Global NEST Journal, Vol 22, No 4, pp 492-501

Tea disease spot recognition based on image feature points extraction and matching

Chen J.[1] and He Q.[2]*

[1]Key Lab of Machine Learning and Computational Intelligence, College of Mathematics and Information Sciences, Hebei University, Baoding 071002, China

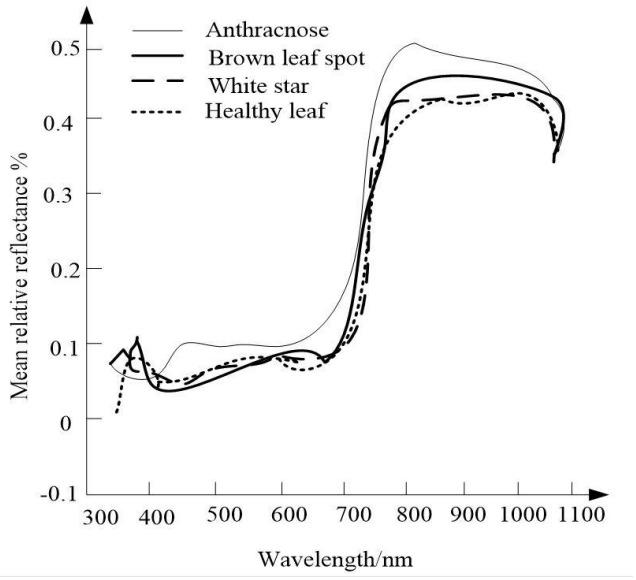
[2]School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Received: 01/07/2020, Accepted: 06/08/2020, Available online: 21/10/2020

*to whom all correspondence should be addressed: e-mail: heq_us@126.com

https://doi.org/10.30955/gnj.003375

Graphical abstract



Abstract

Objective: adaptation of tea disease spot recognition method based on image feature points extraction and matching, can achieve fast and accurate recognition of tea disease spot, and provide reference for remote sensing disease monitoring with plant protection UAV at low altitude. Methods: hyperspectral imaging technology and image processing technology are used to collect hyperspectral images of 4 kinds of samples, such as anthracnose, brown leaf spot disease, tea white star and healthy leaf, and the extracted spectral reflectance of sensitive waveband of interest region is regarded as spectral feature. Twice principal component analysis is performed to obtain the second principal component image after the second principal component analysis as the feature image. Computer processing technology and pattern recognition technology are utilized to analyze characteristic parameters of tea disease spot, like color, shape and texture. On the basis of the feature that tone of tea disease spot in HSI (hue-saturation-intensity) color space is different. H component image is employed to extract tea disease spot, so as to get the tea disease spot image. n total, 12 colors parameters, 11 shape parameters and 8 texture parameters of each disease spot region are extracted. 20 feature parameters of strong classification ability are selected using variance and principal component analysis method to form classification eigenvectors. 35 eigenvectors of tea non-disease spot and 35 eigenvectors of tea disease spot tea are randomly selected to compose the training set. Fischer discriminant function is utilized for accurate recognition of tea disease spot based on the training set. Results: information of disease spot of the leaves with different types of disease has obvious features in the PC2 image. The features of PC2 image is regarded as the feature image for disease spot extraction later. The segmentation effect of anthracnose and tea white star is better. Images of disease spot region of different types before and after segmentation are obtained. The correct recognition rate of tea disease test set is 90%, and the correct recognition rate of non-disease test set is 95%. Conclusion: the proposed method can accurately identify tea disease spot, and provide technical support for monitoring the disease of tea at low altitude.

Keywords: Extraction, feature points, image, matching, recognition, tea disease spot.

1. Introduction

Tea as a perennial cash crop, with function of antibacterial, anti-inflammatory, radiation protection, adjusting blood lipid and others, is popular among consumers. However, the disease problem has greatly affected the quality and production of tea (Ge *et al.*, 2017; Li *et al.*, 2018). The key to solving the problem is how to find the disease accurately and timely and to prevent it. At present, the methods for detecting plant diseases include sensory evaluation, physical and chemical tests, and conventional machine vision (Arbab, 2016). The misjudgment rate of these methods is high, and the recognition takes a long time and the cost is high, which cannot meet the requirements of fast and efficient. Therefore, finding a fast and efficient

recognition method is of great significance to agricultural plant protection (Cai and Li, 2018; Johnstone *et al.*, 2016; Singh *et al.*, 2018).

In recent years, hyperspectral imaging technology has shown great superiority in non-destructive testing, agricultural product classification, and safety assessment and so on. Some achievements have been made at home and abroad (Hu Wang and Yang, 2014; Kamel and Sana, 2019), for example, spectral reflectance of visible light and near-infrared band is used to diagnose early yellow rust in wheat. Li *et al.* (2008) identified the potato surface defects based on the two-color spaces and achieved significant results. Leckie *et al.* used the spectral images of seven characteristic bands in visible and near-infrared region to detect the pine aphid infestation. Feng Lei and others utilized the hyperspectral image data of the eggplant leaves to extract the feature images under the three characteristic bands. Based on the least square support vector machine, an identification model is constructed, and the accuracy of the model is 97.5%. Wu Di and others adopted multi-spectral imaging technology to detect tea disease spot (Wang *et al.*, 2018). According to the study (Latifi Fassnacht and Schumann, 2014; Rezaei *et al.*, 2020), there are two sensitive bands 742-794 nm and 1374-2500 nm for the degree of anthracnose stress in tea plants. Based on the first-order differential vegetation index (Rg-Rr)/(Rg+Rr), it has a good effect on the prediction of the damage degree of anthracnose.

At present, the research on tea plant diseases and insect pests involves a relatively single species of pests and diseases. In this paper, tea disease spot recognition method is proposed based on the image feature points extraction and matching. he most common, more harmful three kinds of diseases for the canopy surface of Yuhuatea tree are tested. The spectral characteristics and image characteristics are selected and combined with a variety of algorithms to identify and classify the disease spot, so as to provide a reference for the detection of ultra-low altitude remote sensing diseases of agricultural plant protection UAV.

2. Materials and methods

2.1. Materials

The research object used in this experiment is Nanjing Yuhua tea leaves. The tea leave is collected from the tea production base located at Pingshan Forest Park in Liuhe District of Nanjing (longitude 118°, latitude 32°). Some diseased leaves and healthy leaves with length between 50-60 mm are selected. The leaves are collected at 9:00 am in sunny day, ambient temperature is 19℃ and relative air humidity is 45%. After collection, the samples are placed in sealed bags, and then stored in micro-refrigerator at 0℃ and immediately sent to laboratory for testing (Mohammad *et al.*, 2020; Wang Wang and Liu, 2015). Before the data collection, the plant protection researchers further confirm that 80 samples of anthracnose, 72 samples of red spot, 80 samples of tea white star and 60 healthy samples are screened out.

2.2. Hyperspectral imaging system

The ISUZU OPTICS hyperspectral imaging system is used for experiments, as shown in Figure 1. The system mainly includes: hyperspectral image spectrometer (ImSpector V10E), CCD camera (GEV-B1621M), two 150 W optical fiber halogen lamps, electric displacement console, dark box (1200 mm × 500 mm × 1400 mm), control Box, a high-performance computer and so on. Hyperspectral camera spectral range is 358-1021 nm, and spectral resolution is 2.8 nm.

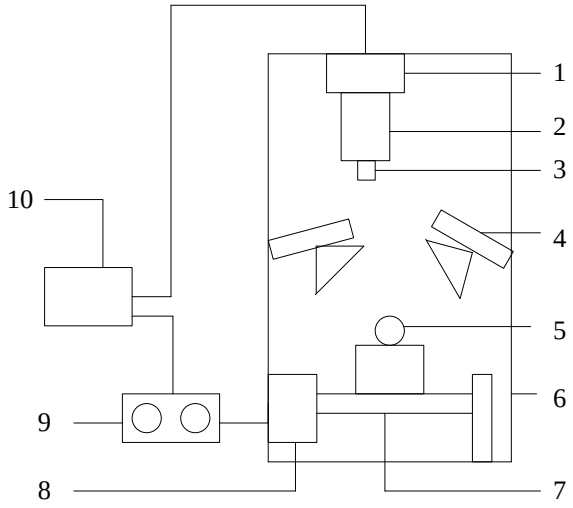


Figure 1. Hyperspectyal imaging system. 1. Camera 2. Spectrometer 3. Shot 4. Halogen light source 5. Tea sample 6. Black Box 7. Electric displacement platform 8. Stepper motor 9. Mobile platform controller 10. Computer

2.3. Hyperspectral data processing

Hyperspectral data processing is based on the software platform ENVI 5.3 (Exelis Visual Information Solutions, USA), Excel 2010 and Matlab 2016a (MathWorks, USA). The data processing hardware conditions are: 16 GB RAM, Intel (R) Coer (TM) i5-6500 CPU.

2.3.1. Hyperspectral image correction

In order to ensure the accuracy of the spectral data and eliminate the noise interference in the middle of the acquisition, the original hyperspectral image is corrected according to equation (1) to obtain a corrected hyperspectral image R:

$$R = \frac{I_s - D}{W - D} \tag{1}$$

Where R is the corrected hyperspectral image; I_s is the original hyperspectral image acquired by the hyperspectral system (Zhang Xu and Chen, 2016); D is the all black calibration image acquired with hyperspectral data (reflectivity is close to 0); W is all white calibration image acquired with hyperspectral data (reflectivity is close to 99%).

2.3.2. The relative spectral reflectance of the region of interest

Since each pixel in the hyperspectral image corresponds to a full-band spectral information, the area of 200 pixels by

200 pixels near the tip of the main vein with respect to the average distribution of the disease spot is selected as region of interest (ROI). In this study, the average spectral reflectance of ROI of four kinds of samples is selected from 80 leaves of anthracnose, 72 leaves of red disease spot, 80 leaves of tea white star and 60 healthy leaves respectively (Ismail *et al.*, 2019; Wang Nielsen and Nash, 2015), as shown in Figure 2.

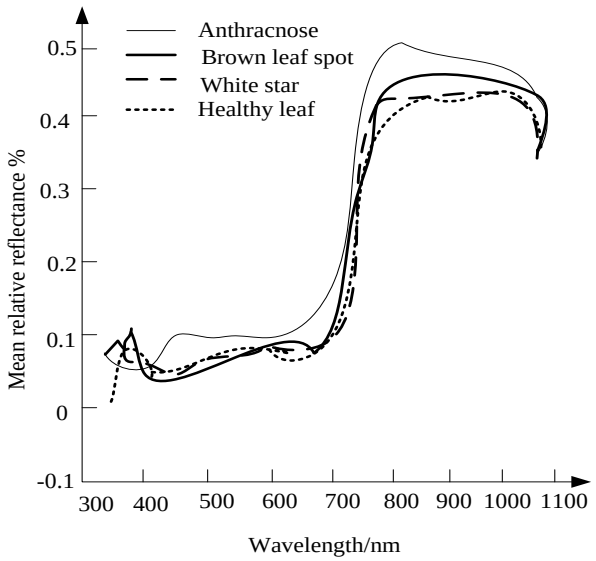


Figure 2. Relative spectral reflectance curve of samples

As can be seen from Figure 2, the hyperspectral data has relatively large noises in the near-ultraviolet and violet bands of 358-400 nm and in the near-infrared band of 980-1021 nm. In order to improve the accuracy of data processing and reduce the noise interference, referring to spectral data processing method, 135 wavebands without start and end are collected (Emir and Bekun, 2019; Nwankwoala, 2019). The effective wavelength range finally obtained for spectral analysis is 430-950 nm, a total of 481 wavebands.

2.3.3. Spectral feature analysis and eigenvector extraction

The texture feature describes the surface properties of the scene corresponding to the image area. By observing the difference of texture features between the diseased area and the non-diseased area of the four types of samples, the texture feature is taken as one of the characteristics for identifying the disease. The gray-level co-occurrence matrix (GLCM) is a commonly used method to describe the texture by studying the spatial correlation of gray scales (Parvin *et al.*, 2019; Sun Liu and Ji, 2014). The gray level co-occurrence matrix $G(g_1, g_2)$ can be represented by equation (2).

$$G(g_1, g_2) = \frac{\#\{[(x_1, y_1), (x_2, y_2)]] \in S \mid f(x_1, y_1) = g_1 \,\&\, f(x_2, y_2) = g_2\}}{\#S} \quad (2)$$

Where S denotes a set of pixel pairs with a specific spatial relationship in the target area, the right-type numerator indicates the number of pixel pairs with gray values g_1 and g_2 in a specific spatial distribution, (x_1, y_1) and (x_2, y_2) denote 2 pixels of distance δ, # is the number of elements in the collection.

The ROI images of pre-treated diseased leaves and healthy leaves are used to calculate five co-occurrence matrix features of energy, contrast, correlation, smoothness and entropy in MALAB. 20 texture feature values are obtained at 0°, 45°, 90° and 135°direction with the distance of 2, Figure 3 shows the mean value of texture features of the sample in four directions.

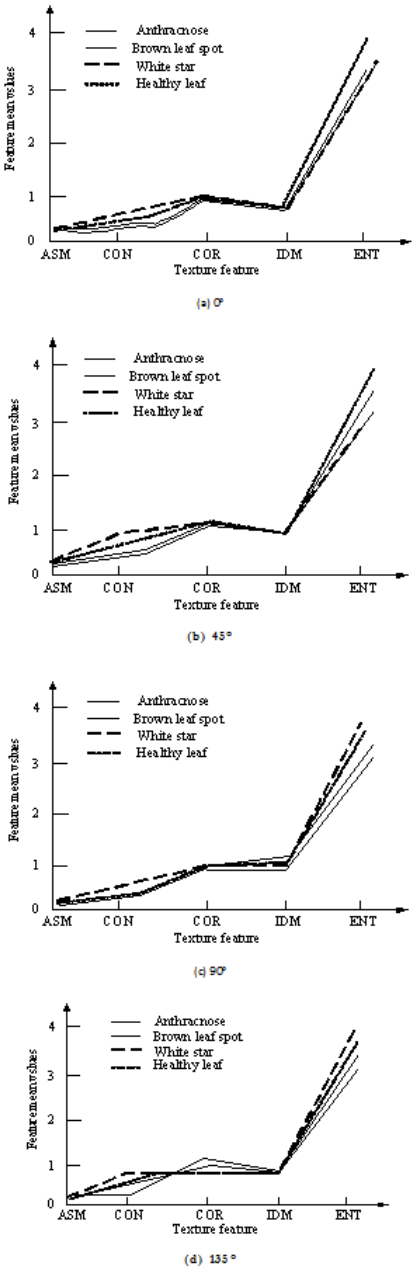


Figure 3. Texture feature mean values of samples in four directions. Note: ASM stands for energy, CON stands for contrast, COR stands for correlation, IDM stands for stationarity, ENT stands for entropy

As can be seen from Figure 3, the contrast in the four directions is good for classifying the four types of samples. Compared with the other three types, the tea white star region has a deeper texture groove (Liu Zhang and Deng,

2015), and the visual effect is clear. Samples with anthracnose have higher total energy values than the other three types, which reflects that the anthracnose region has coarse texture. Because the thicker the texture, the greater the energy (Ademila *et al.*, 2019; Narendranathan *et al.*, 2019). In terms of the level of stationarity, the difference between the four types of samples is not significant. Samples with anthracnose shows the smallest entropy, indicating that the non-uniformity of anthracnose leaf images is larger.

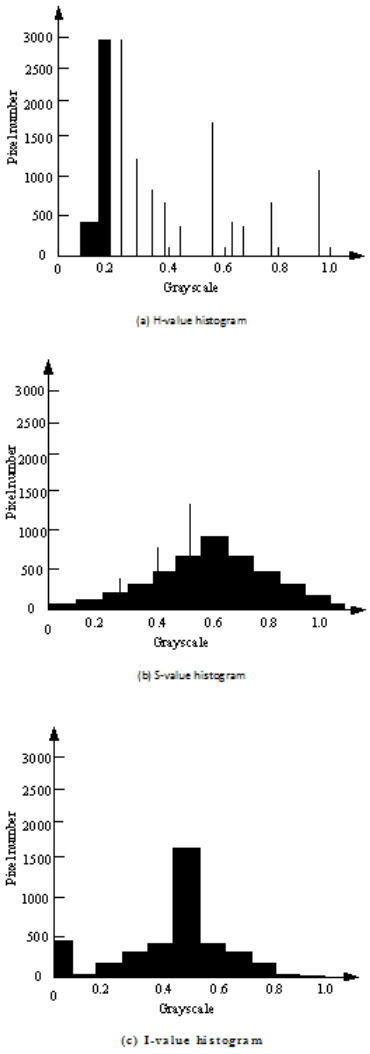


Figure 4. HIS space three-component gray histogram

As can be seen from Figure 3, the contrast in the four directions is good for classifying the four types of samples. Compared with the other three types, the tea white star region has a deeper texture groove (Liu Zhang and Deng, 2015), and the visual effect is clear. Samples with anthracnose have higher total energy values than the other three types, which reflects that the anthracnose region has coarse texture. Because the thicker the texture, the greater the energy (Ademila *et al.*, 2019; Narendranathan *et al.*, 2019). In terms of the level of stationarity, the difference between the four types of samples is not significant. Samples with anthracnose shows the smallest entropy, indicating that the non-uniformity of anthracnose leaf images is larger.

Color moment is a method of describing color distribution by calculating moments. The distribution of color information is mainly concentrated in lower order moment. The first order moment describes the average color and the second order moment describes the color variance. The third order moment describes the color's offset. This experiment calculates the first to third order moments of a single channel of the grayscale image and is expressed by equation (3), (4) and (5).

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} P_{ij} \tag{3}$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(P_{ij}-\mu_i)^2\right)^{1/2} \tag{4}$$

$$\varsigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(P_{ij}-\mu_i)^3\right)^{1/3} \tag{5}$$

where, P_{ij} is the jth color component of the ith pixel and N is the number of pixels. Distribution of the first order moment μ_i, the second order moment σ_i, and the third order moment ς_i reflect the average intensity, inhomogeneity and asymmetry of color.

Spectral information on the surface of tea reflects its internal biochemical composition information. After the leaves are affected by disease, chlorophyll shortages and reduction of moisture content at the disease region are caused, and the spectral reflectance in the visible light band shows a great difference (Hamedianfar *et al.*, 2014; Mohd Hanapiah *et al.*, 2020). In the spectral curve, reflectance of a green light region of 490-560 nm and a red light region of 620-780 nm increases, whereas the reflectance decreases in the near-infrared region. Through observation and analysis, the spectral reflectance of the four types of samples at 560, 640 and 780 nm is taken as the spectral feature of the image.

2.4. Image processing and disease recognition

2.4.1. Leaf image preprocessing

The pathogenesis of tea disease spot is the appearance of round or irregular spots on the lower leaves, then the spot gradually expands. The edge is dark brown and center is gray. In order to facilitate the segmentation of the disease spot image of a complete leaf image (Alvarez-Betancourt and Garcia-Silvente, 2016), the captured image needs to be preprocessed to eliminate noise generated by dust, light, and remove the non-treated leaf section. The improved Grabcut algorithm is used to extract the leaves, the leaves are separated from the background, and the disease spot images are denoised by the median filtering algorithm.

2.4.2. Disease spot extraction

After the pretreatment of the complete tea disease spot image, MATLAB8.0 image processing technology is adopted to transform the image into HSI color space to get H, S, I component diagram. HSI space three-compo-nent grayscale histogram is shown in Figure 4, and S, I component grayscale histograms are normal distribution, and the H-value of pixel in H component image is

concentrated in a region, so the H component image is used for disease spot extraction. The H component image is binarized by using the feature of concentrated disease spot color in the HSI space, and the H value concentration area is 0 and the remaining area is 1, so as to obtain the background. By multiplying the background with the pre-processed image, the green leaf can be removed and the disease spot can be extracted (Zhao Ding and Zhao, 2016). The treatment effect is shown in Figure 5.

Through the above method, the diseased leaf image is processed to obtain a clear disease spot image, but the color, shape and texture parameters of the disease spot image also need to be separately extracted for disease recognition.

Figure 5. Effect pictures of extractive pathological part
(a) Preprocessed image; (b) Background; (c) Color Spot)

(1) Color is the most prominent feature of color image. It is insensitive to the change of direction and size of image, and is not influenced by image rotation and translation (Fischer Ye and Groene, 2014). The commonly used color space is RGB, HSV, HSI, YCbCr, etc., different color space describes the color differently. In RGB color space, R, G and B are the stimulation values of the three primary colors, respectively. The HSI color space is based on the human visual system and hue, saturation and intensity are used to describe color (Wang *et al.*, 2018). HSI color space and the principle of human visual system are similar, and can eliminate the impact of light on the image, so the HSI color space and RGB color space can be utilized together. The mean and variance and other 12 statistical features (CRa, CGa, CBa, CHa, CIa, CSa, CRv, CGv, CBv, CHv, CIv, CSv) of the color components R, G, B and H, I, S in the RGB and HSI color space of the disease spot image are calculated (Angeli Befera and Peyrat, 2014) are regarded as the color feature parameters of disease spot image.

(2) The shape of disease spot is the feature that is not influenced by external factors (such as light, shadow, dust, etc.), but only depends on the disease itself. The disease spot image is processed with morphology algorithm to obtain the contours and regions of the disease spot. A total of 11 statistical features such as circularity Sc, squareness Sa, eccentricity Se, shape complexity Ss and seven invariant moments (H1, H2, H3, H4, H5, H6 and H7) are selected as the shape feature parameter, the calculation of seven invariant moments is based on "a practical calculation method of invariant moments" (Kassouk Thouret and Gupta, 2014). The remaining parameters are calculated as equation (6):

$$S_c = 4\pi \times \frac{A}{P^2}; S_a = \frac{A}{A_R}; S_e = \frac{L_{short}}{L_{long}}; S_s = \frac{P}{A^2} \quad (6)$$

Where: A is the disease spot area, cm^2; P is the disease spot circumference, cm; A_R is the minimum circumscribed rectangle area, cm^2; L_{long} is the length of the major axis of the disease spot lesion, cm; l is the length of the minor axis of the disease spot, cm.

(3) The texture features include important information about the arrangement of the surface structures of the disease spot and the relationship between the disease spot and the surrounding environment (Sun Wei and Zhang, 2014). A total of eight statistic features, includes correlation mean value T_c, energy mean value T_e, mean value of inertia distance T_i, entropy mean value T_h, correlation variance T_{cv}, energy variance T_{ev}, mean value of inertia moment T_{iv} and variance of entropy T_{hv} of the disease spot region are calculated with gray-level co-occurrence matrix and used as texture feature parameters. The mean value of correlation, energy, moment of inertia, entropy, etc., are calculated as equation (7), (8), (9) and (10):

$$T_c = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{ijP(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \quad (7)$$

$$T_e = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \left[P(i,j)^2\right] \quad (8)$$

$$T_i = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i-y)^2 P(i,j) \quad (9)$$

$$T_h = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \mathrm{P}(i-y)\log_2\left[P(i,j)\right] \quad (10)$$

The variance of correlation, energy, moment of inertia, entropy, etc. can be obtained from the corresponding mean values, and 31 characteristic parameters of the disease spot image can be obtained from the aspects of color, shape and texture, thereby forming the classification eigenvector L1 of the leaf with disease spot.

2.5. Feature optimization of disease spot

After feature extraction, the eigenvector of each disease spot sample has 31 characteristic parameters including color, shape and texture features (Cheng *et al.*, 2019). Due to the large number of characteristic parameters and the unequal contribution of each characteristic parameter to the disease classification model or recognition algorithm, in order to effectively implement the disease recognition, it is necessary to exclude the characteristic parameters with strong correlation and select the characteristic parameters that can best reflect the nature of the classification.

The variance is used to analyze the characteristic parameters. If the variance of a characteristic parameter is small, the fluctuation of characteristic is small, that is, this characteristic parameter can reflect the better disease ability. At the same time, the principal component analysis (Sánchezcarnero Ojedazujar and Rodríguezpérez, 2014) is used to eliminate the redundant feature parameters that have little effect on the discriminant results. In the calculation, the common factor analysis in SPSS19.0 is used

for dimensionality reduction. The 20 feature parameters of most significant contribution are screened from L1 to form the optimized classification eigenvector L2, that is, the 10 color feature parameters, the mean value C_{Ra} of R the mean value C_{Ga} of G, the variance C_{Bv} of B, the variance C_{sv} of S, the variance C_{Iv} of I, five invariant moment of shape parameters, H_1, H_4, H_5, H_6, H_7, and five texture feature parameters, mean value of correlation T_c, mean value of energy T_e, mean value of inertia moment T_i, mean value of entropy T_h and variance of inertia moment T_{iv}.

Table 1. Cumulative contribution rate of principal components (%)

Project	Anthracnose	Brown leaf spot	White star	Healthy leaf
PC1	97.32	98.24	98.42	99.43
PC2	99.8	99.44	99.79	99.9
PC3	99.88	99.73	99.87	99.96
PC4	99.96	99.89	99.94	99.97

Table 2. Discrimination results of Fisher discriminant functions

Name	Sample		Correct number		Correct rate/%	
	Training set	Testing set	Training set	Testing set	Training set	Testing set
Tea disease	35	20	30	18	85.7	90
Non tea disease	35	20	31	19	88.6	95

Table 3. Without feature optimized discriminant functions

Name	Sample		Correct number		Correct rate/%	
	Training set	Testing set	Training set	Testing set	Training set	Testing set
Tea disease	35	20	32	15	91.4	75
Non tea disease	35	20	35	17	100	85

Among them, the color feature optimizes mean value C_{Ha} and variance C_{Hv} of hue H, because the color features expressed in the RGB color space are similar to the hue in the HIS color space. The shape feature retains 5 moment invariants, and the moment invariants are based on the non-linear combination of low order (second-order and third-order) normalized moment, which is a statistical feature of the image, and has the same nature for transla-tion, rotation, scale change. The texture feature excludes correlation variance T_{cv}, energy variance T_{ev}, entropy variance T_{hv} of low correlation with the target results.

Therefore, the eigenvector of the jth sample is written as equation (11):

$$L2_j = (C_{Raj}, C_{Gaj}, C_{Baj}, C_{Saj}, C_{Iaj}, C_{Rvj}, C_{Gvj}, C_{Bvj}, C_{Ivj}, H_{1j}, H_{4j}, H_{5j}, H_{6j}, H_{7j}, \mathrm{T}_{ej}, \mathrm{T}_{cj}, T_{ij}, T_{hj}, T_{ivj}), j = 1,2,...,n \quad (11)$$

Where n is the number of samples.

The obtained 110 samples are optimized. Among them, 55 are eigenvectors of non-diseased tea leaf and 55 are eigenvectors of diseased tea leaf.

2.6. Disease recognition

From the 110 samples after the feature optimization, 35 eigenvectors of non-diseased tea leaf and 35 eigenvectors of diseased tea leaf are randomly selected. A total of 70 eigenvectors forms the training set, and the remaining 40 samples are used as the test set. The training set is used to construct the Fisher's discriminant function equation (Rotundi Rietmeijer and Ferrari, 2014), and the test set is classified and identified.

3. Results

3.1. Detection of cumulative contribution rate of main component

The hyperspectral image acquired in this experiment is a three-dimensional data cube. Compared with two-dimensional image and one-dimensional spectrum, the hyperspectral image has rich bands and high spectral resolution and contains many information, but the correlation of adjacent bands is strong, and data redundancy is high, reducing the accuracy and speed of post-processing.

Principal component analysis can effectively remove the correlation between data and is often used for dimensionality reduction. In this study, PCA is used to reduce the dimension of the hyperspectral image in the 430-950 nm band after removing the noise (Cai Tang and Li, 2015) to obtain the principal component (PC) image. According to the covariance contribution rate to determine the PC image. As can be seen from Table 1, the cumulative contribution rate of the first four principal components of the sample reaches more than 99.89% while the contribution rate of PC1 reaches more than 97%, which can best represent the original information of the image.

In the experiment, the first principal component analysis for the band of 430-950 nm takes longer time and slower calculation speed, so the second principal component analysis is performed based on the characteristic wavelength found by PC1. PC1 is the result of the linear combination of 481 bands. The weight coefficients of linear combinations are compared, and the maximum weight coefficients of the four types of samples are corresponding to 762,700,721,719 nm. The second principal component analysis is performed on the four characteristic wavelengths that are selected, and Figure 6 shows the PC image after the second principal component analysis of the sample.

As can be seen from Figure 6, although the contribution rate of PCI images is the largest, the contrast between the diseased region and the non-diseased region is insignificant, which is unfavorable for disease spot extraction (Chen, 2016). Anthracnose and white star are

highlighted in the PC2 images, brown leaf spot region in the PC2 image has a clear distinction with the normal part. In PC3 images, there are significant differences between brown leaf spot and white star spot and the non-diseased region, but the grayscale of anthracnose spot region and non-diseased region are similar (Khoo *et al.*, 2019). Double images appear in some samples of PC4 images, the difference between various types of diseased region and non-diseased regionis small. Through observation and analysis, the disease spot information of different kinds of diseased leaves has obvious characteristics in PC2 images, and PC2 images are finally determined as characteristic images for disease spot extraction in the later period.

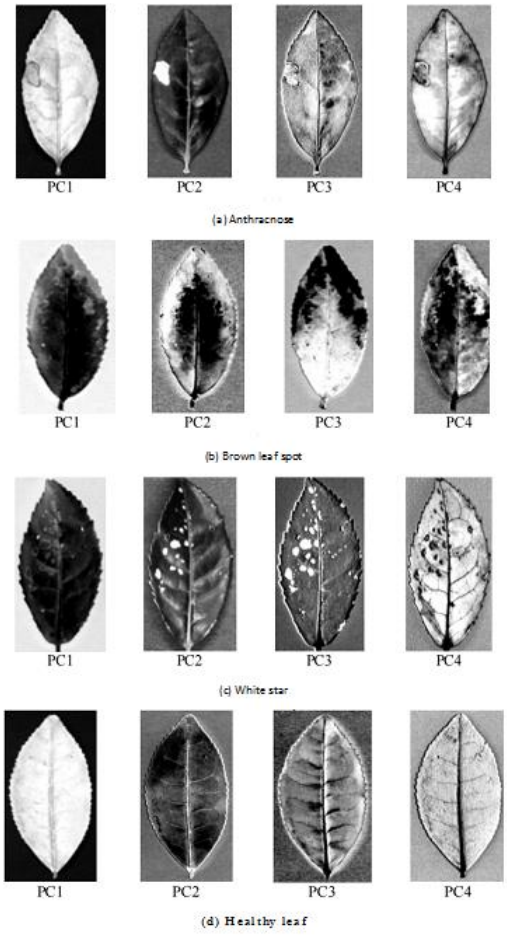


Figure 6. Four principal component images after second PCA

3.2. Segmentation results of disease spot region

Image segmentation is a key step in image processing. Accurate segmentation of disease spot region greatly affects the performance of later feature extraction and algorithm verification. In order to ensure the accuracy of the data processing in the later period, the maximum inter-class variance method, also called Otsu algorithm, is adopted, whose basic principle is to divide the gray value of the grayscale image into two parts with maximum variance by the optimal threshold. Figure 7 shows the images before and after segmentation for different types of disease spot region.

It can be concluded that segmentation for anthracnose and white star spot are better, but there is a case of erroneous segmentation of the tiny area at the root of veins and petiole. The difference of the gray value between the brown leaf spot and the edge of leaf is small, and the segmentation result is general. Because the reflected light is not uniform due to the irregular surface of the sample, and the gray scale of the root, the petiole and the edge are similar to the disease spot during image processing. The resulting binarized image is treated by removing the segmentation region of area less than 50 and holes in the disease spot are filled, which are then processed with the PC2 image to finally obtain an image containing only the disease spot region.

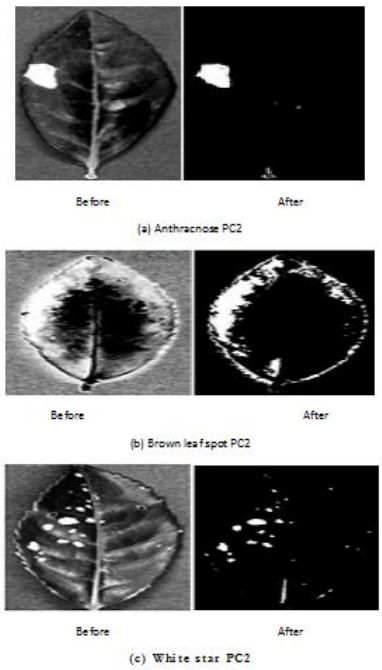


Figure 7. Different type of disease spot before and after image segentation

3.3. Fisher's discrimination

Table 2 shows the classification results of the test set by Fisher's discriminant function. As can be seen from the table, the correct recognition rate of the tea disease test set is 90% and the correct recognition rate of the tea non-disease test set is 95%.

4. Discussion

Analysis of the discriminant process indicates that two of the diseased tea leaf samples are misidentified as non-diseased tea leaf and one of non-diseased tea leaf samples is misjudged as diseased tea leaf. The reason for the miscarriage of justice is that the non-diseased tea leaf contains other diseases (Huang, 2015). Therefore, under the premise of not changing the recognition object, in order to improve the recognition accuracy, the influence of different classification eigenvectors and different training sample numbers on the recognition result needs to be considered.

If the 31 characteristic parameters in L1 are directly used to construct the Fisher's discriminant function and

recognize the diseases, the recognition results are shown in Table 3.

Table 4. Different Training sample identification results

Training sample	Correct rate/%	
	Tea disease	Non tea disease
70	90	95
50	63.3	83.3
30	62.5	77.5

Table 5. Detection results of feature vectors based on different algorithms

Group	Type	Sample Set	Sample number	Correct identification number		Recognition rate/%	
				BP	SVM	BP	SVM
Color features and texture features	Anthracnose	Training set	50	36	30	72	60
		Testing set	30	17	18	56.67	60
	Brown leaf spot	Training set	48	42	34	87.5	78.33
		Testing set	24	18	17	75	70.83
	White star	Training set	50	33	35	66	70
		Testing set	30	17	15	56.67	50
	Healthy leaf	Training set	40	25	27	62.5	67.5
		Testing set	20	13	14	65	70
Color features and texture features and spectral features	Anthracnose	Training set	50	48	48	96	96
		Testing set	30	27	26	90	86.67
	Brown leaf spot	Training set	48	46	44	95.83	91.67
		Testing set	24	22	22	91.67	91.67
	White star	Training set	50	46	45	92	90
		Testing set	30	26	25	86.67	83.33
	Healthy leaf	Training set	40	39	38	97.5	95
		Testing set	20	18	17	90	85

As can be seen from Table 3, the correct recognition rate of the test set by Fisher's discriminant function without feature optimization is 75% (diseased-tea leaf) and 85% (non-diseased tea leaf), respectively, which are obviously lower than that in Table 2. The reason of low recognition rate is that the existence of redundant characteristic parameters in L1, which adversely affect the establishment of Fisher's discriminant function, thus affecting the recognition result. This indicates that there are 1 sets of classification eigenvectors which have the best discriminant results in disease spot features, and need to be screened out by feature optimization.

Considering the influence of different training samples on the recognition results, 30, 50 and 70 samples are used respectively for training. The recognition results are shown in Table 4.

As can be seen from Table 4, reduction of 20 training samples results in that the recognition accuracy drops sharply to 63.3%, thus the number of training samples can have a significant impact on the correct recognition rate. The fewer training samples, the lower the correct recognition rate. During the experiment, it is found that the correct recognition rate of non-diseased tea leaf is above 70% for different training samples, which is due to the fact that the non-diseased tea leaf test set contains other diseases similar to tea leaf diseases. The fewer training samples, the more easily confused between tea diseases and similar diseases.

To test the validity of the spectral features for classification, this paper attempts to test the model with two eigenvector combinations. First-order moment, second-order moment, third-order moment of single-channel and energy, contrast, correlation, stationaryness, entropy in 0°, 45°, 90°, 135° directions; first-order moment, second-order moment, third-order moment of single-channel and energy, contrast, correlation, stationaryness, entropy in 0°, 45°, 90°, 135° directions and relative spectral reflectance at 560, 640, 780 nm. Combination 1 is color feature+texture feature, and combination 2 is color feature+texture feature+spectral feature.

For two groups of eigenvectors, two kinds of models are used to identify the disease spot of the sample. A total of 188 samples, including 50 anthracnose samples, 48 brown leaf spot samples, 50 white star samples and 40 healthy leaf samples are randomly selected as the training set. The remaining 104 are used as the test set and tested with two algorithms. The test results are shown in Table 5.

The test results in Table 5 show that the eigenvector combination of color features and texture features has generally lower recognition rate for the samples, and the classification efficiency of the two algorithms is slightly higher for the brown leaf spot than that for the other samples. The difference of feature values of energy, contrast, correlation, smoothness, entropy and other texture features are not very obvious, which can easily lead to misrecognition. This is an important reason why the

recognition rate of eigenvector combination 1 is generally low. The eigenvector combination 2 of color feature, texture feature and spectral feature, has higher recognition rate of 4 kinds of samples than that of combination 1; the recognition rate of 4 kinds of sample test sets by BP neural network and support vector machine is higher than 83%, and the recognition rate of two algorithms to the brown leaf spot is above 90%. The spectral reflectance of different leaves is obviously different. The spectral characteristics of relative spectral reflectance of 560, 640 and 780 nm in combination 2 can distinct four types of samples easily.

5. Conclusions

In this paper, the spectral information and image information fusion technology are utilized to identify the eigenvector of disease recognition and establish a rapid recognition model of tea disease for the rapid recognition of diseases. The research shows that the average recognition rate of four kinds of samples is 89.59% and 86.67 %. This indicates that the spectral characteristics of the relative spectral reflectance at 560, 640 and 780 nm are significant for the classification of tea diseases. Hyperspectral imaging technology can be used to identify the tea disease spot efficiently. However, when it is affected by light and background under natural conditions, the recognition efficiency is low, and further research is needed to provide valuable information for disease monitoring of UAV remote sensing. According to the experimental results, it is found that the Fisher's discriminant function established by this study has a good classification effect, and the recognition rate of tea leaf disease is up to 90%. It shows that the tea leaf disease recognition method based on the extraction and matching of image feature points proposed in the paper is effective and feasible and can be used for disease recognition in actual production.

In this study, the following conclusions are obtained through the discussion and analysis: (1) utilizing the characteristics of the H-component chromatogram of the diseased leaf image in the HSI color space, the disease spot can be isolated by multiplying the binarized H-component image and the original image. (2) The principal component analysis is used to optimize the features, and the optimal classification eigenvector including 10 color feature parameters, 5 shape feature parameters and 5 texture feature parameters is selected to provide a comprehensive description of the disease spot. (3) By comparing the recognition results of different training samples, we can see that Fisher's discriminant function should be established after sufficient training samples are collected. The Fisher's discriminant function established in this paper can accurately identify the test samples by 90%. It shows that the use of computer vision technology can realize the recognition of tea diseases and provide some theoretical basis and practical technology for agricultural experts and scholars to study tea and other crops.

Acknowledgments

This work is supported by Natural Science Foundation of Hebei Province (No. F2016201161), and National Natural Science Foundation of China (No. 61473111), and the special fund for one university in one province, and Research Fund of Beijing University of Civil Engineering and Architecture (No. KYJJ2017017).

Authors' Contributions

Junfen Chen: Modelling and optimization, writing of manuscript; Qiang He: Proposal of the research topic, experiments, modelling, writing of manuscript.

References

Ademila O., Okpoli C.C. and Ehinmitan D. (2019), Geological And Lithological Mapping of Part of Igarra Schist Belt Using Integrated Geophysical Methods, *Earth Sciences Pakistan*, 1–9.

Alvarez-Betancourt Y. and Garcia-Silvente M. (2016), A keypoints-based feature extraction method for iris recognition under variable image quality conditions, *Knowledge-Based Systems*, **92**, 169–182.

Angeli S., Befera N. and Peyrat J.M. (2014), A high-resolution cardiovascular magnetic resonance diffusion tensor map from ex-vivo C57BL/6 murine hearts, *Journal of Cardiovascular Magnetic Resonance,* **16**, 1–14.

Arbab H. (2016), A fabrication method for non-integrated parabolic mirror based on laser spot image processing and plumbs line, *Journal of Optics*, **45**, 158–166.

Cai L.N., Tang D.L. and Li C.Y. (2015), An investigation of spatial variation of suspended sediment concentration induced by a bay bridge based on Landsat TM and OLI data, *Advances in Space Research*, **56**, 293–303.

Cai W. and Li G. (2018), The drivers of eco-innovation and its impact on performance: Evidence from China, *Journal of Cleaner Production*, **176**, 110–118.

Chen F. (2016), Optimization of color image information feature retrieval in digital image library, *Computer Simulation*, **33**, 430–433.

Cheng Y.W., Ng K.H., Lam S.S., Lim J.W., Wongsakulphasatch S., Witoon T. and Cheng C.K. (2019), Syngas from catalytic steam reforming of palm oil mill effluent: An optimization study, *International Journal of Hydrogen Energy*, **44**, 9220–9236.

Emir F. and Bekun F.V. (2019), Energy intensity, carbon emissions, renewable energy, and economic growth nexus: New insights from Romania, *Energy & Environment*, **30**, 427–443.

Fischer G., Ye S.Y. and Groene J.B. (2014), A possible influence of the Great White Spot on Saturn kilometric radiation periodicity, *Annales Geophysicae*, **32**, 1463–1476.

Ge S., Liu Z., Furuta Y. and Peng W. (2017), Characteristics of activated carbon remove sulfur particles against smog, *Saudi Journal of Biological Science*s, **24**(6), 1370–1374.

Hamedianfar A., Shafri H.Z.M. and Mansor S. (2014), Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data, *International Journal of Remote Sensing*, **35**, 1876–1899.

Hu Y.P., Wang Z.Y. and Yang X.P. (2014), Hand vein recognition based on the connection lines of reference point and feature point, *Infrared Physics and Technology*, **62**, 110–114.

Huang X.H. (2015), Research and Simulation of 3D image motion feature registration, *Computer simulation*, **32**, 206–209.

Ismail N.M., Ismail A.F., Mustafa A., Zulhairun A.K., Aziz F., Bolong N. and Razali A.R. (2019), Polymer Clay Nanocomposites for Gas Separation: A Review, *Environmental Contaminants Reviews*, **2**, 1–5.

Johnstone E., Raymond J., Olsen M.J. and Driscoll N. (2016), Morphological Expressions of Coastal Cliff Erosion Processes in San Diego County, *Journal of Coastal Research,* 174–184.

Kamel K. and Sana B. (2019). Spatial Variability of Soil Erodibility At El Hammam Catchment, Northeast of Algeria, *Environment & Ecosystem Science*, **3**, 17–25.

Kassouk Z., Thouret J.C. and Gupta A. (2014), Object-oriented classification of a high-spatial resolution SPOT5 image for mapping geology and landforms of active volcanoes: Semeru case study. Indonesia, *Geomorphology*, **221**, 18–33.

Khoo S.C., Phang X.Y., Ng C.M., Lim K.L., Lam S.S. and Ma N.L. (2019), Recent technologies for treatment and recycling of used disposable baby diapers, *Process Safety and Environmental Protection*, **123**, 116–129.

Latifi H., Fassnacht F.E. and Schumann B. (2014), Object-based extraction of bark beetle (Ipstypographus L.) infestations using multi-date LANDSAT and SPOT satellite imagery, *Progress in Physical Geography*, **38**, 755–785.

Li Z., Han C. and Gu T. (2018), Economics of biomass gasification: a review of the current status, *Energy Sources Part B Economics Planning & Policy*, **13**, 137–140.

Liu L., Zhang W. and Deng C. (2015), BriGuard: a lightweight indoor intrusion detection system based on infrared light spot displacement, *Iet Science Measurement and Technology*, **9**, 306–314.

Mohammad S.S., Mohiuddin K.M., Laith K.T.A., Hassan M.N., Akter R., Hossain M.S. and Niuz Morshed Khan Md. (2020), Effect Of Bio-Nematicide And Bau-Biofungicide Against Root-Knot (Meloidogyne Spp.) Of Soybean, *Malaysian Journal of Sustainable Agriculture*, **4**, 44–48.

Mohd Hanapiah M.F., Saad S. and Ahmad Z. (2020), Hydrodynamic Modelling In Inshore Reef Area Within Kuantan Coastal Region, *Journal Clean Was*, **4**, 1–7.

Narendranathan S.K., Sudhagar K. and Karthikeyan R. (2019), Optimization of engine operating parameters suitable for punnai oil application in CI engine using Grey relational method, *Energy & Environment*, **30**, 732–751.

Nwankwoala H.O (2019), Geoethics As an Emerging Discipline: Perspectives, Ethical Challenges and Prospects, *Earth Sciences Malaysia*, **3**, 1–8.

Parvin M., Chowdhury H. and Majumder A.K. (2019), Water Supply And Sanitation Facilities Of Some Selected Schools In Khagrachhari, *Water Conservation and Management*, **3**, 14–17.

Rezaei A., Hassani H., Moarefvand P. and Golmohammadi A. (2020), Lithological mapping in Sangan region in North-east Iran using ASTER satellite data and image processing methods, *Geology, Ecology, and Landscapes*, **4**, 59–70.

Rotundi A., Rietmeijer F.J.M. and Ferrari M. (2014), Two refractory Wild 2 terminal particles from a carrot-shaped track characterized combining MIR/FIR/Raman microspectroscopy and FE-SEM/EDS analyses, *Meteoritics and Planetary Science*, **49**, 550–575.

Sánchezcarnero N., Ojedazujar J. and Rodríguezpérez D. (2014), Assessment of different models for bathymetry calculation using SPOT multispectral images in a high-turbidity area: the mouth of the Guadiana Estuary, *International Journal of Remote Sensing*, **35**, 493–514.

Singh K., Gupta N. and Dhingra M. (2018), Effect of temperature regimes, seed priming and priming duration on germination and seedling growth on American cotton, *Journal of Environmental Biology*, **39**, 83–91.

Sun C., Liu T. and Ji C. (2014), Evaluation and analysis the chalkiness of connected rice kernels based on image processing technology and support vector machine, *Journal of Cereal Science*, **60**, 426–432.

Sun H., Wei J. and Zhang J. (2014), A comparison of disease severity measurements using image analysis and visual estimates using a category scale for genetic analysis of resistance to bacterial spot in tomato, *European Journal of Plant Pathology*, **139**, 125–136.

Wang H., Zhong H. and Bo G. (2018), Existing forms and changes of nitrogen inside of horizontal subsurface constructed wetlands, *Environmental Science and Pollution Research*, **25**, 771–781.

Wang S., Xu Z., Pan Y., Liu H. and Yang S. (2018), Experimental study on pressure field of combined well pattern by water flooding, *Kuwait Journal of Science*, **45**, 115–120.

Wang V.Y., Nielsen P.M. and Nash M.P. (2015), Image-Based Predictive Modeling of Heart Mechanics, *Annual Review of Biomedical Engineering*, **17**, 351–83.

Wang Z., Wang Z. and Liu H. (2015), Scale-invariant feature matching based on pairs of feature points, *Iet Computer Vision*, **9**(6), 789–796.

Zhang Y., Xu W. and Chen S. (2016), The extraction of spot signal in Shack–Hartmann wave front sensor based on sparse representation, *Optics Communications*, **371**, 76–82.

Zhao Y., Ding Y. and Zhao X.Y. (2016), Image quality assessment based on complementary local feature extraction and quantification, *Electronics Letters,* **52**, 1849–1851.