzed for the five input leaf images. The figure 7 shows the clustering segmentations results for five input leaf samples.



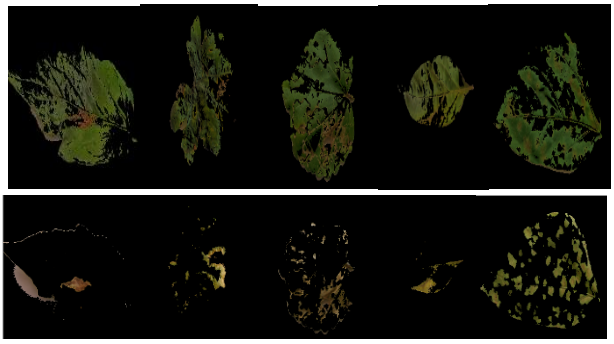


Fig.7 Segmentation results using Clustering techniques a) K-means clustering b) Improved k-means clustering c) Fuzzy c-means clustering d) Improved Fuzzy C-means

The performance evaluation for five input images using RI, BDE and VOI tabulation is given below

TABLE II. PERFORMANCE EVALUATION FOR SEGMENTATION

METHODS	RI	BDE	VOI	LEAF IMAGES
K-means	0.9898	0.2109	0.1678	Beans
Improved K-means	0.9725	0.136	0.1598	
FCM	0.9981	0.1534	0.0227	
Improved FCM	0.9987	0.0234	0.1743	
K-means	0.9861	0.2199	0.1624	Bitter Gourd
Improved K-means	0.9755	0.1224	0.1532	
FCM	0.9942	0.1524	0.0254	
Improved FCM	0.9982	0.0123	0.0743	
K-means	0.9854	0.2152	0.1673	Chilly
Improved K-means	0.9791	0.1363	0.1532	
FCM	0.9984	0.1549	0.0225	
Improved FCM	0.9985	0.0675	0.1342	
K-means	0.9883	0.2121	0.1628	Lady finger
Improved K-means	0.9735	0.145	0.1511	
FCM	0.9951	0.1555	0.0267	
Improved FCM	0.9964	0.0281	0.1324	
K-means	0.9811	0.2109	0.1678	Tomato
Improved K-means	0.9725	0.1363	0.1511	
FCM	0.9883	0.1549	0.0267	
Improved FCM	0.9912	0.0743	0.132	

From the Table III. The average value for Rand Index(RI), Boundary Displacement Error (BDE) and Variation Of Index(VOI) are taken from all the four segmentation results. The average performance analysis table is given below

TABLE III. AVERAGE CALCULATION OF PERFORMANCE ANALYSIS

METHODS	RI	VOI	BDE
K-means	0.97516	0.16562	0.2138
Improved K	0.98614	0.15368	0.1352
FCM	0.9912	0.0248	0.1242
Improved FCM	0.9966	0.04112	0.12944

If the value of RI is higher and BDE, VOI values are lower, then that segmentation approach is better. From the above Table III. the performance analysis chart is drawn. It is given below

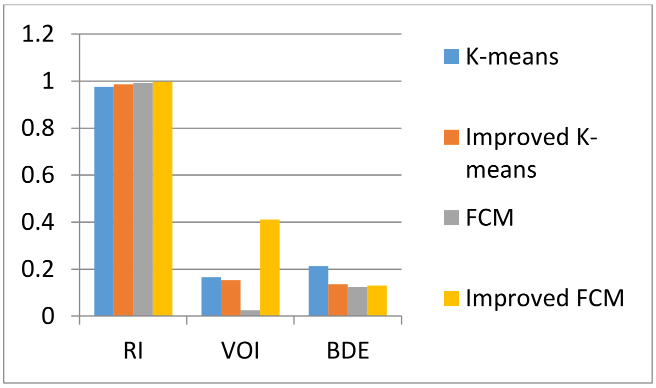


Fig.8 Performance analysis chart for segmentation

The figure 8 average performance analysis chart reveals that the rand index of improved fuzzy c-mean is higher than others and also the boundary displacement error and variation of information are lower than others. This indicates that the improved FCM segmentation approach is better based on these three parameters. In k-means it is difficult to predict the centroid value and overlap of data occurs. And it does not work well with clusters of different size and different density. In FCM it consumes more time and provides somewhat improved result. Here the RI value is higher than other values. So Improved Fuzzy C-means clustering technique is better.

IV. CONCLUSION AND FUTUREWORK

In this paper the clustering algorithm is proposed for the image segmentation. The clustering techniques such as K-means, improved K-means and Fuzzy C-means and improved fuzzy C-means clustering were tested for five different leaf images such as bean, bitter gourd, chilly, lady finger and tomato leaves. The performance of proposed algorithms is measured using segmentation parameters RI, BDE and VOI. With these three parameters the performance is analyzed and based on the analysis the Improved FCM segmentation approach provides better result.

An extension of this work will focus on developing hybrid algorithms for better segmentation result. The severity of the disease in the leaf images can be calculated and grouped into various classes based on the percentage of disease affected in the leaf image samples.

ACKNOWLEDGMENT

I sincerely thank our principal Dr.K. Ramar and management of our college for providing full support and encouragement for preparing this paper.

REFERENCES

[1] B.Sathya and R.Manavalan, "Image Segmentation by clustering methods: performance analysis," *International journal of computer applications (0975-8887), Vol 29-No.11, 2011*.

[2] P.Revathi and M.Hemalatha, "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques," *International Conference on Emerging Trends in Science, Engineering and Technology, 2012*.

[3] N.Valliammal and Dr.S.N.Geethalakshmi.,"Leaf Image Segmentation Based On the Combination of Wavelet Transform and K Means Clustering," *(IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 1, No. 3, 2012.*

[4] J. Fan and D. K. Y. Yau, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE on Image Processing, 10, vol. 10, no. 10, pp. 1454-1466,2001.*

[5] K.Muthukannan et al.," Extraction of Disease Portion from Plant Leaves and Its Severity Measurement Using Image Segmentation" *International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.1 (2015) pp. 337-343*

[6] Sanjay B. Patil et al., "Leaf Disease Severity Measurement using Image Processing", *International Journal of Engineering and Technology Vol.3 (5), 2011, 297-301.*

[7] S. Wesolkowski and P. Fieguth, "A Markov random fields model for hybrid edge-and region-based colour image segmentation," in Proc. *Canadian Conference on Electrical and Computer Engineering,. 2002.*

[8] Gao ronghua et al., "Nearest Neighbor recognition of cucumber disease images based on kd-tree", *Information technology journal 12(23): 7385-7390, 2013.*

[9] Al-Bashish, D., M.Braik, and S. Bani-Ahmad."Detection and classification of leaf diseases using K-means-based segmentation and neural networks based classification". *Information Technology Journal, vol.10 (2): 267-275, December 2011.*

[10] Al-Hiary, H et al. "Fast and accurate detection and classification of plant diseases". *International Journal of Computer Applications, 17(1): 31-38, 2011.*

[11] Bradley, P.S., Fayyad, U.M., Refining initial points for Kmeans clustering. In: Sharlik, J. (Ed.), Proc. 15th Internat. Conf. on Machine Learning (ICML98). Morgan Kaufmann, San Francisco, CA, (1998) 91-99 621.

[12] Au Dimitrios, Moshou, and Cedeic Bravo,"Automatic detection yellow rust in wheat using reflectance measurements and neural networks," *Elsevier, pp.173-188, 2004.*

[13] K.S.Ravichandran and B. Ananthi," Color Skin Segmentation Using K-Means Cluster" *International Journal of Computational and Applied Mathematics ISSN 1819-4966 Volume 4 Number 2 (2009), pp. 153–157.*

[14] Bock H. and Poole G.H ,"Plant disease severity estimate visually and by Hyper spectral imaging," *Plant Science, pp.59-10, 2010.*

[15] Hai Gao, Wan-Chi and San-Chi), "Improved Techniques for Automatic Image Segmentation", *IEEE transactions on circuits and systems for video technology, vol. 11, NO. 12, 2011.*

[16] Jain J. K. and Rastogi.R"Application of image processing in Biology and Agriculture", *Nuclear India, pp.12-13, 1996.*

[17] Jun pang and zhong-ying Bai, "Automatic Segmentation of Crop leaf spot disease images by integrating local threshold and seeded region growing," *978-1-61284-881-5/11/$26.00@2011 IEEE,2011.*

[18] Kridsakron Auynirndronkool and Varinthon Jarnkoon,"Analysis of economic crop reflectance by field spectral signature, case study sugarcane", *Journal of plant physiol, pp.1-9, 2008.*

[19] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 929–944, Jun. 2007*.

[20] William W. Hargrave and Crosslacy D. A."Video digitizer for the rapid measurement of leaf area lost due to Herbivorous insect", *Journal of Entomological Society of America, pp.591- 598, 1998.*

[21] F. Ge, S.Wang, and T. Liu, "New benachmark for image segmentation evaluation," *J. Elect. Imag., vol. 16, no. 3, Jul.–Sep. 2007.*

[22] Dilpreet Kaur and Yadwinder Kaur "Intelligent Medical Image Segmentation Using FCM, GA and PSO" International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6089-6093.

[23] Liang pang et al., "A improved clustering analysis method based on Fuzzy C-means algorithm by adding PSO algorithm", Springer-verlag berlin Heidelberg 2012, part I, LNCS 7208, pp 231.