Image Recognition of Wheat Disease Based on RBF Support Vector Machine*

Lianzhong Liu[1], Wu Zhang[1], Shuangbao Shu[2], Xiu Jin[1]
[1]School of Information and Computer Science, Anhui Agricultural University, Hefei, China
[2]School of Instrument Science and Opto-electronic Engineering, Hefei University of Technology, Hefei, China
lzliu@ahau.edu.cn, zhangwu@ahau.edu.cn, shu@hfut.edu.cn, jinxiu123@ahau.edu.cn

Abstract - The paper proposes an image recognition method of wheat disease. Image background is first removed by image segmentation using green feature of wheat leaf to obtain only disease pixels from original leaf image. Then disease features are calculated through 3 schemes: 1) mean values of R, G, B; 2) normalized mean values of R, G, B; 3) green ratios of R/G, B/G. Using disease features as input, image samples are trained and recognized using multi-class RBF SVM. The method has been tested on healthy leaves and leaves infected by leaf powdery mildew, stripe rust, leaf rust and leaf blight. The result shows normalized R, G, B achieved the best recognition rate up to 96%, and the overall recognition rate decreases dramatically while including more disease types in samples.

Index Terms - Plant disease, Computer vision, Image processing, Support Vector Machine.

I. Introduction

Wheat is the second important crops in China, but diseases occurred on wheat has caused severe yield reduction every year. Quick and accurate determination of wheat disease is essential to the early field control, so as to improve the quality and yield. In general, the diseased parts of a crop and the normal parts are usually different in color. For example, the healthy leaf is usually green, and the diseased leaf is probablly in yellow, brown or white [1].

Computer vision(CV) technology has been widely used in many fields[2]. By using computer vision, the visual information of disease, such as color,shape and texture, can be obtained immediately without destroying the leaf. Up to now, many researchers and experts have done extensive works on it [2-4].

The paper proposes an disease diagnosis algorithm of wheat based on Computer Vision, which provides the decision support for early detection and fast treatment on wheat disease.

II. Materials and Methods

The experimental samples are collected from healthy leaves and 6 types of diseased leaves: powdery mildew, leaf blade stripe rust, leaf rust, leaf blight, yellow dwarf and yellow mosaic. Firstly, the disease images are cut manually from the original images, 30 samples for each type, with total of 210 image samples. The samples are saved into JPEG format files. In order to retain the characteristic of disease, the size of each sample image should be not less than 30*30 pixels.

According to the principle of computer vision, the overall process of the diagnosis algorithm for wheat disease is designed, as shown in Fig. 1.

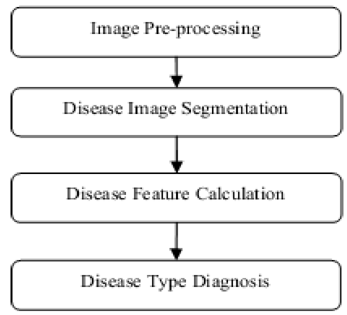

Fig.1 Overall Process of Image Diagnosis Algorithm for Wheat Disease

Firstly, the leaf image is pre-processed, such as enhances the disease pixels by wavelet transformation and filters the noise interference in the image. Then, the image segmentation is performed to remove background pixels in the image to obtain the disease image including only the diseased pixels. Further more, statistical calculations of the disease image are made to get the feature parameters of each disease. Finally, the disease feature parameters are used for the diagnosis of wheat disease to determine the disease type.

A. Image preprocessing

The purpose of image preprocessing is to enhance the diseased area and filter the noise interference in the original image, improving the stability of the image features which will be used for the subsequent disease diagnosis.

Wavelet transformation has the multi-resolution attribute that it is considered as a group of bandpass filters in signal filtering [5]. As it is very irregular, the diseased area usually contains the high frequency components in the image. The original image is transformed by Daubechies wavelet transformation, decomposed into high and low frequencies, filtered the low frequency components,and restored by inverse wavelet transformation. By this way, the diseased area in the image is enhanced, and the low frequency interference of background image is also eliminated.

B. Disease image segmentation

Image segmentation is to separate the image into disease image which contains only disease pixels, it directly affects the

* Supported by the Key Technology Projects of Anhui Province, China (NO:1201a0301008)

accuracy of feature extraction and disease diagnosis[6]. Found from the analysis of image samples of wheat disease, healthy leaves are normally pure green, and the diseased area is usually yellow, brown or white, which is not green. Therefore, image segmentation based on the features of disease color can separate the diseased area effectively.

According to this thought, an algorithm is acquired for disease image segmentation. By mathematical morphology processing with the leaf image, pixels of the healthy area are set to black, and pixels of diseased area are unchanged.

C. Disease Feature Calculation

Image features, such as color, shape and texture, can all be used for plant disease diagnosis [7]. The color feature is most intuitive and simple for calculation. In this paper, according to different statistic patterns of R, G, B vlaues of each pixel in the disease image, three calculation schemes are proposed and used to get disease parameters for further diagnosis comparation.

Scheme I: mean RGB values. In the segmented image which contains only disease pixels, R, G, B values are calculated respectively, as follows:

$$\overline{R}=\sum_{i=1}^{N}R_i/Total,\ \overline{G}=\sum_{i=1}^{N}G_i/Total,\ \overline{B}=\sum_{i=1}^{N}B_i/Total \tag{1}$$

$\overline{R}$, $\overline{G}$, $\overline{B}$ are mean values of R, G, B in the image; *Total* is the total count of disease pixels, N is the number of image pixels. $\overline{R}$, $\overline{G}$, $\overline{B}$ are used as feature parameters of each disease.

Scheme II: normalized mean RGB values. While capturing disease image, the light intensity in different environment changes the amplitude of R, G, and B values, using color mean values as disease feature directly would lead to misjudgment. By the analysis of experimental image samples, it is found that the value of $\overline{R}$, $\overline{G}$, or $\overline{B}$ changes proportionally as the light intensity increased or decreased. The normalized $\overline{R}$, $\overline{G}$, $\overline{B}$ are defined as:

$$\begin{aligned} \text{R}&=\overline{R}/(\overline{R}+\overline{G}+\overline{B}) \\ \text{G}&=\overline{R}/(\overline{R}+\overline{G}+\overline{B}) \\ \text{B}&=\overline{R}/(\overline{R}+\overline{G}+\overline{B}) \end{aligned} \tag{2}$$

R, G, B are the disease parameters of which are insensitive to the change of light intensity.

Scheme III: green mean ratio. There is a certain correlation in an image among $\overline{R}$, $\overline{G}$ and $\overline{B}$. In order to reduce the number of parameters and the insensitivity to light variation, the ratios of $\overline{R}$, $\overline{B}$ to $\overline{G}$ are defined as disease parameter s:

$$r=\overline{R}/\overline{G},\qquad b=\overline{B}/\overline{G} \tag{3}$$

Using r and b as disease parameters can improve the diagnosis speed.

D. Disease Type Diagnosis

Support Vector Machine (SVM) is a classification method based on statistical learning theory. It has good generalization ability which can compromise between model complexity and learning ability under finite samples. The main advantages of SVM are : it can obtain current optimal solution under finite samples; it can obtain the global optimal solution without falling into local optimums that normal algorithms have; it transforms nonlinear problems into linear problems in a higher dimension space, and the algorithm complexity is irrelated with space dimension.

For linearly separable training samples $(x_i,\ y_i)$, there exists a hyperplane which can separate two-type samples without error. When the interval of the two types is maximum, the hyperplane is then the optimal classification plane.

Optimal classification function can be calculated by quadratic programming:

$$f(x)=Sgn\left\{\sum_{i=1}^{l}y_i\alpha_i^*(x_i\cdot x)+b^*\right\} \tag{4}$$

where $Sgn\{\ \}$ is the signal function, x_i the input of training samples, y_i the correspondant output, l the number of samples, α_i^* the Lagrange coefficient, and b^* the bias value. For linear inseparable case, a non slack variable and a penalty factor are needed to obtain global optimal hyperplane.

For nonlinear separable training samples, the input vectors are mapped to a feature space with higher dimension. The optimal hyperplane is constructed in the feature space.This method is called the support vector machine method. While a certain kernel function is defined, the classification function becomes:

$$f(x)=Sgn\left\{\sum_{i=1}^{l}y_i\alpha_i^*K(x,\ x_i)+b^*\right\} \tag{5}$$

where $K(x,\ x_i)$ is the kernel function. The following kernel functions are commonly used:

Linear kernel function: $K(x,\ x_i)=x\cdot x_i$

Polynomial kernel function $K(x,\ x_i)=(s(x\cdot x_i)+c)^d$

Radial basis kernel function $K(x,\ x_i)=\exp\left\{-\frac{\|x-x_i\|}{2\sigma^2}\right\}$

Sigmoid kernel function $K(x,\ x_i)=\tanh(\nu(x\cdot x_i)+c)$ (6)

where x_i is the sample vector of the support vector, x the predictor vector, c、d、σ、ν the corresponding kernel parameters.

Support vector machine has been used well in the diagnosis of crop disease. In the four SVM kernel functions, RBF kernel function has the strongest adaptability to samples, it can maximize the ability of mapping linearly inseparable samples to linear separable samples in a higher dimension space. In this paper, the support vector machine based on radial basis kernel function is adopted in the wheat disease classification. First, according to formula (1-3), the disease parameters of known samples are calculated. Then, the parameters are sent to SVM for training. Finally, the diagnosis of unknown samples is made to determine the disease type from the output of SVM.

III . Experimental Results

OpenCV is a computer vision library provided by Intel. It consists of a group of C functions and a small amount of C++ classes, in which many general algorithms in image processing and computer vision have been realized[8]. Visual C++ is used to realize the diagnosis algorithm with OpenCV library.

A. Image preprocessing and disease image segmentation

Preprocessing and segmentation of the disease image samples are made. The result is shown in Fig. 2, where a is a healthy leaf, b powdery mildew leaf, c stripe rust leaf , d rust blade leaf, e blight leaf, and f to j correspondant disease images. As can be seen from the result, the disease area is mostly seperated from the leaf image.

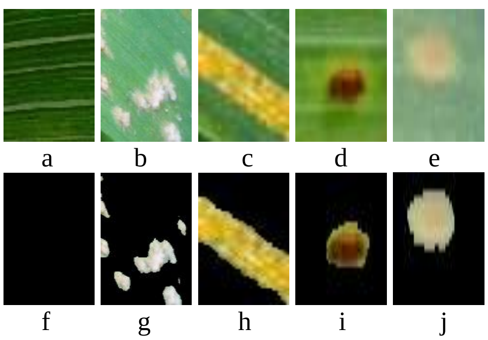

Fig.2 Segmentation of Disease Image from Wheat Leaf

B. Disease parameter calculation

With the disease image, the disease parameters are calculated. TABLE I shows the parameters from Fig. 2.

TABLE I Values of Disease Parameters from Fig. 2

Disease Parameter	*Healthy leaf*	*Powdery mildew*	*Stripe rust*	*Leaf rust*	*Ye Kubing*
$\overline{R}$	0	188	61	15	130
$\overline{G}$	0	204	176	91	179
$\overline{B}$	0	199	197	115	182
R	0	0.318	0.141	0.071	0.264
G	0	0.345	0.405	0.412	0.364
B	0	0.337	0.454	0.517	0.372
R	0	9.241	3.480	1.725	7.266
B	0	9.794	11.24	12.58	10.16

C. Disease type diagnosis

Many factors affect the accuracy of diagnosis. The influence of parameters selection, number of training samples and increasing disease types are discussed.

(1) Influence of disease parameters on accuracy rate. 15 images for each disease are randomly picked as training samples, the remaining 15 images as test samples. The diagnosis is made by one of the three schemes seperately, the result is shown in TABLE II. The number shown in the table means the number of samples that are correctly recognized.

TABLE II Average Accuracy Rate Using Different Color Feature Scheme

Color feature	*Health y leaf*	*Powdery mildew*	*Stripe rust*	*Leaf rust*	*Ye Kubing*	*Accuracy rate*
I	15	15	8	13	13	85.33%
II	15	14	7	8	14	94.67%
III	15	14	7	8	14	77.33%

It is found from Table 2 that scheme II has the highest accuracy, scheme III has the lowest accuracy,and scheme I is between scheme II and scheme III.

(2) Influence of training samples on accuracy rate. Using different number of trainning samples for disease diagnosis, the result is obtained as shown in Fig. 3.

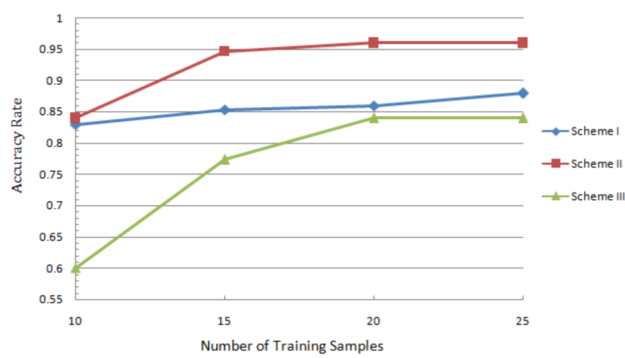

Fig.3 Influence of Training Samples on Accuracy Rate on the Three Schemes

As seen from the graph, for scheme I, the accuracy rate is above 80% when using 10 training samples, and increases with the increase of training samples; for scheme II, the accuracy rate is above 90% when using 15 training samples, and increases with the increase of training samples; for scheme III, the accuracy rate is under 60% when using 10 training samples, but increases rapidly with the increase of training samples. The three schemes all reach high accuracy rate when using 20 training samples, but increases very slow or even decrease with the further increase of training samples. This shows that SVM is more suitable for small training samples. Thus, while using SVM for disease diagnosis, the appropriate number of training sample is about 15 to 20.

(3) Influence of increasing disease types on accuracy rate. Add disease samples of yellow dwarf disease and yellow mosaic disease, and repeat the whole test. Pick 30 samples each, with 15 samples for training, others for diagnosis. The result obtained is shown in TABLE III.

TABLE III Diagnosis Result of Increasing Wheat Diseases on the Three Schemes

Color feature	*Powdery mildew*	*Stripe rust*	*Leaf rust*	*Ye Kubing*	*yellow dwarf*	*yellow mosaic*	*Average accuracy rate*
I	13	7	13	13	5	1	63.81%
II	14	12	11	13	5	5	71.43%
III	14	7	5	13	5	3	59.05%

Comparing with TABLE II and TABLE III, the three schemes have higher accuracy rate while recongnizing 4 disease samples, but drop significantly while using 6 disease samples. This is because less disease parameters lead to similar features of different diseases that can not effectively distinguish more diseases. More feature parameters, such as shape and texture of the disease, should be added to recognize more diseases.

IV . Conclusions

By using the computer vision technology, a diagnosis algorithm of wheat disease is proposed based on color feature and RBF SVM. Experimental result shows that the diagnosis of the wheat disease using color feature achieved good result. Among the three schemes of disease parameters, normalized R, G, B is the best scheme, with the average accuracy rate up to 96%.

In order to improve the applicability of the algorithm, further improvement is needed, especially shape and texture features of the disease should be added to achieve more reliable diagnosis.

References

[1] Sena J D G, Pinto F A C, Queiroz D M,et al. Fall armyworm damaged maize plant identification using digital images. Biosystems Engineering 85(3) (2003) 449-454

[2] MIGUEL P, RUBÉN O, IGNACIO I, et al. New method to assess barley nitrogen nutrition status based on image colour analysis: Comparison with SPAD-502.Computers and Electronics in Agriculture 65(2) (2009) 213-218

[3] CAMARGO A, SMITH J S. Image pattern classification for the identification of disease causing agents in plants. Computers and Electronics in Agriculture, 2009,66(2): 121-125

[4] ALBERTO T, XAVIER P B, GONZALO P,et al. A new vision-based approach to differential spraying in precision agriculture. Computers and Electronics in Agriculture 60(2)(2008)144-155

[5] CHANG S G,YU B,MARTIN V. Adaptive wavelet thresholding for image denoising and compression.IEEE Transactions on Image Processing 9(9)(2000)1532-1546.

[6] LIE J,LYSAKER M,TAI X C.A variant of the level set method and applications to image segmentation. Math Computation 75(255)(2006)1155-1174.

[7] HAIM P, JOSEPH F, IAN J. A study of Gaussian mixture models of color and texture features for image classification and segmentation. Pattern Recognition 39(4)(2006) 695-706

[8] OpenCV library [EB/OL]. http://sourceforge.net/projects/opencvlibrary, 2012/12/01