

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Design of Automatic Recognition of Cucumber Disease Image

Peng Guo, Tonghai Liu and Naixiang Li
School of Computer and Information Engineering,
Tianjin Agricultural University, Tianjin, 300384, China

Abstract: An automatic recognition method for cucumber disease images is presented. Threshold for image segmentation was generated with 2 dimensional maximum entropy principle and optimized with differential evolution algorithm. With threshold values generated, we segmented cucumber disease images and picked up the lesion with maximum area from segmentation results as representative lesion. Then we analyzed representative lesions of disease images and extracted their color features and texture features. Bayes classifier model for cucumber disease recognition was built with information of color features and texture features of representative lesions. We applied our method to recognition of cucumber powdery mildew leaf image, anthracnose leaf image, botrytis leaf image and downy mildew of leaf image which were collected under natural conditions. Experimental results show that automatic recognition approach for cucumber disease image can recognize cucumber disease images without human interaction and high performance of our method.

Key words: Image recognition, image segmentation, automatic threshold generation, bayes classifier, cucumber disease

INTRODUCTION

Agricultural disease image recognition with computer vision technology can supply fast, accurate, efficient diagnosis of diseases. Disease image recognition based on image color statistical features (Cen *et al.*, 2007) classify disease sample images with image color statistical features. Crop disease is characterized in lesion, method using image color statistics processes whole pixels of the leaf and requires a large amount of computation. Image segmented with Otsu algorithm and combination of BP neural network with color and texture features of the lesion can realize image recognition of crop diseases (Zhang *et al.*, 2012; Jia and Ji, 2013). Plant disease recognition can also be realized with K-means clustering in high dimensional space with texture features and color features of disease image (Wang *et al.*, 2009). As for the method which segmented lesion from crop disease image, there are some disease lesions in an image, selecting the most representative lesion, analyzing its image features and applying them to disease classification has excellent application prospect.

We propose an automatic disease image recognition method and apply our method to classification of cucumber diseases images. Threshold for image segmentation is generated with 2 D maximum entropy (Zhou *et al.*, 2001) and optimized with Differential

Evolution algorithm (DE) (Storn and Price, 1997), representative lesion is selected from image segmentation result and Bayes classifier is used to judge disease types. We test our method with recognition of cucumber powdery mildew leaf image, anthracnose leaf image, botrytis leaf image and Downy mildew of leaf image which were collected under natural conditions, our method can operate without human interaction.

AUTOMATIC IMAGE SEGMENTATION

DE is an evolution algorithm based on population difference, its mutation are results of Individuals arithmetic combination (Storn and Price, 1997). DE has been applied in many fields because it is easy, simple and robust. Maximum entropy (Zhou *et al.*, 2001) selects threshold under condition of maximum edge area entropy after computation of 2D entropy based on gray-gradient co-occurrence matrix. To obtain optimized threshold, DE is used to optimize threshold generated with 2D maximum entropy because threshold generated with only maximum entropy skew the results.

Fitness computation: Fitness function values determine selection of individuals in evolutionary algorithm. In accordance with ref (Zhang *et al.*, 2007), the optimal of threshold vector $\{s,t\}$ is defined as:

$$\varphi(s^*, t^*) = \max_{\substack{s=0,1,2,\dots,L-1 \\ t=0,1,2,\dots,L-1}} \{\varphi(s,t)\}$$

we define the fitness function as:

$$\text{fitness} = \frac{1}{\varphi(s^*, t^*)} \quad (1)$$

Individual encoding: In our approach, individuals are encoded with tuples (i, j). Because there is only two digits in tuples (i, j), we make individuals with n tuples, individuals are organized as:

$$(i_1, j_1, i_2, j_2, \dots, i_n, j_n) \quad (2)$$

where, i is gray level of pixel (x, y) and j is neighborhood average gray level of (x, y). In our experiments, n is set to 8. In initialization, individuals are randomly generated integers belong to (0, L-1) and these trial integers are selected uniquely. L is the maximum gray level.

Termination criterion: We set maximum iteration is 100 or the result has not changed for generations.

Image segmentation with threshold algorithm:

- Step 1:** Reading image into Img and translating it into grayscale, converting image form RGB to HSI format to obtain H component, S component and I component, setting t to 1, setting neighborhood size k = 3
- Step 2:** Setting parameter crossover rate CR, scale factor F, population size NP and termination criterion
- Step 3:** Computing two-dimensional histogram joint probability, p_{ij}
- Step 4:** Generating initial population as (2)
- Step 5:** Evolution operation of DE
 - Mutation operation of DE
 - Crossover operation of DE
 - Separating (i, j)'s from individuals and computing each (i, j)'s fitness as (1)
 - Selection operation of DE
- Step 6:** If iteration reaches to termination criterion
 - Outputting best values of (i, j), exit iteration of DE
 - Else if $t < 30$
 - $t = t+1$, return to step 5
- Step 7:** Computing average value of (i, j) of 30 DE results as threshold value, Thre

- Step 8:** Segmenting I component with Thre to obtain binary image, tmp
- Step 9:** Morphological processing on tmp
- Step 10:** Logical AND operation on binary image, tmp and disease leaf image, Img
- Step 11:** Outputting segmentation result

REPRESENTATIVE LESION AND FEATURE EXTRACTION

Representative lesion extraction: We obtain disease lesions after image segmentation, but some disconnected lesions result in difficulty for image features analysis, therefore selection of representative lesion is required. The process to obtain representative lesion is:

- Binarizing lesion obtained from image segmentation
- Obtaining row size and column size of leaf lesion
- Defining size of window
- Setting border of image, pixels of area outside image are set to 0
- Marking all lesions with different symbols
- For every lesion, computing the smallest rectangle that contains a single lesion area
- Finding the rectangle with the largest area as representative area
- Obtaining representative lesion after logical operation with representative area and initial disease leaf image

Feature extraction: Image features include color, texture and shape, for leaf diseases with single lesion such as gray mold, ideal shape features can be obtained after image segmentation, but for those with vague border such as powdery mildew, it is difficult to obtain ideal shape feature, therefore, we use color feature and texture feature in image recognition.

Image color feature: Color moment is an effective and simple method for color features (Stricker and Orengo, 1995) and three order moments of color are enough to express image. We select the first order and the second order moment to convey information because there are negative values in the third order moments, Bayes classifier can not process the negative directly. Moments of color used are defined as:

- The first order moment, μ_i is:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N q_{ij} \quad (i = 1, 2, 3) \quad (3)$$

- The second order moment, σ_i is:

$$\sigma_i = \left[\sum_{j=1}^N (q_{ij} - \mu_i)^2 \right]^{\frac{1}{2}} \quad (i = 1, 2, 3) \quad (4)$$

where q_{ij} is the component i of pixel j in an image and $i = 1, 2, 3$ represent red, green and blue, respectively, N means the number of pixels in an image.

Image texture feature: Texture involves distribution of gray value of pixel in an image which is homogeneous property in the image. Gray tone spatial-dependence matrices is constructed with pixel pairs which satisfies specific shift relations and specific gray value. We select Angular second moment, Inverse difference moment and Contrast from statistical functions based on gray tone spatial-dependence matrices defined by Haralick *et al.* (1973), there are:

- Angular second moment:

$$E = \sum_{i,j} p(i, j)^2 \quad (5)$$

- Contrast:

$$C = \sum_{i,j} (i - j)^2 p(i, j) \quad (6)$$

- Inverse difference moment:

$$I = \sum_{i,j} \frac{1}{1 + (i - j)^2} p(i, j) \quad (7)$$

where, $P(i, j)$ is gray level of pixel (i, j) .

BAYES CLASSIFIER FOR DISEASE IMAGE RECOGNITION

Naive Bayes classifier is a kind of simple classifier based on Bayes theorem (Rish 2001). With lesion image features, we build classifier model for disease image recognition based on Bayes theorem. Supposing data of color features or texture features of crop disease image fit normal distribution, Bayes disease classifier is built as:

$$P(d_i | c_i, t_i) = \frac{p(d_i) \times P(c_i | d_i) \times P(t_i | d_i)}{\text{evidence}} \quad (8)$$

where, d_i is the i 's disease and c_i is the color feature of disease d_i , t_i is texture feature of disease d_i , evidence is normal constant and is defined as:

$$\text{evidence} = \sum_{i=1}^N P(d_i) \times P(c_i | d_i) \times P(t_i | d_i)$$

N is the number of types of disease.

Given training data, $\{(c_{1s}, t_{1s}), (c_{2s}, t_{2s}), \dots, (c_{ns}, t_{ns})\}$, we can obtain parameters of disease:

$$\mu_c = \frac{1}{n} \sum_{i=1}^n c_i \quad (9)$$

$$\mu_t = \frac{1}{n} \sum_{i=1}^n t_i \quad (10)$$

$$s_c = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (c_i - \mu_c)^2} \quad (11)$$

$$s_t = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (t_i - \mu_t)^2} \quad (12)$$

With color feature and texture feature pair, (x, y) to be recognized, the probability of disease d_i is:

$$p(x | d_i) = \frac{1}{\sqrt{2\pi s_c^2}} e^{-\frac{(x - \mu_c)^2}{2s_c^2}} \quad (13)$$

$$p(y | d_i) = \frac{1}{\sqrt{2\pi s_t^2}} e^{-\frac{(y - \mu_t)^2}{2s_t^2}} \quad (14)$$

Then we can obtain normal constant, evidence and posterior probability, $P(d_i | c_i, t_i)$, the type of disease is the disease having maximal value of all $P(d_i | c_i, t_i)$:

$$\text{classify}(d_1, d_2, \dots, d_i) = \underset{i}{\operatorname{argmax}} \{P(d_i | c_i, t_i)\} \quad (15)$$

EXPERIMENTS

Parameter setting: In our experiments, F is set to 0.9 and CR is 0.5, the size of population is 50, maximal iteration 100. The scheme of DE we used is DE/rand/1 strategy.

Disease image segmentation: We test our approach with 4 kinds of cucumber leaf disease images in experiments, they are downy mildew leaf as shown in Fig. 1, anthracnose leaf in Fig. 4, gray mold leaf in Fig. 7 and powdery mildew in Fig. 10.

After segmented with threshold generated automatically, the binary lesion areas are: downy mildew leaf as shown in Fig. 2, anthracnose leaf in Fig. 5, gray mold leaf in Fig. 8 and powdery mildew in Fig. 11. Segmented lesions are downy mildew as shown in Fig. 3, anthracnose in Fig. 6, gray mold in Fig. 9 and powdery mildew in Fig. 12. Representative lesions of the 4 disease images are shown in Fig. 13-16, respectively.



Fig. 1: Cucumber downy mildew leaf



Fig. 4: Cucumber anthracnose leaf

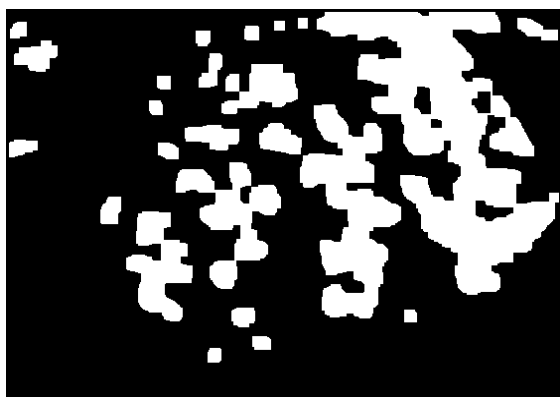


Fig. 2: Binary lesion of downy mildew

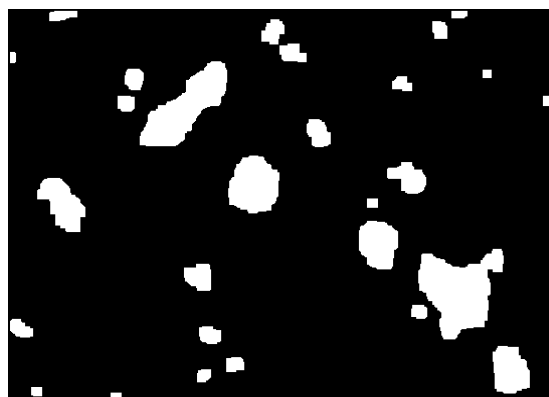


Fig. 5: Binary lesion of anthracnose

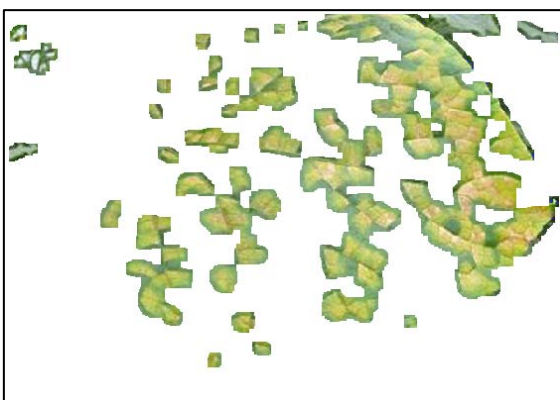


Fig. 3: Lesion of downy mildew

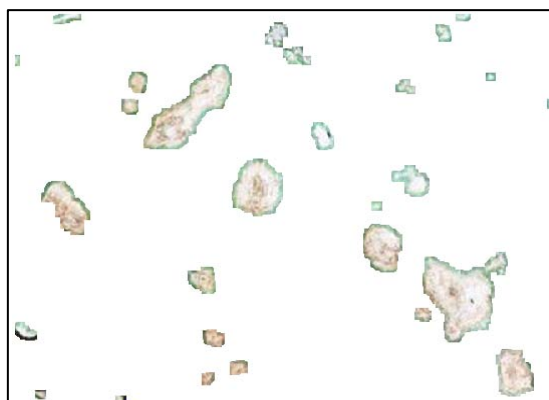


Fig. 6: Lesion of anthracnose

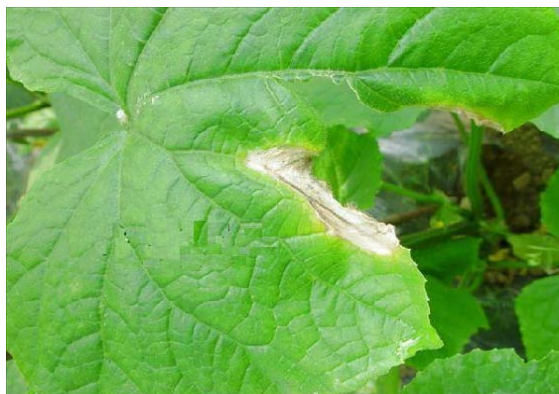


Fig. 7: Cucumber gray mold leaf



Fig. 10: Cucumber powdery mildew leaf



Fig. 8: Binary lesion of gray mold



Fig. 11: Binary lesion of powdery mildew

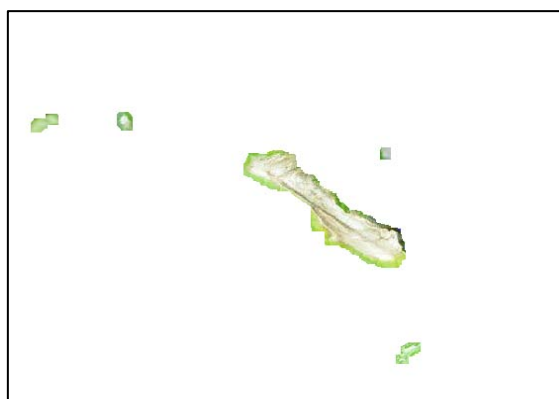


Fig. 9: Lesion of gray mold



Fig. 12: Lesion of powdery mildew



Fig. 13: Representative lesion of downy mildew



Fig. 15: Representative lesion of gray mold

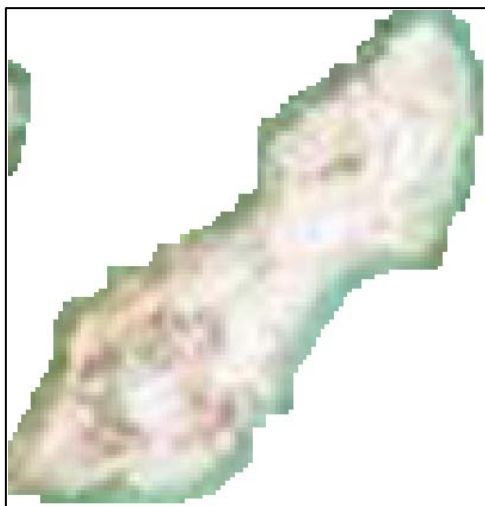


Fig. 14: Representative lesion of anthracnose

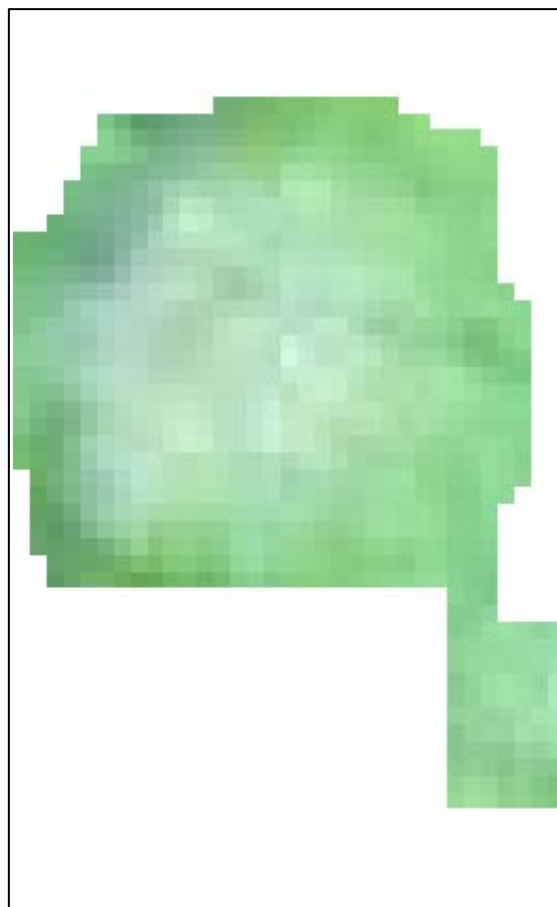


Fig. 16: Representative lesion of powdery mildew

Table 1: Cucumber disease image recognition base on color feature

Image to be recognized		Recognition				Detection accuracy (%)
Diseases	No. of images	Powdery mildew	Gray mold	Downy mildew	Anthracnose	
Powdery mildew	80	68	8	2	2	85.0
Gray mold	45	8	35	0	2	77.8
Downy mildew	100	10	2	85	3	85.0
Anthracnose	60	3	5	4	48	80.0

Table 2: Cucumber disease image recognition base on texture feature

Image to be recognized		Recognition				Detection accuracy (%)
Diseases	No. of images	Powdery mildew	Gray mold	Downy mildew	Anthracnose	
Powdery mildew	80	55	12	10	3	68.8
Gray mold	45	5	30	2	8	66.7
Downy mildew	100	7	4	83	6	83.0
Anthracnose	60	2	6	5	47	78.3

Table 3: Cucumber disease image recognition base on color feature and texture feature

Image to be recognized		Recognition				Detection accuracy (%)
Diseases	No. of images	Powdery mildew	Gray mold	Downy mildew	Anthracnose	
Powdery mildew	80	71	6	0	3	88.8
Gray mold	45	4	38	0	3	84.4
Downy mildew	100	2	1	94	3	94.0
Anthracnose	60	2	2	4	52	86.7

Disease image recognition: Experiments of cucumber disease image recognition with Bayes classifier include recognition on color feature, recognition on texture feature and recognition on both color and texture. We have collected 100 powdery mildew cucumber leaf images, 65 gray mold cucumber leaf images, 80 anthracnose cucumber leaf images and 120 downy mildew leaf images. Before classification, we take 20 samples for every disease to obtain means and standard deviation of every image feature of the 4 diseases. With means and standard deviations obtained, 285 images are analyzed and results are shown in Table 1-3, respectively.

From Table 1-3, we can find that results on color and texture feature are higher than those on color features and on texture features. For powdery mildew, detection accuracy on color is 85.0% and that on texture is 68.8%, result on texture and color is 88.8%. Accuracies of anthracnose are 80, 78.3 and 86.7%. The highest accuracy among the 4 diseases is downy mildew, whose accuracy are 85.0, 83.0 and 94% and the lowest one is gray mold, the results are 77.8, 66.7, 84.4%, respectively. For different disease in Table 3, accuracy of downy mildew is the highest which is 94% and gray mold is the lowest which is 84.4%. From these 3 tables, we can find that recognition accuracy on color and texture are better than those on single features.

CONCLUSION

We presented an automatic recognition method of cucumber disease image, cucumber disease images collected under natural condition are segmented with threshold which are generated with 2D maximum entropy

optimized with Differential Evolution. Representative lesions are selected from segmented result and analyzed to obtain color features and texture features. In classification, we select bayes classifier to recognize plant disease images, the recognition can be run without human intervention and our method show high performance in experiments.

On image feature used in this study, we only use color moment and three texture features. But for different image, different color feature or texture feature will result in different recognition accuracy, selecting more suitable feature to improve accuracy is the job to carry out in the future. Classifier we used is the simplest one in Bayes classifier family, Bayesian network with automated inference and automatic learning is a powerful reasoning tool in artificial intelligence, improving intelligentization of crop disease image recognition with inference function of Bayesian network is our next research in the future.

ACKNOWLEDGMENT

This study is supported by Tianjin University Science and Technology Development Fund Projects of China (20120811), Chinese National Spark Program project (2011GA610012).

REFERENCES

- Cen, Z.X., B.J. Li, Y.X. Shi, H.Y. Huang, J. Liu, N.F. Liao and J. Feng, 2007. Discrimination of cucumber anthracnose and cucumber brown speck base on color image statistical characteristics. Acta Hort. Sin., 34: 1425-1430.

- Haralick, R.M., K. Shanmugam and I. Dinstein, 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybernet.*, SMC-3: 610-621.
- Jia, J. and H. Ji, 2013. Recognition for cucumber disease based on leaf spot shape and neural network. *Trans. Chin. Soc. Agric. Eng.*, 29: 115-121.
- Rish, I., 2001. An empirical study of the naive bayes classifier. *Proceedings of the IJCAI-01 Workshop on Empirical Methods in Artificial Intelligence*, August 4, 2001, Seattle, USA., pp: 41-46.
- Storn, R. and K. Price, 1997. Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.*, 11: 341-359.
- Stricker, M.A. and M. Orengo, 1995. Similarity of color images. *Proceedings of the Storage and Retrieval for Image and Video Databases*, February 9, 1995, San Jose, CA., USA., pp: 381-392.
- Wang, S., D. He, W. Li and Y. Wang, 2009. Plant leaf disease recognition based on kernel k-means clustering algorithm. *Trans. Chin. Soc. Agric. Mach.*, 40: 152-155.
- Zhang, H.L., J.S. Song and X.Y. Zhai, 2007. A 2D maximum entropy based self adaptive threshold segmentation algorithm for SAR image processing. *Electron. Opt. Control*, 14: 63-65, 69.
- Zhang, J.H., L.J. Qi, R.H. Ji, H. Nang and S. Huang, 2012. Cotton diseases identification based on rough sets and BP neural network. *Trans. Chin. Soc. Agric. Eng.*, 7: 161-167.
- Zhou, D.L., Q. Pan, H.C. Zhang and G.Z. Dai, 2001. Maximum entropy thresholding algorithm. *J. Software*, 12: 1420-1422.