The Journal of Animal & Plant Sciences, 25 (3 Suppl. 1) 2015 Special Issue Page: 42-45
ISSN: 1018-7081

PLANT DISEASE RECOGNITION BASED ON PLANT LEAF IMAGE

S. W. Zhang* Y. J. Shang and L. Wang

SIAS International University, Xin Zheng Zheng Zhou, Henan 451150, P. R. China
*Corresponding author:wjdw716@163.com

ABSTRACT

Plant disease automatic detection is an important research topic as it has been proved useful in monitoring large crop fields, and thus automatically detects the leaf disease symptoms as soon as they appear in plant leaves. In this paper, a plant disease recognition method is proposed based on plant leaf images. First, the spot is segmented, and the disease feature vector is extracted. Then, the extracted features are provided for the K-nearest-neighbor classifier to recognize the plant diseases. Experimental results show the effectiveness of the proposed approach.

Key words: Leaf image segmentation, Plant disease identification, Feature extraction.

INTRODUCTION

As known, plants are very important for human beings. The photosynthesis of plants can maintain the balance of carbon dioxide and oxygen in the atmosphere. At the same time, plants are important resources of food and some products, and they also play a vital role in water conservation, inhibiting desertification and improving climate. However, the plant diseases many cause significant reduction in both quality and quantity of agricultural products (Ananthi and Varthini, 2012; Wang *et al*., 2008; Camargo and Smith, 2009; Arivazhagan*et al*., 2013). In 1943, in north eastern India, it is estimated that the outbreak of the rice helminthosporiose caused a heavy loss of food grains and death of a million people (Ananthi and Varthini, 2012). In 2007, in Georgia (USA), it is estimated that the plant disease losses was about $539.74 million, about $185 million was spent to control the diseases, and the rest was the value of damage caused by the diseases. So plant disease resistance and management are crucial to the reliable production of food. In fact, about 80% to 90% of disease on the plant is appeared on its leaves. So we interest in the plant leaf rather than whole plant. There are many leaf based plant disease recognition methods (Sabine Bauer *et al*., 2011; Al-Bashish*et al*., 2011; Al-Hiary*et al*.,2011; Dheeb*et al*., 2010; Arivazhagan*et al*., 2013). But, there is not an effective method because of the complexity of color and shape of the disease leaves. In this paper, a disease recognition method is proposed and the major steps of plant disease identification are introduced.

Disease leaf image preprocessing: In our disease recognition method, to ensure the high disease recognition accuracy, we need to segment the disease leaf imageand obtain the spot images before extracting the recognition features, and the ultimate goal of segmentation is to separate leaf from the back-ground and obtain the spot of leaf. We require that the picture of leaves must be against light and untextured background. Taking into account both the software running speed and the segmentation and recognition results, we adopt threshold segmentation (Haralick*et al*., 1985; Otsu, 1979) in this process, after that we remove small bad regions and the stem of the leaves. The area of the leaf is cropped out. This segmentation method is based on color pixels, and it works well for our system to be an interactive application. At last, the segmentation method is used to extract the diseased region and the plant disease is graded by calculating the quotient of disease spot and leaf areas. An optimal threshold value for plant disease leaf segmentation can be obtained by weighted Parzen-window. This method can reduce the computational burden and storage requirements without degrading the final segmentation and disease recognition results.

The segmentation approach is introduced based on the color pixels. The RGB space is converted firstly to gray,

$$Gray = 0.299R + 0.587G + 0.114B \quad (1)$$

where*R*, *G* and *B* are three matrices of different color channels of the original disease leaf image, and the *Gray* is the matrix of gray image.

Threshold technique is an important and efficient technique for disease leaf and spot image segmentation. First, we convert the *RGB* space into gray, and then binarize the image according to the gray level based on a clip-level (or a threshold value), and selection of threshold value for each image is performed automatically without human intervention. According to the experiment, the best threshold value *M* is set to the mean value of each image matrix, so we define

$$B(x,y) = \begin{cases} 1, & \text{if } g(x,y) > M \\ 0, & \text{if } g(x,y) \le M \end{cases} \quad (2)$$

where x, y are the coordinates of the threshold value point, $g(x,y)$ are points of the gray level image pixels, $B(x, y)$ are points of binary level image, M is the mean value of the disease leaf image data.

By human observation, there is no uniform layout, width or length for leaf stems, which usually failures. Stem may increase the inter-class scatter but increase the intra-class scatter, simultaneously, but it is no benefit for the disease identify, usually results in disease recognition failures. Therefore, we decide to remove the stems in the image preprocessing. And at this part, for standardize the result, after filtering to eliminate the influence of noises; we remove the stem after the threshold segmentation through the methods called Otsu (1979) and. Tophat transformation (Zhong-Qiu*et al*., 2013) can determine the set of all thin structures that are protrude from the leaf, and it works by designing a structuring disc element with a diameter longer than the width of stem. For a binary image B and structuring element S, the Tophat method is

$$T(B) = B - B \oplus S \qquad (3)$$

where $\oplus$ denotes the opening operation, i.e., an erosion followed by a dilation.

Disease leaf image feature extraction: Feature extraction for plant disease leaf image identification faces many challenges such as rotational invariance, scale invariance, and so on, which are also common in other CBIR applications. After segmentation, we have a single leaf spot which is segmented from its background and normal part, and then we need excellent features for disease identification. However the algorithm should be as simple as possible to meet the requirements of real-time, and we can use integral measures to compute the features at a boundary point (Manay*et al*., 2006), so our experiment is designed with several little features. We compute the color (Huang *et al*., 2010), HSV (Ma *et al*., 2009) and texture features for experiments and made a choice after the comparison among them.

(1) Color features. Color (Huang *et al*., 2010) is a widely used feature. Many methods have been proposed to describe color feature, among which, the color moments have the lowest dimension and lower computational complexity, and easily to be extracted. So, it is more suitable for disease leaf image retrieval. The color can be defined by 3 or more channels, but we restrict it to the namely, Hue (H), Saturation (S) and Value (V) space. The color moments are calculated for each of these channels in an image. And the three color moments can then be defined as:

$$E_i = \sum_{j=1}^{N} \frac{1}{N} P_{ij},\ v_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(P_{ij}-E_i)^2}, S_i = \sqrt[3]{\frac{1}{N}\sum_{j=1}^{N}(P_{ij}-E_i)^3} \ (4)$$

where$N$ is the feature dimensionality, E_i , v_i , S_i are Mean, Standard deviation and Skewness respectively; P_{ij} is defined as the i th color channel at the j th image pixel.

HSV is a perception-oriented nonlinear color model, in which, color signals can be expressed as three kinds of attributes (Manay*et al*., 2006; Huang *et al*., 2010; Ma *et al*., 2009). Where, H refers to the wavelength of the light which is reflected from an object or comes through it; S refers to the color depth, which is measured in percentage, ranging from 0 to 100% (full saturation); V is indicated in percentage, ranging from 0 (black) to 100% (white). The color feature extraction processes are as follows, convert the RGB image to HSV image, and divide it to 3×3 grids, and divide the HSV image to three components (H, S, V) at first; then get the histogram equalized of the H, S and V components of each grid. The method can lead to better structure views in images, and conducive to subsequent treatment; calculate three moments for the H, S and V histogram respectively. Based on the color model of substantial analysis, we convert the RGB image to HSV image, and divide color into 32 parts; saturation and intensity are divided into 8 parts separately; then 48-dimensional feature vector can be set through histogram statistics, noted as L1.

For a plant disease leaf spot, the color changes with the season, but the shape and texture relatively remains un-changed. Thus, the shape and texture based features are more suitable for the disease leaf identification.

(2) Shape features. The classifying features of the crop disease spot are extracted by Eq.(5), which are represented by eccentricity S_{ecc} , roundness S_{cir} , complexity S_{com} and shape S_{fac} , respectively.

$$S_{ecc} = \frac{L_{long}}{L_{short}}, S_{cir} = \frac{R_{incride}}{R_{excride}}, S_{com} = \frac{P^2}{Area}, S_{fac} = \frac{4® \cdot Area}{P} \qquad (5)$$

where L_{long} , L_{short} are the length of long and short axis, respectively, $R_{incricle}$, $R_{excricle}$ are radius of the Circle and circumcircle of disease spot area; P is the circumference of a lesion, area is the square of the spot.

Thus, the shape feature vector is formed as $L_2 = [S_{ecc}, S_{cir}, S_{com}, S_{fac}]$.

(3) Texture features. The disease images texture feature vector L_3 includes 5 features, i.e., energy, contrast, moment of inertia, correlation and entropy of lesion area, which are computed as follows:

$$T_{ene} = \sum_{i=0}^{255}\sum_{j=0}^{255} p(i,j)^2, T_{con} = \sum_{i=0}^{255}\sum_{j=0}^{255}(i-j)^2 p(i,j),$$

$$T_{inv} = \sum_{i=0}^{255}\sum_{j=0}^{255}\frac{p(i,j)}{1+(i-j)^2}, T_{cor} = \sum_{i=0}^{255}\sum_{j=0}^{255}\frac{ijp(i,j) - «_x «_y}{\pm_x \pm_y},$$

$$T_{ent} = \sum_{i=0}^{255}\sum_{j=0}^{255} p(i,j)\log_2[p(i,j)] \qquad (6)$$

where $\pm_x, \pm_y$ and $«_x, «_y$ are variances and mean values in the x and y components, respectively; $p(i, j)$ is the normalized gray-level co-occurrence matrix; i and j are the pixel gray value of the disease spot image.

These above statistical features extracted from the color, shape and texture of each spot image constitute the classification feature vector of disease leaf image, as [L_1, L_2, L_3]. When plant disease recognition based on many extracted features is carried out, it is better to reduce the dimensions of the obtained feature data firstly and then to conduct plant disease recognition using the optimal classifier.

Disease leaf recognition process: First, the plant disease leaf images are acquired using a digital camera. Then the image preprocessing and pattern recognition methods are applied to the acquired images to extract useful features for further disease recognition. Several classifiers are used to classify the images according to the specific problem at hand. Fig.1 shows the basic procedure of the proposed method for disease recognition in this paper.

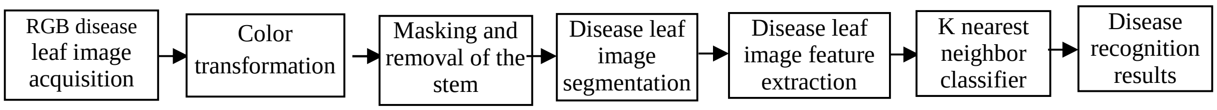

Fig.1 Flowchart of the proposed method for disease recognition.

In the initial step, the RGB images of all the leaf samples are picked up. The step-by-step procedure of the proposed system:

(1) Acquisition and preprocessing the disease leaf RGB image;
(2) Transforming the input disease leaf image from RGB to HSV format space;
(3) Masking the green-pixels and remove the stem and the masked green pixels;
(4) Segmenting the spot components, obtain the useful spot;
(5) Extracting the recognition features of color, shape and texture;
(6) Configuring the K nearest neighbor classifier for the disease recognition.

Experiment results: The original plant disease leaf images are preprocessed before extracting the classifying features and constructing templates and dimensional reduction, i.e., convert the input image from RGB to HSV format; mask the green-pixels; removal of masked green pixels; segment the spot components; obtain the useful spots. And all preprocessed images are cropped and normalized (in scale and orientation) and resized to 32×32 pixels by histogram equilibrium with 255 gray levels per pixel and with the white background. Finally, each processed leaf image is represented as point in vector space. Fig.2 shows the maize disease leaf image and its spot. We select 100 images of the maize disease leaves from 5 kinds of maize disease; each kind of disease has 20 images.

After mapping the RGB components of the input image to the threshold image, the co-occurrence features are extracted. The shape and texture features for the leaves are extracted and compared with the corresponding feature values stored in the feature library.

All feature vectors of the 100 maize disease leaf images are randomly divided into training set and test set. We perform 3-fold-cross-validation strategy, i.e., two-thirds of leaf images of each kind of disease are selected to form the training set, while the remaining images are used as testing set. The disease recognition experiments

Fig.2 Disease spot segmentation results. Left is original image, right is the segmentation result

are done using Matlab function k-nearest neighbor classifier: [class_predictions]=knn(traindata, trainlabels, k, testdata) to produce the class labels on the test data. This disease recognition experiment procedure is repeated 50 times. The average classification rate of the maximum recognition rate in every experiment is recorded as the average classification accuracy. The standard deviation over the 150 trials is counted. To verify effectively the proposed method, we compare its results with that of the three methods which are based on color and texture (Color and Texture, CT) (Al-Bashish*et al*., 2011), Principal Component Analysis and Neural Networks (PCA+NN) (Haiguang*et al*., 2012) and Bayesian(Zhao *et al*., 2007). In CT method, the classifying features are extracted and the diseases are recognized by neural networks. In PCA+NN method, the classifying features are reduced by PCA, and the diseases are classified by NN. In Bayesian method, the classifying features are selected by Bayesian. Table 1 shows the recognition rates and variances versus the number of training samples by the four methods.

From Table 1, the experimental results show that the proposed method is effective. The classification rate is above 90%.

In the experiments, the reasons for misclassification of the plant disease are concluded as follows: the symptoms of the diseased plant leaves vary at the beginning, tiny, dark brown to black spots, while, at later time, they have the phenomena of withered leaf, black or part leaf deletion, etc. To improve the plant disease identification rate at various stages, we need to increase the training samples and extract the effective features from leaf color, leaf shape and leaf texture.

Table1. The recognition results of maize disease leaves by four methods.

Training samplesMethod	10	15	18
CT	82.47±2.63	82.30±2.36	89.27±2.56
PCA+NN	80.37±2.43	81.22±2.30	86.95±2.88
Bayesian	78.44±2.56	84.48±2.37	89.25±2.76
The proposed method	82.14±2.75	85.65±2.59	90.30±2.49

Conclusions: In this study, a plant disease based leaf recognition method was introduced in this paper. The proposed algorithm was varsities on five kinds of maize diseases. The experimental results indicate the proposed method can recognize and classify the plant diseases with high recognition rate.

Acknowledgment: This work was supported by the grant of the basic and frontier technology research projects of Henan Province (142300410309&142300410853), also supported by the Henan provincial education department Foundation under grant No. 15A520101, and the grants of the introduction of talent project of SIAS University (2012YJRC01, 2012YJRC02).

REFERENCES

Al-Bashish,D., M. Braik andS. Bani-Ahmad (2011). Detection and classification of leaf diseases using K-means-based segmentation and neural networks based classification. Inform. Technol. J.10: 267-275.

Al-Hiary,H., S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh(2011). Fast and Accurate Detection and Classification of Plant Diseases. International J. Computer Applications (0975-8887),17(1):31-39.

Ananthi,S. and S. V. Varthini(2012). Detection and classification of plant leaf diseases. IJREAS, 2(2):763-773.

Arivazhagan,S., R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini(2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Ananthi Agric Eng Int,CIGR J.,15(1): 211-217

Arivazhagan,S., R. N.Shebiah, S.Ananthi and S.V.Varthini(2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR J.,15(1):211-217.

Camargo, A. and J. S. Smith(2009). An image processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering,102(1):9-21.

Dheeb, A.B., B.Malik and S.Bani-Ahmad(2010). A Framework for Detection and Classification of Plant Leaf and Stem Diseases. International Conference on Signal and Image Processing, 113-118.

Haiguang, W., L. Guanlin, M. Zhanhong and L.Xiaolong (2012). Image Recognition of Crop Diseases Based on Principal Component Analysis and Neural Networks. 2012 8th International Conference on Natural Computation (ICNC):246-251

Haralick,M. and L. G. Shapiro(1985). Image segmentation techniques. Computer vision, graphics, and image processing. 29 (1): 100–132.

Huang,Z. C., P. P. Chan, W. W. Ng and D. S. Yeung(2010). Content-based image retrieval using color moment and gabor texture feature, Machine Learning and Cybernetics (ICMLC), 2010 International Conference on. 2: 719–724.

Ma ,J. Q (2009). Content-based image retrieval with HSV color space and texture features. Web Information Systems and Mining, 2009. WISM International Conference on. 61–63.

Manay,S., D. Cremers, B.W. Hong, A. J. Yezzi and S. Soatto(2006). Integral invariants for shape matching. Pattern Analysis and Machine Intelligence.28 (10): 1602–1618.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. 9: 62–66.

Sabine Bauer,D., K. Filip and F. Wolfgang (2011). The Potential of Automatic Methods of Classification to identify Leaf diseases from Multispectral images. Precision Agric,12:361-377.

Wang, X., M.Zhang, J. Zhu and S. Geng(2008). Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN). International J. Remote Sensing, 29 (6): 1693–1706.

Zhao Y.X., K.R.Wang and Z.Y. Bai (2007). The application of Bayesian method in maize leaf disease image. Computer Engineering and Applications, 43(5): 193-195. (in Chinese with English abstract).

Zhong-Qiu.Z., M. Lin-Hai and W.Jing (2013).ApLeaf: An Android-based Plant Leaf Identification System. Neurocomputing.00: 1–11.