WEED IDENTIFICATION FROM SEEDLING CABBAGES USING VISIBLE AND NEAR-INFRARED SPECTRUM ANALYSIS

W. Deng[1], Y. Huang[2], C. Zhao[1], X. Wang[1]

1. *Beijing Research Center of Intelligent Equipment for Agriculture*, *Beijing Academy of Agriculture and forestry Science, Beijing* 100097, *China*

2. *United States Department of Agriculture, Agricultural Research Service, Crop Production Systems Research Unit, Stoneville, Mississippi, USA*

ABSTRACT

Target identification is one of the main research content and also a key point in precision crop protection. The main purpose of the study is to choose the characteristic wavelengths (CW for short) to classify the cabbages and the weeds at their seedling stage using different data analysis methods. Using a handheld full-spectrum FieldSpec-FR, the canopies of the seedling plants, cabbage '8398, cabbage 'zhonggan', Barnyard grass, green foxtail, goosegrass, crabgrass, and small quinoa,, at three- & four-week growth were measured in the range of wavelength 350 ~ 2500nm. In Unscrambler Data Analysis software system, the Principal Component Analysis (PCA) was applied respectively to extract CWs. Then plants were classified by means of Bayes discriminant analysis method with the chosen CWs as variable. The results showed that (1) According to the load factors and its changing rate of PCs corresponding to the spectral wavelengths, the CWs which were sensitive to plant identification were extracted respectively as 567, 667, 715, 1345, 1402, 1725, 1925, and 2015 nm for the first stage and 567, 667, 745, 1345, 1402, 1545, 1725, and 1925nm for the second stage. among the each 8 CWs of two stages, just two of the CWs were different, which indicated that the changes of spectral characteristics at different growth stages of cabbages have little influence on identification of cabbages and weeds. (2) The corresponding spectral data of the 8 CWs extracted from the data at the first stage were taken as the input variables of the model which was built up using Bayes discriminant analysis to classify two varieties of cabbages and five kinds of weeds. The correct classification rates for the training and testing sets were respectively 90.7% and 84.3%. When the two varieties of cabbages were regarded as the same category, using the analysis method the correct classification rates of the training and testing sets were respectively 95% and 100%, which indicated that different varieties of cabbages owned similar the spectral features.

Keywords: Weed identification; spectrum analysis; visible and near-infrared; Bayes; seedling cabbage

INTRODUCTION

According to the research report from the United Nations Food and Agriculture Organization (FAO) in August 2009, weeds should be regarded as farmers' No. 1 natural enemy. The report said that according to a leading environmental research organization, Land Care of New Zealand, weeds cause about $95 billion every year in the lost food production at global level, compared with $85 billion for pathogens, $46 billion for insects and $2.4 billion for vertebrates (excluding humans). Of the $95 billion, $70 billion are estimated to be lost in developing countries (FAO, 2009). In China, the crop yield losses annually caused by weeds sum up to about 10% of the gross grain output (Tang, 2010). Facing the severity of the crop loses caused by weeds, it is urgent for us to seek highly efficient methods for effective weed control. The chemical weeding method commonly adopted at present has provoked a lot of problems, such as excessive pesticide residues, growing number of pesticide-resistant weeds, destruction of the ecological environment, and quality and safety of agricultural products (Thompson et al., 1991). Therefore, people have gradually realized that it is critical to have a method which could not only control the growth of weeds, but also decrease the use of herbicides, and hence prevent from excessive herbicide application.

In previous studies, many scholars, at home or abroad, have done extensive research on automatic identification technologies of weeds and made a great number of achievements. Koger, et al, who analyzed the hyperspectral reflectance of soybean, morning glory, and soil at the two- and four-leaf stages of weed growth using Wavelet method, detected weeds in bean seedlings and the correct identification rates were 83% and 81% (Koger et al., 2003). Jurado-Exposito et al, who distinguished sunflowers, wheat, and seven kinds of broadleaf weeds in seedlings using near-infrared spectroscopy, found that just the near infrared spectroscopy within 750~950nm was enough to identify these plants (Jurado-Exposito et al, 2003). Staughter et al, who distinguished Solanum weeds and tomato using the spectral reflectance in visible and near-infrared wavebands and with the methods of narrow-band hyperspectral modeling and typical discriminant analysis, found that the spectral absorbance data of weeds and tomatoes at the wavelength rang of 2120 ~ 2320nm offered the best classification accuracy (100%), narrow-band hyperspectral models of the data in the visible range also got good classification results (95%), while broad-band models based on color information provided 75% correct classification rate (Slaughter, 2004). Thenkabail et al, who studied how to select the optimum wavebands for classifying plants (shrubs and weeds) and crops (corn) in the range of 400 ~ 2500 nm waveband, found that 90% correct classification rate could be obtained by modeling using 13~22 wavebands selected from the original 168 wavebands using PCA and stepwise discriminant analysis, which accuracy was increased by 9 ~ 43% compared with modeling using ETM (Landsat Enhanced Thematic Mapper), plus brandband data (Thenkabail et al., 2004). Piron et al, who classified seven kinds of weeds in carrot fields under artificial lighting conditions using the visible and near-infrared multispectral device, found that overall correct classification rate was 72% when three optimal wavebands, 450, 550, 700 nm were selected using method of exhaustion and used to establish identification models (Piron et al., 2008). Mao et al, who measured the spectral reflectance of wheat, shepherd's purse, and small quinoa in the wavelength range of 700 ~ 1100nm using a Fourier transform infrared (FTIR) spectrometer, extracted 7 charactersitics wavelengths, 686, 708, 722, 795, 929, 956, 1122 nm using stepwise discriminant analysis and achieved 97% correct identification rate through establishing the model of identifying wheat and weeds (Mao et al., 2005). Deng et al, who measured the canopies' spectral data of plants of corns, crabgrass, and barnyard grass, and set up the bi-classification of corns and weeds using SVMs methods respectively with kernel functions, like linear, radial basis, polynomial, and Multilayer Perceptron, after data preprocessing, found that one-to-one SVM multiclass model based on voting mechanism and SVM method to identify corn, crabgrass, and barnyard grass in seedlings could achieve 80% correct identification rate (Deng et al., 2009). Deng et al, who compared three kinds of methods, Support Vector Machine (SVM), Decision Tree (DT), and Radial Basis Function Neural Network (RBF-NN) for modeling and classifying the spectral data of weeds (Dchinochloa crasgalli, and Echinochloa crusgalli) and corns in seedlings in the corn fields, found that the SVM method provided an 81.58% correct classification rate while the methods of DT and RBF-NN provided the rates of 63.16% and 52.63%, respectively. This indicates that among the three methods, SVM has the highest accuracy in identification of corn and weeds in the fields in the case of limited samples. SVM could provide a method to build a real time tool to identify crop and weeds with high accuracy in practice (Deng et al., 2011). Chen et al, who measured the spectral reflectance of leaves of rice, cotton, barnyard grass and Cephalanoplos indoors in the range of 350 ~ 2500 nm wavelength using ASD spectroradiometer, extracted characteristic wavelengths using stepwise discriminant method, and then classified these plants using the Discrim processing, found that monocotyledons, like rice and barnyard grass, could be accurately classified using the five extracted CWs, 375, 465, 585, 705, 1035 nm, in which the correct identification rate reached 100%; dicotyledons, like cotton and Cephalanoplos, could also be accurately classified using the three extracted CWs, 383, 415 and 435nm, in which the correct identification rate also reached 100% (Chen et al., 2009).

In the previous studies, most of the weed identifications were specific to crops like corn, wheat, rice, and cotton, few to vegetables. However, vegetables, especially dicotyledonous vegetables, are important economic crops in China, which are widely cultivated throughout the country with wide cultivated area and large inputs of labor cost, therefore the study on weed identification in vegetable fields has considerable social and economic benefits and practical significance. In addition, there have been a large number of researches on spectral identification of weeds, but few researches on whether the possible changes of spectral characteristics caused by their changing metabolism at the different growth periods for the same plant would affect the consistency of spectral identification of crops and weeds, which should be studied.

In this study, two varieties of cabbages, ‘No. 8398’ and cabbage ‘Zhonggan No. 11’, were selected as the representatives of vegetables, and five varieties of weeds, Barnyard grass, green foxtail, goosegrass, crabgrass, and small quinoa, which are commonly-seen annual gramineous plants in cabbage fields with strong adaptability, wide coverage, fast multiplying, and causing inestimable harm to crops, were chosen as the studied weed targets. The spectral reflectance of canopies of the selected two kinds of cabbages and 5 kinds of weeds was collected in the wavelength range of 350 ~ 2500 nm at the seedling growth stages of the 35th and 50th days, respectively. The data were modeling and classified using clustering analysis based on PCA and Bayes discriminant analysis.

2 Materials and methods

2.1 Experimental material

Two kinds of cabbages used in the study were cabbage ‘8398, cabbage ‘zhonggan 11’ (short for ‘8398’ and ZG 11), which seeds were provided by the Institute of Vegetables, the Chinese Academy of Agricultural Sciences. Five kinds of weeds were barnyard grass, green foxtail, goosegrass, crabgrass, small quinoa, which seeds were offered by China Agricultural University, College of Agronomy and Biotechnology. The two kinds of cabbages and five kinds of weeds were separately planted in flowerpots in a greenhouse of Beijing Research Center for Information Technology in Agricultural on 23 March, 2012. Each kind of plants was grown in 30 flowerpots, so the total number of the flowerpots was 210 for the seven kinds of plants.

The instrument for collecting spectral data was the Analytical Spectral Device Full Range FieldSpec Pro. (short for ASD), a handheld FieldSpec spectroradiometer (ASD, Boulder, Colorado, USA). The measuring range of the spectroradiometer is 350 ~ 2500 nm, within which the spectral resolution is 1.4 nm in the range of 350 ~ 1000nm and 2 nm in the range of 1000 ~ 2500 nm. the field of view (FOV) of the measuring probe is 25°.

2.2 Data acquisition

The spectral data of the 210 flowerpots of plant canopies mentioned above were collected outdoor in the test field of Beijing Research Center for Information Technology in Agricultural during 10:30 am. ~ 14:30 pm. on 28 April and 13 May, 2012, respectively corresponding to two growing periods of the plants.

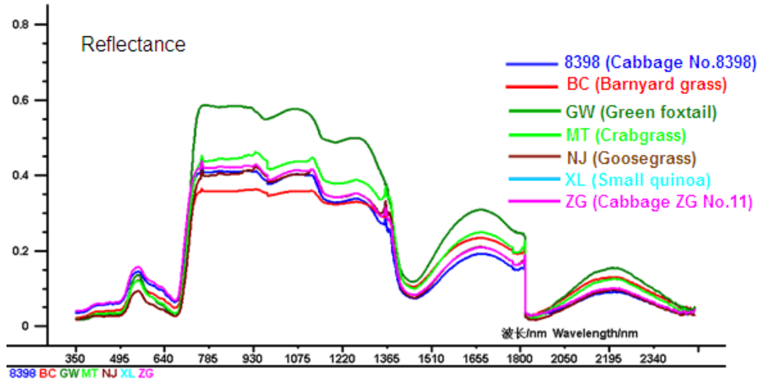
After calibrating using a whiteboard, the fiber optic probe of the ASD spectroradiometer was placed vertically above the plant canopy and began to measure the data. And the whiteboard should be calibrated once each 10-15 minutes depending on weather condition. In order not to affect the reflectivity of the plants, the operator must dress dark. The spectroradiometer was set as an output datum was obtained from the average of 10 times' measurements. The collected spectral data were first displayed and converted to text file format using a software allocated in ASD ViewSpectro Pro. Through imported to Microsoft EXCEL, the text-files were transformed to matrixes which were then guided into the Unscrambler and SAS software for further data processing.

For the first time of collecting, namely the data measurement on 28 April, 2012, each pot of plants was measured for three times so that the total number of the obtained spectral data was 90 for each kind of plants (30 flowerpots for each plant) and 630 for all the 7 kinds of plants. For the second time of measurement, namely on 13 May, 2012, each pot was measured for five times so that the total number of the collected spectral data was 150 for each kind of plant and 1050 for all the 7 kinds of plants.

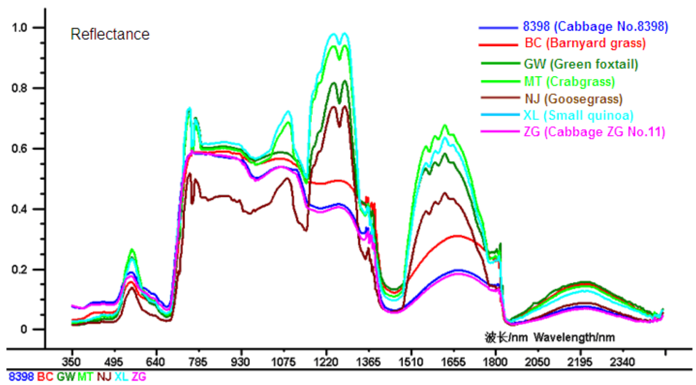
2.3 Observation and sampling of spectral data

In order to reduce the random errors which always accompanied with the spectral signal during the process of data acquisition, it needs to average the spectral data for each pot of plants. Since there were 3 data for each pot of plant at the first time measurement and 5 data for each pot of plants at the second time, the original data, 90 spectral data for the first time and 150 data for the second time, were respectively averaged to 30 data.

In order to clearly observe the differences of spectral reflectance of types of plants, the 30 spectral data of each plant were averaged and the responding curves of average spectral reflectance for 7 kinds of plants were shown in Fig.1. It is shown in Fig.1 that for the spectral curves at the first measurement stage, the spectral reflectivity of green foxtail in the range of 700 ~ 1800 nm is obviously higher than that of other kinds of plants, while the spectral reflectivity of crabgrass ranks the second. In the range of 750 ~ 1100 nm, the reflectance of Barnyard grass is lower than that of other kinds of plants, while the spectral reflectivity of cabbages is centered and the spectral curves of two kinds of cabbages almost superimpose. For the spectral curves at the second measurement stage, the spectral reflectance curves of goosegrass make an obvious distinction from that of other plants and further separate from the spectral curves of other plants in the whole wavelength range. It is seen that the spectral curves of cabbages are relatively gentle while spectral curves of weeds fluctuate more, which feature might be used to identify cabbages and weeds. Overall, although the spectral samples of cabbages and weeds can be roughly divided into two categories through observing the spectral curves, the spatial distributions of some samples seriously overlap, which make it difficult to exactly distinguish the type of each sample. In order to accurately classify cabbages and weeds, quantitative discriminant models should be established for and precise and deep research.



(a) Curves of average spectral reflectance at the first stage



(b) Curves of average spectral reflectance at the second stage

Fig.1 Curves of average spectral reflectance for the 7 kinds of plants

2.4 Principal Component Clustering Analysis

On the basis of the pre-processing mentioned above, Principal Component Analysis method was applied to classify the data of cabbages and weeds. Using the acronyms of Chinese Pin Yin, the seven kinds of plants are marked as follows: Cabbage 8398 as 8398, Cabbage Zhonggan No.11 as ZG 11, Barnyard grass as BC, green foxtail as GW, goosegrass as NJ, crabgrass as MT, and small quinoa as XL. For each kind of plant, 20 sample data were randomly selected as the training sample set, the other 10 data as the predicting sample set. In Unscrambler software system, the Full Cross Validation methods in PCA were used to extract the principal components which were then used to establish the classification models for the plants. The analysis process is started by extracting 20 principal components from the spectral data. And then the outliers were repeatedly excluded through considering the spatial aggregation conditions and spatial position of all the sample points in scoring graphs of the run results depending on the principle of maximization of distance between classes and minimization of distance within a class. Last, the appropriate principal components were decided according to the cumulative credibility of each PC and re-establish the classification model for observing the situation of clustering all the plants.

2.5 Extraction of characteristic wavelength

In order to find out the characteristic wavelength (CW) sensitive to identification of cabbages and weeds, it needs to analyze the score of each PC, accumulative confidence level, and loading diagrams resulted from the former PCA and the relationship between PC and original wavebands was expressed through loading graph. According to the loading graph of wavelength variable responding to the optimum PCs obtained from the former analysis, the wavelengths greatly (positive and negative) correlating with PCs were selected as the characteristic wavelengths sensitive to the identification of various of plants and with higher correlation for establishing the identification models. The loading coefficients of the selected wavelengths were used to reflect the importance of the wavelengths to PCs.

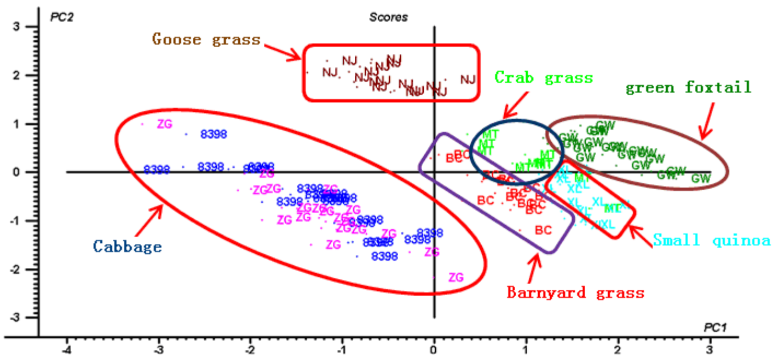
2.6 Bayes Classification Modeling

Setting the eight CWs extracted from the data at the first stages through PCA method as the input variables, discrimination function was established based on Bayes criterion and used to discriminate two cabbages and five kinds of weeds. During the process of specific implementation, 7 kinds of plants were separately labeled using categorical variables as Y-8398 (cabbage No. 8398), ZG (cabbage Zhonggan the 11th), BC (barnyard grass), GW (Setaria viridis), MT (crabgrass), NJ (Eleusine indica), and XL (small quinoa). For each kind of plant, two-thirds of the obtained samples were randomly selected as the training group (140 samples) so that all the samples for the 7 kinds of plants were divided into two groups, training sample and testing sample. Then depending on the data of categorical variables and 8 CWs, the discrimination model was established. In order to verify the reliability and robustness of the established model, the left one-third of sample (70 samples) were regarded as the testing group and taken as the input of the model to observe the correct classification rate.

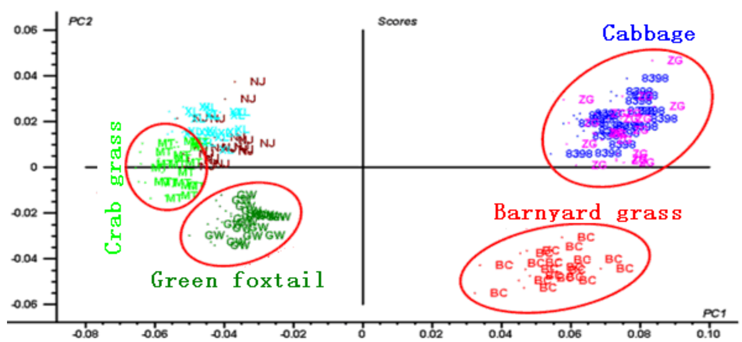
3 Results and discussion

3.1 Clustering based on PCA

The corresponding clustering results are shown in fig.2. In the case of optimum preprocessing, the score plot of PC 1 and PC 2 of the training set is shown in Fig.2, in which the horizontal axis presents the score value of the first PC and the vertical axis is the score value of the second PC. It can be seen from Fig.2(a) that the data sample of cabbages mainly concentrate in the second and third quadrant, goose grass samples mainly distribute in the above area of the second quadrant, crabgrass and green foxtail samples are in above area of the first quadrant, whereas barnyard grass and small quinoa principally focus near the horizontal axis in the fourth quadrant. It can be found from Fig.2(b) that cabbage samples closely distributing in the first quadrant show a good degree of aggregation which also indicates that the two varieties of cabbages can be regarded as the same category. As well in a good degree of aggregation, all the barnyard grass samples closely gather in the fourth quadrant and all the green foxtail in third quadrant. Although crabgrass samples distribute in both the second and third quadrant, the aggregating degree is still high. The samples of small quinoa and goosegrass slightly loosely gather together in the second quadrant and even partly appearing confusion, however on the whole the better concentrations are shown within same category. Therefore, it illustrates that PC1 and PC2 have better contribution to clustering cabbages and weeds. the synthetic method of principal component analysis (PCA) and clustering analysis can not only to a large extent reduce the data dimension but also greatly express the features of original data without losing the effective information.



(a) The clustering graph of the plants and weeds at the first stage



(b) The clustering graph of the plants and weeds at the second stage

Fig.2 The clustering graphs of PCs obtained after optimum preprocessing

3.2 Extracting CWs

The loading diagrams corresponding to the optimum PCs obtained from the processing of spectral data in the first and second stage are respectively shown in Fig.3 and 4, in which just the loading diagram for the first PC is list here because the number of the optimum PCs for the data in the first stage, e.g. 10 optimum PCs. In the loading diagrams, the horizontal coordinate represents the wavelength and the vertical coordinate is the load factor (i.e. the correlation between wavelength and plant species) of each wavelength, wherein, the larger the absolute value of the corresponding load factor of a wavelength variable is, the stronger the correlation between the PC and the corresponding load factor is, and the more sensitive to the discrimination of the plant species.

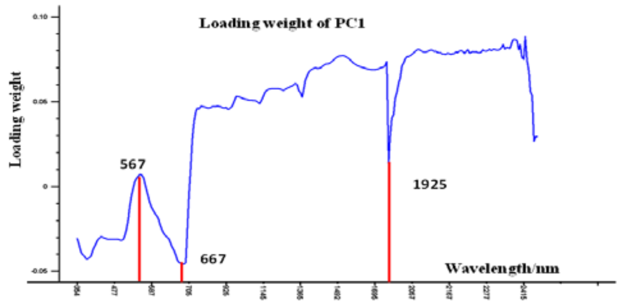
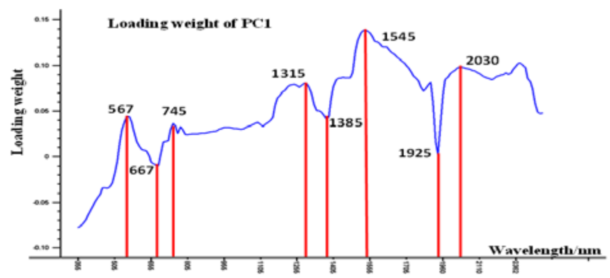
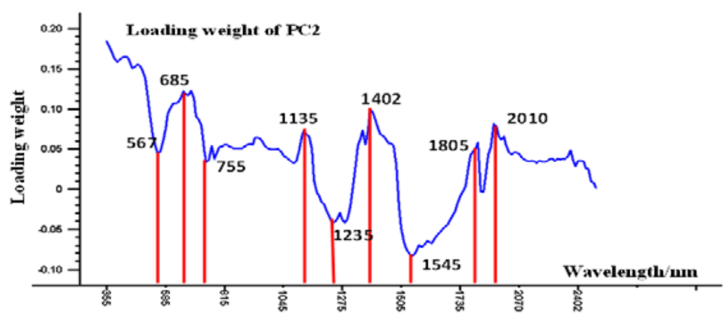


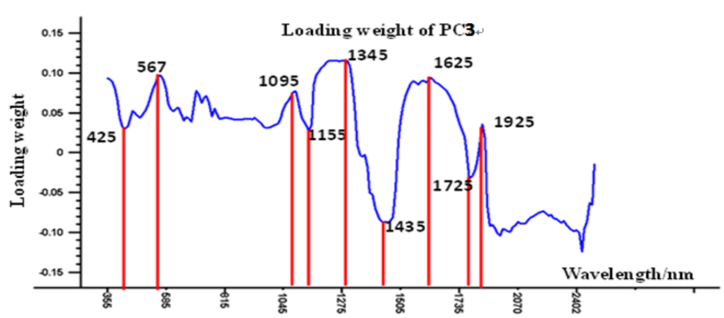
Fig.3 Loading graph of the PCs extracted from the data in the first stage



(a) the corresponding loading diagram of PC1



(b) the corresponding loading diagram of PC2



(c) the corresponding loading diagram of PC3

Fig.4 Loading graph of the PCs extracted from the data in the second stage

It could be found from the loading diagrams that obvious crests and troughs present at some wavelengths and the rates of change of corresponding load factors appear local maximum. These wavelengths are likely to play a decisive role in the identification of cabbages and weeds. Whereby, the corresponding CWs for the two stages' plants were selected from the loading diagrams and shown in Table 1. The CWs selected from the spectral data in the first and second stages were respectively 16 and 23.

Although the amount of data were already greatly reduced relative to the original collected data, the data size is still relatively large with regard to the later instrumental development of agricultural machinery. There it is in need of further selection of CWs. The selecting principle is to start with the first PC by ranking the order of wavelengths according to the absolute value of the corresponding load factors and selecting the wavelengths at which the absolute value of load was large and greatly obvious crests and troughs was presenting in loading diagrams. As the result, the CWs selected from the spectral data were 567, 667, 715, 1345, 1402, 1725, 1925, and 2015 nm for the first stage, 567, 667, 745, 1345, 1402, 1545, 1725, and 1925 nm for the second stage. In order to verify the effect of the selected CWs, it is still necessary to establish the identification models using these CWs and analyze the correct classification rate.

Table 1. The CWs selected from the loading diagrams of PC

Testing stage	Characteristic Wavelengths (CWs) / (nm)
The first stage	552, 567, 602, 607, 667, 715, 725, 1345, 1402, 1447, 1725, 1925, 1945, 1955, 2015, 2072
The second stage	425, 567, 667, 685, 745, 755, 1095, 1135, 1155, 1235, 1315, 1345, 1385, 1402, 1435, 1535, 1545, 1625, 1725, 1805, 1815, 1925, 2030

In addition, among the each 8 CWs which were extracted respectively from the spectral data of cabbages and weeds at two growth stages sensitive to the identification of cabbages and weeds, just two of them were different, which indicated that the change of spectral features with the growth of cabbages and weeds had little influence on the identification of them. Hence, it is feasible to take use of spectral characteristics to precisely control weeds in cabbage fields.

3.3 Bayes classification

After programming based on discriminant analysis, submitting the codes, and running the program, the discriminant functions of various models were obtained and are shown as Equation (1), by which the frequency numbers of each training sample for 7 kinds of plants judged into various categories are exhibited in Table 2 and the misjudged rates of 7 kinds of plants are shown as Table 3.

$$\begin{cases} BC = -49.89 + 75.11x_1 + 976.62x_2 - 247.88x_3 + 8.52x_4 + 184.98x_5 + 360.82x_6 + 115.39x_7 - 464.08x_8 \\ GW = -68.91 + 1116x_1 + 745.44x_2 - 1112x_3 + 677.47x_4 - 436.09x_5 + 222.38x_6 + 44.97x_7 - 216.67x_8 \\ MT = -45.98 + 213.48x_1 + 668.17x_2 - 309.73x_3 + 300.99x_4 + 149.24x_5 - 24.16x_6 + 114.15x_7 - 285.32x_8 \\ NJ = -41.12 + 187.37x_1 + 725.07x_2 - 422.14x_3 + 663.92x_4 - 98.88x_5 - 338.86x_6 + 127.71x_7 - 76.97x_8 \\ XL = -93.59 - 3026x_1 + 1839x_2 + 1459x_3 - 329.18x_4 + 78.99x_5 + 600.50x_6 - 3.27x_7 - 331.96x_8 \\ Y - 8398 = -83.34 - 1803x_1 + 1380x_2 + 1344x_3 + 39.46x_4 - 148.77x_5 - 358.52x_6 + 87.37x_7 + 218.22x_8 \\ ZG = -88.29 - 1874x_1 + 1398x_2 + 1432x_3 - 118.03x_4 - 87.21x_5 - 211.36x_6 + 54.09x_7 + 172.79x_8 \end{cases} \quad (1)$$

Table 2. the frequency number of 7 varieties of training samples judged into various categories in Bayes analysis

Crab grass

	Categories / Samples	Barnyrad grass	Green foxtail	Crab grass	Goose grass	Small quinoa	Cabbage ‘8398’	Cabbage ‘zhonggan 11’
Training samples	Barnyard grass	19	0	1	0	0	0	0
	Green foxtial	0	20	0	0	0	0	0
	Crabgrass	0	0	19	1	0	0	0
	Goose grass	0	0	2	18	0	0	0
	Small quinoa	0	0	0	0	20	0	0
	Cabbage ‘8398’	0	0	0	0	0	15	5
	Cabbage ‘zhonggan 11’	0	0	0	0	0	4	16
	Overall samples	19	20	22	19	20	19	21
Predicting samples	Barnyard grass	10	0	0	0	0	0	0
	Green foxtial	0	10	0	0	0	0	0
	Crabgrass	1	0	8	1	0	0	0
	Goose grass	0	0	0	10	0	0	0
	Small quinoa	0	0	0	0	10	0	0
	Cabbage ‘8398’	0	0	0	0	0	7	3
	Cabbage ‘zhonggan 11’	0	0	0	0	0	6	4
	Overall samples	11	10	8	11	10	13	7

Among the results of classifying the training sets, 5 Cabbage ‘8398’ samples were misjudged as Cabbage ‘Zhonggan No.11’ and 4 ‘Zhonggan No.11’ as ‘ 8398’, which was maybe because of the reason that they all belong to the species of cabbage with the consistent internal structural organization and almost the same appearance. From this point, it is apparent that different varieties of cabbages can be considered as the same category. In addition, 1 sample of barnyard grass was mistaken as crabgrass, 1 crabgrass be mistaken as goose grass, 2 samples of crabgrass were misjudged as crabgrass, which is perhaps because they are all monocotyledonous weeds with similar internal structure and composition. Total of 13 samples in the training set were falsely adjusted and the misjudged rate was 0.0929, namely, the overall correct classification rate was 90.7%.

For the classification results of the testing set, 3 Cabbage '8398' were mistaken as Cabbage 'Zhonggan No.11' and 6 Cabbage 'Zhonggan No.11' were mistaken as '8398'. The more, 1 crabgrass be mistaken as goosegrass, 1 crabgrass is misidentified as barnyardgrass. Total of 11 samples were classified wrong so the misjudged rate is 0.1571, in other words, the correct identification rate is 84.3%.

Table 3. Rates of erroneous identification for the 7 kinds of samples in Bayes discriminant analysis

Categories	prior probability	Rates of erroneous identification of the training sample set	Rates of erroneous identification of the predicting sample set
Barnyard grass	0.1429	0.0500	0.0000
Green foxtial	0.1429	0.0000	0.0000
Crabgrass	0.1429	0.0500	0.2000
Goose grass	0.1429	0.1000	0.0000
Small quinoa	0.1429	0.0000	0.0000
Cabbage '8398'	0.1429	0.2500	0.3000
Cabbage 'zhonggan 11'	0.1429	0.2000	0.6000
Overall samples	1.0000	0.0929	0.1571

In order to verify the similarity of the spectral characteristics of different varieties of cabbages, cabbage '8398' and 'Zhonggan No.11' were combined into one category. The spectral data of the 8 CWs which had been used earlier were still made as the input variables. All varieties of plants were labeled by categorical variables as GL (cabbage), BC (barnyard grass), GW (Setaria viridis), MT (crabgrass), NJ (Eleusine indica) and XL (small quinoa). Repeating the previous operation, the discriminant functions were obtained and shown as Equation (2) and the classification results are shown in Table 4 and 5.

As to the classification results of the training set, for green foxtails one was mistaken as crabgrass and one other as goose grass; for crabgrass one was mistaken as Barnyard grass and three of it were mistaken as goose grass; for goose grass, one of it was mistaken as barnyard grass and all the other were correctly classified. Overall, the error rate was 0.05, namely, the correct classification rate was 95%. As to the testing set, all the samples were correctly classified so its correct classification rate was 100%. Compared with the former correct classification rate when the two varieties of cabbages were regarded as two categories, the correct classification rate when the two varieties of cabbages were regarded as the same category had been greatly raised, which indicated that different varieties of cabbages owns similar spectral characteristics. Therefore, when precision chemical applications and other agricultural mechanical operation are implemented through spectral identification of cabbages and weeds in fields

$$\begin{cases} BC = -50.96 + 3412x_1 - 3703x_2 - 1259x_3 + 1986x_4 + 301.55x_5 - 400.99x_6 + 341.52x_7 - 79.76x_8 \\ GW = -74.97 + 626.77x_1 - 2181x_2 - 1144x_3 + 1988x_4 + 246x_5 - 27.37x_6 + 89.15x_7 - 30.93x_8 \\ MT = -53.05 + 2481x_1 - 3611x_2 - 906.51x_3 + 1878x_4 + 356.56x_5 - 300.68x_6 + 176.24x_7 - 27x_8 \\ NJ = -42.19 + 3054x_1 - 3533x_2 - 1284x_3 + 1688x_4 + 280.05x_5 - 146.62x_6 + 75.17x_7 - 1.76x_8 \\ XL = -72.26 - 5909x_1 + 7404x_2 - 1905x_3 + 585.01x_4 + 241.69x_5 - 821.22x_6 + 1190x_7 - 356.42x_8 \\ GL = -73.44 + 4979x_1 - 5095x_2 - 744.15x_3 + 2056x_4 + 59.29x_5 + 981.29x_6 - 1335x_7 + 374.43x_8 \end{cases} \quad (2)$$

Table 4. Frequency number of 6 kinds of samples classified into responding categories in Bayes analysis

Samples	Categories	Barnyrad grass	Green foxtail	Crab grass	Goose grass	Small quinoa	Cabbage
Training samples	Barnyard grass	20	0	0	0	0	0
	Green foxtial	0	18	1	1	0	0
	Crabgrass	1	0	16	3	0	0
	Goose grass	1	0	0	19	0	0
	Small quinoa	0	0	0	0	20	0
	Cabbage	0	0	0	0	0	40
	Overall samples	22	18	17	23	20	40
Predicting samples	Barnyard grass	10	0	0	0	0	0
	Green foxtial	0	10	0	0	0	0
	Crabgrass	0	0	10	0	0	0
	Goose grass	0	0	0	10	0	0
	Small quinoa	0	0	0	0	10	0
	Cabbage	0	0	0	0	0	20
	Overall samples	10	10	10	10	10	20

Table 5. misjudged rates of the 6 kinds of samples in Bayes discriminant

Categories	prior probability	Rates of erroneous identification of the training sample set	Rates of erroneous identification of the predicting sample set
Barnyard grass	0.1429	0.0000	0.0000
Green foxtial	0.1429	0.1000	0.0000
Crabgrass	0.1429	0.2000	0.0000
Goose grass	0.1429	0.0500	0.0000
Small quinoa	0.1429	0.0000	0.0000
Cabbage	0.2857	0.0000	0.0000
Overall samples	1.0000	0.0500	0.0000

4 Conclusions

(1) According to the load factors and its changing rate of PCs corresponding to the spectral wavelengths, the CWs which were sensitive to plant identification were extracted respectively as 567, 667, 715, 1345, 1402, 1725, 1925, and 2015 nm for the first stage and 567, 667, 745, 1345, 1402, 1545, 1725, and 1925nm for the second stage. In addition, among the each 8 CWs of two stages, just two of the CWs were different, which indicated that the changes of spectral characteristics at different growth stages of cabbages have little influence on identification of cabbages and weeds. Therefore, it is reliable to make use of spectral features to control weeds in cabbage fields.

(2) The corresponding spectral data of the 8 CWs extracted from the data at the first stage were taken as the input variables of the model which was built up using Bayes discriminant analysis to classify two varieties of cabbages and five kinds of weeds. The correct classification rates for the training and testing sets were respectively 90.7% and 84.3%. When the two varieties of cabbages were regarded as the same category, using the analysis method the correct classification rates of the training and testing sets were respectively 95% and 100%, which indicated that different varieties of cabbages owned similar the spectral features. Therefore combining different varieties of cabbages as the same category could greatly improve the correct classification rate compared with the condition in which two varieties of cabbages were seen as different categories.

ACKNOWLEDGEMENTS

This research was supported by the National High Technology Research and Development Program of China (863 Program) (No. 2012AA101904). The authors acknowledge the National Experimental Station of Precision Agriculture of China.

REFERENCE

Chen, S., Li, Y., Mao, H., et al. (2009). Research on distinguishing weed from crop using spectrum analysis technology. *Spectroscopy and Spectral Analysis*, 29(2): 463-466. (In Chinese with English abstract)

Deng, W., Zhang, L., He, X., et al. (2009). SVM-based spectral recognition of corn and weeds at seedling stage in fields. *Spectroscopy and Spectral Analysis*, 29(7): 1906-1910. (In Chinese with English abstract)

Deng, W., Huang, Y., Zhao, C., Chen, L., Meng, Z. (2011). Comparison of SVM, RBF-NN, and DT for crop and weed identification based on spectral measurement over corn fields. *International Agricultural Engineering Journal*, 20(1):11-19.

FAO. 2009. The lurking menace of weeds. Available at: http://www.fao.org/news/story/en/item/29402/icode/

Jurado-Expósito, M., López-Granados, F., Atenciano, S., Garcıa-Torres, L., González-Andujar, J. L. (2003). Discrimination of weed seedlings, wheat (Triticum aestivum) stubble and sunflower (Helianthus annuus) by near-infrared reflectance spectroscopy (NIRS). *Crop Protection*, 22(10): 1177-1180.

Koger, C. H., Bruce, L. M., Shaw, D. R., Reddy, K. N. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (Ipomoea lacunosa) in soybean (Glycine max). *Remote sensing of environment*, 86(1): 108-119.

Mao, W., Wang, Y., Wang, Y., Zhang, X. (2005). Spectrum analysis of crop and weeds at seedling. *Spectroscopy and Spectral Analysis*, 25(6): 984-987. (In Chinese with English abstract)

Piron, A., Leemans, V., Kleynen, O., Lebeau, F., Destain, M. F. (2008). Selection of the most efficient wavelength bands for discriminating weeds from crop. *Computers and Electronics in Agriculture*, 62(2): 141-148.

Slaughter, D. C., Lanini, W. T., Giles, D. K. (2004). Discriminating weeds from processing tomato plants using visible and near-infrared spectroscopy. *Transactions of the ASAE*, 47(6): 1907-1911.

Tang, J. (2010). Research on Weed Detection and Navigation Parameters Acquisition of Pesticide Spraying Robot. Yang Ling: North-west Agriculture and Forestry University. (In Chinese with English abstract)

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote sensing of environment*, 91(3): 354-376.

Thompson, J. F., Stafford, J. V., Miller, P. C. H. (1991). Potential for automatic weed detection and selective herbicide application. *Crop Protection*, 10(4): 254-259.