Detecting Leaf Spots in Cucumber Crop Using Fuzzy Clustering Algorithm

Mohammed El-Helly
Central Lab. For Agricultural Expert System (CLAES)
Agriculture Research Center (ARC)
helly@mail.claes.sci.eg

Hoda Onsi
Information Technology Dept.
Faculty of Computers &Information
Cairo University

Ahmed Rafea
Computer Science Dept.
American University in Cairo

Salwa El-Gammal
Computer Science Dept.
Faculty of Computers &Information
Cairo University

Abstract

Leaf spots are indicative of crop diseases, where leaf batches are usually examined manually and subject to expert opinion. In this paper we present a segmentation technique for identifying the leaf batches in a cucumber crop using fuzzy clustering algorithm (FCM). Adapting the FCM algorithm parameters based on fuzzy entropy, partition coefficient, and compactness measures was used for choosing the optimal cluster number. It was found that, the adapted FCM technique has a better detection of spots when using a window selection. Experimental results have demonstrated the effectiveness and superiority of the algorithm after an extensive set of color images was tested.

1. Introduction

Image segmentation is the first step in image analysis and pattern recognition. It is a critical and essential component of image analysis and/or pattern recognition system. It is one of the most difficult tasks in image processing, and determines the quality of the final result of analysis. Image segmentation is the process of partitioning an image into disjoint regions such that each region is homogeneous, according to a certain criterion, such as intensity or color, and no union of any two adjacent regions is homogeneous. The problem has been broadly investigated by scientists using both classical [15,20,23,30,32,37] and fuzzy based techniques [3-12, 38-48]. The classical segmentation approaches take crisp decisions about the regions. Nevertheless, the regions in an image are not always crisply defined, and uncertainty can arise within each level of image processing. Fuzzy set theory provides a mechanism to represent and manipulate uncertainty and ambiguity. The more important advantage of a fuzzy methodology lies in the membership function that provides a natural means to model the uncertainty in an image. Segmentation can be viewed as a clustering problem. One problem with traditional clustering techniques is that there are only two values, either 1 or 0 to indicate to what degree a data point belongs to a cluster. This requires well-defined boundaries between clusters, which is not the usual case for the real image. Most plant images present overlapping gray-scale intensities for different tissues. In particular, borders between tissues are not clearly defined and memberships in the boundary regions are intrinsically fuzzy. Therefore fuzzy clustering turns out to be particularly suitable for the segmentation of plant image. One widely used algorithm is the fuzzy c-mean (FCM) algorithm. In this paper we apply FCM algorithm and adapt its parameters to meet our application. Section 2 presents materials and methods. Results of the case study are presented in Section3. Conclusions are presented in Section4.

2. Materials and methods

2.1 Case Description

The image data sets in this paper relate to cucumber crop. Some of them exist in the literature, these include the disorders, Gummy stem blight, Pesticide injury, Scab, Phosphorus def., High temperature, and White fly. The other sets were captured at central lab for agricultural expert system by high resolution 3-CCD color camera (DSC-P1 Cyber-shot, Sony) with 3.3 million-pixel sensitivity, 3X optical zoom lens, and auto focus illuminator light, Focal Length 8 - 24mm. The camera was placed at about 60mm from top of the leaves. The image from the camera is digitized into a 24-bit image with resolution 720 x 540 pixels. The 90 defected images of cucumber are taken in cucumber green house to cover three categories of disorders, mainly powdery mildew, downy mildew, and leafminer. Figure1 represents examples of those images.

Figure1: Samples of defected images

2.2 Fuzzy c-mean Clustering

Fuzzy c-mean clustering is a simple unsupervised learning method, which can be used, for data grouping. The FCM algorithm is the best known and most widely used fuzzy clustering technique. It was first presented by Dunn [8], further developed by Bezdek [2], and subsequently revised by Rouben[36], Trauwaert[46], Goth [11], Gu [14], and Xie [47]. However, Bezdek’s FCM remains the most commonly used [47]. The basic mathematical foundation of FCM is as follows:
Let X = $\{x_1, \ldots, x_n\}$ be a data set. Let C be a positive integer greater than one. A partition of X into C clusters can be presented by mutually disjoint sets $X_1, \ldots, X_c$ such that $X_1 \cup \ldots \cup X_c = X$ or equivalently by the indicator functions $\mu_1, \ldots \mu_c$ Such that

$\mu_i(x)=1$ if x is in X_i otherwise $\mu_i(x)=0$ for all $i = 1,\ldots C$. This is known as hard c-partition. The well-known hard c-mean is an iterative algorithm to minimize the objective function $\boldsymbol{J_{HCM}}$ defined as:

$$J_{HCM}(\mu, a) = \sum_{i=1}^{C}\sum_{j=1}^{n}\mu(x_j)\left\|x_j - a_i\right\|^2 \tag{1}$$

Where $\boldsymbol{a_1},\ldots\ldots\ldots\ldots \boldsymbol{a_C}$ is the cluster centers.
The fuzzy extension allows $\mu_i(x)$ to be a membership function in a fuzzy sets μ_i on X, the degree of membership is between [0, 1] such that

$$\sum_{i=1}^{C}\mu(x_j) = 1.\forall x \in X \tag{2}$$

In this case $\{\mu_1,\ldots\ldots\ldots\ldots \mu_C\}$ is called fuzzy c-partition on X. Thus the fuzzy c-mean (objective function) $\boldsymbol{J_{FCM}}$ becomes

$$J_{FCM}(\mu, a) = \sum_{i=1}^{C}\sum_{j=1}^{n}\mu(x_j)^m\left\|x_j - a_i\right\|^2 \tag{3}$$

Where *m* is a number bigger than one to present the fuzziness. The FCM clustering algorithm is an iteration through the necessary conditions for minimizing $\boldsymbol{J_{FCM}}$ with the following equations

$$\mu_i(x_j) = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\|x_j - a_i\right\|^2}{\left\|x_j - a_k\right\|^2}\right)^{\frac{1}{m-1}}} \tag{4}$$

$$a_i = \frac{\sum_{j=1}^{n}\left(\mu_i(x_j)\right)^m x_j}{\sum_{j=1}^{n}\left(\mu_i(x_j)\right)^m} \tag{5}$$

Where $i = 1, \ldots\ldots\ldots, C$ $\quad j = 1, \ldots\ldots\ldots, n$

The FCM algorithm is shown in Figure 2

Initialize $V=\{v_1,\ldots. v_C\}$
Initialize convergence criteria
Initialize degree of fuzziness $m >= 1$
Repeat

- $V^{previous} = V$
- Compute membership function using Eq. 4
- Update Cluster centers using Eq 5

Until $\sum_{i=1}^{c}\left\|V_i^{previous} - V_i\right\| < \varepsilon$

Figure 2:Fuzzy c-mean algorithm

2.3 Adaptation of FCM

The segmentation of defected plant images involves partitioning the image space into different cluster regions with similar intensity image values. The success of applying FCM to fit the segmentation problem depends mainly on adapting the input parameters value [24,25]. As a consequence, if the parameters are assigned an improper value, the clustering results in a partitioning scheme that is not optimal for the specific data set and that leads to a wrong decision. These parameters include, the feature of the data set, the optimal number of cluster, and the degree of fuzziness.

2.3.1 Feature of the Data Set

We considered the image intensity as the main feature of the data set, since the defected part of the leaves has high intensity. The intensity image is defined as the average of the three components r, g, b. So the cluster with high intensity center represents the defected part of the leaves. The clustering algorithm does not directly incorporate spatial information and which makes it is sensitive to the noise and intensity in-homogeneities [9]. This lack of spatial information can be overcome by choosing the mean computed in a neighborhood of 5x5 pixels. In addition, the x, y coordinates of the pixels are used as a second and third feature respectively for spatial information.

2.3.2 Optimal Cluster Number

The problem of deciding the optimal number of clusters has been the subject of many research efforts [33,45]. Several cluster validation measures have been developed [5]. We will describe and use three of these measures: partition coefficient [2,10,25], partition entropy [2,10,25], compactness and separation [2,10,25]. The partition coefficient measures the closeness of all input samples to their corresponding cluster centers, as defined by:

$$F(U,c) = \frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^2 \qquad \textbf{(6)}$$

The optimum choice by this measure is given by:

$$\max\{F(U,c)\} \qquad c = 2,\ldots,n-1 \qquad \textbf{(7)}$$

The partition entropy measures the average amount of the information contained [10] as defined by:

$$H(U,c) = -\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}\log(\mu_{ik}) \qquad \textbf{(8)}$$

The optimum choice by this measure is given by:

$$\min\{H(U,c)\} \qquad c = 2,\ldots,n-1 \qquad \textbf{(9)}$$

When all membership values are closed to 0.5 - which represent the high degree of fuzziness of the clusters- then the entropy gets large indicating poor cluster results. On the other hand, if all membership values are close to 0 or 1, then the entropy gets small indicating good clustering results [5,25]. The compactness and separation is a ratio between the average distance of the input samples to the corresponding cluster centers and the minimum distance between cluster centers [5]. It is defined as:

$$S(U,c) = \frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^{2}\left\|X_k - V_i\right\|^2}{\min_{i,j}\left\|V_i - V_j\right\|^2} \quad \textbf{(10)}$$

The optimum choice by this measure is given by:

$$\min\{S(U,c)\} \qquad c = 2,...,n-1 \quad \textbf{(11)}$$

Good cluster results should make all input samples as close to their cluster center as possible, and all cluster centers separated as far as possible [5]. For the compact and well separated clusters the value of S should be minimum [25]. The results for applying those methods on our defected images to get the optimal clusters are shown in subsection 3.1

2.3.3 The degree of fuzziness

The value of fuzzy exponent *m* remains, however, to be determined. The parameter *m* controls the amount of fuzziness; the performance of FCM is dependent on a good choice of this parameter. Bezdek *et al*. [1] suggested that, this value should be in the range 1.5 to 3 to give good results. It was noted, however, that there was no strong theoretical justification or empirical evidence for this choice. Cannon *et al*. [31] also noted that " no theoretical basis for choosing a good value is available", and suggested 1.1 to 5 were "typically reported as the most useful range of values". McBrateney and Moore [27] also investigated the choice of *m* and found that a value of approximately 2 was optimal. Whereas Choe and Jordan [6] found that the algorithm is insensitive to the range from 8 to 30. Cannon and Jacobs [4] found that, for image applications, the range between 1.1 to 2.5 "has proved adequate for all practical purposes". In subsection 3.2, the experimental results of applying different values of the fuzzy exponential *m* over the following sets {1.1,1.5,2,2.2, 3,8,15} of values, are demonstrated.

3 Results

Through this section, experimental results of adapting the most important parameters of FCM will be discussed as well as the output of applying the adapted FCM to our data sets

3.1 Optimal Cluster Number Results

According to the analysis of the optimal number of clusters using the partition coefficient, partition entropy, and compactness and separation measures, the FCM algorithm was applied for the number of clusters between 4 and 8 clusters. It was found that it is not useful to specify less than 4 clusters and more than 8 clusters for this dataset. Since fewer than 4 clusters give an under segmentation and over than 8 clusters give an over segmentation. The results are summarized in table 1, and plotted in Figures 3,4,5. As shown from Figures, the optimal number of clusters is 4.

Clusters	Powdery (Class1)			Downey (Class2)			Leafminer (Class3)		
	F(U,c)	H(U,c)	S(U,c)	F(U,c)	H(U,c)	S(U,c)	F(U,c)	H(U,c)	S(U,c)
4	**0.805**	**0.538**	**0.083**	**0.837**	**0.455**	**0.074**	**0.841**	**0.441**	**0.086**
5	0.77	0.658	0.126	0.805	0.56	0.107	0.794	0.577	0.13
6	0.743	0.757	0.139	0.781	0.647	0.12	0.759	0.688	0.246
7	0.71	0.866	0.167	0.755	0.647	0.133	0.722	0.798	0.423
8	0.667	0.997	0.264	0.721	0.84	0.249	0.696	0.886	0.672

Table1: Partition&Entropy&Compactness measures for the three Classes

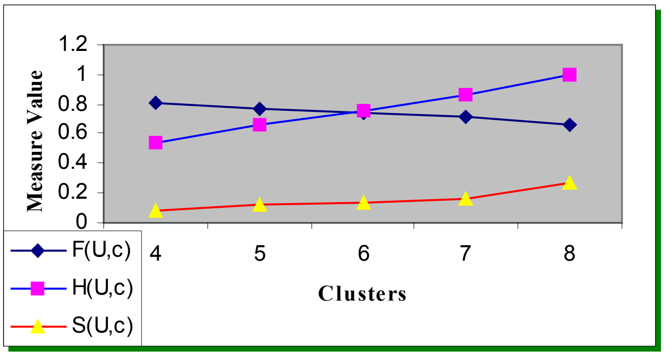

Figure3: Fuzzy measures for Class1

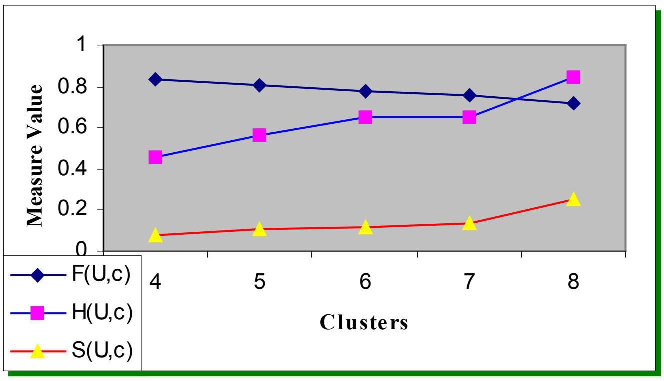

Figure4: Fuzzy measures for Class2

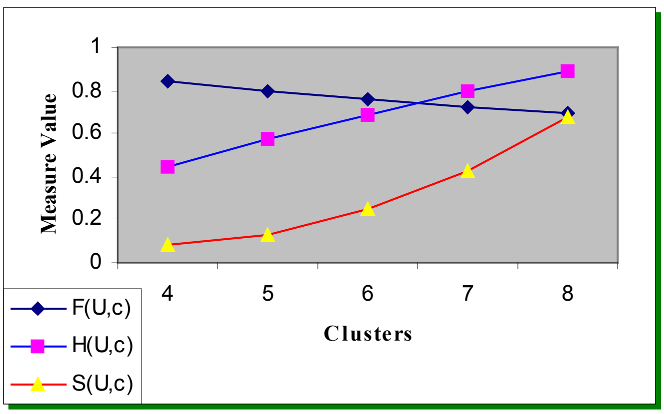

Figure5: Fuzzy measures for Class3

3.2 Degree of Fuzziness Results

As discussed in Subsection 2.3.3 there is no accurate decision about the exponential value *m*. from the previous research the most recommended values between 1.1, 15. Thus through this subsection we present the experimental results of applying these set of values. Figure 6 shows the results of applying the previous set of values on our three classes of dataset. As shown, the values 1.1to 2.2 for *m* do not cause a significant difference in noise. On the other hand, values of *m* higher than 2.2 have more noise. So we choose the number 2 as the optimum number of *m* for our problem since the calculation will be easier than using 1.1 and 1.5.

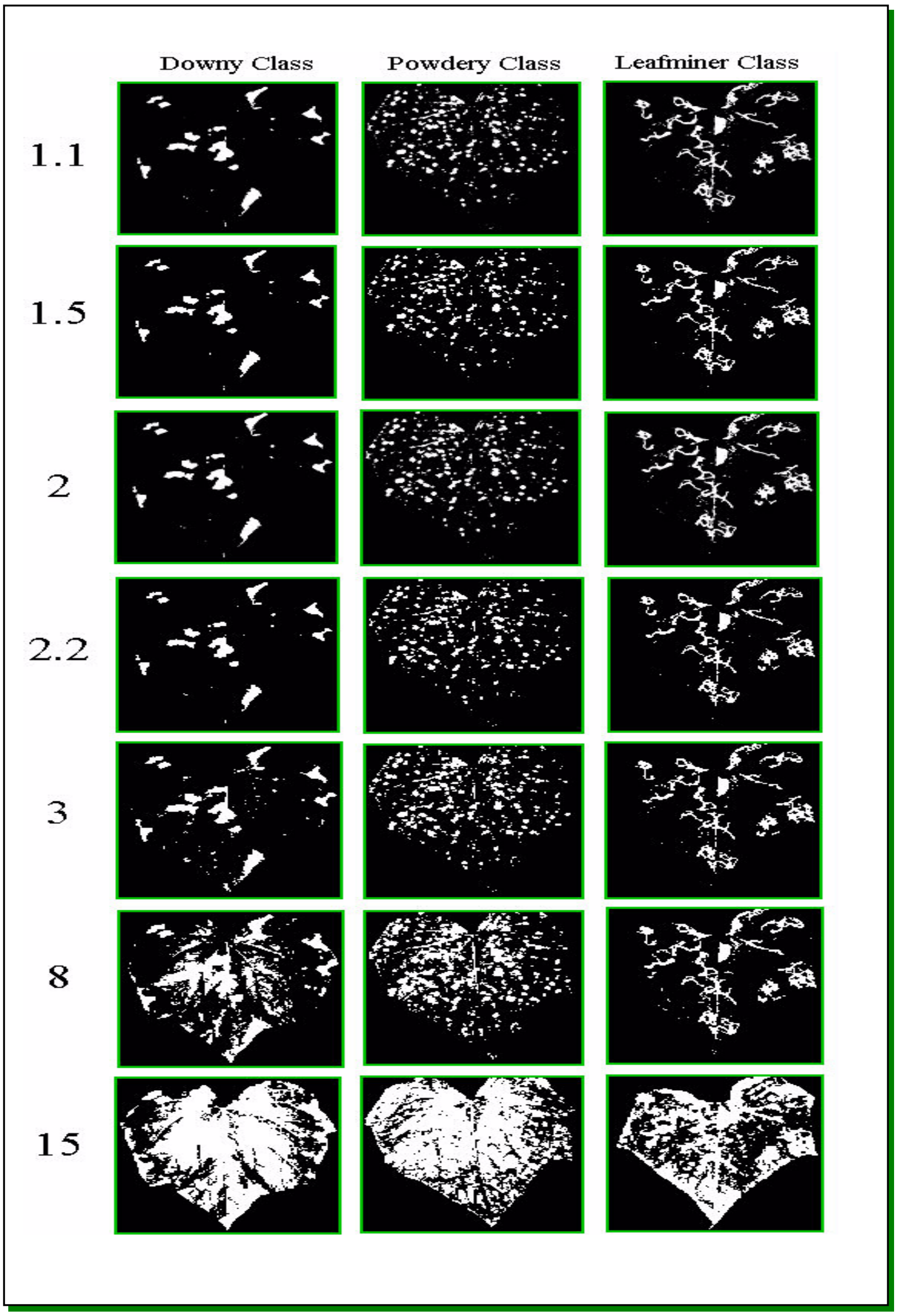

Figure6: Comparison between different values of *m*

3.3 Results of FCM

Through this subsection, experimental result of applying FCM on the literature images and our set of images will be demonstrated. Figure 7 shows the original literature images and the results of segmentation.

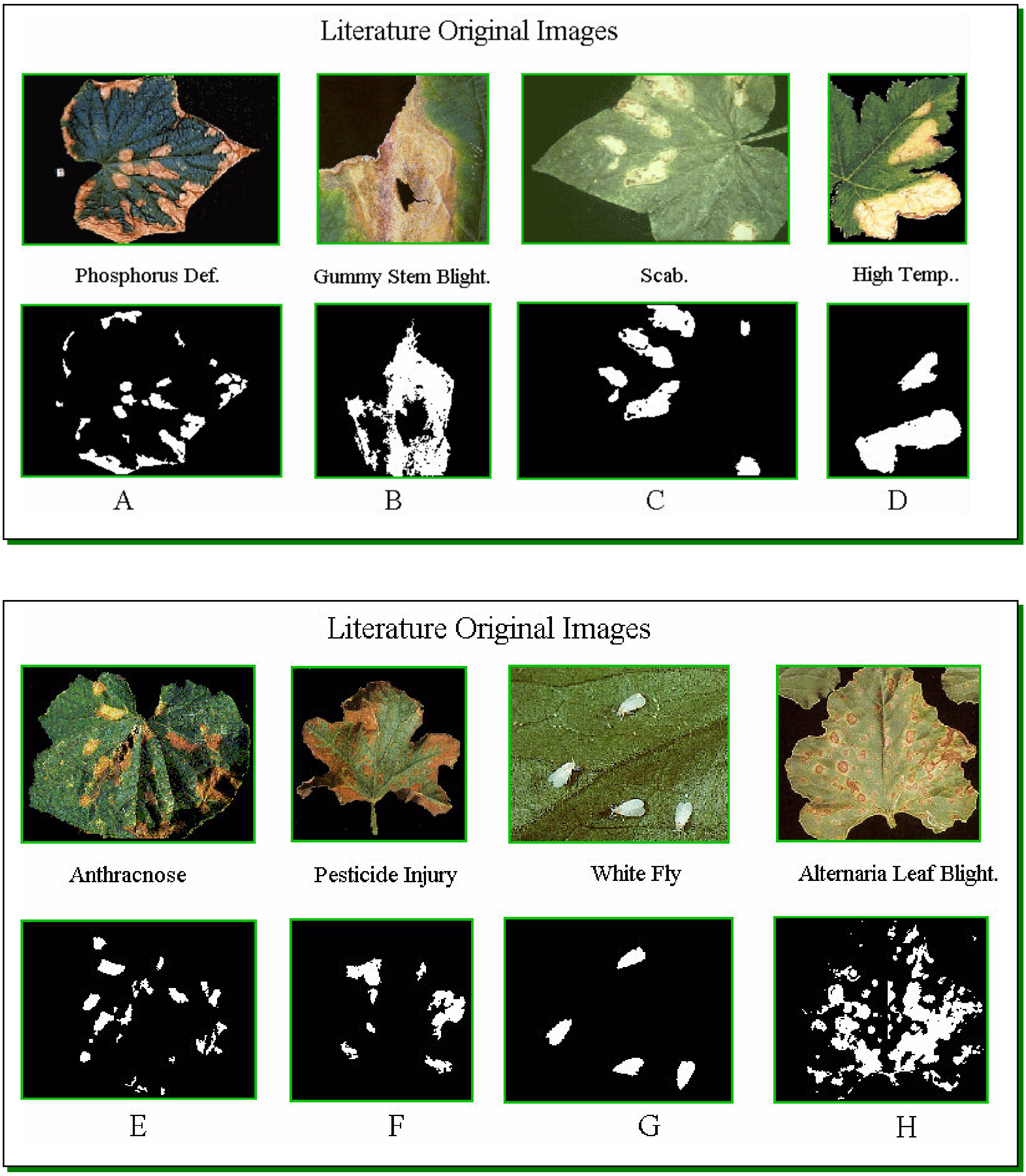

Figure7: Segmentation Results on some Literature Images

Figure 8,9,10 show some images that are taken from central Lab. for agricultural expert systems (CLAES), for leafminer, downey milew, and powdery mildew and the segmentation results.

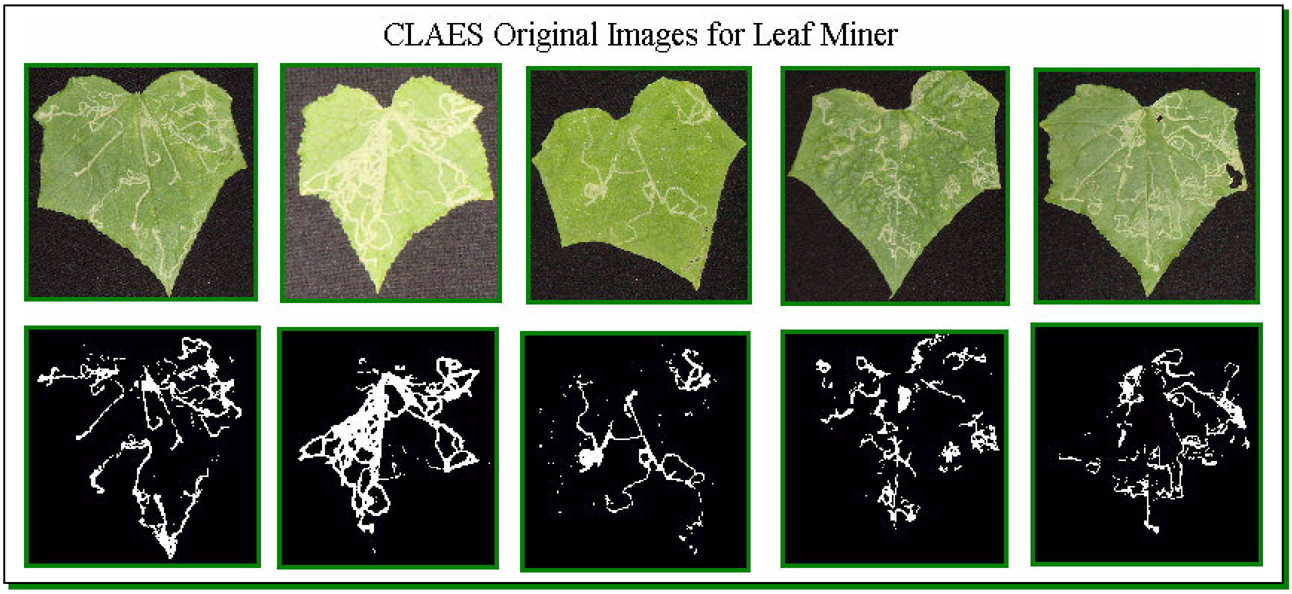

Figure8: Original and segmented CLAES images for Leafminer

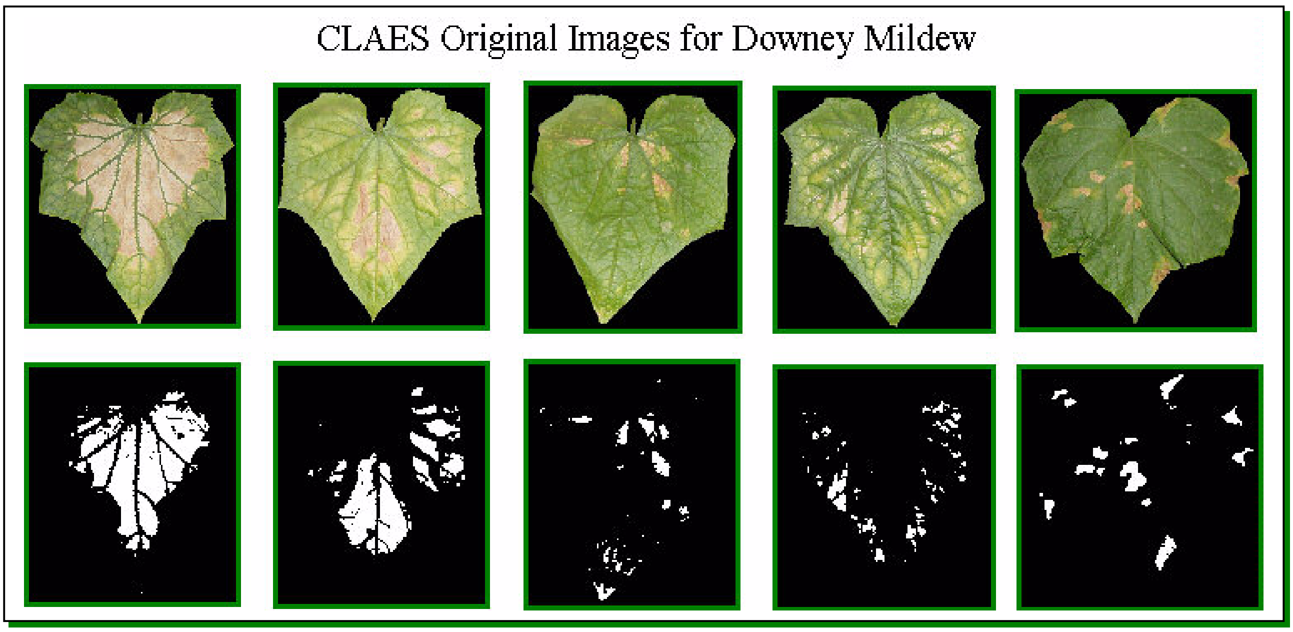

Figure9: Original and segmented CLAES images for Downey

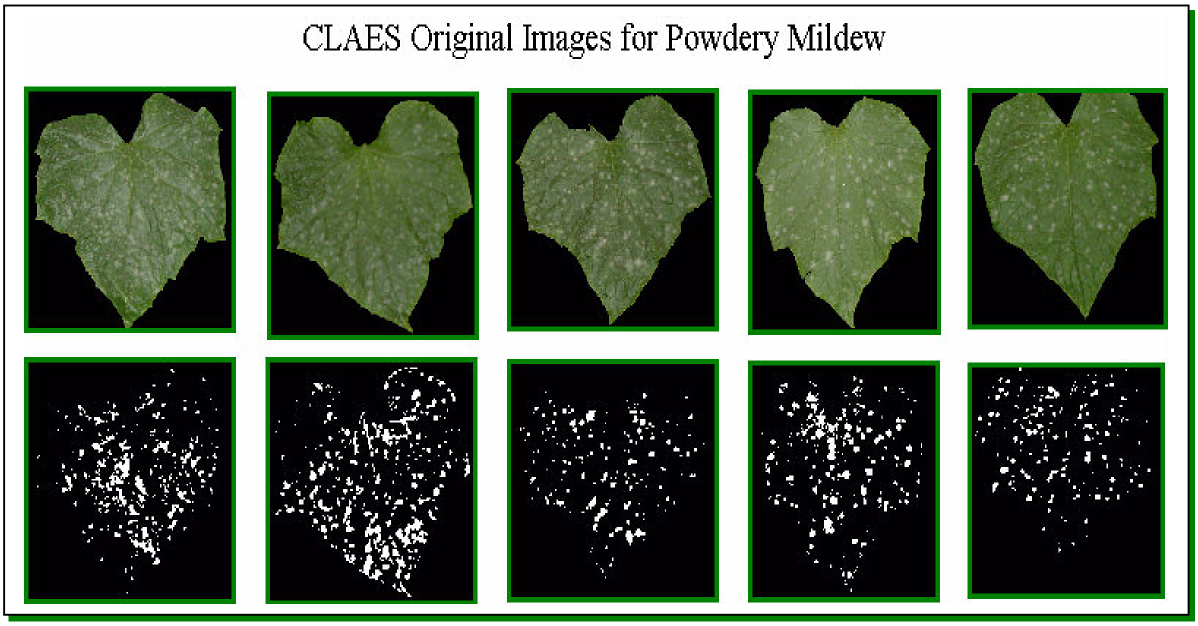

Figure10: Original and segmented CLAES images for Powdery

Figure 11 demonstrates the result of segmentation after selecting only part of the defected image.

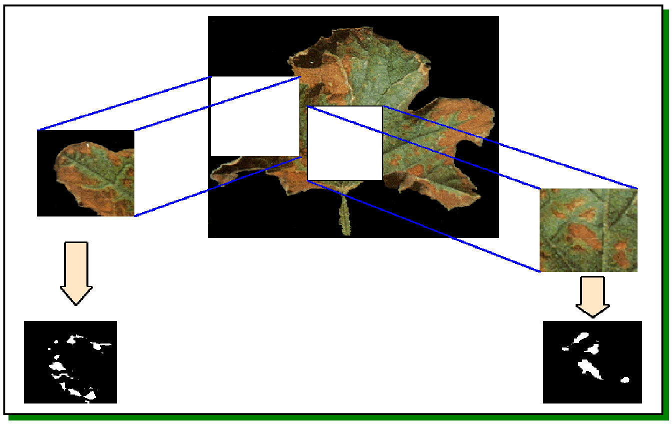

Figure11: Window Selection Segmentation

4 Conclusions

In this paper, we have implemented FCM as a clustering algorithm for segmenting the leaf spots in cucumber crop. The algorithm was successfully applied in two types of defected images, literature images, and our own acquired images. We also described several methods for choosing optimal cluster numbers and experiments for choosing the degree of fuzziness. Our experiments show that, the optimal cluster number for leaf spots problems is 4, and the degree of fuzziness is 2. Those parameters give accurate results for segmenting the spots. This method can be used effectively to detect the spots. The feature of these spots can be extended to another phase for delivery to a diagnostic system.

Also the segmentation accuracy was improved by using spatial information from neighboring pixels and by incorporating the x, y coordinates of the pixels. From experiments we have determined that applying the clustering algorithm to a full size image with a large number of pixels results in excessively long processing time. The segmentation was found to be more accurate when we considered only the defected part of the leaf. Therefore, we used the window segmentation for two reasons: The first one was to enhance the detection of the segmentation specially when the spots where too small. The second one was to reduce the processing time.

In general, we recommend using FCM for segmenting the plant images to detect the leaf abnormalities, and hence can be diagnosed and controlled.

Reference

[1] Bezdek J.C., R. Ehrlich, W. Full. (1984*), FCM: The Fuzzy c-mean Clustering Algorithm*, Computers and Geoscience, 10, pp.191-203.

[2] Bezdek, J. C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, NewYork, NY.

[3] Bloch I. (1993), Fuzzy Connectivity and Mathematical Morphology, Pattern Recognition Letters, Vol. 14, 483-488.

[4] Cannon R.L. and C.Jacobs. (1984), Multispectral Pixel Classification with fuzzy Objective Functions, Center for Automation Research, Technical Report CAR-TR51, University of Maryland, College Park.

[5] Chi Z., Yan H., Pahm T. (1996), Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition, World Scientific.

[6] Choe H., J.B. Jordan. (1992), On the optimal Choice of parameters in a Fuzzy c-mean Algorithm, Proceedings of the IEEE International Conference On Fuzzy Systems, pp.349-354.

[7] De Luca and S. Termini.(1972), A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Set Theory, Information and Control, 20, 301-312.

[8] Dunn, J. C. (1974), Well-separated clusters and the optimal fuzzy partitions, Journal of Cybernetics, 4: 95-104.

[9] Dzung L., Chenyang X., Jerry L. (2001), A Survey of Current Methods In Medical Image Segmentation, Technical Report JHU/ECE 99-01, The Johns HopkinsUniversity.

[10] Erhan G. (2000), A new Clustering Algorithm For Segmentation of Magnetic Resonance Images, University of Florida.

[11] Gath and A. B. Geva. (1989), Unsupervised Optimal Fuzzy Clustering, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-11, No. 7, 773-781.

[12] Gath, I., Geva A. B. (1989), Unsupervised Optimal fuzzy clustering, IEEE Transaction on Pattern Analysis Machine Intelligence, 11(7): 773-781.

[13] Gonzaliz RC., Woods RE. (2002), Digital image processing, Pearson Education. Inc.

[14] Gu. T., Dubuission B. (1990), Similarity of classes and fuzzy clustering, Fuzzy Sets and Systems, 34: 213-221.

[15] H. Chen. (1981), On the Statistical Image Segmentation Techniques, Proceedings of IEEE Conference on Pattern Recognition and Image Processing, 262-266.

[16] J. C. Bezdek, P. F. Castelaz, (1993), Prototype Classification and Feature Selection with Fuzzy Sets, Pattern Recognition Letters, Vol. 14, 483-488.

[17] J. K. Udupa, S. Samarasekera. (1996), Fuzzy Connectedness and Object Definition: Theory, Algorithms and Applications in Image Segmentation, Graphical Models and Image Processing, Vol. 58, No. 3, 246-261.

[18] J. M. Keller and C. L. Carpenter. (1990), Image Segmentation in the Presence of Uncertainty, International Journal of Intelligent Systems, Vol. SMC-15, 193-208.

[19] J. M. Keller, M. R. Gray and J. A. Givens. (1985), A Fuzzy K-Nearest Neighbor Algorithm, IEEE Trans. Syst. Sci. and Cybern., Vol. SMC-15, 580-585.

[20] K. S. Fu and J. K. Mui. (1981), A Survey on Image Segmentation, Pattern Recognition, 13, 3-16.

[21] K. Tsuda, M. Minoh, K. Ikeda. (1996), Extracting Straight Lines by Sequential Fuzzy Clustering, Pattern Recognition Letters, Vol. 17, 643-649, 1996.

[22] L. K. Huang and M. J. Wang. (1995), Image Thresholding by Minimizing the Measures of Fuzziness, Pattern Recognition, Vol. 28, No. 1, 41-51.

[23] L. Spirkovska. (1993), A Summary of Image Segmentation Techniques, NASA Technical Memorandum 104022, June.

[24] M. Halkidi, Y. Batistakis, M. Vazirgiannis. (2002), Cluster Validity Methods: Part I, SIGMOD Record, June.

[25] M. Halkidi, Y. Batistakis, M. Vazirgiannis. (2002), Cluster Validity Methods: Part II, SIGMOD Record, September.

[26] M. P. Windham. (1983), Geometrical Fuzzy Clustering Algorithms, Fuzzy Sets and Systems, Vol. 10, 271-279.

[27] McBratney A.B. and A.W.Moore (1985. Application of Fuzzy Sets to Climatic Classification. Agricultural and Forest Meteorology, 35, pp. 165-185.

[28] N. Chun, H. S. Yang. (1996), Robust Image Segmentation Using Genetic Algorithm With A Fuzzy Measure, Pattern Recognition, Vol. 29, No.7, 1195-1211.

[29] N. R. Pal and S. K. Pal. (1989), Object-background Segmentation Using New Definition of Entropy, IEEE Proceedings, Vol. 136, Part E, No. 4, 284-295, July.

[30] P. K. Sahoo, et al. (1988), A Survey of Thresholding Techniques, Computer Vision, Graphics and Image Processing , 41, 233-260.

[31] R. L. Cannon, J. V. Dave, J. C. Bezdek. (1986), Efficient Implementation of the Fuzzy c-mean Clustering Algorithms, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 2, 249-255.

[32] R. M. Haralick, L. G. Shapiro. (1985), Image Segmentation Techniques, Computer Vision, Graphics and Image Processing, 29, 100-132.

[33] Rezaee, R, Lelieveldt, B.P.F., Reiber, J.H.C. (1998), A new cluster validity index for fuzzy c-mean, Pattern Recognition Letters, 1, pp. 237-246.

[34] Rosenfeld. (1979) Fuzzy Digital Topology, Information Control, Vol. 40, No. 1, 76-87.

[35] Rosenfeld. (1984), The Fuzzy Geometry of Image Subsets, Pattern Recognition Letters, Vol. 2, 311-317.

[36] Rouben, M. (1982), fuzzy clustering algorithms and their clustering validity, European Journal of Operational Research, 10: 294-301.

[37] S. K. Pal et al. (1993), A Review on Image Segmentation Techniques, Pattern Recognition, 29, 1277-1294.

[38] S. K. Pal, A. Rosenfeld. (1988), Image Enhancement and Thresholding by Optimization of Fuzzy Compactness, Pattern Recognition Letters, Vol. 7, 77-86.

[39] S. K. Pal, R. A. King. (1983), On Edge Detection of X-Ray Images Using Fuzzy Sets, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No.1, 69-77.

[40] S. K. Pal, R. A. King. (1993), Prototype Classification and Feature Selection with Fuzzy Sets, Electronics Letters, Vol. 16, No. 10, 376-378.

[41] S. K. Pal. (1992), Image Segmentation Using Fuzzy Correlation, Information Sciences, 62, 223-250.

[42] S. Tamura, S. Higuchi, K. Tanaka. (1971), Pattern Classification Based on Fuzzy Relations, IEEE Trans. on System, Man, Cybern., Vol. SMC-1, No. 1, 61 – 66.

[43] S. Z. Selim, M. A. Ismail (1986), On the Local Optimality of the Fuzzy Isodata Clustering Algorithm, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No.2, 284-288, Mar.

[44] T. L. Huntsberger, C. Rangarajan, S. N. Jayaramamurphy. (1993), Representation of the Uncertainty in Computer Vision Using Fuzzy Sets, IEEE Trans. on Comput., Vol. C-35, No. 2, 145-156.

[45] Theodoridis S., Koutroubas. (1999), Pattern recognition, Academic Press.

[46] Trauwaert, E. (1985), On the Meaning of Dunn's partition coefficient for Fuzzy Clusters, International Working Paper, Vrije Universiteit Brussels.

[47] Tsuen, Ho, Hsu. (2000), An Application of Fuzzy Clustering in Group-Positioning Analysis, Proc. Natl. Sci, Counc. ROC(C),Vol. 10, No.2, pp.157-167.

[48] X. L. Xie, G. Beni. (1991), A Validity Measure for Fuzzy Clustering, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, No. 8, 841-847.

[49] W. S. Kim and R. H. Park. (1996), Color Image Palette Construction Based on the HSI Color System for Minimizing the Reconstruction Error, IEEE International Conference on Image Processing, C, 1041-1044.