

A PSO MODEL FOR DISEASE PATTERN DETECTION ON LEAF SURFACES

KANTHAN MUTHUKANNAN^{✉,1} AND PITCHAI LATHA²

¹Department of ECE, Einstein College of Engineering, Tamilnadu, India; ²Department of CSE, Government College of Engineering, Tirunelveli, Tamilnadu, India

e-mail: maha_muthukannan@yahoo.com, plathamuthuraj@gmail.com

(Received October 3, 2014; revised January 16, 2015; revised April 24, 2015; accepted May 8, 2015)

ABSTRACT

The main objective of this paper is to segment the disease affected portion of a plant leaf and extract the hybrid features for better classification of different disease patterns. A new approach named as Particle Swarm Optimization (PSO) is proposed for image segmentation. PSO is an automatic unsupervised efficient algorithm which is used for better segmentation and better feature extraction. Features extracted after segmentation are important for disease classification so that the hybrid feature extraction components controls the accuracy of classification for different diseases. The approach named as Hybrid Feature Extraction (HFE), which has three components namely color, texture and shape based features. The performance of the preprocessing result was compared and the best result was taken for image segmentation using PSO. Then the hybrid feature parameters were extracted from the gray level co-occurrence matrices of different leaves. The proposed method was tested on different images of disease affected leaves, and the experimental results exhibit its effectiveness.

Keywords: disease pattern, particle swarm optimization, performance analysis, segmentation, texture feature.

INTRODUCTION

Many diseases affect the plant and its leaf all over the world, including India which reduces the production of food. It causes a significant impact on rice quality and yield. India is an agricultural country wherein most of the population depends on agriculture and is one of the major domains which decides economy of the nation. The quality and quantity of the agricultural production is affected by environmental parameters like rain, temperature and other weather parameters which are beyond control of human beings. In addition to environmental parameters like rain and temperature, diseases on crop are a major factor which affects production quality and quantity of crop yield. Hence disease management is key issue in agriculture. For management of disease, it needs to be detected at earlier stage so as to treat it properly and control its spread. Because of advances in the technologies, nowadays it is possible to use the images of diseased leaf to detect the particular type of disease. This can be achieved by extracting features from the images which can be further used with classification algorithms or content based image retrieval systems stated by Chaudhary *et al.* (2012).

Nowadays almost all of these tasks are processed manually or with distinct software packages. It is not only tremendous amount of work but also suffers from two majors, firstly extreme computation times and secondly, subjectiveness expanding from different individuals. Hence to conduct high throughput experiments, plant biologists need an efficient computer software to automatically extract and analyze significant features stated by Jabal *et al.* (2013). Kadir *et al.* (2011), proposed as far as the leaf of the plant is considered, the significant features can be obtained by color, texture and shape of the leaf stated by

Gurjar and Gulhane (2011) describe Eigen feature regularization and extraction technique in which three different diseases are considered for detection. This system has more accuracy, when compared to other feature detection techniques. With this method about 90% of detection of red spot, *i.e.*, fungal disease which affects the plant leaf is detected.

Al Bashish *et al.* (2010), proposed an image processing based work which consists of the following main steps: In the first step the acquired images are segmented using the K-means techniques and secondly the segmented images are passed through a pre-trai-

ned neural network. In this paper, diagnosis system for grape leaf diseases is proposed and it is composed of three main parts: Firstly grape leaf color extraction from complex background, secondly grape leaf disease color extraction and finally the classification of grape leaf disease. Eventhough there are some limitations like extracting ambiguous color pixels from the background of the image, the system demonstrates very promising performance for any agricultural product analysis.

Phadikar *et al.* (2013), developed an automated classification system based on the morphological changes caused by brown spot and the leaf blast diseases of rice plant. To classify the diseases radial distribution of the hue from the center to the boundary of the spot images has been used as feature by using Bayes' and SVM Classifier. The feature extraction for classification of rice leaf diseases is processed in the following steps: Firstly the image is acquired from the fields of diseased rice leaves. Secondly preprocess the images to remove noise from the damaged leaf and then enhance the quality of image by using the mean filtering technique. Thirdly Otsu's segmentation algorithm was applied to extract the infected portion of the image, and then radial hue distribution vectors of the segmented regions computed which are used as feature vectors. Here classification is performed in two different phases. In the first phase an uninfected and the diseased leaves are classified based on the number of peaks in the histogram. In the second phase the leaf diseases are classified by Bayes' classifier. This system gives 68.1% and 79.5% accuracies for SVM and Bayes'-classifier based system respectively.

An automated image segmentation system that can detect diseases and extract the features present in a leaf depends only on the color image. The problems of image segmentation and grouping remain great challenges for computer vision. Gonzalez and Woods (2007) stated that color is one of the most widely used features. Color features can be obtained by various methods like color histogram, color correlogram, color R moment and color structure descriptor. The color moment method has the lowest feature

vector dimension and lower computational complexity stated by Ford and Roberts (1998). Hence, it can be considered as suitable parameter to generate feature vectors which can be used further for classification purpose.

This paper is organized in a way that the below section presents the description of the proposed methods. The next section deals with a new segmentation technique PSO, follows with the feature extraction of the segmented leaf and the results for different disease patterns of the leaf. Finally, this paper summarizes the conclusions for this research work

DESCRIPTION OF THE METHODS

The proposed method can be described in Fig. 1. In detail, the disease affected image is acquired from the environment. Secondly, the image is resized and it is filtered by using Gaussian filter. Resizing is used to improve the performance of the upcoming processing, *i.e.*, it reduces the computational time. Then, the disease affected portion is segmented using particle swarm optimization based color image segmentation technique. Next it creates the co-occurrence matrix for the segmented area of the affected portion of the leaf and at last it extracts the hybrid features analyzing the disease features for different disease pattern for the different leaf.

IMAGE DATA COLLECTION

Many diseases and disorders can affect most probably all plant leafs during its growth. The most common diseases which are considered in this paper are late blight, septoria leaf spot, down mildew, blast and rust teak. The image acquisition process also identifies whether the acquired image is affected or not by calculating the intensity values of the leaf spot using 'improfile' command. Improfile computes the intensity values along a line or a multiline path in an image. Improfile selects equally spaced points along the specified path, and then uses interpolation to find the intensity value for each point. Improfile works with grayscale images and RGB images.

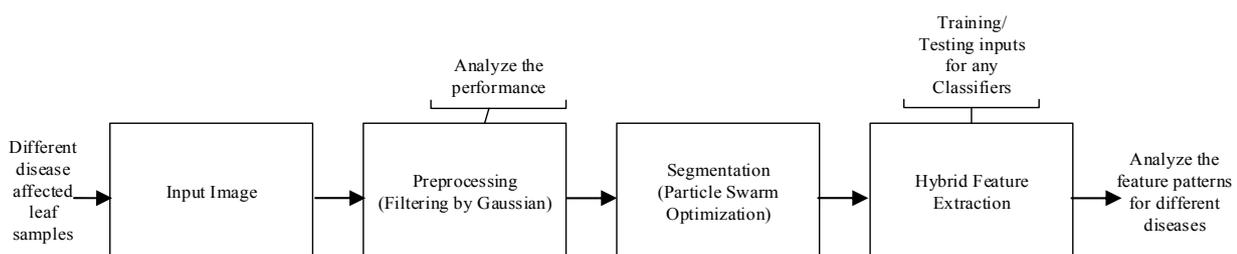


Fig. 1. Proposed Description Methods-PSO based segmentation and hybrid feature extraction for different diseases.

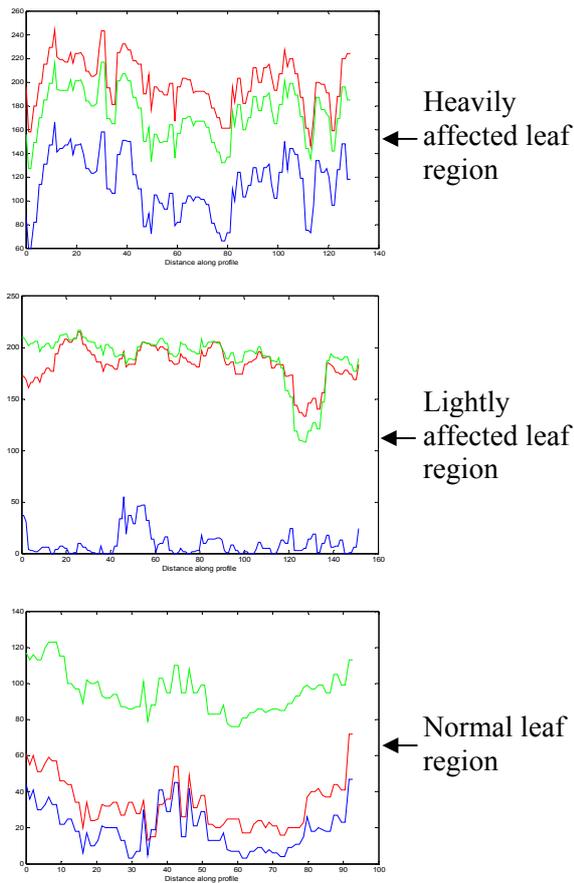


Fig. 2. Comparison of the Intensity values of different affected leaf portion in RGB plane.

The Fig. 2 shows the intensity values of the different affected portion in leaf surfaces. Based on the intensity values, it can be classified into three such as heavily affected region, lightly affected region and normal leaf region. Table 1 shows the intensity values of different color planes of disease affected portion of the leaf.

The intensity values of red color varies from 140 to 240 for the affected leaf region, the intensity values for green color varies from 120 to 220 for affected leaf region and intensity value for blue color is from 50 to 160.

Table 1. Intensity values of different color planes for leaf disease.

Items	Red	Green	Blue
Heavily affected	140-240	120-220	50-160
Lightly affected	140-210	100-210	0-50
Normal	20-80	90-120	5-50

IMAGE FILTERING

Anami *et al.* (2011) stated that the images obtained during the image acquisition step may not be suitable for further image processing steps because of lighting variations, noise, poor resolution, unwanted background, climatic conditions, etc. The noise is inevitable during transmission or any other processing steps. So, it must be removed before any further image analysis steps. $H = \text{FSPECIAL}(\text{TYPE})$ creates a two-dimensional filter ‘H’ of the specified type. Woering (2009) stated that the possible filter types are: averaging filter, circular averaging filter Gaussian low pass filter, Laplacian filter, Laplacian operator; similarly: Gaussian filter, motion filter, Prewitt horizontal edge-emphasizing filter, Sobel horizontal edge-emphasizing filter and unsharp contrast enhancement filter.

In this work, the FSPECIAL (Gaussian) filter is used to remove the unwanted information of speckles. Depending on the type of filter the parameter value may vary.

$$h_g(n_1, n_2) = e^{-\frac{n_1^2 + n_2^2}{2\sigma^2}}, \quad (1)$$

$$h(n_1, n_2) = \frac{h_g(n_1, n_2)}{\sum n_1 \sum n_2 h_g}, \quad (2)$$

Eq. 2 is used to select the size of the filter mask using Eq. 1. The size of Gaussian low pass filter is denoted by HSIZE with standard deviation SIGMA (positive). HSIZE can be a vector specifying the number of rows and columns in H or a scalar, in which case H is a square matrix. The default HSIZE is [3 3], the default SIGMA is 0.5 n.

Table 2. Comparison of PSO based segmentation performance with different parameters.

PSNR		MAX. ERROR	
Before filtering	After filtering	Before filtering	After filtering
16.7783	29.60528	144	77
15.91573	29.31211	139	65
15.66395	30.91939	147	75
18.53243	28.65575	148	111
17.74416	27.02016	138	106
18.39126	30.31561	137	96
17.74502	28.00379	133	77
20.4242	28.89249	103	62
18.95516	30.19277	115	70
16.16058	30.29904	145	50
20.94817	41.83114	105	40
17.74612	29.94361	147	49
16.91681	28.9304	141	74
17.321	27.885	116	88
17.3826	30.06132	141	77
16.60841	29.05485	175	93
17.60873	33.30894	122	72
20.85571	50.09587	113	9
21.88938	43.9068	91	13
19.51193	31.35436	116	78
17.48875	29.27354	126	81
16.26635	29.70152	162	68
18.05872	30.86602	126	79
16.38633	27.36198	158	78

Table 2 shows the importance of preprocessing using filters in which the filtering performance values are clearly mentioned for before filtering and after the process of filtering.

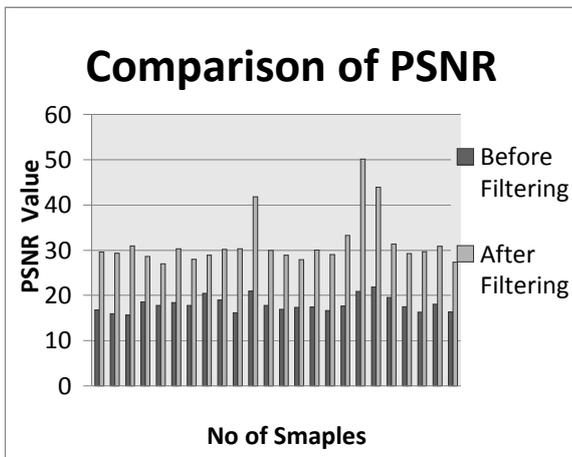


Fig. 3. Comparison of PSNR before and after filtering for different disease affected leaf samples.

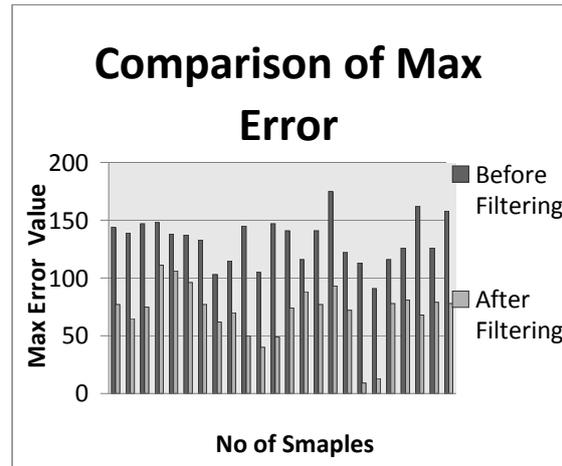


Fig. 4. Comparison of maximum Error before and after filtering for different disease affected leaf samples.

Fig. 3 and Fig. 4 shows the filtering performance for different disease affected leaf samples. Here, Gaussian filter gives much improved performance which is analyzed by PSNR and maximum error value. The plot also shows the comparison of the individual performance of the PSNR and max error and these results are very much helpful for the feature extraction and disease classification in the affected leaf image.

PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Eberhart and Kennedy (1995), inspired by social behavior of bird flocking or fish schooling. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. In every iteration, each particle is updated by following two "best" values. The first one is the best fitness solution and it stores the fitness value. This value is called "pbest". The next "best" value that is tracked by the particle swarm optimizer obtained by any particle in the population. This best value is a global best and called as "gbest". When a particle takes part in the population as its topological neighbors, the best value is a local best and is called as "Lbest". After finding the two best values, the particle updates its velocity and positions with the following equations.

$$v[] = v[] + c1 * rand() * (pbest[] - present[]) + c2 * rand() * (gbest[] - present[]), \quad (3)$$

$$present[] = present[] + v[], \quad (4)$$

$v[]$ is the particle velocity, $present[]$ is the current particle (solution). $pbest[]$ and $gbest[]$ are defined as stated before. $rand()$ is a random number between (0,1). $c1, c2$ are learning factors.

The pseudocode procedure for PSO is as follows:

```

For each particle
Initialize particle
END
Do
For each particle
Calculate fitness value
If the fitness value is greater than the best
    fitness value in history
Set the current value as pbest
END

Choose the particle with the best fitness value of
all the particles as the gbest
For each particle
Calculate particle velocity (eqn.3)
Update particle position (eqn.4)
END

While maximum iterations or minimum error
criteria is not attained
    
```

While maximum iterations or minimum error criteria is not attained

Initially, the algorithm partitions the data set into comparatively large number of clusters in order to reduce the complexity in the initial conditions. The binary PSO helps to select the best number of clusters and the centers of the chosen clusters are refined by k-means clustering. Clustering is an unsupervised technique for image segmentation. The PSO is used for assigning each pixel to a cluster. One of the advantage of the proposed method is that user can choose any validity index according to the given data. The fitness function used here is quantitative evaluation function. It deals with segmented images as a set of regions, the target image is divided into a set of regions and not to a set of classes. The PSO parameters are initially set as follows, $V_{min} = -5$, $V_{max} = 5$, population size $N = 150$, $c1 = 0.8$, $c2 = 0.8$ and $w = 1.2$. Here, the index value is set as $d = 4$ for all images.

PARAMETER FOR DISEASE PATTERN ON LEAF SAMPLES

Most of the gray level co-occurrence matrix (GLCM) texture calculations used in remote sensing was syste-

matized in a series of papers by Robert Haralick in the 1970's. When the GLCM is generated, there are a total of 14 texture features that could be computed from the GLCM, such as contrast, variance, sum average, etc. The four common texture features discussed here are contrast, correlation, energy, and homogeneity. Contrast is used to measure the local variations. Correlation is used to measure probability of occurrence for a pair of specific pixels. Energy is also known as uniformity of ASM (angular second moment) which is the sum of squared elements from the GLCM. Homogeneity is to measure the distribution of elements in the GLCM with respect to the diagonal.

The direction for generation of GLCM is shown in Fig. 5 and it is a two-dimensional histogram in which the (i,j) th element is the frequency of event i co-occurs with event j . A co-occurrence matrix is specified by the relative frequencies $P(i, j, d, \theta)$ in which two pixels, separated by distance d , occur in a direction specified by the angle θ , one with gray level i and the other with gray level j . A co-occurrence matrix is therefore a function of distance d , angle θ and grayscales i and j .

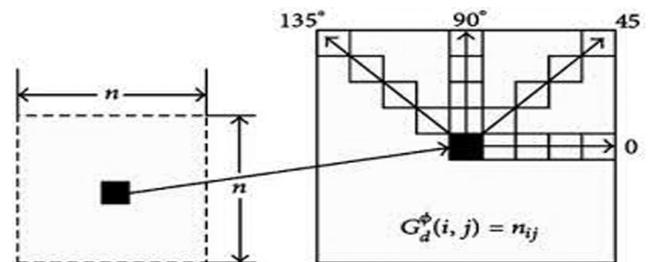


Fig. 5. Direction for generation of GLCM.

Kadir *et al.* (2012) stated that the importance of features in leaf classification. In this research, there are four parametric features extracted based on color, texture and shape of the leaf disease region by using the following mathematical relations.

- 1) Contrast returns a measure of the intensity. The range of contrast is between a pixel and its neighbor over the entire image.

Range = $[0, (size(G,1)-1)^2]$. Contrast is '0' for constant image. Where, G is the co-occurrence matrix as given below.

$$Contrast = \sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \tag{5}$$

- 2) Correlation returns a measure by which how a correlated pixel is its neighbor over the entire image. Range = $[-1, 1]$. Correlation is '1' for a perfectly positively correlated image and '-1' for

a negatively correlated image. If correlation is not a number (NaN), then the image is a constant image. Correlation is computed as follows,

$$Correlation = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - m_r)(j - m_c)P_{ij}}{\sigma_r \sigma_c} \right]. \quad (6)$$

- 3) Energy is the sum of squared elements in G. Range = [0 1]. Energy is 1 for a constant image. Energy is computed as follows,

$$Energy = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j}^2}. \quad (7)$$

- 4) Homogeneity- A value that measures the closeness of the distribution of elements in the G to the diagonal of G. Range = [0 1]. Homogeneity is 1 for a diagonal G. Homogeneity is computed as follows,

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{(1 + |i - j|)} \quad (8)$$

RESULTS

This section deals the results and statistical analysis of the experiment done based on the proposed methods. Also, it shows the different patterns of the various diseases which affect the different plant leaves on a 256x256 pixels image. Fig. 6 shows the segmented results of late blight disease pattern in a tomato leaf, Fig. 7 shows the blast disease in a rice leaf and Fig. 8 shows the late blight in a potato leaf respectively.

From the segmented results, the particular disease affected portion in the leaf can be easily identified. Here, the cluster validity index chosen is 4. Similarly, the results were taken and analyzed for $d = 2$ and $d = 3$ for different diseases such as blast and late blight. Table 3 shows the features of different diseases and its average value for four samples. The average value of late blight disease for contrast, homogeneity, energy and correlation is 0.146844, 0.975526, 0.49948 and 0.920768 respectively. Similarly, the average value of rust teak disease for contrast, homogeneity, energy and correlation is 0.266736, 0.977563, 0.645706 and 0.921863 respectively. From these results, the feature varies from one particular disease to another disease.

So, these results are helpful for better classification. Fig. 9 shows the comparison plot for different disease based on the hybrid features.

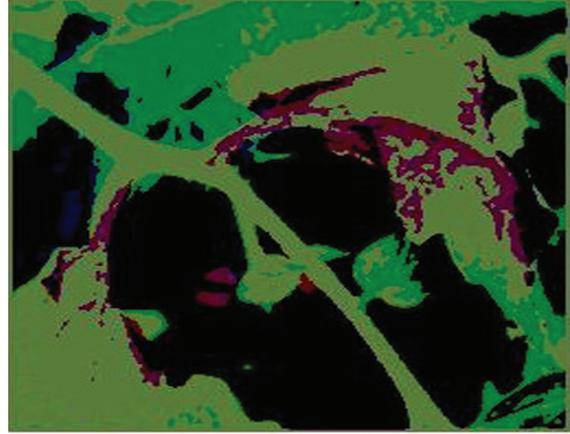


Fig. 6. Late blight disease pattern on tomato leaf image.

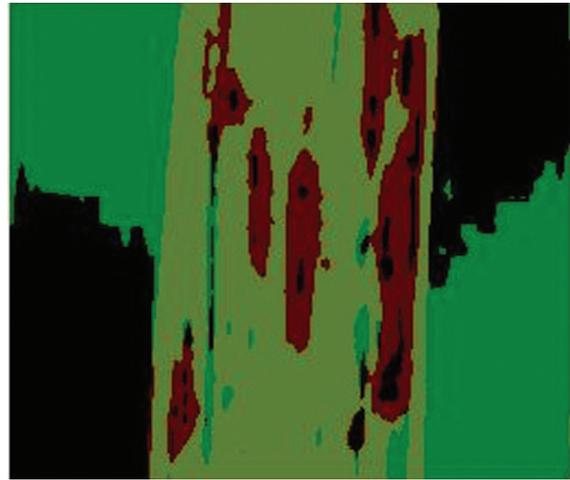


Fig. 7. Blast disease pattern on rice leaf image.



Fig. 8. Late blight disease pattern on potato leaf image.

Table 3. Extraction of different features after PSO based segmentation of disease affected leafs for the distance $d = 4$.

Name of the Disease	Contrast	Homogeneity	Energy	Correlation
Late Blight Tomato	0.187990	0.968668	0.523607	0.891110
Average	0.146844	0.975526	0.499480	0.920768
SPL	0.302206	0.974816	0.488743	0.929871
Average	0.236581	0.980285	0.678468	0.911174
Late Blight Potato	0.576287	0.951976	0.443158	0.871140
Average	0.500735	0.958272	0.466228	0.884385
Late Blight Potato	0.316544	0.973621	0.466104	0.929651
Average	0.567739	0.952688	0.464151	0.867682
Downy Mildew	0.478125	0.960156	0.471642	0.888873
Average	1.919363	0.904032	0.394852	0.759877
Downy Mildew	0.820443	0.947624	0.449187	0.861521
Average	0.489568	0.959203	0.496823	0.879585
Downy Mildew	0.179963	0.970006	0.475451	0.906577
Average	0.486949	0.959421	0.552597	0.863448
Blast	1.039706	0.948015	0.439261	0.870031
Average	0.549046	0.959161	0.491033	0.879910
Blast	0.235616	0.980365	0.531000	0.940971
Average	0.189430	0.984214	0.671295	0.931685
Rust teak	0.297794	0.975184	0.475340	0.932838
Average	0.039093	0.993484	0.798113	0.949153
Rust teak	0.190483	0.983312	0.618937	0.938662
Average	0.010049	0.998325	0.832645	0.984760
Rust teak	0.470680	0.960777	0.457642	0.893875
Average	0.157721	0.986857	0.756614	0.922516
Rust teak	0.428493	0.964292	0.535924	0.886299
Average	0.266736	0.977563	0.645706	0.921863

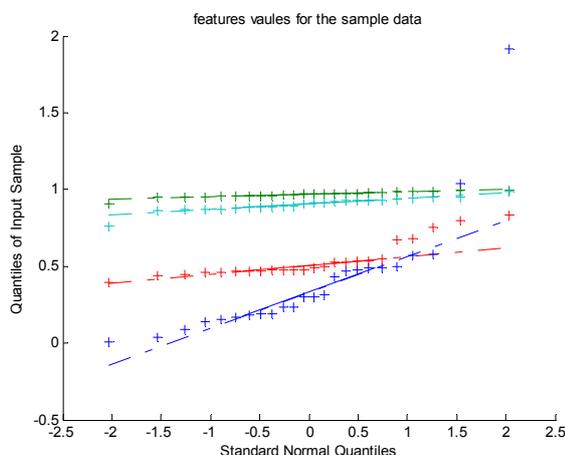


Fig. 9. Analysis of features for different disease patterns.

Fig. 9 gives the excellent results for different disease patterns classification using hybrid feature extraction. Also, it may be possible for accurate classification of plant leaf diseases infected by virus, bacteria and fungus.

CONCLUSION

In this paper, PSO based image segmentation was considered for disease affected portion segmentation. Then, the hybrid features were extracted from the affected portion by means of gray level co-occurrence matrices of different diseases. Before the process of segmentation and feature extraction the preprocessing was done and the results were analyzed with the parameters such as PSNR and Maxerror. In this implementation the texture features such as contrast, correlation,

homogeneity and energy was considered and analyzed for different disease affected plant leaves. From the analysis, the hybrid feature extraction approach is helpful for plant leaf disease classification in terms of minimization of misclassification and also improves the correct prediction classification accuracy. Finally, the proposed hybrid feature extraction approach may be suitable for better classification for various diseases affecting the plant leaf.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable suggestions which help in improving the quality of this paper. We would like to thank the International Rice research Institute and Tamilnadu Agricultural University. We would like to thank Dr. K. Ramar and Prof. A. Amudavanan for his continuous encouragement and support. We would like to thank Dr. Meg McGrath for the use of his images.

REFERENCES

- Al Bashish D, Braik M, Bani-Ahmad S (2010). A framework for detection and classification of plant leaf and stem diseases. In: ICSIP 2010. Proceedings of the International Conference on Signal and Image Processing 2010, Dec 15-17; Chennai, India, 118–33.
- Anami BS, Pujari JD, Yakkundimath R (2011). Identification and classification of normal and affected agriculture/horticulture produce based on combined color and texture feature extraction. *IJCAES* 1:356–60.
- Chaudhary P, Chaudhari AK, Cheeran AN, Godara S (2012). Color transform based approach for disease spot detection on plant leaf. *IJCST* 2:65–70.
- Eberhart RC, Kennedy J (1995). A new optimizer using particle swarm theory. In: ISMMHS 1995. Proceedings of the Sixth International Symposium on Micro Machine and Human Science 1995, Nov/Dec 27-1; Nagoya, Japan, 39–43.
- Ford A, Roberts A (1998). Colour space conversions. Technical report, Poynton, August 11, 9–17.
- Gonzalez RC, Woods RE (2007). Digital image processing (second edition). New Jersey: Prentice Hall, 39–42.
- Gurjar AA, Gulhane VA (2011). Disease detection on cotton leaves by eigen feature regularization and extraction technique. *IJECSCSE* 1:1-4.
- Jabal MFAB, Hamid S, Shuib S, Ahmad I (2013). Leaf features extraction and recognition approaches to classify plant. *JCS* 9:1295–304.
- Kadir A, Nugroho LE, Susanto A, Santosa PI (2011). Leaf classification using shape, color and texture features. *IJCTT* July to August:225–30.
- Kadir A, Nugroho LE, Susanto A, Santosa PI (2012). Performance improvement of leaf identification system using principal component system. *IJAST* 44:113–24.
- Phadikar S, Sil J, Das AK (2013). Rice diseases classification using feature selection and rule generation techniques. *IJCEA* 90:76–85.
- Woering R (2009). Design of a video processing algorithm for detection of a soccer ball with arbitrary color pattern. Traineeship report. Eindhoven, 7–11.