

## Research Article

# Plant Diseases Recognition Based on Image Processing Technology

Guiling Sun , Xinglong Jia , and Tianyu Geng

Nankai University, College of Electronic Information and Optical Engineering, Tianjin 300350, China

Correspondence should be addressed to Guiling Sun; [sungl@nankai.edu.cn](mailto:sungl@nankai.edu.cn)

Received 23 November 2017; Revised 12 March 2018; Accepted 10 April 2018; Published 21 May 2018

Academic Editor: Jose R. C. Piqueira

Copyright © 2018 Guiling Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new image recognition system based on multiple linear regression is proposed. Particularly, there are a number of innovations in image segmentation and recognition system. In image segmentation, an improved histogram segmentation method which can calculate threshold automatically and accurately is proposed. Meanwhile, the regional growth method and true color image processing are combined with this system to improve the accuracy and intelligence. While creating the recognition system, multiple linear regression and image feature extraction are utilized. After evaluating the results of different image training libraries, the system is proved to have effective image recognition ability, high precision, and reliability.

## 1. Introduction

The prevention and control of plant disease have always been widely discussed because plants are exposed to outer environment and are highly prone to diseases. Normally, the accurate and rapid diagnosis of disease plays an important role in controlling plant disease, since useful protection measures are often implemented after correct diagnosis [1].

This system is based on image processing technology and uses MATLAB as the main processing tool. Besides, digital image processing, mathematical statistics, plant pathology, and other relative fields are also considered. Comparing to the traditional image recognition, there are plenty of innovations in image segmentation and system construction. To strengthen the division of the lesion, users have different creative interactive options to meet their own needs. Meanwhile, linear regression model can be used in various types of plant disease.

The remainder of this paper is organized as follows. In Section 2, traditional threshold segmentation methods are introduced. And there are four parts in Section 3 describing this system in detail: improved histogram segmentation method, Disease Recognition System Based on Multiple Linear Regression, multiselective interactive image segmentation methods, and overall process of the disease recognition system. Section 4 shows the results and analysis

of this recognition system, and summary is concluded in Section 5.

## 2. Traditional Threshold Segmentation Methods

Threshold can never be ignored in image processing. Iterative Method, Otsu Method, and 2-Mode Method are the most common threshold segmentation methods. This section introduces these traditional methods.

*2.1. Iterative Method.* The Iterative Method can calculate the threshold in a certain extent automatically. For the iterative process, the Iterative Method includes a prior knowledge concerning the image and noise statistics. And the optimal segmentation threshold can be found by continuously reducing the gray scale mean [2].

*2.2. Otsu Method.* Given the split threshold of foreground and background  $T$ , the foreground image ratio  $w_0$ , the average gray scale  $u_0$ , background image ratio  $w_1$ , and the average gray scale  $u_1$ , the total gray scale of the image is  $u_T = w_0 \times u_0 + w_1 \times u_1$ . When  $T$  makes the variance value  $\sigma^2 = w_0 \times (u_0 - u_T)^2 + w_1 \times (u_1 - u_T)^2$  maximum, it becomes the best segmentation threshold [3].

TABLE 1: The ratio of three methods ( $10^{-3}$ ).

	Minor (15)	Medium (15)	Serious (15)	Average
Improved Method	0.76051	1.23364	0.8973	0.963817
Iterative Method	2.2246	2.85	2.57	2.5482
Otsu Method	3.42	3.61	3.38	3.47

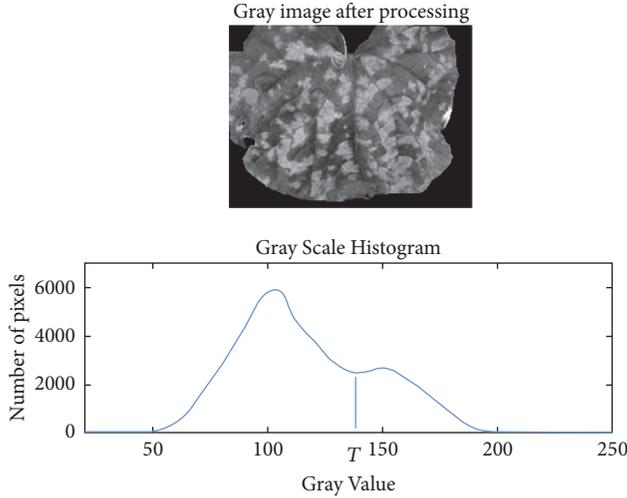


FIGURE 1: The histogram in 2-Mode Method.

2.3. *2-Mode Method.* The image is often composed of normal foliage and diseased area, so the histogram of gray scale can be regarded as two normal distribution functions, which is shown in Figure 1.

Select the trough, that is,  $T$  position in Figure 1; split the image into two parts, and the result  $g(x, y)$  is

$$g(x, y) = \begin{cases} t_0 & f(x, y) < T \\ t_1 & f(x, y) \geq T. \end{cases} \quad (1)$$

$T$  is the segmentation threshold, usually  $t_0 = 0$  (black),  $t_1 = 1$  (white) [4].

### 3. Image Recognition System Based on Multiple Linear Regression

In most practical situations, however, traditional methods are not able to be the most appropriate choice. It is unwise to ignore difference between actual images and hypothesized data. Therefore, improved methods and a new recognition system based on multiple linear regression are created. This section introduces the characteristics of this system.

3.1. *Improved Histogram Segmentation Method (Calculate Threshold Automatically).* Traditional 2-Mode Method needs to set the threshold manually. As the user has huge task burden, it lacks identification efficiency.

This proposed segmentation method can automatically determine the threshold. It can greatly reduce the user's task burden and optimize the image segmentation process. The

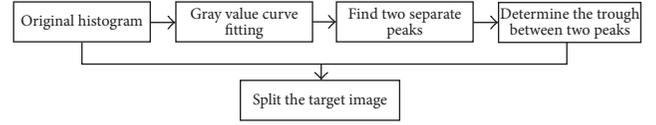


FIGURE 2: The process of the improved histogram segmentation.

processing details are as follows; the flow diagram of the improved histogram segmentation is shown in Figure 2

(1) Perform median filtering operation on grayscale image in 5-by-5 neighborhood, smooth value with Robust Loess (quadratic fit) and specified span of moving average (17). These preprocessing operations are designed to provide fit histogram for segmentation.

(2) Limit selection range (100–190 of 255 pixels) and minimum peak interval (10 pixels), because there are some interferential fluctuations on histogram that need to be filtered out.

(3) Pick the maximum height (between peak and trough) and locate this most suitable trough, which is the adaptive threshold  $T$ .

(4) Segment image according to the threshold  $T$  [5].

Comparing with other threshold segmentation methods, this improved histogram segmentation method is more accurate, as shown in Figure 3.

In Figure 3, sample plants come from different disease severities, minor, medium, and serious disasters. It represents visual results of segmentation directly.

$$\text{Ratio} = \frac{\text{MeanS}}{\text{MeanO}}. \quad (2)$$

Extract Green matrix in RGB color image and calculate mean of this two-dimensional matrix. MeanS is average of green pixels of the segmented image, MeanO represents green pixel mean of original image. The ratio is smaller, the more parts of normal foliage are excluded.

Table 1 reflects result of segmentation indirectly by comparing the change ratio of green pixel values, 15 samples for each severity. In the table, the ratio of the improved method is smaller than the other two methods.

From Figure 3 and Table 1, new histogram segmentation method works better in filtering out normal foliage comparing with Iterative Method and Otsu Method, even in different severities. Both of the two methods remain a little more normal foliage parts after segmentation. This is mainly because Iterative Method is greatly influenced by the overall grayscale of the image; it is weak in distinguishing the subtle pattern and will cause division errors. Although Otsu Method is more stable and practical, it will have problems while handling gray statistic, especially when the ratio of the target

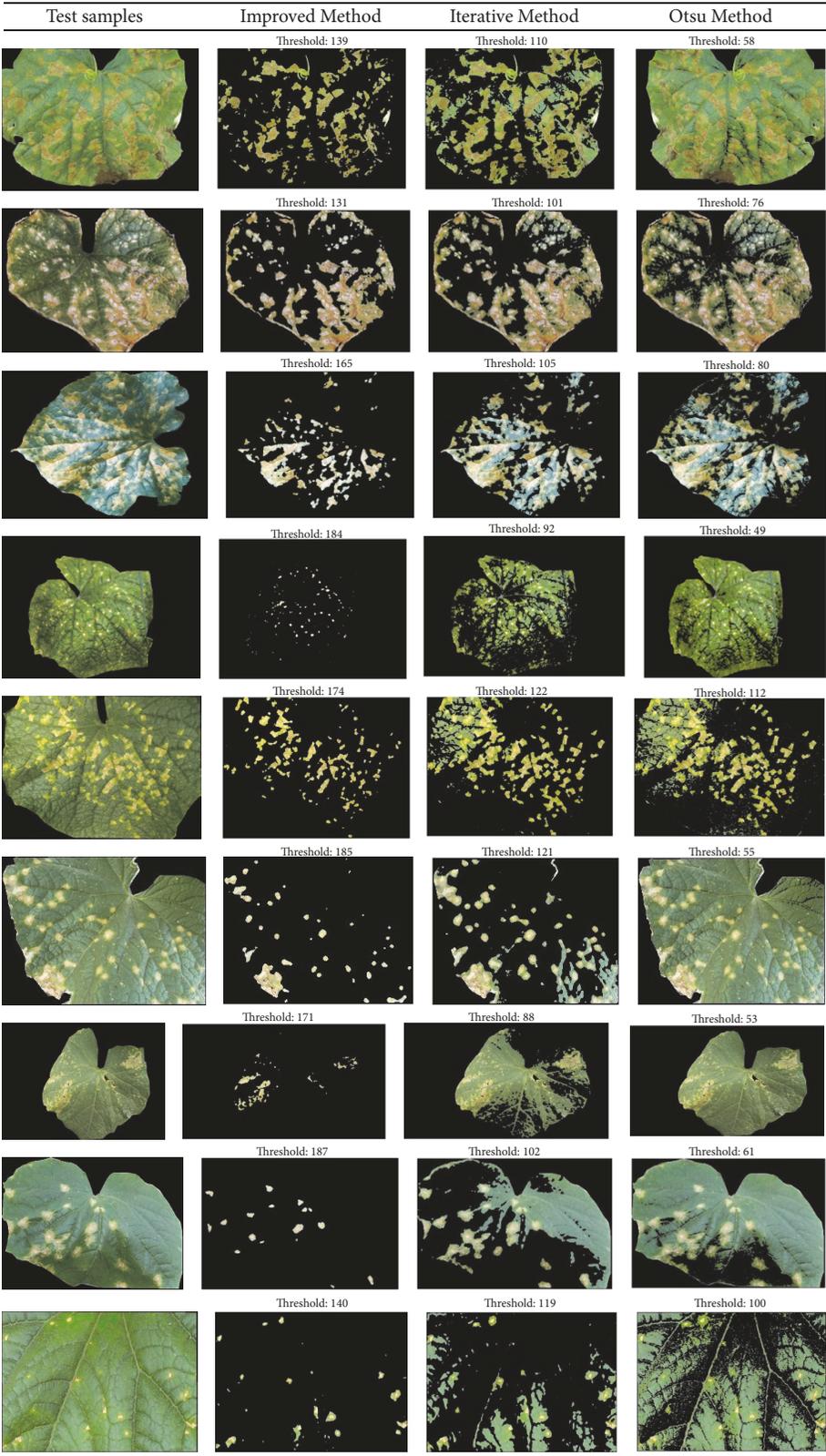


FIGURE 3: Segmentation results of each threshold method.

TABLE 2: Coefficients and confidence intervals.

Coefficients	Coefficient estimation	Coefficients confidence intervals
$\beta_0$	-816.1596	[-3122.8 1490.4]
$\beta_1$	-323.6432	[-634.8063 -12.4801]
$\beta_2$	-9.7902	[-126.8978 107.3174]
$\beta_3$	-91.8584	[-399.8541 216.1372]
$\beta_4$	10.4528	[-3.4562 24.3619]
$\beta_5$	-0.0038	[-0.0151 0.0074]
$\beta_6$	-19.4105	[-664.0947 625.2737]
$\beta_7$	993.5927	[-580.5174 2567.7]
$\beta_8$	-72.9863	[-428.3056 282.3329]
$\beta_9$	1689.2	[-118.7379 3497.2]
$\beta_{10}$	65.8747	[-909.1611 1040.9]
$\beta_{11}$	37.0236	[-219.3019 293.3492]

lesion area to the background area is very small, so it cannot be used in this system either.

In conclusion, the improved histogram segmentation method is more suitable for lesion segmentation in this system with great advantages, such as fast, efficient, and accurate.

**3.2. Disease Recognition System Based on Multiple Linear Regression.** In this system, a total of 11 features are extracted from three aspects: color, texture, and shape. In color aspect, three characteristic values (Hue, Saturation, and Value) are extracted to represent the color features of the lesion [6]; in texture part, energy and homogeneity are selected by using the gray level cooccurrence matrix. Meanwhile, the statistical matrix is also used to collect four characteristics (smoothness, third-order moment, consistency, and entropy) [7]; for the shape, this system chooses the degree of rectification and density. All the extracted 11 characteristic parameters still remain stable no matter how the image changes in rotation, translation, and scale. Thus, they are representative and comprehensive.

The selected characteristics are regarded as 11 independent variables  $x_1, x_2, \dots, x_{11}$ , and the dependent variable  $y$  is the severity of the plant disease. Therefore, the regression model is

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{11} x_{11} + \varepsilon \quad (3)$$

$\beta_0, \beta_1, \dots, \beta_{11}$  are the required coefficients [8].

In this system, plant diseases are divided into four categories, including normal situation, minor disasters, medium disasters, and serious disasters. And a score system, from 0 to 100, is used to assess the severity of disease. The higher the score is, the more serious the disease is. 0–25 scores are normal situation, 25–50 scores are minor disasters, 50–75 scores are medium disasters, and 75–100 scores are serious disasters. Each situation has 10 images, the training library has a total of 40 images.

There are 11 selected eigenvalues of the 40 training images showing in Figure 4; the database has 40 columns, which are 40 images, and 11 lines, which are 11 characteristic parameters.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.4457	0.4361	0.4577	0.4379	0.4142	0.4223	0.4370	0.4825	0.4167	0.4478	0.3550	0.3621	0.3774	0.37
0.7617	0.8266	0.9298	0.8887	0.8114	0.8480	0.7980	0.9365	0.8028	0.8884	0.6440	0.7396	0.7178	0.73
0.1921	0.1425	0.0651	0.1076	0.1526	0.1039	0.1564	0.0601	0.1203	0.1341	0.2388	0.2340	0.2318	0.24
0.0367	0.1053	0.0255	0.0788	0.0954	0.0435	0.1177	0.0041	0.1472	0.0170	0.0837	0.1405	0.0848	0.13
30.6642	26.2702	0.8496	22.5051	47.5825	22.6450	27.6200	1.9233	126.0868	17.9709	3.5157e+03	138.7717	38.0805	189.22
1.4677	3.6050	0.0830	1.9938	4.0386	0.7904	4.4275	0.0438	4.3735	1.2138	0.7189	5.7909	6.2426	11.96
16.7630	23.7735	3.7804	15.5961	27.0224	10.8437	26.6498	2.3728	26.9822	13.6842	36.3796	31.2648	33.7228	50.25
0.9945	0.9982	0.9349	0.9959	0.9986	0.9916	0.9986	0.8482	0.9986	0.9947	0.9992	0.9990	0.9991	0.99
0.0034	0.0058	1.7892e-04	0.0025	0.0086	0.0012	0.0073	5.8447e-05	0.0073	0.0023	0.0160	0.0102	0.0126	0.02
0.9469	0.8085	0.9901	0.8919	0.8039	0.9460	0.7986	0.9954	0.7623	0.9436	0.8360	0.8336	0.8293	0.63
0.2973	0.8975	0.0593	0.5184	0.9329	0.2915	0.9705	0.0327	1.1051	0.2974	0.9087	0.8874	0.8722	1.75

FIGURE 4: Selected eigenvalue of the training images.

Put the 11 features into multiple linear regression models. The coefficients and confidence intervals are calculated via using the least squares estimation algorithm.

From Figure 5(a), the 14th point is an anomaly and exceeds the credible range. Therefore, we remove this point, and then relinear regression of 39 normal points helps to obtain the correct coefficients and confidence intervals, as shown in Table 2.

Finally, after eliminating the residua, the obtained multiple regression system model is evaluated and analyzed to test the accuracy of the recognition system.

### 3.3. Multiselective Interactive Image Segmentation Methods

**3.3.1. Color Cutting Based on True Color Image Processing.** Color cutting is to use a sphere to determine a color of the cluster; extract the following:

$$S_i = \begin{cases} C & \sum_{j=1}^3 (r_j - a_j)^2 > R_0^2 \\ r_i & \text{Other} \end{cases} \quad i = 1, 2, 3 \quad (4)$$

$R_0$  is the radius of the color clustered sphere,  $a_j$  ( $j = 1, 2, 3$ ) is the color component of the center of the sphere, and  $C$  represents a fixed color element.

This system combines true color image processing with disease identification to help users to extract special color lesions. First, select the color of the target lesion as a pixel from the graph, and then stratify color according to the RGB mean of each selected pixel  $3 \times 3$  neighborhood domain, and

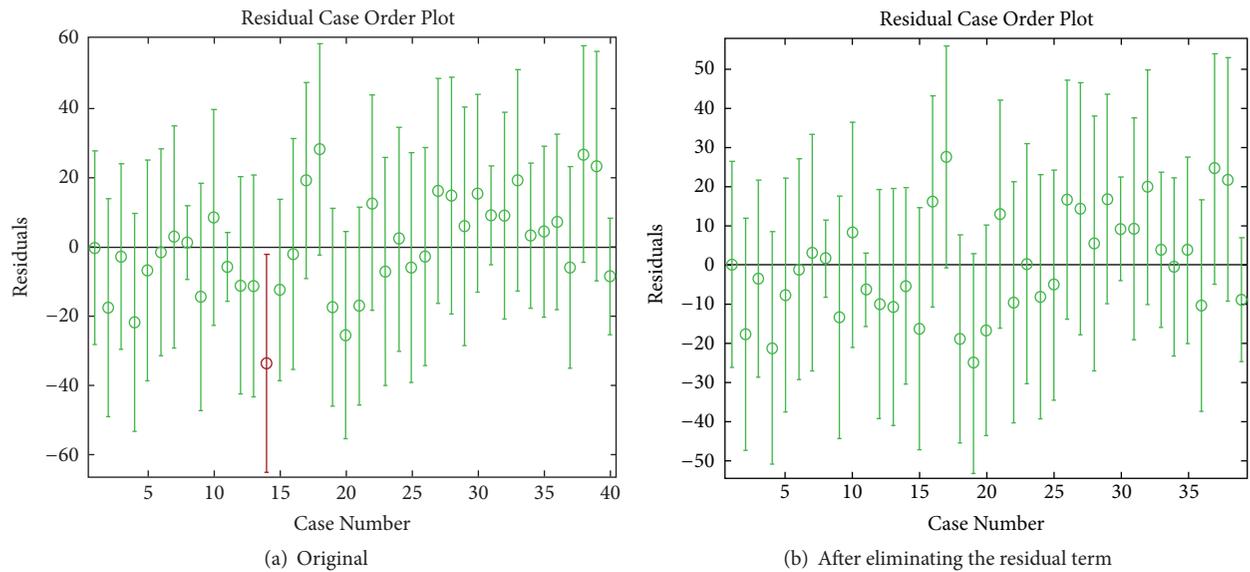


FIGURE 5: Coefficients and confidence intervals before and after eliminating residual items.

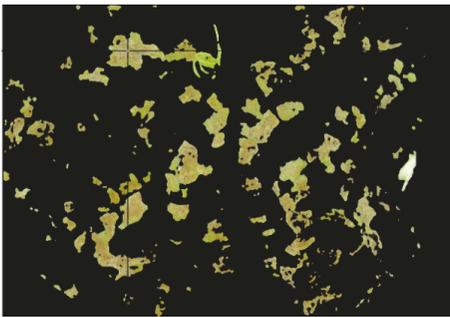


FIGURE 6: Target selection interface in color cutting.

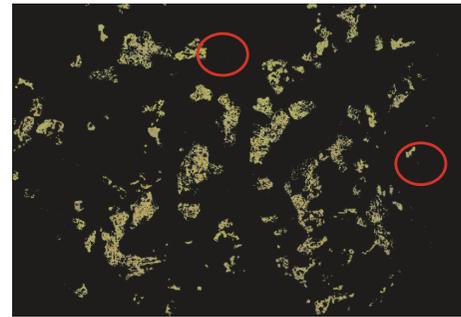


FIGURE 7: Lesion image after color cutting.

set the appropriate color clustering sphere radius ( $R_0 = 70$ ) to filter out useless stains [9]. As shown in Figure 6, the user chooses the certain color parts. After color cutting, as shown in Figure 7, the target color part is retained, the uncorrelated parts are filtered, and the lesions are more accurate.

**3.3.2. Multipoint Selection Based on Regional Growth Method.** Regional growth method refers to the method of continually deriving the pixel seed dot regions with spatial domain similarity and finally growing a region [10].

In this system, user selects a number of points; then the system will use the regional growth method only to split out these areas of the lesion which encircle these specific points, as shown in Figure 8.

**3.4. The Overall Process of the Disease Recognition System.** The system is based on image processing technology and MATLAB as the main processing tool. It is divided into four stages: image preprocessing, image segmentation, feature extraction, and regression model. The detailed workflow of the system can be summarized as follows:

(1) Select edge to cut the image and separate leaf from the complex background.

(2) Filter out image noise by using spatial domain image denoising.

(3) Create improved histogram segmentation method to separate the lesion from normal foliage.

(4) Depending on different purposes, user can choose color cutting based on true color image processing and multipoint selection based on regional growth method to extract certain disease parts.

(5) Extract the lesion characteristics from color, texture, and shape, three aspects.

(6) Establish the linear regression model and the disease identification system.

Step (1) and Step (2) belong to the preprocessing part, filtering out irrelevant image information. Step (3) and Step (4) belong to the image segmentation part, separating the lesions accurately. Step (5) belongs to the feature extraction part, extracting the characteristic parameters from various aspects of the lesion. Step (6) belongs to the regression model part, establishing multiple linear regression model to determine the type of disease.

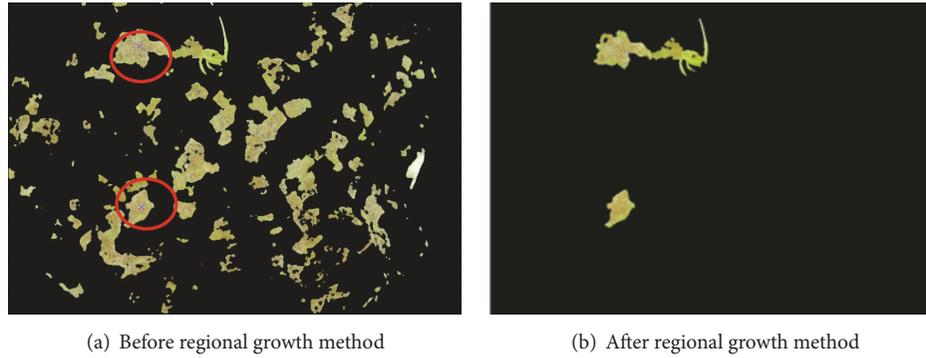


FIGURE 8: The lesion image before and after regional growth method.

TABLE 3: The recognition accuracy of images inside (40) and outside (10) the training library.

Image	Correct	Wrong	Accuracy
Inside			
Normal situation (10)	9	1	90%
Minor disaster (10)	8	2	80%
Medium disaster (10)	8	2	80%
Serious disaster (10)	7	3	70%
Outside			
Random situation (10)	6	4	60%

TABLE 4: The recognition accuracy of images inside (60) and outside (10) the training library.

Image	Correct	Wrong	Accuracy
Inside			
Normal situation (15)	14	1	93.33%
Minor disaster (15)	13	2	86.67%
Medium disaster (15)	12	3	80%
Serious disaster (15)	12	3	80%
Outside			
Random situation (10)	9	1	90%

#### 4. Results and Analysis

The authentication system has the following: (1) server-side configuration: 2.9 GHz Intel Core i7, Intel HD Graphics 630 and Radeon Pro 560, 16 GB memory, MacOS Version 10.13.3; (2) tool platform: MATLAB R2017b.

After the establishment of multiple linear regression model, the images from training libraries are placed into the multiple linear regression model; then the disease recognition system is constructed by using the least squares method. Use the images inside and outside the training libraries to test the accuracy of the system.

The results of the discrimination are shown in Table 3

From Table 3, as the disease situation becomes gradually complex, the number of errors gradually increased, because when the disease becomes more serious, the characteristic parameters will become more complex, which will make the results instable.

Increase the number of the training images and collect the new accuracy results; the results are shown in Table 4.

Compare the multiple linear regression recognition system with the traditional minimum distance method [11] with the same training data. The results are shown in Table 5

The results of multiple regression system can discriminate the severity of plant diseases better. Meanwhile, as the increase of the training images, the results are more accurate, which proves the accuracy and good potential of this system.

At the same time, due to the arbitrariness of the independent and dependent variables, user can also choose different parameters to establish new regression models. For example, the random variables can be only relevant to several specific diseases which makes it more accurate to distinguish these diseases. Besides, the selection of the characteristic parameters depends on user, so it is easy to highlight the different lesion characteristics to meet different requirement.

TABLE 5: Comparison of accuracy of multiple linear regression and minimum distance method.

Image	Multiple linear regression	Minimum distance method
Inside		
Normal situation (15)	93.30%	60.00%
Minor disaster (15)	86.67%	73.33%
Medium disaster (15)	80.00%	73.33%
Serious disaster (15)	80.00%	80.00%
Outside		
Random situation (10)	90.00%	67.66%

In conclusion, the recognition system in this paper has great applicability and modifiability.

## 5. Summary

This paper improves image segmentation and disease recognition system. An improved histogram segmentation method is proposed; this method can find appropriate threshold automatically rather than manually, which is more scientific, reliable, and efficient. Meanwhile, the linear regression model can be modified easily by changing the independent and dependent variables; it has accuracy, applicability, and greater potential.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] V. Pooja, R. Das, and V. Kanchana, "Identification of plant leaf diseases using image processing techniques," in *Proceedings of the 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, pp. 130–133, Chennai, April 2017.
- [2] P. L. Cheong and S. D. Morgera, "Iterative methods for restoring noisy images," *Institute of Electrical and Electronics Engineers Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 4, pp. 580–585, 1989.
- [3] Y. S. Tang, D. H. Xia, G. Y. Zhang, L. N. Ge, and X. Y. Yan, "The detection method of lane line based on the improved Otsu threshold segmentation," *Applied Mechanics and Materials*, vol. 741, pp. 354–358, 2015.
- [4] Y. Geng, *Study on Crop Disease Diagnosis Based on Image Recognition*, University of Science and Technology of China, Anhui, China, 2009.
- [5] Q. Wang, *Image Histogram Features and Its Application*, University of Science and Technology of China, 2014.
- [6] R. Pydipati, T. F. Burks, and W. S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," *Computers and Electronics in Agriculture*, vol. 52, no. 1-2, pp. 49–59, 2006.
- [7] S. Huang, "A novel method of stone surface texture image recognition," in *Proceedings of 2016 IEEE International Conference on Signal and Image Processing (ICSIP '16)*, IEEE, Beijing, China, 2016.
- [8] F. Peng, D.-L. Zhou, M. Long, and X.-M. Sun, "Discrimination of natural images and computer generated graphics based on multi-fractal and regression analysis," *AEUE - International Journal of Electronics and Communications*, vol. 71, pp. 72–81, 2017.
- [9] M. Riccardi, G. Mele, C. Pulvento, A. Lavini, R. D'Andria, and S.-E. Jacobsen, "Non-destructive evaluation of chlorophyll content in quinoa and amaranth leaves by simple and multiple regression analysis of RGB image components," *Photosynthesis Research*, vol. 120, no. 3, pp. 263–272, 2014.
- [10] F. Y. Shih and S. Cheng, "Automatic seeded region growing for color image segmentation," *Image and Vision Computing*, vol. 23, no. 10, pp. 877–886, 2005.
- [11] J. Su, R. Han, L. Shi, and X. Qian, "Improved minimum distance discrimination method used in image analysis of fabric wear resistance," *Applied Mechanics and Materials*, vol. 263-266, no. 1, pp. 421–426, 2013.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

