Comparison of Features Extraction Algorithms Used in the Diagnosis of Plant Diseases

Mohammed A.Hussein

mohammed.alsraf@gmail.com
Amel H.Abbas

dr.amelhussein2017@uomustansiriyah.edu.iq2

Computer Science/ College of Science/ University of Mustansiriyah

Abstract

The detection of diseases affecting plant is very important as it relates to the issue of food security, which is a very serious threat to human life. The system of diagnosis of diseases involves a series of steps starting with the acquisition of images through the pre-processing, segmentation and then features extraction that is our subject finally the process of classification. Features extraction is a very important process in any diagnostic system where we can compare this stage to the spine in this type of system. It is known that the reason behind this great importance of this stage is that the process of extracting features greatly affects the work and accuracy of classification. Proper selection of the right features leads to high accuracy in the system diagnostics and vice versa. The proposed system collect images of different crop (Rice, cotton and tomato) disease, we will enter the images of cropping them , then Re-size the images to fixed size, then improve the image through Fuzzy histogram equalization (FHE) , then perform image segmentation using color based K-means and finally compare the methods of features extraction (Percentage of Leaf Area Infected (PI),Texture-Based Features, Color Moments, Features obtained by Color Co-occurrence Method and Shape based Features) we found that the use of 4 methods together (Percentage of Leaf Area Infected (PI),Texture-Based Features, Color Moments and Shape based Features) produce excellent result..

Keywords: Plant Leaf, Feature Extraction, Texture, Color Moments, and Segmentation.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

Introduction

There are many diseases that affect crops and lead to significant production losses, which threaten the issue of food security. Human visual examination with the naked eye is the way most widely used and common. This method gives a large room for error depending on where the farmers trying to detect the disease through visual inspection as a big chance of error in some cases resorting to experts, this thing needs a lot of time, effort and money. Another problem in Iraq, said most of the crop fields is located in rural areas, which requires farmers to go long distances to find experts [16]. Image processing gives accuracy, high-speed, do not require large sums of money have been spent and time-consuming as in the case brought experts [17]

Features Extraction is an important task and key in the plant disease detection algorithm .The extraction of features simplifies the amount of data required to describe the large amount of data contained in the image. The large amount of information consumes memory and processor time therefore we need the smallest possible quantity of information able to describe the picture as best as possible This can be done by using features extraction . Given the Importance of features extraction in the process of classification of plant diseases, there is a lot of research that focused on the details of this subject as [7],[13],[14] etc. There are many ways to extract features, in this research the focus was on comparing the most important methods used in this topic

1- Percentage of Leaf Area Infected (PI)
2- Texture-Based Features Gray Level Co-Occurrence Matrix (GLCM)
3- Color Moments(CM)
4- Features obtained by Color Co-occurrence Method (CCM)
5- Shape based Features (SF)

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

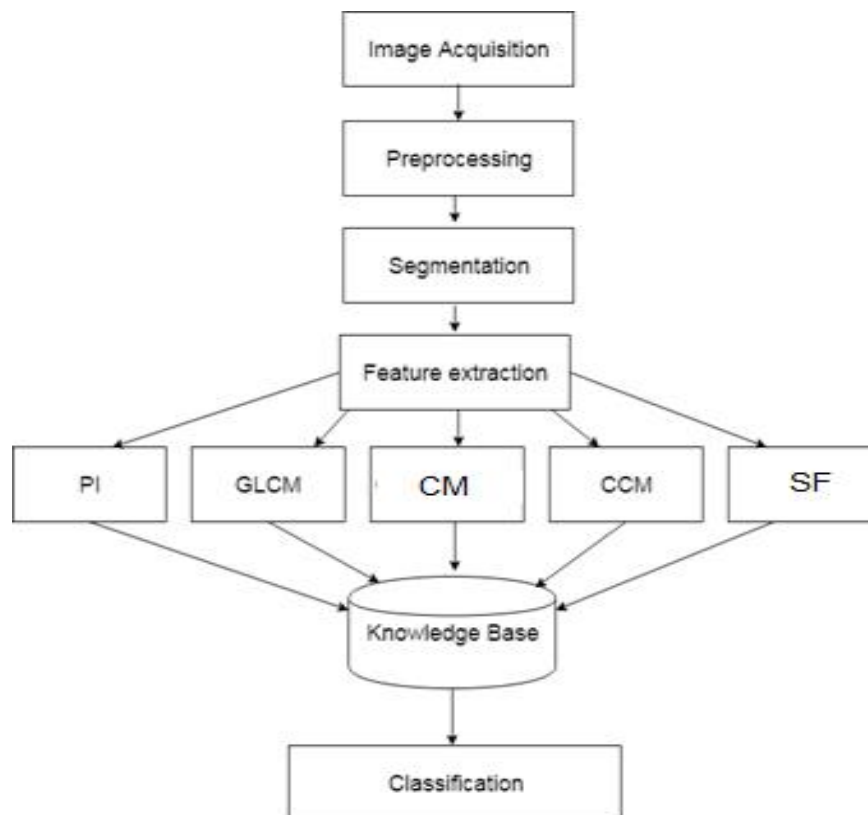The steps followed for the proposed comparison process appear in Figure 1

Figure (1). Research Block Diagram

Image Acquisition

This paper uses the most important agricultural products in Iraq (rice, cotton and tomatoes) the images were obtained from the Internet. Twelve images were used in this paper; with two images per disease. All images obtained are RGB Color space and jpg format. Two diseases were used for each of the three crop types. Figure 2 shows some examples of this image.
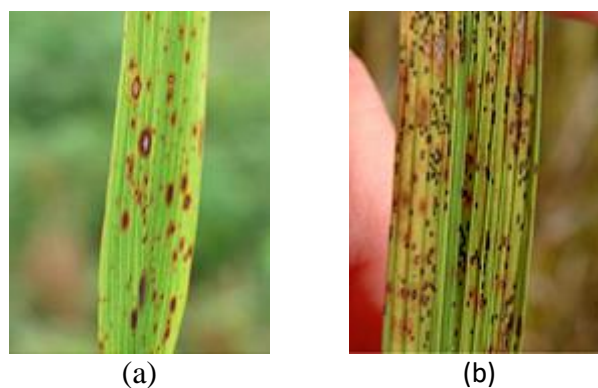
(a)   (b)

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

(c)   (d)
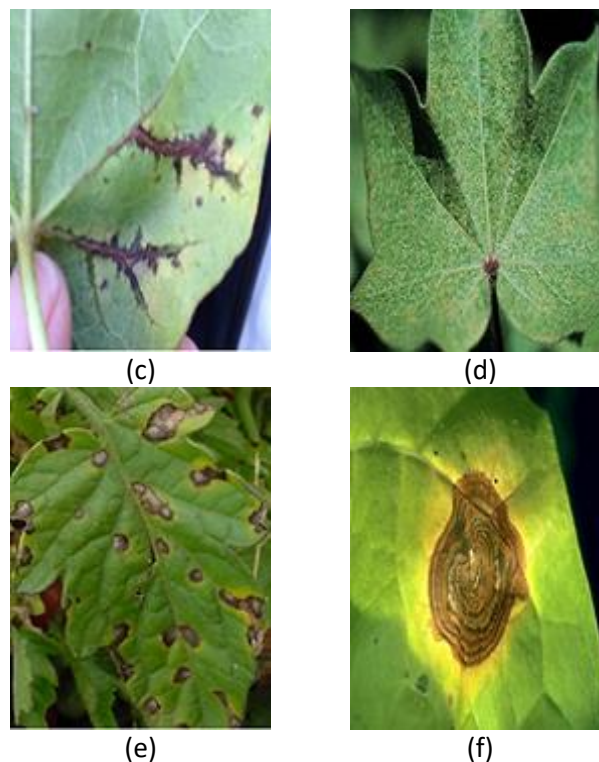
(e)   (f)

Figure.(2) (a) Rice brown spot,(b)Rice leaf smut,(c) Cotton bacterial blight,(d) Cotton spider mite ,(e) Tomato septoria leaf spot,(f) Tomato target spot

Preprocessing

In the pre-processing phase, a set of operations will be performed in order to prepare the image for the segmentation process, this operation include

3.1 Crop the image:

The captured image containing about 30% of the infected plant leaf information and the remaining 70% of the rest of the information is not important because it represents the background. This background is unnecessary consumption of memory, and also in the treatment time in the CPU during the process of retail segmentation in order to gain efficiency in the storage and speed of processing is important ,we deduct the portion of the image through a process of cropping as shown in the Figure3.

3.2 Image Resize :

Resize all images that will be used to a fixed size (300*400). This fixed size was used for all imported images because the accuracy of the feature extraction process is affected if the images are of different sizes.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

(a)   (b)

(c)   (d)

(e)   (f)

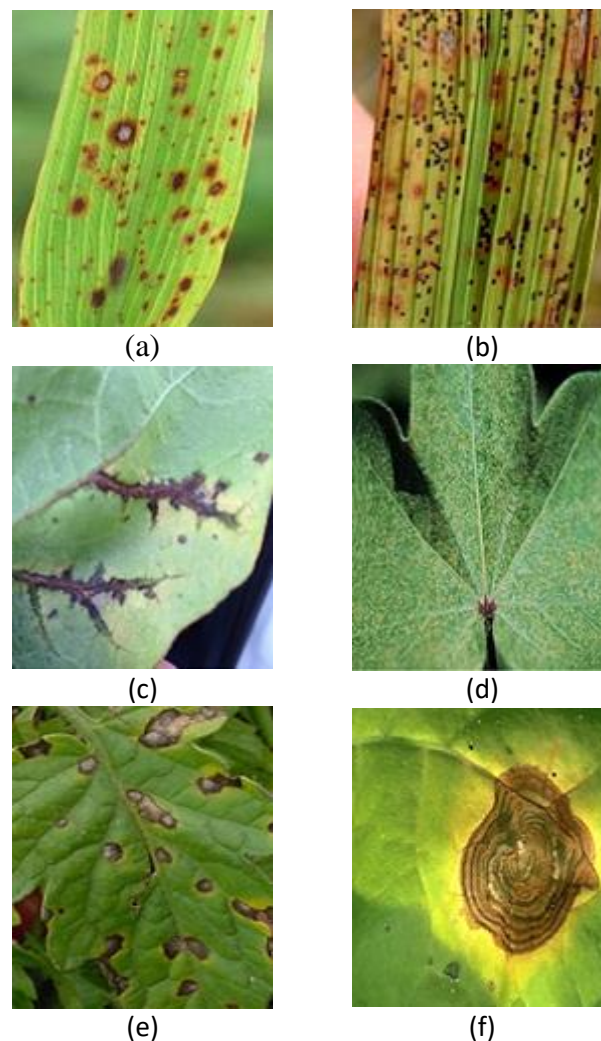**Figure.(3) ((a),(b) rice diseases after resize and crop,(c) ,(d) cotton diseases after resize and crop ,(e),(f)tomato diseases after resize and crop**

3.3 Image enhancement:

For enhance the image the paper using Fuzzy histogram equalization using the equation (1) as shown in Figure 4. Fuzzy histogram equalization (FHE) is proposed for image contrast enhancement [1]. The FHE contains two periods. First, fuzzy histogram is computed based on weird set theory to manage the inexactness of grey level values in an improved way in comparison to classical clean histograms. Second stage, the fuzzy histogram is divided into two sub histograms based on the median value of the initial image and then equalizes them independently to preserve image brightness computed as

$$F \longleftarrow h(i) + \sum_i \sum_j \mu_{F(i,j)}, \quad (1)$$

Where h(i) is a sequence of real numbers

$h(i), i \in (0, 1 \ldots, L-1),$ here $h(i)$, $\mu_{F(i,j)}$ is the fuzzy membership function

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

(a)   (b)

(c)   (d)

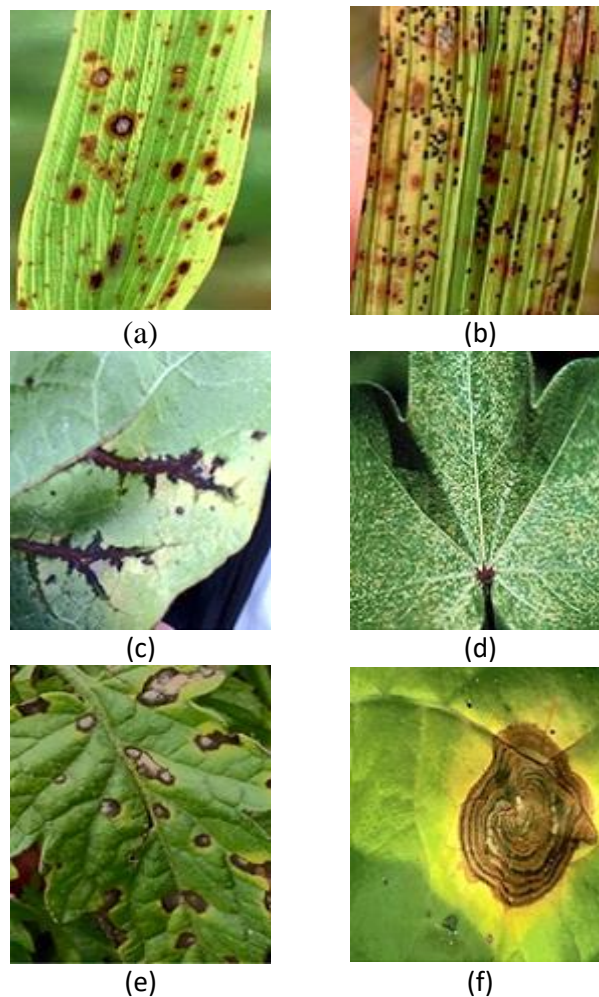(e)   (f)

Figure.(4) ((a),(b) rice diseases after FHE,(c) ,(d) cotton diseases after FHE,(e),(f) tomato diseases after FHE

Segmentation

Image segmentation is an important and fundamental process where the process of features extraction depends on the process of segmentation. The process of image segmentation   is to separate the region of interest (ROI) from the rest of the parts of the Image. this research used the color based K-means segmentation algorithm using the equation (2) as shown in the Fig 5 adopted on the color because the plant diseases can clearly distinguish them through the colors and this facilitates the process of separating the affected parts from the rest of the parts of the image. K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem [3]. K-Means Clustering Algorithm is used to segment the leaf image into one cluster [4, 6] or more than one [5]

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

(2)

Where '$\|x_i - v_j\|$' is the euclidean distance between x_i and v_j.

'C_i' is the number of data points in ith cluster.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

'C' is the number of cluster centers.

(a)   (b)

(c)   (d)

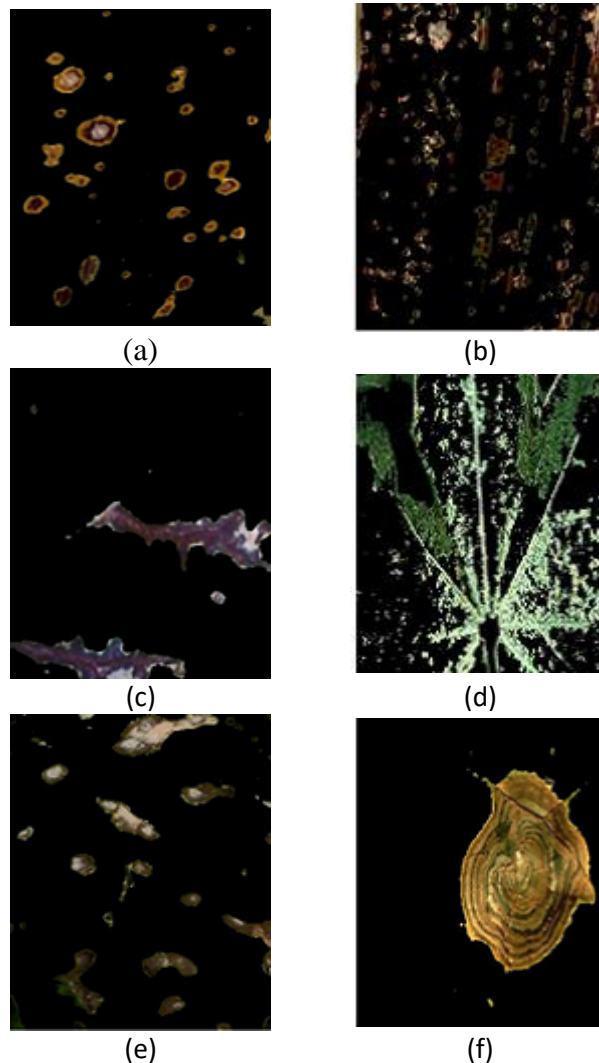(e)   (f)

Figure.5 ((a),(b) rice diseases after K-mean segmentation ,(c) ,(d) cotton diseases after K-mean segmentation,(e),(f) tomato diseases after K-mean segmentation.

Features Extraction

5.1 Percentage of Leaf Area Infected (PI):

Image Segmentation by the methods mentioned previously, characterize some pixels in the image as that of the diseased part of the leaf, as the others lay in the undispersed section. Thus we can analyze values of portion of diseased part and the full total section of the leaf . Then, we estimate ratio of area attacked, using Equation (3) as [7].

$$PI = \left(\frac{A_D}{A_L}\right) \times 100$$

(3)

Where (AD) is the diseased part, (AL) is the leaf total area.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

This value of PI has been used as one of the features in many of the previous researches we mention some of them. The following are the scientific names of the agricultural crops that have been worked on Cassava [7], Pomegranate [8] and Sugarcane [9] leaf images.

5.2 Texture-Based Features:

In case of feature extraction based on texture of the image, we used gray level co-occurrence matrix [10, 11]. The co-occurrence matrix C (i,j) counts the number of co-occurrence of pixels with gray-levels i and j respectively, at a given distance d. The matrix [12] is given by:

$$C(i,j) = cord \begin{cases} ((x_1,y_1),(x_2,y_2)) \in XY * XY \\ f(x_1,y_1) = i, f(x_2,y_2) = j \\ (x_2,y_2) = (x_1,y_1) + (d\cos\theta, d\sin\theta) \\ 0 < i,j < N \end{cases}$$

(4)

Where d is the distance defined in polar coordinates (d,) with discrete length and orientation. takes values 0, 45, 90, 135, 180, 225, 270 and 315. Cord {} represents the number of elements present in the set.

Equations (5), (6), (7) and (8) represent the features extracted from GLCM matrix [13].

Contrast feature

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2$$

(5)

Correlation feature

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2}$$

(6)

Energy feature

$$Energy = \sum_{i,j=0}^{N-1} (P_{ij})^2$$

(7)

Homogeneity feature

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$$

(8)

Where:

Pij= Element i,j of the normalized symmetrical GLCM

N   =   Number of gray levels

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

μ = the GLCM mean (being an estimate of the intensity of all pixels in the relationships that contributed to the GLCM),

$$\mu = \sum_{i,j=0}^{N-1} i P_{ij}$$

$\sigma 2$ = the variance of the intensities of all reference pixels in the relationships that contributed to the GLCM, calculated as:

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{ij} (i - \mu)^2$$

5.3 Color Moments:

In order to extract the features based on color [28], we have used the 4 color moments. (Mean, Standard Deviation, Kurtosis and Skewness) as [14]

Mean

$$Mean(\mu) = \sum_{N}^{j=1} \frac{1}{N} P_{ij} \tag{9}$$

Standard Deviation

$$Std.Deviation(\sigma) = \sqrt{(\frac{1}{N} \sum_{j=1}^{N} (P_{ij} - E_i)^2)} \tag{10}$$ Kurtosis

$$\mathbf{Kurt}[X] = \frac{\mu_4}{\sigma^4} = \frac{\mathbf{E}[(X - \mu)^4]}{(\mathbf{E}[(X - \mu)^2])^2}, \tag{11}$$

Skewness

$$Skewness(s) = \sqrt[3]{(\frac{1}{N} \sum_{j=1}^{N} (P_{ij} - E_i)^3)} \tag{12}$$

5.4 Color Co-occurrence:

using the same texture feature but not gray image its use Hue, Saturation, Value (HSV) color image H component only.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

5.5 Shape Features:

There are various processes and techniques for shape representation that are summarized in [15]. Normally shape descriptors such as the number of the area of an object, no of objects, centroid of object, perimeter are important characteristics to describe the object shape and It will be used in this research**.**

Compare Results of Features Extraction

 In this research, we used images of three types of crops, two types of diseases per crop two images for each disease. The research compares five types of feature extraction methods (PI) Table 1 ,GLCM Table 2, Color Moments Table 3, Color Co-occurrence Method Table 4 and Shape based Features Table 5.

1-   Percentage of Leaf Area Infected (PI)

Table 1. Percentage of Leaf Area Infected (PI) results

.Image NO	PI
Rice Brown Spot	
1	6.1586
2	5.0381
Rice Leaf Smut	
1	10.8805
2	17.7698
Cotton Bacterial Blight	
1	8.6376
2	8.989
Cotton Spider Mite	
1	47.1679
2	66.5836
Tomato Septoria Leaf Spot	
1	8.746
2	14.4644
Tomato Target Spot	
1	21.4469
2	10.4172

2-   Texture-Based Features

For more information about the Conference please visit the websites:
 http://www.ihsciconf.org/conf/
 www.ihsciconf.org

Table 2. Texture-Based Features (Contrast, Correlation, Energy, Homogeneity) results

Image NO.	Con.	Cor.	Ener.	Hom.
Rice Brown Spot				
1	0.08	0.89	0.86	0.98
2	0.07	0.88	0.86	0.99
Rice Leaf Smut				
1	0.24	0.8	0.72	0.95
2	0.31	0.88	0.62	0.94
Cotton Bacterial Blight				
1	0.08	0.92	0.79	0.99
2	0.18	0.9	0.84	0.98
Cotton Spider Mite				
1	2.33	0.76	0.38	0.84
2	1.6	0.86	0.29	0.84
Tomato Septoria Leaf Spot				
1	0.07	0.95	0.79	0.98
2	0.44	0.83	0.74	0.94
Tomato Target Spot				
1	0.21	0.94	0.59	0.95
2	0.06	0.97	0.82	0.98

For more information about the Conference please visit the websites:
 http://www.ihsciconf.org/conf/
www.ihsciconf.org

3- Color Moments

Table 3. Color Moments Features (Skewness, Kurtosis, StandardDeviation, Mean)

Image NO.	Skewn	Kurt	SD	Mean
Rice Brown Spot				
1	4.5139	24.0356	23.36	6.1586
2	4.5248	23.8195	19.677	5.0381
Rice Leaf Smut				
1	3.5172	16.5848	28.26	10.881
2	3.3723	15.8211	39.89	17.77
Cotton Bacterial Blight				
1	3.9337	20.4906	27.33	8.6376
2	4.6906	26.9211	34.11	8.989
Cotton Spider Mite				
1	1.3519	3.3453	72.71	47.168
2	0.7062	1.9852	79.24	66.584
Tomato Septoria Leaf Spot				
1	3.9957	19.621	28.3	8.746
2	3.5136	15.4118	39.37	14.464
Tomato Target Spot				
1	2.2133	7.09	45.86	21.447
2	4.114	20.4903	34.26	10.417

For more information about the Conference please visit the websites:
 http://www.ihsciconf.org/conf/
www.ihsciconf.org

4-Features obtained by Color Co-occurrence Method

Table(4). Features obtained by Color Co-occurrence Method for hue component only(Homogeneity, Energy, Correlation, Contrast)

Image NO.	Hom	Ener	Cor	Con
Rice Brown Spot				
1	0.956	0.71	0.81	0.757
2	0.952	0.6873	0.78	0.7752
Rice Leaf Smut				
1	0.894	0.434	0.75	1.675
2	0.88	0.332	0.75	1.982
Cotton Bacterial Blight				
1	0.979	0.749	0.94	0.395
2	0.974	0.828	0.81	0.423
Cotton Spider Mite				
1	0.89	0.297	0.69	1.405
2	0.924	0.39	0.75	0.926
Tomato Septoria Leaf Spot				
1	0.954	0.611	0.82	0.642
2	0.893	0.337	0.75	1.507
Tomato Target Spot				
1	0.975	0.799	0.81	0.33
2	0.972	0.694	0.82	0.458

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

5-Shape based Features

Table(5). Features obtained by Shape (Centroid, Number of Object,Object, Piremeter ,Area) results

Image .NO	(Cent(x,y		obj.number	Pirm	Area
Rice Brown Spot					
1	148.68	192.07	369	3269	10468
2	185.96	182.72	268	2620	9391.1
Rice Leaf Smut					
1	118.19	213.26	877	9636	23624
2	107.3	218.29	1240	14181	34355
Cotton Bacterial Blight					
1	209.29	237.28	193	2174	14195
2	162.8	234.97	341	2793	11032
Cotton Spider Mite					
1	164.82	230.72	2106	15025	53336
2	154.48	188.47	1547	12773	68006
Tomato Septoria Leaf Spot					
1	108	210.45	274	3852	15501
2	141.42	229.41	818	10266	27024
Tomato Target Spot					
1	150.77	264.45	205	1533	29170
2	186.76	164.62	110	3112	21604

For more information about the Conference please visit the websites:
 http://www.ihsciconf.org/conf/
www.ihsciconf.org

Discussion of Results

Results were discussed in each method based on the number of features that give good results and we mean the good results of those features that give a clear difference between the two diseases for the same crop if the method gives good results in all the features in both two images say this way is 100% good, but if the way gives good results in three features out of four we say this method is good by 75% and so on.

7.1 Rice disease

1. The PI feature gave us good results by **100%**

2. The GLCM Features gave us good results by **75%**
3. The Color Moment Features gave us good results by **75%**
4. Color Co-occurrence Features gave us good results by **50%**
5. The Shape Features gave us good results by **75%**

7.2 Cotton disease

1. The PI feature gave us good results by **100%**
2. The GLCM Features gave us good results by **75%**
3. The Color Moment Features gave us good results by **100%**
4. Color Co-occurrence Features gave us good results by **50%**
5. The Shape Features gave us good results by **75%**

7.3 Tomato disease

1. The PI feature gave us good results by **100%**
2. The GLCM Features gave us good results by **50%**
3. The Color Moment Features gave us good results by **50%**
4. Color Co-occurrence Features gave us good results by **50%**
5. The Shape Features gave us good results by **75%**

7.4 All diseases

In order to calculate the accuracy of this method in all diseases in general, we collect their proportions and divide them into three (number of crops)
Results in all diseases were as follows PI (100%) in all disease, GLCM (66.6%) ,color moment (66.6%), 4.   Color Co-occurrence (50%), Shape Features (75%)

Conclusion

In this research found the use of the following mix of features PI, GLCM, color moment, shape features give excellent results in all types of plants that were taken in this research.

Future Work

Comparisons between other methods have never been used in this type of research.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

References

[1]. V. Magudeeswaran and Ravichandran C,(2013)," Fuzzy Logic-Based Histogram Equalization for Image Contrast Enhancement", Mathematical Problems in Engineering, 013 (2013), Article ID 891864, 10 pages.

[2]. Sheet D, Garud.H, Suveer A, Chatterjee J and Mahadevappa M, (2010), "Brightness Preserving Dynamic Fuzzy Histogram Equalization", IEEE Trans Consumer Electronics, 56, no. 4, 2475 – 2480.

[3]. Majumdar D, Kumar D, Chakraborty A and Dutta D,(2014)," DETECTION & DIAGNOSIS OF PLANT LEAF DISEASE USING INTEGRATED IMAGE PROCESSING APPROACH ", International Journal of Computer Engineering and Applications, VI –III.

[4] S, Sannaki, V. Rajpurohit, V. Nargund, A. Kumar ,and P, Yallur (2010), " Leaf Disease Grading by Machine Vision and Fuzzy Logic" . Int. J. Comp. Tech. Appl., 2 (5), 1709-1716:2229-609.

[5]. Bashish D., Braik M., Bani-Ahmad S,(2010), "A Framework for Detection and Classification of Plant Leaf and Stem Diseases" IEEE.

[6]. M. Badnake, P. Deshmukh,(2012), "Infected Leaf Analysis and Comparison by Otsu Threshold and k-means Clustering", International Journal of Advanced Research in Computer Science and Software Engineering, 2, 3.

[7] K. Powbunthorn, W. Abudullakasim, Unartngam J,(2012), "Assessment of the severity of Brown Leaf Spot Disease in Cassava using Image Analysis" , The International conference of the Thai Society of Agricultural Engineering.

[8] S. Sannaki, V. Rajpurohit, Nargund, V. Kumar A, Yallur P ,(2010) , "Leaf Disease Grading by Machine Vision and Fuzzy Logic" . Int. J. Comp. Tech. Appl., 2 (5) 1709-1716:2229-609.

[9]. S.Patil , S. Bodhe ,(2011), "LEAF DISEASE SEVERITY MEASUREMENT USING IMAGE PROCESSING". International Journal of Engineering and Technology ,.3 (5), 297-301.

[10]. D. Majumder, B. Chanda,(2007), "Digital Image Processing and Analysis", Prentince Hall of India Private Limited.

[11]. R, Haralick. K. Shanmugam. and Dinstein.I, (1973),"Textual features for image classification", IEEE Trans. Syst. Man Cybern., 3,.6, 610–621, Nov. 1973.

[12]. Zhang, Zhang F,(2008) ,"Congress on Image and Signal Processing", IEEE computer society, 773-776.

[13]. K. Xu and Y. Li. ,(2005),"An image search approach based on local main color feature and texture feature", Journal of Xi'an Shiyou University (Natural Science Edition), 20 (2): 77 - 79.

[14]. Yu H, Li M, Zhang H, Feng J,(2002), "Color Texture Moments for Content-Based Image Retrieval", Proc. IEEE Intl Conf. on Image Processing, pp. 929-932.

[15]. S.Patil, S. Bodhe,(2011) ,"LEAF DISEASE SEVERITY MEASUREMENT USING IMAGE PROCESSING",International Journal of Engineering and Technology,.3 (5), 297-301.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org

[16]. M.Saade,(2012), "Iraq Agriculture sector note" ,Food and Agriculture Organization of the United Nations (FAO).

[17]. S. Pramod, A. Sushil, Dhanashree S, Omkar D, Utkarsha G,(2013)," Automatic Detection and Classification of Plant Disease through Image Processing",International Journal of Advanced Research in Computer Science and Software Engineering, 3, 7.

For more information about the Conference please visit the websites:
http://www.ihsciconf.org/conf/
www.ihsciconf.org