Application of support vector machine for detecting rice diseases using shape and color texture features

Qing Yao, Zexin Guan, Yingfeng Zhou
College of Informatics and Electronics
Zhejiang Sci-Tech University
Hangzhou, China
e-mail: qingyaozstu@gmail.com

Jian Tang, Yang Hu, Baojun Yang
China National Rice Research Institute
Hangzhou, China
e-mail: tangjian@mail.hz.zj.cn

Abstract—**For detecting rice disease early and accurately, we presented an application of image processing techniques and Support Vector Machine (SVM) for detecting rice diseases. Rice disease spots were segmented and their shape and texture features were extracted. The SVM method was employed to classify rice bacterial leaf blight, rice sheath blight and rice blast. The results showed that SVM could effectively detect and classify these disease spots to an accuracy of 97.2%.**

Keywords-rice diseases spots; image processing; texture features; support vector machine

I. Introduction

Rice is one of the most important food crops in the world. Rice is often infested by some diseases caused by some pathogens including fungi, bacteria and viruses. These diseases result in great economic losses to farmers every year [1]. It is important that disease symptoms are identified early and accurately for undertaking disease control actions or remedial measures. The symptoms of rice diseases, such as rice bacterial leaf blight, rice sheath blight and rice blast, appear initially as spots around the infected areas. So the detection of these diseases mainly relies on their spots. At present, farmers or technicians mainly rely on their experiences, the guide books and experts to identify these diseases. It is suitable for detecting some familiar and typical diseases. However, some different diseases can cause similar spots and the same diseases can cause different spots because of different rice varieties and local conditions. It increases the complexity of identifying rice diseases. Misidentification usually leads to some incorrect control measurements, such as indiscriminate and untimely use of pesticides.

For detecting crop diseases early and accurately, many researchers have used image processing and classification techniques for detecting diseases of agricultural or economic crops [2-9]. The general approaches of these researches are similar. First, disease images are acquired using cameras or scanners. Second, disease spots are segmented from background. Third, color, shape or texture features are extracted. Finally, classification methods, such as neural networks, Bayesian classifier method, K-nearest neighbor, support vector machine, are used to classify disease images. For example, Pydipati et al. (2006) extracted color texture features of citrus leaves with greasy spot, melanose, normal and scab to identify the four different classes. Discriminant analysis using textural features in the color co-occurrence method achieved classification accuracies of over 95% for all class [6]. Zhao et al (2007) used Bayesian classifier to classify five kinds of maize leaf diseases images. The result shows that the precision of maize disease identifying is higher than 83% and Bayesian classifier was effective for classifying maize diseases [7]. Huang (2007) applied back-propagation neural network classifier and gray level co-occurrence matrix (GLCM) of texture features of the lesion area to detect three diseases of the Phalaenopsis seedling. The accuracy of 89.6% in classifying these lesions was achieved [8]. Tian et al (2007) extracted texture, shape and color features of three kinds of grape leaf diseases and studied the recognition accuracy and efficiency of grape disease by Support Vector Machine (SVM) and their results showed that classification performance of Support Vector Machine was better than that of neural networks [9]. In these techniques, the segmentation of disease spots, the selection of features and identification method are important to accuracy rate of detecting diseases.

In our study, an accurate and automatic method based on image processing and Support Vector Machine was developed to detect three rice diseases including rice bacterial leaf blight (RBLB), rice sheath blight (RSB) and rice blast (RB).

II. Materials and Methods

A. Image acquisition

The images of three rice diseases (RBLB, RSB and RB) were acquired using a CCD color camera (Nikon D80, zoom lens 18-200mm, F3.5-5.6, 10.2 Mega pixels) in the rice field of China National Rice Research Institute, located in Fuyang, Zhejiang, China (30° North Latitude, 120° East Longitude) (Fig. 1). Seventy-two samples of each disease images were collected. These images were stored in uncompressed JPEG format.

B. Preprocessing of rice disease image

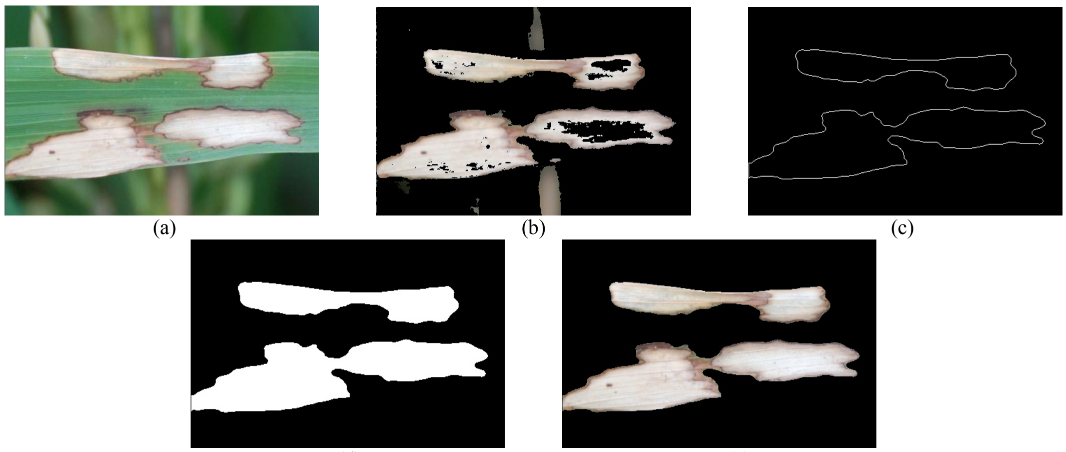
For reducing the computational burden and standardizing the image resolution, the resolution of all images was reduced to 800 pixels×600 pixels. Image processing was performed using Visual C++ 2005 programming.

Because all images were acquired in rice field, it was unavoidable that dewdrop, insects' excrement and dust might

978-0-7695-3655-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICEC.2009.73

(a) (b) (c)

Figure 1. Three infected rice leaves: (a) RBLB, (b) RSB, (c) RB.



(a) (b) (c)

(d) (e)

Figure 2. Segmentation of RSB spots: (a) Original image of RSB spot, (b) Image after segmentation by changed threshold values, (c) Outline of RSB spots, (d) Binary image of RSB spots, (e) RSB spots by segmented.

appear in the images. They were regarded as the image noises which would disturb the segmentation and the feature extraction of disease spots, so they must be removed or weakened before any further image analysis. In our study, a median filter of color image with 3*3 rectangle filter window was used to original rice disease images for removing or weakening these noises.

C. Segmentation of rice disease spots

For obtaining integrated rice disease spots, we developed an algorithm based on spot color and outline to segment disease spot area from rice leaf. The algorithm was described as follows:

First, the images were transformed from a red, green, blue (RGB) color representation to y_1 and y_2 representation. y_1 and y_2 represented two color function defined as

$$\begin{cases} y_1 = 2g - r - b \\ y_2 = 2r - g - b \end{cases} \tag{1}$$

Where r, g and b represented pixel values of red, green, blue (RGB) color of image respectively.

Second, in y_1 and y_2 representation, the method of maximum classes square error (Otsu method [10]) was used to segment disease spots from rice leaf and two threshold values (T_1, T_2) were automatically obtained according to Otsu method. Preliminary investigations showed the edge of some disease spot was removed using Otsu method. However, the edge of disease spot was important to identify the disease category. For keeping the edge of disease spot in segmentation, we changed the threshold values and gave the pixel of the disease spot by

$$p_s = \begin{cases} (0, 0, 0) & y_1 > T_1 + 10 \text{ and } y_2 < T_2 - 10 \\ p_r & y_1 < T_1 + 10 \text{ and } y_2 > T_2 - 10 \end{cases} \tag{2}$$

Where p_r is the pixel of original image.

The disease spot area was described by

$$A_1 = \{p_s, p_s \neq (0, 0, 0)\} \tag{3}$$

In the disease spot area, some holes might be found, but it could be filled by the following method. If some noises were still kept in the image, they could be removed by comparing noises area with the disease spot area.

Finally, the edge detection of area A_1 was completed by compressing horizontal, vertical, and diagonal segments [11]. We could obtain the binary image of disease spot according to the edge of disease spot A_1 and this area was represented by A_2. The pixel of disease spot could be calculation by a copy of original image according to formula

$$p_d = \begin{cases} p_r & p_r \in A_2 \\ (0, 0, 0) & p_r \notin A_2 \end{cases} \tag{4}$$

The disease spot of RBLB was segmented by above method in Fig.2.

D. Features extraction

In image processing, image features usually included color, shape and texture features. Because the color features are influenced largely by outside light, we selected shape and color texture features of disease spot as characteristic values of classification.

The area (A) and perimeter (P) of disease spots were calculated from the binary image of disease spot. The minimum enclosing rectangle (MER) of a disease spot were obtained by the method of rotating image with same angle [11]. The long axis length and short axis length of MER represented the length (l) and width (w) of a disease spot. The shape features including rectangularity, compactness, elongation and roundness were calculated using area, perimeter, MER of disease spot (Table 1).

We adopt texture features from the gray level co-occurrence matrix (GLCM, Haralick et al., 1973, [12]) to classify three rice diseases. Firstly, we transformed the image from a red, green, blue (RGB) representation to a hue, saturation, value (HSV) color representation according to formula (5).

$$\begin{cases} v = \max(r,g,b) \\ s = \begin{cases} (v - \min(r,g,b)*255)/v & v \neq 0 \\ 0 & v = 0 \end{cases} \\ h = \begin{cases} (g-b)*60/s & v = r \\ 180 + (b-r)*60/s & v = g \\ 240 + (r-g)*60/s & v = b \end{cases} \\ \text{if } h < 0 \text{ then } h = h + 3 = 60 \end{cases} \tag{5}$$

Where r, g and b represent pixel values of red, green, blue (RGB) color of image respectively.

Then, a GLCM with $N \times N$ square matrix was generated from HSV spatial system, where N was the number of different gray levels in an image. An element $p(i, j, d, \theta)$ of a GLCM of an image represented the probability of two gray-level pixels, where i was the gray level at location (x, y), and j represented the gray level of other pixel at a distance d and an orientation angle of θ from location (x, y). In our study, d=1, N=64, "θ" was selected orientation angles 0º, 45º, 90º, 135º.

Five texture features including contrast, uniformity, entropy, inverse difference and linearity correlation in every orientation angle were extracted from GLCM of each image (Table 1). The total 60 texture feature values from 4 orientation angles and 3 spatial (H, S, V) systems were obtained in each image.

E. Support Vector Machine

Support vector machine (SVM) is a kind of new machine learning methods based on statistical learning theory, which was put forward by Vapnik in 90s' [13]. Most traditional classification models are based on the empirical risk minimization principle. SVM implements the structural risk minimization principle which seeks to minimize the training error and a confidence interval term. A number of applications showed that SVM hold the better classification ability in dealing with small sample, nonlinearity and high dimensionality pattern recognition.

The main idea of SVM is to find a decision surface (H) determined by certain points of the training set, termed support vectors between two point classes. This surface divides the training data (x_i, y_i) without mistake, that is, all points of the same class are divided in the same side while the minimum distance between either of the two classes and this surface is maximal margin (Fig. 3). This surface can be obtained from the solution of a problem of quadratic programming [14].

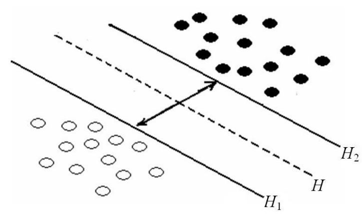


Figure 3. The optimal plane of SVM in linearly separable condition

If points are linearly separable, the function of this surface is estimated by

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\right) \quad (x_i, y_i) \in R^N \times \{-1,1\} \tag{6}$$

Where α_i^* is a Lagrange multiplier, b^* is the bias.

If the classes will not be linearly separable, the function of this surface is given by

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i^* y_i k(x, y) + b^*\right) \tag{7}$$

Where $k(x, y)$ is a kernel function.

Some commonly used kernel functions were defined as follows.

The linear kernel function

$$k(x, y) = x \cdot y \tag{8}$$

The polynomial kernel function

$$k(x, y) = (1 + x \cdot y)^q, q = 1,2,..., N \tag{9}$$

The radial basis kernel function

$$k(x, y) = \exp(-\| x - y \|)^2 \tag{10}$$

The program of SVM was written in Visual C++ 2005.

F. Classification of diseases

The 72 disease spot images from each disease were divided randomly into two groups according to 1:1 divide. One group was used as the training samples and the other group was used as the testing samples.

After completing the extraction of shape and color texture feature from disease spot images, we developed three classification models. Model 1 consisted of 4 shape and 60 texture features. Model 2 consisted of 60 texture features. Model 3 consisted of 4 shape features. The numbers of goal vector (1, 2, 3) represented RBLB, RSB, RB in order.

SVM was used to classy the three diseases, where radial basis kernel function was selected.

III. Results and Discussion

SVM was used to test the accuracies of classifying three diseases. The classification results showed in Table 2. The overall classification accuracy of model 1 was 97.2%.

The overall classification accuracy of model 2 dropped to 88.0%. This decrease in performance was due to some similar texture between RBLB and RSB. The 11.1% of RB spots were incorrectly classified into the RSB.

If only using shape features of model 3, the overall classification accuracy dropped to 82.4%. The shape of RBLB was long and clearly different from RSB and RB. It couldn't be classified into the other two diseases. So the reason of the low classification accuracy of model 3 was that the shape features of RSB and RB spots had instability.

So it was suggested that using shape and texture features could achieve high classification accuracy of the three diseases.

IV. CONCLUSION

A study was completed to detect and classifying rice diseases, including RBLB, RSB and RB. All disease images were acquired by CCD digital camera in rice field. Rice disease spots were segmented efficiently according to color and outline of disease spots. Four shape features (rectangularity, compactness, elongation and roundness) and 60 texture features (contrast, uniformity, entropy, inverse difference and linearity correlation) of disease spot from 4 orientation angles (0º, 45º, 90º, 135º) and 3 spatial (H.S.V) were extracted. The SVM method was developed with 108 training samples data and evaluated with 108 testing samples data of disease spot images. The rice diseases can be detected and classified efficiently using shape and color texture features based on SVM. This method could be applied to classification of other crop diseases.

ACKNOWLEDGMENT

This work was supported by National Science Foundation China Program Grant (NSFC-30600398) and Special Fund for Basic Expenditure for Scientific & Research of Central Non-profit Scientific Research Institutions (2007RG016).

TABLE I. MATHEMATICAL FORMULATION OF SHAPE AND COLOR TEXTURE FEATURES

Features	Formula
Rectangularity	A / A_{MER}
Compactness	$4 \times \pi \times A / P^2$
Elongation	w / l
Roundness	$4 \times A / (\pi \times l^2)$
Contrast	$\sum_i \sum_j (i-j)^2 p(i,j,d,\theta)$
Uniformity	$\sum_i \sum_j p(i,j,d,\theta)^2$
Entropy	$-\sum_i \sum_j p(i,j,d,\theta) \ln p(i,j,d,\theta)$
Inverse difference	$\sum_i \sum_j p(i,j,d,\theta) \Big/ \left[1+(i-j)^2\right]$
Linearity correlation	$\left[\sum_i \sum_j \left((ij)\, p(i,j,d,\theta) - \mu_x \mu_y\right)\right] \Big/ \sigma_x \sigma_y$ $\mu_x = \sum_i i \sum_j p(i,j,d,\theta) \quad \mu_y = \sum_j j \sum_i p(i,j,d,\theta)$ $\sigma_x = \sum_i (i-\mu_x)^2 \sum_j p(i,j,d,\theta) \quad \sigma_y = \sum_j (j-\mu_y)^2 \sum_i p(i,j,d,\theta)$

TABLE II. CLASSIFICATION RESULTS FROM TESTING DATA FOR MODEL 1, MODEL 2 AND MODEL 3

Disease types	Number of sample classified into three disease types											
	Model 1				Model 2				Model 3			
	RBLB	RSB	RB	Total (%)	RBLB	RSB	RB	Total (%)	RBLB	RSB	RB	Total (%)
RBLB	35	1	0	97.2	32	2	2	88.9	36	0	0	100
RSB	1	34	1	94.4	5	31	0	86.1	5	22	9	61.1
RB	0	0	36	100.0	0	4	32	88.9	0	5	31	86.1
Total	36	35	37	97.2	37	37	34	88.0	41	27	40	82.4

REFERENCES

[1] M.L.Gullino, J. Fletcher, A. Gamliel and J.P. Stack, *Crop Biosecurity*. Netherlands: Springer, 2008, pp. 25-36 .

[2] T.F. Burks, S.A.Shearer and F.A. Payne, "Classification of weed species using color texture features and discriminant analysis," Trans. ASAE, vol.43, no.2, pp.441–448, Apr. 2000.

[3] J.A. Marchant, and C.M.Onyango, "Comparison of a Bayesian classifier with a multilayer feed-forward neural network using the example of plant/weed/soil discrimination," Comput. Electron. Agric., vol. 39, no.1, pp. 3–22, Apr. 2003.

[4] D. Moshou, C. Bravo, J. West, S. Wahlen, A. McCartney and H. Ramon, "Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks," Comput. Electron. Agric., vol. 44, no.3, pp. 173–188, Sept. 2004.

[5] P.M. Granitto, P.F. Verdes and H.A. Ceccatto, "Large-scale investigation of weed seed identification by machine vision," Comput. Electron. Agric., vol. 47, no.1, pp.15–24, Apr. 2005.

[6] R. Pydipati, T.F. Burks and W.S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," Comput. Electron. Agric., vol.52, no.1, pp.49-59, June 2006.

[7] Y.X. Zhao, K.R. Wang, Z.Y. Bai, S.K. Li, R.Z. Xie and S.J. Gao, "Bayesian classifier method on maize leaf disease identifying based images," Comput. Engin. Applic., vol.43, no.5, pp.193-195, Feb. 2007.

[8] K.Y. Huang, "Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features," Comput. Electron. Agric., vol.57, no.1, pp.3-11, May 2007.

[9] Y.W. Tian, T.L. Li, C.H. Li, Z.L. Piao, G.K. Sun and B. Wang, "Method for recognition of grape disease based on support vector machine," Trans. CSAE, vol.23, no.6, pp.175-180, June 2007.

[10] N. Otsu, "A threshold selection method from gray-level histogram," IEEE Trans. Syst., Man and Cybern., vol.9, no.1, pp.62-66, Jan.1979.

[11] G. Bradski, and A. Kaehler, *Learning OpenCV: computer vision with the OpenCV library*. California: O'Reilly Media, Inc., 2008, pp. 238-250 .

[12] R.M. Haralick, K. Shanmugam and I. Dinstein, "Textual features for image classification," IEEE Trans. Syst. Man Cybern., vol.3, no.6, pp. 610–621, Nov. 1973.

[13] Vapnik,V., *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 1995 .

[14] S.H. Yue, P. Li and P.Y. Hao, "SVM classification: Its contents and challenges," Appl. Math. J. Chinese Univ. Ser. B, vol.18, no.3, pp. 332-334, Sept. 2003.