Research on Cucumber Downy Mildew Detection System based on SVM Classification Algorithm

Bingyu Zhou[1, a] , Jingwen Xu[1,b*] , Junfang Zhao[2c], Aiwen Li[1,d] and Qiuyu Xia [1,e]

[1]College of Resources, Sichuan Agricultural University, Chengdu 611130, China

[2]Chinese Academy of Meteorological Sciences, CMA, Beijing 100081, P.R. China

[a]email:bing9yu27@163.com, *[b]Corresponding author: x.j.w@163.com

Keywords: cucumber downy mildew; support vector machine; feature extraction; detection

Abstract. Cucumber, a common economic crop, occupies a large proportion of vegetable cultivation in China. Plant diseases and insect pests, especially the cucumber downy mildew, are important causes for the decrease in the yield of cucumbers. In order to reduce the losses caused by pests and diseases and achieve rapid automatic identification of plant diseases and insect pests, this paper studies machine vision system and disease image detection with support vector machine (SVM) classification algorithm, taking cucumber downy mildew for example. This paper carries out a method study in image acquisition, image preprocessing, feature parameter extraction, and pattern recognition, which obtains satisfactory results. The accuracy of cucumber downy mildew detection reaches 90%, significantly higher than that of artificial recognition.

Introduction

Nowadays, when rapid modern agricultural development has solved the problem of natural environmental limit, diseases are still the biggest threat for greenhouse vegetables cultivation. However, there might be delay and misjudgement of the causes and varieties of the crops' diseases only through visual observation and cultivating experience, which will lead to lower product quality. Therefore, information technology is necessary to realize automatic information acquisition of crop growth, simulating human visual system to replace the human monitoring and identification of crop diseases.

Developed countries have started the researches to realize quality detection of agricultural products using image recognition since 1970s. In 1980s, researches in this field were refined to all aspects of application, including early identification of plant species, crack detection of grain surface, seed quality grading, seed shape recognition, surface damage detection, etc. [1]. In China, the application of computer image processing technology to agriculture has started from early 1990s, mainly applied to information acquisition of crop growth and quality inspection and sorting and classification of agricultural products. The research of image processing technology in the intelligent recognition of vegetable diseases has become popular in recent 20 years [2].

Image sensor and recognition system of plant diseases and insect pests can simulate human vision system to realize intelligent identification of plant diseases and insect pests of vegetables through rapid computer operation.

Cucumber downy mildew, a common major disease for cucumbers planted outdoors or in greenhouse [3]. The disease is swift and damaging. Early prevention and treatment have to be taken to reduce the losses caused by downy mildew. Machine vision system is an automatic, accurate and real-time method to effectively detect cucumber downy mildew.

This paper carrying out a research of computer image processing technology with support vector machine (SVM) classification algorithm, taking cucumber downy mildew for example, is significant for automatic diagnosis of crop diseases and insect pests.

Materials and data

Image source and image acquisition device

The images used in this research are from the following three parts:

(1) The experimental farm of Sichuan Agricultural University. Digital camera is Fuji FinePixS2995 with 1/2.3 inch CCD sensor and 14 million effective pixels.

(2) Vegetable pest control maps and literatures. By looking up maps and paper materials, obtain cucumber downy mildew image information with a color scanner.

(3) Agricultural map websites and SouTu image database engine. Download online picture resources from China's agriculture hospital website (http://www.nyyy.cc/photo/) color spectrum, China NongYeWang agricultural pattern plate (http://tupu.zgny.com.cn/), Baidu picture (http://image.baidu.com/), Google images (http://images.google.com.hk/), etc.

Plant disease image acquisition and sample preparation

Captured images are screened to remove those with serious interference. Images with blur, shadow, or reflections, are saved to an image processing computer. In Windows photo viewer, 320 sample images with the size of 60*60 are obtained with screenshots software HyperSnap-DX5 issued by Hyperionics Technology from the USA, including 100 normal samples and 220 negative samples. 80 normal samples and 200 negative samples are used for the training of the classifier, and the remaining 40 samples for testing.

Research methods

Image preprocessing

Due to the imperfect tools and instruments for image acquisition and light exposure, all digital images have a certain degree of noise and interference, and may even distort and cover the essence of things. Therefore, original captured images have to be preprocessed to strengthen useful information and improve the quality of the sample image, paving the way for feature extraction and classification recognition of the target.

In this research, image preprocessing mainly include: scale normalization, gray-scale transformation, image filtering and noise reduction, contrast enhancement, image scab segmentation, and edge detection of scab.

The introduction of SVM

After expert system, machine learning has become another important research field of artificial intelligence application, and also one of the core research topics of artificial intelligence and neural computation, increasingly widely used in machine vision, pattern recognition, feature extraction, and regression analysis, etc. At present, a new machine learning method called support vector machine (Supportvectormachines, SVM) [4], brought up by Vapnik[5] in the 1990s, has been developed based on statistical learning theory. It is based on structural risk minimization principle, i.e. the decision rules obtained through a small number of training samples can still get small errors for individual samples. With good generalization performance and classification accuracy, higher learning generalization ability, less parameter setting, and faster operation speed, it is better than neural network model in many ways and can solve learning problems effectively.

Extraction of characteristic parameters of cucumber downy mildew pathological leaf

Identified target characteristics can be divided into three broad categories: physical characteristics, mathematical characteristics, and structural characteristics. Physical characteristics and structural features are easy to be accepted by people's sense organs, convenient for people's intuitive identification of objects, while mathematical characteristics are abstract, not easy to understand and know. However, computer automatic identification system has much better ability in dealing with abstract mathematical characteristics. As a result, mathematical characteristic statistics including mean value, variance, and eigenvalues of covariance matrix can facilitate computer automatic recognition of objects [6] and describe the characteristics of the image or image region through image segmentation with gray level histogram. The change of leaf color and appearance caused by diseases, is an important visualized characteristic. Cucumber disease identification with computer image processing technology to simulate human eye recognition mechanism is the reason why we choose color as characteristic parameters. This research adopts RGB model and HSV model to describe the color feature of images in cucumber disease area. Texture, another important visual clue, which can describe the surface properties of image or corresponding photographic field in image area, is a

common characteristic but difficult to describe. There are four texture feature description methods: statistical method, geometrical method, model method, and signal processing method.

The normalization of characteristic parameters

As different characteristics correspond to different physical quantities, the physical significance and value range of the characteristic values vary with the algorithm based on different characteristics, which is not convenient for the analysis and comparison under the same system. Solution to this problem is to normalize the characteristics. This research selects 11 characteristic parameters, including gray level statistics Gr; RGB color characteristics of r, g, b; HSV color features, tonal H, saturation S, brightness V; gray level co-occurrence matrix texture feature, energy E, correlation C, inertia moment I, entropy H1. All characteristic parameters are in m*n matrix; therefore, each characteristic parameter is to be generalized to a value, and all the values constitute a feature vector. This article describes characteristic values with average number and unbiased standard deviation so that each training sample will be converted into a vector of 1*22.

Research results and analysis

Among 320 samples in total, 280 are training samples, and the other 40 are test samples, which are distinguished by SVM. In this article, the maximum and minimum characteristic values of normal samples are set as threshold values, and those without the thresholds are judged to be negative samples.

The related conclusion is as the following Table 1 and Table 2:

Table 1 Classification results of linear kernel function without Gr eigenvalues

Image category	test samples	correct	error	correct rate	error rate
With downy mildew	20	12	2	60%	10%
No downy mildew	20	18	8	90%	90%
Average				75%	25%

Table 2 Test results of complete eigenvalues

Image category	test samples	correct	error	correct rate	error rate
With downy mildew	20	17	1	85%	5%
No downy mildew	20	19	3	95%	15%
Average				90%	10%

The results indicate that:

(1) With gray level statistic characteristic parameters, downy mildew recognition accuracy rate has increased from 60% to 85%, the disease-free detection accuracy rate has increased from 75% to 95%, and the overall recognition accuracy rate has increased from 75% to 90%, which shows that the correct selection of characteristic parameters can significantly improve the downy mildew recognition accuracy rate.

(2) In terms of training samples quantity, there are 80 normal samples and 200 negative samples. In Table 1 and Table 2, the detection accuracy rates of downy mildew are 60% and 85% respectively, and the downy mildew-free detection accuracy rate are 90% and 95% respectively. Obviously, the number of training samples directly affects the accuracy of detection.

(3) In terms of the error rate, with gray statistic characteristic value, the error rate for test samples with downy mildew drops 50%, and that for disease-free samples drops 62.5%, which indicates that Gr characteristic value has a more significant effect in reducing the error rate for disease-free images.

Conclusion

Currently, automatic identification of plant diseases and insect pests is at the research stage of artificial test environment, and there are many problems, including slow recognition, low accuracy, weak adaptability to illumination, and limited recognition methods and technology. It is a long way to go to realize the practical application. The identification accuracy is not high because on the one hand, it is difficult to accurately extract the edge of the image by color as the color of disease area is similar to that of the image background; on the other hand, map recognition methods such as neural network and support vector machine (SVM) do not have abundant samples. The algorithm of image edge detection, which combines features such as color and shape, is one of the keys to solve the problems; secondly, build a database of plant diseases and insect pests characteristics covering major characteristic values of size, color, shape and texture, which provides an accurate portrayal of the complex information of plant diseases and insect pests; thirdly, build a digital dictionary of plant diseases and insect pests, which can identify plant diseases and insect pests by looking up corresponding items automatically.

To develop modern green agriculture, higher requirement is raised for the prevention, rapid detection and removal of plant diseases and insects pests. Automatic and intelligent detection systems are especially necessary for the facility agriculture or large-scale agricultural factories.

Therefore, the study of automatic detection and identification system of plant diseases and insect pests is of great practical value and broad application prospects. Meanwhile, with the development of theoretical research, it will take long time and great effort to constantly adjust automated recognition model and improve the simulation precision of forecasting model and its fitting degree with practical production.

References

[1] SuPing Sang. The Development of Corn Seed Morphological Image Processing Technology and Its Application in Variety Identification[D]. Jinzhong: Shanxi Agricultural University, 2007.

[2] Zhanwu Peng, Xue Wang, Xiuli Si, etc. A Study on Image Preprocessing and Feature Extraction Methods of Intelligent Identification of Greenhouse Vegetable Diseases[J]. Journal of Jilin Agricultural University, 2007, 29 (2): 219-223.

[3] Jiangliang Li, Liujian Liu. Rapid Diagnosis and the Countermeasures of Cucumber Downy Mildew[J]. Journal of Vegetables, 2007 (3): 28-29.

[4] Naiyang Deng, Yingjie Tian. A New Method in Data Mining: Support Vector Machine[M]. Beijing: Science Press, 2004.

[5] Vapinkvn. The Nature of Statistical Learning Theory[M]. New York: Springer Verlag, 2000.

[6] Suhua Liu. A Study on the Classification and Identification of Stored Product Insects Based on Fuzzy Theory[D]. Zhengzhou: zhengzhou University, 2003.