Classification of Rice Leaf Diseases Based on Morphological Changes

S. Phadikar, J. Sil, and A. K. Das

***Abstract*—Rice is widely cultivated economical crop in the world. During cultivation the earliest and accurate diagnosis of the rice plant diseases able to reduce the damage, resulting environment protection and better return. In the work, an automated system has been developed to classify the leaf brown spot and the leaf blast diseases of rice plant based on the morphological changes of the plants caused by the diseases. Radial distribution of the hue from the center to the boundary of the spot images has been used as features to classify the diseases by Bayes' and SVM Classifier. The system has been validated using 1000 test spot images of infected rice leaves collected from the field, gives 79.5% and 68.1% accuracies for Bayes' and SVM Classifier based system respectively.**

***Index Terms*—Rice leaf disease, Bayes classifier, leaf blast, brown spot, support vector machine, radial hue distribution.**

I. Introduction

Rice (Oryzae Sativa) is considered as the main crop in the east India and believed to be the second central crop after wheat, in the world. Rice blast caused by the fungus Pricularia Grisea [1] - [3] one of the major rice diseases occur in most of the rice growing areas in the world. It causes severe problem, especially in the areas where the resistant varieties are not available and the environment condition is favorable towards out brake of the disease. The damages caused by the Blast, depend on the degree of severity of the disease. Another important rice disease Leaf brown spot caused by the fungus Bypolaris Oryza [1], [2], [4] is visible throughout the growing season. Brown spots generally become severe when rice is grown on silicon-deficient soils.

New trends of research in agriculture aim towards the development of the disease resistant variant of seed using gene technology, which increase productivity and food quality at reduced expenditure [4]. With the advancement of information technology, remote sensing techniques have been used in the field of crop management described in [5] – [7]. A relation among the ground disease index and remote sensing data is established in [8] to classify the diseases. Very recently, image processing techniques are applied to detect the rice diseases automatically [9] In the paper, an automated system has been designed using image processing and pattern recognition techniques to distinguish the leaf blast and the brown spot diseases based on the morphological changes caused by the diseases. The accuracy of the proposed system is compared with the existing classifier, which shows comparable results.

The paper, is organized in four Section II describes the process of image acquisition and feature extraction. Section III describes the classification process. Section IV represents results and discussion and the paper is concluded in section V.

Manuscript received February 24, 2012; revised April 26, 2012.

S. Phadikar is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@ boulder.nist.gov).

J. Sil was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar. colostate.edu).

A. K. Das is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

II. Image acquisition and Features Extraction

The images are collected from different parts of East Midnapur, a major rice producing district of South Bengal, India and the features for classification of the rice leaf diseases has been extracted through execution of the following steps:-

a) Acquire the images of the rice leaf of infected diseased plants from the field.
b) Preprocess the images to remove noise from the infected portion of the leaf image.
c) For a diseased leaf, execute steps (d), (e)
d) Extract the infected portion of the image by segmentation.
e) Compute radial hue distribution vectors of the segmented regions, which are used as feature vectors.

A. Image Acquisition

The images are grabbed using Nikon COOLPIX P4 digital camera in macro mode. The samples of the leaves affected by Diseases (blast, brown spot) and the uninfected rice leaves are shown in Fig 1.

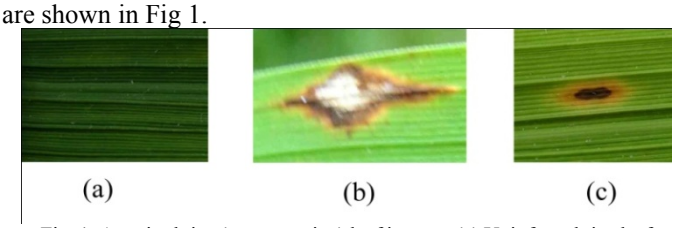

Fig. 1. Acquired rice (oryzae sativa) leaf images (a) Uninfected rice leaf (b) Rice leaf infected by blast(pricularia grisea) (c) Rice leaf infected by brown spot (bypolaris oryza)

B. Preprocessing

The automated system first classifies the acquired image either as an uninfected or a diseased leaf. For an uninfected leaf the distribution of color is nearly uniform while for the diseased leaf the distribution is not uniform because the pixel values of the diseased leaves varies widely and differs from the normal intensity value of pixels. In order to distinguish between the uninfected and diseased leaves, the aim of the

preprocessing step is to reduce the difference among the intensity value of pixels of the uninfected image while increase the difference among the pixels of the diseased image. The processed gray image has been obtained using (1)

$$MG=2\times(G-(0.75\times R-0.25\times B)) \quad (1)$$

where, R, G and B are Red, Green and Blue plane respectively in the RGB color image. The processed images are shown in Fig. 2.

The quality of the image is enhanced further applying the[mean filtering technique [10] and the resultant images are shown in Fig. 3.

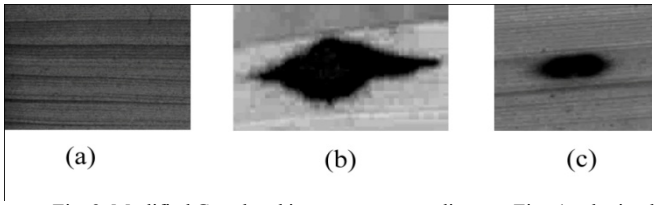

Fig. 2. Modified Gray level images corresponding to Fig. 1. obtained by applying the equation 2*(G-(0.75*R-0.25*B))

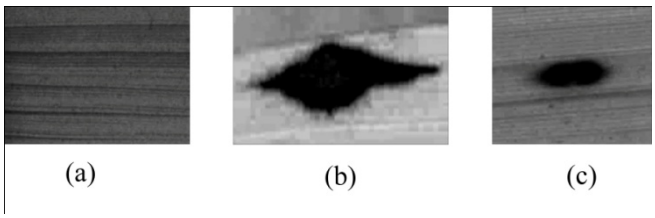

Fig. 3. Enhanced images corresponding to Fig. 2. obtained by applying the mean filter of size 3X3

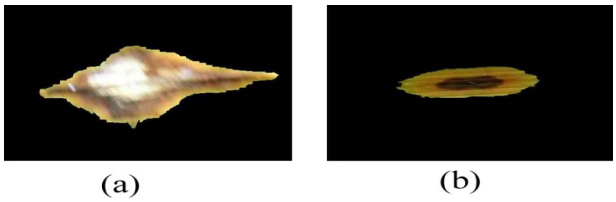

Fig. 4.(a) and (b) are the segmented images of Fig. 1(b) and (c) respectively, obtained by applying Otsu's segmentation algorithm in the Hue plane and taking the color values corresponding to the segmented region.

A. Image Segmentation

Image segmentation is the process of subdividing the image into its constituent parts [10] of interest. There are several methods [10], [11] of image segmentation like thresholding [12] - [14], region based [15]-[18] methods, morphological operation [10] and soft computing techniques [12], [19], [20]. The selection of the algorithm generally depends on the problem being solved. Here, Otsu's threshold based segmentation algorithm [13] has been applied which maximizes the between class variance parameter to detect the threshold value.

Hue plane of the image has been used for the threshold selection. Using that threshold value the image has been segmented into two parts. Fig. 4 (a) and (b) show the segmented images correspond to Fig. 1 (b) and (c) respectively obtained by taking the color values corresponding to the black pixel of the segmented images.

Fig. 5 (a) and (b) are the segmented images of Figure 1 (b) and (c) respectively, obtained by applying Otsu's segmentation algorithm in the Hue plane and taking the color values corresponding to the segmented region.

B. Feature Selection and Extraction

The morphological changes caused by the diseases are studied extensively and it has been observed that the brown spots are oval shaped while the blast spots are elliptical shaped. The shapes are distorted with the severity of the disease. The spots have different color at the center and the boundary. Radial distribution of the hue from the center towards the boundary of the spot has been considered as feature to distinguish the spots since the color of the spots vary from center towards the boundary. Hue represents the true color and the hue distribution values are stored in vectors, denoted as $H = \{h_1, h_2,\ldots\ldots h_8\}$, where h_i is the vector containing the hue value of the spots corresponding to the line number shown in Fig. 5.

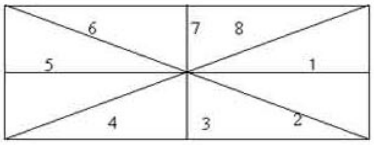

Fig. 5. Eight direction of radial features

III. Classification

Following two different phases of classification is performed.

- Classification of uninfected and diseased leaf
- Classification between diseased leaf

A. Classification of Uninfected and Diseased Leaf

Histogram represents the probability of occurrences of different gray levels in the image. The high probability of a particular gray level corresponds to a peak in the histogram. For an uninfected image, the probability of a particular gray level is very high since all the pixels have similar gray level values. On the other hand, a diseased image exhibits high probability for at least two sufficiently different gray levels, one for the diseased region of pixels and other for the uninfected region of pixels. Therefore, the histogram of an image with single peak is treated as uninfected and more than one peak is treated as the diseased leaf.

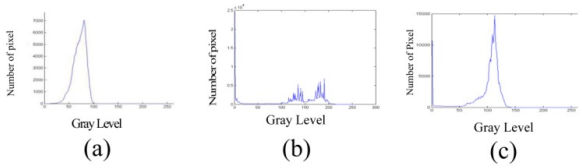

Fig. 6. Histogram of the images corresponding to Fig. 3

First, the number of peak is determined by translating the histogram upward with the factor ($m\times n/250$), where $m\times n$ is the size of the image. It has been observed that in order to reduce the noise, and to obtain clarity of the image size of the infected portion of the leaf must be at least greater than $(1/250)^{th}$ of the original image size. The histograms of the images are shown in Fig. 6. The number of peak is counted

by considering change of direction of the histogram.

B. Classification between Diseased Leaves

Classifiers based on Bayes' theorem and Support Vector Machines (SVMs) are applied for diseased image classification and the performance are compared

1) Bayes' Classifier

Define Bayes' classifier determines the belongingness of a pattern in a particular class depending on the likelihood of the pattern with that class [21]. The probability that a pattern x belongs to class ω_i is denoted by $p(\omega_i/x)$. Assume pattern x actually belongs to class ω_i but the classifier detects that x comes from class ω_j, and so it incurs loss equal to L_{ij}. Since the pattern x may belong to any of the M classes under consideration, the expected loss incurred to assigning x to class ω_j is formulated in (2).

$$r_j(x) = \sum_{i=1}^{M} L_{ij}\, p(\omega_i / x) \tag{2}$$

This is often referred to as the conditional average risk in decision theory terminology. The classifier has M possible categories to classify each pattern. Equation (2) computes the quantities $r_1(x), r_2(x), \ldots\ldots r_M(x)$, for each x, and assign each pattern to the class with the smallest conditional loss, which produces total expected loss with respect to all decision is minimum. The classifier, which minimizes the total expected loss, called Bayes' classifier. From the statistical point of view the Bayes' classifier represents the optimum measure to performance using Bayes' formula, given in (3)

$$p(\omega_i/x) = \frac{p(\omega_i)\,p(x/\omega_i)}{p(x)} \tag{3}$$

Thus the (2) is rewritten as (4)

$$r_j(x) = \frac{1}{p(x)} \sum_{i=1}^{M} L_{ij}\, p(\omega_i) p(x/\omega_i) \tag{4}$$

where $p(x/\omega_i)$ is called the likelihood function of class ω_i. Since $1/p(x)$ is a common factor so dropped from equation (4) and the expression of average loss then reduce to (5).

$$r_j(x) = \sum_{i=1}^{M} L_{ij}\, p(x/\ \omega_i)\, p(\omega_i) \tag{5}$$

Therefore, a pattern x is assigned to a class ω_i if $r_i(x) < r_j(x)$ for $j=1,2,3,\ldots..,M$, where $j \neq i$.

However, a special type of loss function has been used in the paper, which computes loss to zero for the correct decision and same nonzero value for all other erroneous decisions. The loss function is expressed as (6).

$$L_{ij} = 1 - \delta_{ij} \tag{6}$$

where $\delta_{ij}=1$ when $i=j$ and $\delta_{ij}=0$ when $j \neq i$. If we substitute the loss function of equation (5) using (6), it becomes as follows:

$$r_j(x) = \sum_{i=1}^{M} (1-\delta_{ij})\, p(x/\ \omega_i)\, p(\omega_i) = p(x) - p(x/\ \omega_j)\, p(\omega_j) \tag{7}$$

Therefore the Bayes' classifier assigns a particular pattern x to class ω_i if

$$p(x) - p(x/\omega_i)\, p(\omega_i) < p(x) - p(x/\omega_j)\, p(\omega_j)$$

or

$$p(x/\omega_i)\, p(\omega_i) > p(x/\omega_j)\, p(\omega_j), \quad j=1,2,3.....M;\ j \neq i \tag{8}$$

2) Support Vector Machine (SVM)

SVMs [22], [23] are supervised machine learning technique that can be analyzed theoretically using concepts from computational learning theory while being able to achieve good performance when applied to real-world problems. The most attractive feature of the SVM is the maximum-margin hyper plane, the soft margin and the kernel function. For classifying any two linearly separable classes there may exist many separating lines that correctly classify the data. Among these lines the SVM select the line, which maximizes the distance between the separating hyper-planes. To explain it clearly we label the training data $\{x_i, y_i\}$, $i = 1, \ldots\ldots, l$, $y_i \in \{-1, 1\}$, $\mathbf{x}_i \in \mathbf{R}^d$ where l denotes the total no of training sample and d denotes the dimension of the feature vector. Suppose we have some hyper-plane, which separates the positive from the negative examples (a "separating hyper-plane"). The points $\boldsymbol{x}$ which lie on the hyper-plane satisfy $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$, where $\boldsymbol{w}$ is normal to the hyper-plane, $|b|/\|\boldsymbol{w}\|$ is the perpendicular distance from the hyper-plane to the origin, and $\|\boldsymbol{w}\|$ is the Euclidean norm of $\boldsymbol{w}$. Let d_+ (d_-) be the shortest distance from the separating hyper-plane to the closest positive (negative) instance. Define the "margin" of a separating hyper-plane to be $d_+ + d_-$. For the linearly separable case, the support vector algorithm simply looks for the separating hyper-plane with largest margin. This can be formulated as follows: suppose that all the training data satisfy the following constraints:

$$\boldsymbol{x}_i * \boldsymbol{w} + b = +1 \text{ for } y_i = +1 \tag{9}$$

$$\boldsymbol{x}_i * \boldsymbol{w} + b = -1 \text{ for } y_i = -1 \tag{10}$$

These can be combined into one set of inequalities:

$$y_i(\boldsymbol{x}_i * \boldsymbol{w} + b) - 1 \geq 0 \quad \forall i \tag{11}$$

Now consider the points for which the equality of (9) hold (requiring that there exists such a point is equivalent to choosing a scale for $\boldsymbol{w}$ and b) good. These points lie on the hyper-plane $H_1 : \mathbf{x}_i \cdot \boldsymbol{w} + b = 1$ with normal $\boldsymbol{w}$ and perpendicular distance from the origin $|1 - b|/\|\boldsymbol{w}\|$. Similarly, the points for which the equality of (10) holds false, these points lie on the hyper-plane $H_2: \boldsymbol{x}_i \cdot \boldsymbol{w} + b = -1$, with normal $\boldsymbol{w}$, and perpendicular distance from the origin $|-1 - b|/\|\boldsymbol{w}\|$. Hence $d_+ = d_- = 1/\|\boldsymbol{w}\|$ and the margin is simply $2/\|\boldsymbol{w}\|$. Note that H_1 and H_2 are parallel (they have the same normal) and so no training points reside between them. Thus we can find the pair of hyper-planes, which gives the maximum margin by minimizing $\|\boldsymbol{w}\|^2$, subject to constraint (11), as depicted in Fig. 7. Other important characteristic of the SVM is the soft margin that gives the user flexibility to choose the parameter to roughly control the number of examples. Another feature is the kernel function that projects the non-linearly separable data from low-dimensional space to a space of higher dimension so that they may become separable

in the higher dimensional space too.

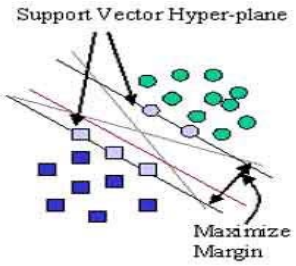

Fig. 7. Hyper-plane separated by the SVM

IV. Result and Discussion

Five hundred samples of each data class are considered to test the system. For the first level classification (i.e. uninfected leaf or a diseased leaf), it has been found that success rate is 92%, 96.4% and 84% respectively for uninfected leaf, leaf with brown spot and leaf with blast, summarized in Table I. Next, the diseased leaves are considered in tenfold way to testing using Bayes' Classifier. Taking first 450 of the data set as training set and remaining 50 as test set the likelihood for both classes has been calculate and then assign the pattern to the class, which have the maximum likelihood value. Again the same procedure is repeated for the first 400 and last 50 of the data set as training set while remaining 50 of the data set is considered as the test set. Thus, 10 different combinations of training and testing data set are applied resulting classification accuracy 79.5%. SVM classification with 10 fold cross validation is carried out for same data set and 68.1% classification accuracy is achieved, which has been shown in Table II.

TABLE I: Result of Classification of Uninfected and Diseased Leaves

Image Type	Success Rate(%)
Normal Leaf Image	92
Brown Spot Image	96.4
Blast Image	84

TABLE II: Result of Classification of Diseased Leaves

Classifier	Number of Samples	Success Rate (%)
Bayes' Classifier	1000	79.5
Support Vector Machine	1000	68.1

V. Conclusion

In the work, an automated system has been developed for detecting two different types of rice diseases. In the first stage, uninfected and the diseased leaves are classified based on the number of peaks in the histogram. Miss classification may occur due to shadow effect and color distortion of aging leaves. In the second level Bayes' classifier and SVM are applied to classify the leaf diseases. Time complexity of the Bayes' classifier is $O(N \times D^2)$ where as for the support vector machine it is $O(D \times N^2)$ where D is the dimension of the feature vector and N is the number of training samples. Since number of samples generally much larger than the dimension of the feature vector, therefore the proposed system is time efficient compare to SVM.

References

[1] S. H. OU. *In: Rice Diseases. Kew Surrey*, England, Commonwealth Mycological Institute, 1972

[2] C. Eric. *In: Report on Study Tour to Examine Symptoms of Rice Diseases in Southern USA and California. Principal Research Scientist NWS Agriculture*, Agricultural Institute, Orange NSW, 2002

[3] R. K. Webster, *In: Rice Blast Disease Identification Guide. Davis*, Dept. of Plant Pathology, University of California, 2000

[4] H. Sato, I. Ando, H. Hirabayash, Y. Takeuchi, S. Arase, J. Kihara, H. Kato,T. Imbe, and H. Nemoto, "QTL Analysis of Brown Spot Resistance in Rice(Oryza Sativa," *Breeding Science*, vol. 58, pp. 93-96, 2008.

[5] H. L. Khatibl, F. Hawels, H. Hamdi, and N. L. Mowelhi. "Spectral Characteristics Curves of Rice Plants Infected by Blast," *IEEE Geoscience and Remote Sensing Symposium*, vol. 2, pp. 526-528, 1993.

[6] T. Kobayashi, E. Kanda, K. Kitada, K. Ishiguro, and Y. Torigoe, "Detection of Rice Panicle Blast with Multispectral Radiometer and the Potential of Using Airborne Multispectral Scanners," *Phytopathology*, vol. 91, pp.316-323, 2001.

[7] P. J. Pinter, Jr. J. L. Hatfield, J. S. Schepers, E. M. Barnes, M. S. Moran, C. S. T. Daughtry, and D. R. Upchurch, "Remote Sensing for Crop Management," *Photogrammetric Engineering and Remote Sensing*, vol. 69, pp. 647-664, 2003.

[8] Z. Qin, M. Zhang, T. Christensen, W. Li, and H. Tang, "Remote Sensing Analysis of Rice Disease Stress for Farm Pest Management Using Wide-band Airborne Data," *IEEE Geoscience and Remote Sensing Symposium*, vol. 4, pp. 2215-2217, 2003.

[9] P. Sanyal, U. Bhattacharya, and S. K. Bandyopadhyay, "Color Texture Analysis of Rice Leaves Diagnosing Deficiency in the Balance of Mineral Levels towards Improvement of Crop Productivity. *IEEE 10th International Conference on Information Technology*, pp. 85-90, 2007.

[10] R. C. Gonzalez and R. E. Woods, *In: Digital Image Processing*. New Delhi, India, Pearson Education., 2007.

[11] H. D. Chen, X. H. Jiang, Y. Sun, and W. J. Li, "Color Image Segmentation: Advances & Prospects," *Pattern Recognition*, vol. 34, pp. 2259-2281, 2001

[12] L. K. Huang and M. J. J. Wang "Image Thresholding by Minimizing the Measure of fuzziness," *Pattern Recognition* , vol. 28, pp. 41-51. 1995.

[13] J. N. Kapur, P. K. Sahoo, and Wong AKC. "A New Method for Gray Level Picture Threshold Using the Entropy of the Histogram," *Graphical Models and Image Processing*, pp. 273-285. 1985.

[14] N. Otsu, "A Threshold Selection Method from Gray Level Histograms" *IEEE Transaction on Systems, Man and Cybernetics,* vol. 9, pp. 62-66., 1979.

[15] N. Ikonomakis, K. N. Plataniotis, and A. N. Venetsanopoulos, "A Region Based Color Image Segmentation Scheme," in *Proc. of Electrical Imaging*, vol. 3653, pp. 1202-1209, 1999.

[16] N. Ikonomakis, K. N. Plataniotis, and A. N. Venetsanopoulos, "Unsupervised Seed Determination for A Region-Based Color Image Segmentation Scheme," *IEEE International Conference on Image Processing*, pp. 537-539, 2000.

[17] F. Y. Shih and S. Cheng, "Automatic Seeded Region Growing For Color Image Segmentation," *Image and Vision Computing* vol. 23, pp. 877-886, 2005.

[18] Y. Qixiang, G. Wen, and Z. Wei. "Color Image Segmentation Using Density Based Clustering," IEEE*, ICASSP*, vol. 3, pp. 345-348, 2003.

[19] Y. Chen and J. Z. Wang, "A Region Based Fuzzy Feature Matching to Content Based Image Retrieval," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1252-1267, 2002.

[20] D. P. Tuna, "Image Segmentation Using Probabilistic Fuzzy C-Means Clustering," *International Conference on Image processing*, pp.722-725, 2001.

[21] T. Touanf and R. C. Gonzalez, "In: Pattern Recognition Principles. London," *UK*, Addition Wesley Publishing Company. pp. 110-1543, 1974.

[22] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery* vol. 2, pp. 1-47, 1998.

[23] C. Nello and J. S. Taylor, *In: An Introduction to Support Vector Machines and Other Kernel-based Learning,* Cambridge University Press. 2000.