The Recognition of Cucumber Disease Based on Image Processing and Support Vector Machine

Tian Youwen
College of Information and Electric Engineering, Shenyang AgriculturalUniversity，Shenyang,110161
youwen_tian10@163.com

Li Tianlai
College of Horticulture, Shenyang Agricultural University，Shenyang,110161,China

Niu Yan
College of Information and Electric Engineering, Shenyang Agricultural University Shenyang,110161,China

Abstract

A new method of recognizing cucumber leaf disease based on computer image processing and Support Vector Machine (SVM) is studied to improve recognition accuracy and efficiency. At first, vector median filter was applied to remove noise of the acquired color images of cucumber disease leaf. Then a method of statistic pattern recognition and mathematics morphology was introduced to segment images of cucumber disease leaf. At last texture, shape and color features of color image of cucumber disease spot on leaf were extracted, and classification method of SVM for recognition of cucumber disease was used. Experimental results indicate that the classification performance by SVM is better than that of neural networks. Recognition correct rate of cucumber disease based on SVM of shape and texture feature is better than that of only using the shape feature. Cucumber disease is recognized more correct and faster based on color feature.

1. Introduction

In the productive process, cucumber often occurs diseases which result in influence on its good quality, effected by various bad biological and abiological factors in environment. Traditional diagnosis in cucumber disease depends on pathogenic identification and infection determination by professional personnel in lab with shortcomings such as slow speed, high cost and poor real-time performance[1]. The cucumber syndrome information can also be acquired by eyes with shortcomings such as slow speed, high strength and strong subjectivity, and some features can't be described quantitatively, which caused deviation in diagnostic result[2]. Cucumber disease has different texture, shape and color feature of disease spot on leaf, caused by different pathogens. So disease spot feature can be extracted by computer image processing. Cucumber disease can be recognized by method of Support Vector Machine.

Support Vector Machine (abbr. SVM), is a new pattern recognition method considering both training error and generalization ability, which has many advantages in solving pattern recognition problems of

978-0-7695-3119-9/08 $25.00 © 2008 IEEE
DOI 10.1109/CISP.2008.29

small set of sample, nonlinearity, high dimension, and 1ocal minimization[3-6].So it had been applied in many problem of agricultural area[7-10]. In this research, preprocessing of original cucumber disease images, segmentation of disease spot and extraction of efficiency feature were done by computer image processing, then cucumber disease was recognized automatically and quantitatively by SVM, for the purpose of remedying the defects of traditional diagnosis, promoting the accuracy and efficiency. It will provide important theory and practical basis for intelligent protection of greenhouse vegetable and variable spraying insecticide, which also has a very important meaning in reducing the gap between developed country and us in protection of plant in automatization, intelligention and digitization, promoting the application of computer image processing and pattern recognition in agricultural engineering area.

2. Image processing of cucumber disease

2.1. Image preprocessing

Caused by collective device, environment and so on, blurry edge of disease spot and spots of disease leaf always occur in color image of cucumber disease. Vector median filter had been taken to enhance image, stress some useful information and get rid of or weaken harmful information.

The method of vector median filter for color image, takes the average of all vectors $X_i=[R_i, G_i, B_i]$ (i=1,2，…，N)in given window to get average vector $\overline{X}$, and calculates the distance between X_i(in given window) and $\overline{X}$, then takes the nearest vector as the output value for window center pixel .

The application of vector median filter for color image enhanced cucumber disease image. The experimental result indicates that vector median filter for color image can not only remove spot and noise but keep detail of edge and don't increase new color. The major cause is that center pixel is replaced by min value of the nearest vector, rather than synthetic vector by R, G, B.

2.2. Segmentation of cucumber disease spot on leaf

The component of cucumber disease image is complex, and the disease spot with the light or deep color was arranged on leaf disorderly. Various diseases had different spot color. As the image of cucumber disease on leaf shown, disease leaf was composed by disease spot and natural part, so segmentation is a two class problem in essence. Classification method by statistic pattern recognition was used to segment the image.

Given discriminant function is linear, the normal formula is

$$g(x)=\mathbf{w}^T x+w_0 \qquad (1)$$

Where x is three dimensions feature vector, w is weight vector, respectively expressed by

$$x=[R\ G\ B]^T,\ w=[w_1,w_2,w_3] \qquad (2)$$

w_0 is threshold value.

Design the classifier as follow steps:

(i) Choose two kinds of training sample sets of disease spot and natural part separately.

(ii) Complete statistic analyse on every kinds of sample in RGB color space, then calculate its m_1, m_2, s_1, s_2 to gain w*.

(iii) According to $y=w^{*T}x$, gain one dimention sample set, and determine classification threshold value y_0.

(iv) Scan image, according to $y=w^{*T}x$, determine projection point y of every pixel.

(v)According to decision rules,

$$y \gtrless y_0 \begin{cases} x\in L_1 \\ x\in L_2 \end{cases}$$,decide and classify.

Above classifier was adopted to segment cucumber disease image of powdery mildew and downy mildew. The segmentation result is better, which can classify

the disease spot on leaf from natural part basically.

There are some lonely spot, stab and small empty hole in cucumber disease image after segmentation. Open and close operation of mathematical morphology was adopted to eliminate noise for the purpose of reducing the influence of later recognition. At the beginning, open operation was introduced to remove lonely spot ,stab, then close operation was introduced to fill small empty hole. Ideal image was gained.

3. Extraction of cucumber disease feature

As the image of cucumber disease shown, various of diseased varieties had different texture, spot shape and color, hence the texture, shape and color feature of disease spot were extracted for study.

3.1. Extraction of color texture feature

The selection of texture feature of cucumber disease for color image is based on chromatology concept of CIE XYZ color space. In CIE XYZ color space, each pixel produces a pair of (x,y).The trace of chromatology for an image I is defined as follows:

$$T(x,y)=\begin{cases}1 & if\ \exists I(i,j)\{i,j)\ produce se(x,y)\\ 0 & otherwise\end{cases} \quad (3)$$

Two dimension distribution of chromatology image I is defined as:

$$D(x,y)=n_k \quad (4)$$

Where n_k = # pixels producing of (x,y).

The (m,1) sets of chromaticity moment are defined respectively as:

$$M_T(m,l)=\sum_{x=1}^{x_s}\sum_{y=1}^{y_s}x^m y^l T(x,y) \quad (5)$$

$$M_D(m,l)=\sum_{x=1}^{x_s}\sum_{y=1}^{y_s}x^m y^l D(x,y) \quad (6)$$

Where m,l =0,1,2……,and x_s, y_s are the discrete dimensions of the x-y space.

Through experimental analysis[11], chromaticity moment set CM55, that is expressed by $M_D(1,0)$、$M_D(0,1)$、$M_D(1,1)$、$M_D(2,1)$、$M_D(1,2)$ and $M_T(1,0)$、$M_T(0,1)$、$M_T(1,1)$、$M_T(2,1)$、$M_T(1,2)$，was selected as the fittest one as feature vector of texture feature of cucumber disease for color image.

3.2. Extraction of shape feature on disease spot

Shape features were normally formed by area, roundness, shape complexity, extending length and concavity, equivalent rate of longer axis and shorter axis of ellipse and so on.

Through the recognition experiment based on assemble of different shape feature of cucumber disease spot, it can conclude that adoption of roundness, shape complexity, extending length and concavity as feature vector for disease spot has better performance[11].

3.3. Extraction of color feature on disease spot

Every spot in CIE L*a*b* color space can be seen as a spot in three dimension of CIE L*a*b*. It can be derived from three basic value x, y, z[12].The differences of two kinds of color can be measured by enclidean distance of them. The color can be compared directly in CIE L*a*b*color space, so it is effective in measure of small color difference .Hence the value of CIE L*a*b* in CIE L*a*b* color space is adopted as the color feature of disease recognition in this research.

4. SVM method

A linear and separable sample set (x_i ,y_i) belonging to separate classes can be separated correctly by hyperplane. The distance between the closest vector to the hyperplane is longer, the generalization ability is better. The longest hyperplane is the optimal hyperplane. Optimal classification function can be gained by quadratic programming as follows [3-6]:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i^* y_i (x_i \bullet x) + b^*) \quad (7)$$

Where α_i* is Lagrange multiplier; b* is threshold.

For linear and nonseparable problem, a non-negative relaxation ξ_i and wrong classification penalty constant C are introduced, and the gained optimal classification function is equivalent to formula (7).

For nonlinear problem, nonlinear problem should be transformed to linear problem in high dimension space by kernel function, then the optimal hyperplane will be gained in invariant space, and this is SVM. Now, the classification function is:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, x) + b^*) \quad (8)$$

Normal kernel function has[5，6]:

(i)Linear kernel function :

$$K(x, y) = x \bullet y \quad (9)$$

(ii)Polynomial kernel function :

$$K(x, y) = (x \bullet y + 1)^q, q = 1, 2, \cdots\cdots \quad (10)$$

(iii)Radial basic kernel function:

$$K(x, y) = \exp(-\frac{\|x - y\|^2}{2\sigma^2}) \quad (11)$$

(iv)Sigmoid kernel function:

$$K(x, y) = \tanh(v(x \bullet y) + c) \quad (12)$$

5. Experimental results and analysis

Forty disease images of cucumber powdery mildew and downy mildew had been collected in field of cucumber research base of Shenyang Agriculture University, and eighty subimages with 120*60 pixel of them had been selected as training sample from every kinds of disease image, then other forty subimages had been selected as testing samples. Then preprocessing and segmentation were done to these subimages.

5.1. Recognition of cucumber disease based on texture feature

Calculate the chromaticity moment set CM55 of every diseased texture of color image after preprocessed.

Four kinds of different kernel function had been compared in experiment, and the recognition experiment had been also done in same sample set and testing set by BP artificial neural net. Table 1 shows the recognition result, in which parameter c=10, ζ=0.001, for polynomial kernel function q=3, for radial basic kernel function σ=3, for Sigmoid kernel function v=3, c=-10. The average of five better results is the recognition result in artificial neural net(ANN).

Table1. **Recognition results of cucumber disease based on texture feature**

Method	Kernel function	Number of support vector	Average correct rate of recognition	Running time（ms）
SVM	Linear	25	100%	47
	Polynomial	43	95%	63
	Radial basic	48	90%	63
	Sigmoid	29	95%	32
ANN			78%	924

The table1 shows that the recognition of SVM has higher correct rate, linear function has the best recognition performance and stability, all kinds of kernel function almost have better recognition rate, seen from running time. While artificial neural net has worse performance not only in correct rate but also in

running time. So linear kernel function is the fittest method for recognition of cucumber disease based on texture feature of diseased spot.

5.2. Cucumber disease recognition based on shape and texture feature

For every image preprocessed and segmented, recognition of downy mildew and powdery mildew had been done in two ways, using shape feature only or using shape feature(roundness, shape complexity, extending length and concavity of disease spot） and texture feature(CM55 chromatology moment) together based on SVM. The result was shown as table 2, in which parameter c=10, ζ=0.001, for polynomial kernel function q=3, for radial basic kernel function σ=3/2, for Sigmoid kernel function v=1/3, c=-10. The result of only using shape feature is on left of"/", the other is on right.

Table2. **Recognition results of cucumber disease based on shape and texture feature**

Kernel function	Number of support vector	Correct rate of recognition		Average rate of recognition	Running time （ms）
		powdery mildew	downy mildew		
linear	30/44	95%/100%	80%/100%	87.5%/100%	78/47
polynomial	41/63	95%/100%	75%/100	80%/100%	63/63
radial basic	35/116	95%/95%	90%/95	92.5%/95%	63/62
Sigmoid	53/129	50%/100%	90%/80	70%/90%	47/31

As shown in table2, SVM with radial basic kernel function has the best performance on recognition of cucumber disease when only using shape feature; SVM with linear kernel function has the best performance when using two features together. Generally speaking, the correct rate of recognition based on shape and texture feature is identical that on texture only, and both of them have 12.5% higher than only using shape feature.

5.3. Recognition of cucumber disease based on color feature.

For every image preprocessed and segmented, color feature(L*、a*、b* belongs to disease spot) had been taken to recognize downy mildew and powdery mildew of cucumber based on SVM. The result shows as table 3, in which parameter c=1, ζ=0.001, for polynomial kernel function q=2, for radial basic kernel function σ=1, for Sigmoid kernel function v=1/3,c=-1.

Table3. **Recognition results of cucumber disease based on color feature**

Kernel function	Number of support vector	Rate of recognition		Average rate of recognition	Running time （ms）
		powdery mildew	downy mildew		
linear	21	100%	100%	100%	62
polynomial	39	100%	100%	100%	62
radial basic	29	100%	100%	100%	47
Sigmoid	97	100%	100%	100%	32

As shown in table 3, the effect of recognition is good and reaches 100%, hence the conspicuous color difference of disease spot between downy mildew and powdery mildew. Seen from running time, Sigmoid kernel function is the fast one. Seen from the number of SVM, linear kernel function is the least one. Considering above factors generally, it can conclude that linear kernel function is fittest for recognition of cucumber disease based on color feature of disease spot.

6. Conclusion

The recognition of cucumber disease was studied by image processing and SVM. The results as follows:

(i) Vector median filter in color image can remove the noise on disease leaf effectively.

(ii) Utilizing of statistic pattern and open - close operation of mathematic morphology can separate the disease spot from natural part correctly with ideal effect.

(iii) Essential feature of cucumber disease image can be shown by color texture based on chromaticity effectively. For small training set of sample, utilizing of SVM model has higher recognition rate and reliably in recognition of cucumber disease ,and has higher recognition rate and faster speed than BP artificial neural net in same condition.

(iv) The experiment result of recognition of cucumber disease by SVM shows that correct recognition rate of both using shape feature and texture feature is higher than that of shape feature only. Cucumber disease is recognized more correct and faster based on color feature.

References

[1] Shaojun Han. General Theory of Plant Protection [M]. Higher Education press, Beijing, 2001:48-52

[2] Zhigang Li,Zetian Fu,Liqin Li. Advance in a Agricultural Plant Protection Technology Based on Machine Vision [J]. Transaction of the Chinese Society for Agriculture Machinery ,2005, 36(8):143-146

[3] Burges C J C. A Totorial on Support Vector Machines for Pattern Recognition[J] . Data Mining and Knowledge Discovery, 1998, 2(2):121--169

[4] Steve R Gunn. Support Vector Machines for Classification and Regression[J]. Technical Report,Southampton:University of Southampton,1998:1--28

[5] Zhaoqi Bian, Xuegong Zhang et al.. Pattern Recognition[M]. Qinghua Uinversity press, Beijing, 2002

[6] Xiangdong Liu,Meilin Zhu,Zhaoqian Chen,Shifu Chen. Support Vector Machine and its Applications in Pattern Recognition [J]. Computer Science ,2003, 30(6):113-117

[7] Youwen Tian, Tianlai Li, Chenghua Li, Zailin Piao, Guokai Sun, Bin Wang. Method for Recognition of Grape Disease Based on Support Vector Machine , Transactions of the Chinese Society of Agricultural Engineering, 2007,23(6):134-136.

[8] Youwen Tian, Lidi Wang, Shuhua Jiang. Recognition of Maize Disease Based on Image Processing and SVM [J]. Chinese Journal of Scientific Instrument . 2006, 26(8):2123-2124

[9] Muhua Liu,Wumao Duan,Jing Li,Huadong Zheng. Determination of Beef Marbling Grade by Image Processing and SVM Classification [J]. Journal of Shenyang Agriculture University , 2005, 36(6):650-654

[10] Xiaobo Zou,Jiewen Zhao. Distinguishing Different Cultivar Apples by Electronic Nose Based on Support Vector Machine[J]. Transactions of the Chinese Society of Agricultural Engineering .2007,27(1):146--149

[11] Youwen Tian. Study on Recognition Method of Typical Plant Disease on Leaf based on Computer Image [D]. Shenyang Univercity, Shenyang ,2005:49-53，66-67

[12] H.D.Cheng, X.H.Jiang, Y.Sun, Jingli Wang. Color Image Segmentation: Advances and Prospects[J]. Patter Recognition, 2001.34: 2259-2281