Research Article
Spectral Feature Extraction for Flue-Cured Virginia Tobacco Leaves of Different Maturity Grades

[1]Yin Liang, [2]Junying Li and [1]Yunwei Zhang
[1]Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology,
[2]Yunnan Academy of Tobacco Agricultural Sciences, Kunming 650500, P.R. China

Abstract: The reflectance spectrum has hundred bands of spectral information. The construction of the right Spectral characteristics in such huge information is important for the Classification of tobacco leaves of different maturity. When implementing the continuum removal method to extract Spectral feature, two results has been shown. One result shows that there are ten spectral features which can be used in Classification research, but only the red edge position and C3 feature performs better. The other result shows that Matching C2 or C3 with other features will boost Class separability. When constructing the eigenvector based on the ten features, it is better to consider them as the component.

INTRODUCTION

An important problem in the reaping of tobacco leaves is to judge the maturity of tobacco leaves. In China, growers tend to judge the maturity of tobacco leaves by appearance and harvest the tobacco leaves in accordance with corresponding standards (JianKang *et al.*, 2010). Even those countries, having made systematic researches on maturity of tobacco leaves already, are found to judge the maturity by sampling and determining biochemical parameters or by using colorimetric method (Folin *et al.*, 2007) as well. The above-mentioned methods of judging the maturity of tobacco leaves are difficult to popularize, especially under the condition of expanding planting area of tobacco, owing to their defects of time-consuming and emerging-consuming. As the investigation of vegetation based on hyper-spectral remote sensing advances, quantifying the maturity of tobacco leaves which means an issue of classifying tobacco leaves of different maturity based on hyper-spectral remote sensing data is available.

Nevertheless, at present, only a few researchers apply hyper-spectral remote sensing technology to judging the field maturity of tobacco leaves. The maturity of tobacco leaves would show an influence on reflectance between the band 503-651 nm, among which the reflectance at the three wavelength positions of 514, 629 and 650 nm, respectively was factor of predicting the maturity of tobacco leaves (Folin *et al.*, 2007). The reflectance between the near-infrared reflectance between the band 760-1,300 nm would show a declining trend along with enhancing maturity of tobacco leaves as well (Folin *et al.*, 2008).

During the research on red edge position of reflectance spectrum of tobacco leaves of different maturity (Xiangyang *et al.*, 2007), the author found that when the red edge moves to the range of 693-695 nm, the maturity of tobacco leaves have achieved the ripe status; And tobacco leaves get over-mature when the red edge moves to 688 nm, the limiting value. In short, the above-mentioned researches present part of spectral features. Among them, red edge position is a valid spectral parameter which used to quantifying the senescence of plants (Niemann, 1995). However, the researchers above didn't research the divisibility of spectral features so that subsequent researches on judging the field maturity of tobacco leaves are possible to be based on poor classification feature, leading to poor performance of maturity classifier of tobacco leaves (Qingxi and Zhang, 2006; Jinbao *et al.*, 2010; Gonzales and Woods, 2003). This study extends the previous researches by virtue of continuum removal. Two purposes need to be achieved: one is to extract more effective spectral features reflecting the maturity of tobacco leaves in the visible light band; the other is to research the divisibility of single spectral feature and high-dimensional eigenvector composed by single spectral feature. By choosing the spectral feature or high-dimensional eigenvector which has better divisibility than others, classification features of the classifier could be optimized.

Corresponding Author: Yin Liang, Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, P.R. China

Table 1: Treatments

Treat	Maturity	After-transplant days (d)	Appearance
M1	Initial mature	90~100	Leaves are light green with a little yellow. Two thirds of main vein and part of branch vein get white. There are quite few mature spots.
M2	Proper mature	110~130	Leaves are yellow-green. Main vein and branch vein get white and bright. There are many mature spots and fewer brown spots.
M3	Over-mature	>130	Leaves are yellow or yellow-white. Main vein and branch vein get white. There are many brown spots. Leaves are damaged.

MATERIALS AND METHODS

The test was conducted in the experimental base of Yunnan Academy of Tobacco Science from Jul. 2012 to Sep. 2012, during which the tobacco was managed in the field of moderate fertilizer according to cultivation standards of high-quality flue-cured Virginia tobacco. Leaf 14, 15, 16 and 17 in the central trunk of Yunyan 87, the tested variety, were observed during the test. Three treatments were set based on appearance of leaves as shown in Table 1, combining the after-transplant days.

Tag the tobacco leaves in the test region to prevent them from being picked and baked. The test begins at the time of the lower part leaves are first time to be harvested. Observing test is conducted every 10 days and ends at the picking and baking of the upper leaves.

Height adjustable bracket, dark panel, white reference and ASD fieldspec$^{@}$3 spectroradiometer were used. During observation, lift the cantilever of bracket to the position of the observed leave, put the blackboard on the cantilever, put the observed leaves on the blackboard flat and fasten the edge with clips. Scan the spectrums at the base, middle and tip of the tagged leaves and taking the mean value as the result. The observation time interval of test is between 9:30 AM and 3:30 PM. The meteorological condition within the time interval of observation is clear weather and sky of no cirrocumulus.

Methods: Conduct continuum removal on reflectance spectrum to obtain normalized absorption spectrum (Kokaly and Clark, 1999). This transformation can remove the reflectance difference of spectrum formed under different light conditions and amplify the tiny absorption feature. The algorithm of continuum removal is realized by programming on the platform of MATLAB R2010a.

The extraction of the spectral features is conducted after the processing of continuum removal. First interpret the diagram of determination coefficient of normalized absorption spectrum and maturity of tobacco leaves and then extract the spectral features these features are expressed as follows:

C1: Red edge position parameter calculated by IG model (Miller and Hare, 1990)

C2: Area of absorption peak in red light band:

$$C_2 = \sum_{\lambda = 550}^{750} r_\lambda \tag{1}$$

r_λ = Value of normalized reflectance

C3: Area ratio between absorption peak of blue-green light band and absorption peak of red light band:

$$C_3 = \left. \sum_{\lambda = 350}^{550} r_\lambda \middle/ \sum_{\lambda = 550}^{750} r_\lambda \right. \tag{2}$$

C4: Maximum absorption depth ratio between absorption peak of blue-green light bands and absorption peak of red light bands:

$$C_4 = \frac{(d_{350}, d_{550})_{max}}{(d_{550}, d_{750})_{max}} \tag{3}$$

d_{350} = Normalized absorption depths at positions of 350 nm

d_{550} = Normalized absorption depths at positions of 550 nm

d_{750} = Normalized absorption depths at positions of 750 nm

C5: Area of left half of absorption peak of red light bands:

$$C_5 = \sum_{\lambda=677}^{750} r_\lambda \tag{4}$$

C6: Maximum absorption depth ratio of absorption small peak on left and right of blue-green light band:

$$C_6 = \frac{(r_{350}, r_{450})_{max}}{(r_{450}, r_{550})_{max}} \tag{5}$$

C7: Absorption depth difference of absorption small peak between left and right of blue-green light band:

$$C_7 = (r_{350}, r_{450})_{max} - (r_{450}, r_{550})_{max} \tag{6}$$

C8: Area ratio of absorption small peak between left and right of blue-green light band:

$$C_8 = \left. \sum_{\lambda = 350}^{449} r_\lambda \middle/ \sum_{\lambda = 450}^{549} r_\lambda \right. \tag{7}$$

C9: Linear combination of normalized reflectance factor at positions of 434, 432 and 369 nm, respectively:

$$c9 = 62.74r_{434} - 59.17r_{432} - 7.33r_{369} + 4.52 \qquad (8)$$

C10: Linear combination of normalized reflectance factor at positions of 696 and 561 nm:

$$c10 = 6.62\,r_{696} - 10.09\,r_{561} + 8.13 \qquad (9)$$

Use Excel 2003 software to make calculation and relevant analysis of spectral features, use SPSS 17.0 software for regression analysis; use Excel 2003 software to draw normalized absorption spectrum of treated average normalized absorption spectrum and category of maturity of samples that have been tagged and use Excel 2003 software to draw the diagram of determination coefficient of maturity after relevant analysis.

The divisibility of the classification features are tested after the extraction of them. Use J-M distance method (Swain and Davis, 1978) to measure the divisibility of one-dimensional spectral features and two-dimensional spectral eigenvector. Choose spectral features and eigenvector enabling JM distance to achieve the maximum value to be spectral features of better divisibility. J-M distance is calculated according to the formulas in reference (Qingxi and Zhang, 2006). The algorithm is realized by programming on the platform of MATLAB R2010a. Judge the divisibility of classification feature on the spectral divisibility principle of J-M (Jinbao *et al*., 2010; Degang *et al*., 2010).

RESULTS AND DISCUSSION

Average normalized absorption spectrum treated for different maturity shows three major differences in the visible light bands dominated by photosynthetic pigment as been seen from Fig. 1. Category 1 difference is characteristic absorption peak between 550 and 750 nm, showing a reducing tendency along with change of the maturity of tobacco leaves; Category 2 difference is absorption feature between 350 and 550 nm, showing that symmetry of two tiny bulges will change along with the maturity of tobacco leaves and when tobacco leaves are immature, the right small peak is lower than the left one; when tobacco leaves are mature, the two small peaks are more symmetrical; when tobacco leaves are over-mature, the right small peak is higher than the left one. Category 3 difference is that area of absorption peak between the band 350 and 550 nm will increase and area of absorption peak between the band 550 and 750 nm will decrease along with enhancement of the maturity of tobacco leaves and likewise, the maximum absorption depth of two absorption peaks show similar characteristic of change.
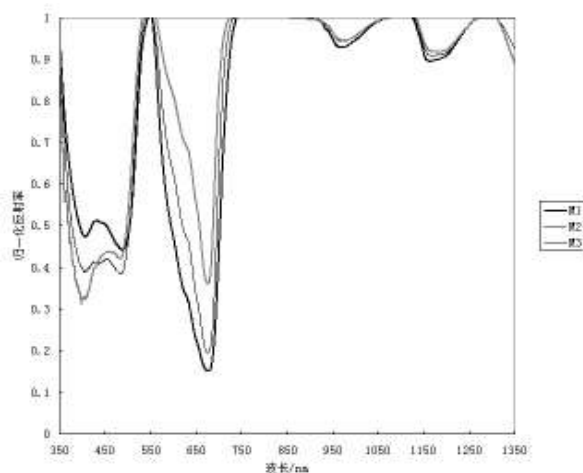

Fig. 1: The difference of average normalization absorption spectrum of samples under different maturity treatment

More specifically, from Fig. 2, with enhancement of the maturity of tobacco leaves, absorption peak between the band 550 and 750 nm keeps changing in shape: the left of absorption peak lifts; the right of absorption peak moves in the direction of short wave; central wavelength of absorption peak keeps unchanged in position but drops greatly in depth, which are close to the result reported in reference (Jing *et al*., 2010). The reason why absorption peak in the band changes in this mode is that pigment molecules of chlorophyll *a* and *b*, having characteristic absorption peak in red light region, are far lower than blue-purple light band in absorption strength (Wenjiang, 2009; Lichtenthaler and Buschmann, 2001) and therefore the central absorption reflectance is difficult to reach the asymptote (Curran, 1989) even under the condition of saturated chlorophyll content, only to find that the characteristic absorption peak is hypersensitive to change of chlorophyll content. The result is that each spectral feature able to reflect characteristic of change of the absorption peak can be used as features of measuring the maturity of tobacco leaves.

There is a big aliasing in absorption feature in the blue-green light short band (between the band 350 and 550 nm) with normalized absorption spectrum of tobacco leaves of different maturity as been seen from Fig. 3.

If spectral parameters such as normalized absorption depth and absorption area calculated based on absolute size and form of value are regarded as spectral features, then the spectral features of good divisibility cannot be obtained in the band; however, if spectral parameters such as symmetry of left and right absorption small peaks calculated on the basis of relative size and form of value are regarded as spectral
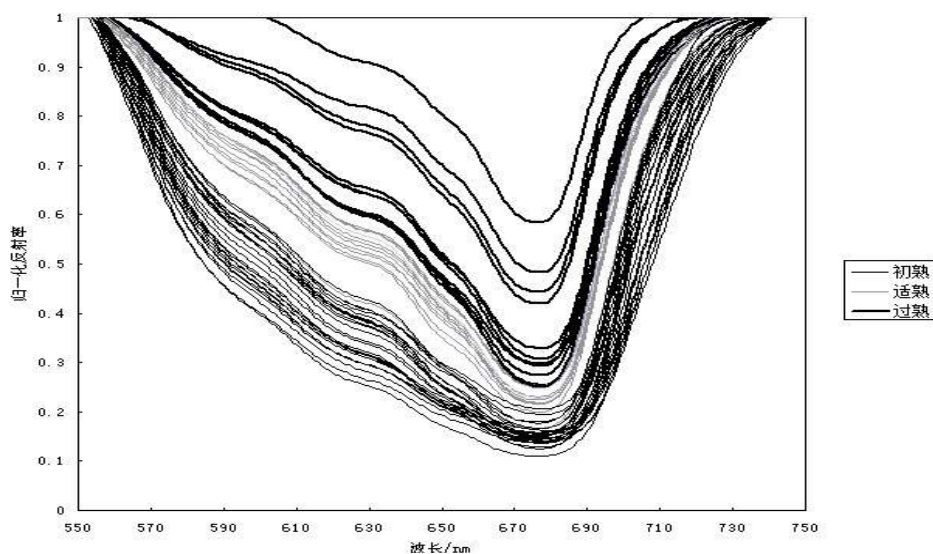
Fig. 2: The change of the absorption feature between the band 550 and 750 nm with the maturity
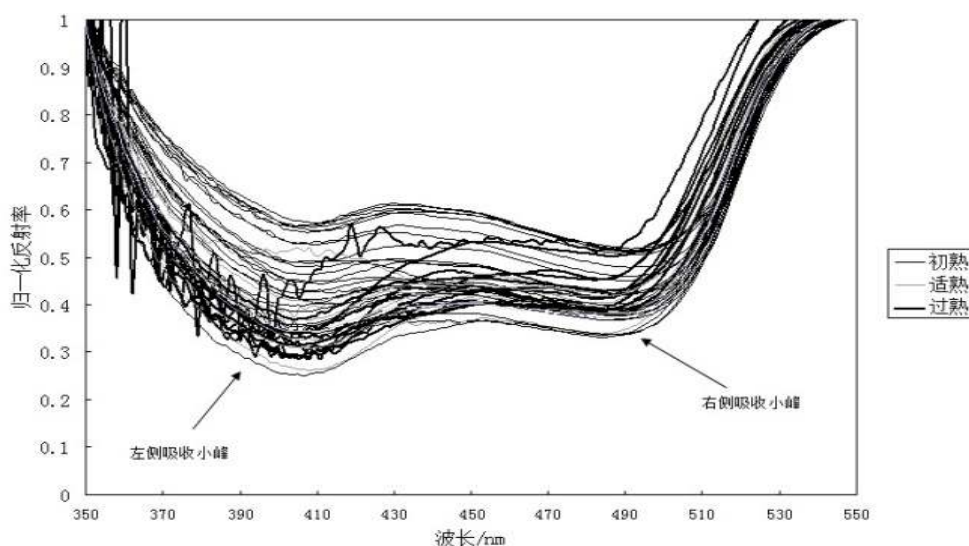

Fig. 3: Variation of the absorption peak within the band of 350~550 nm with the maturity

Table 2: Spectral feature in the bands between 350~550 nm and the J-M distance value

Spectral feature	J-M$_{1,2}$ distance	J-M$_{2,3}$ distance
Normalized absorption depth at position of 405 nm	0.67	0.40
Area of absorption peak between the band 350 and 550 nm	0.61	0.39
Depth ratio of left small peak and right small peak (C6)	0.91	0.78
Depth difference of left small peak and right small peak (C7)	0.89	0.79
Area ratio of left small peak and right small peak (C8)	1.10	0.86

features, the spectral features can lessen aliasing effect among categories and improve the divisibility of spectral features constructed in the band, which is manifested by increasing J-M distance of categories of adjacent maturity as shown in Table 2.

As hyper-spectral inversion of plant disease level requires to search the sensitive bands of disease identification (Keyan *et al.*, 2010). Furthermore, it is required to determine the sensitive bands of the

maturity before solving the problem of classifying the maturity of tobacco leaves.

Use diagram of determination coefficient to determine the sensitive bands of the maturity of tobacco leaves as shown in Fig. 4.

The figure indicates that the maturity of tobacco leaves affects mainly the normalized reflectance in blue-purple light region (between the band 350 and 450 nm), green light region (between the band 510 and 550
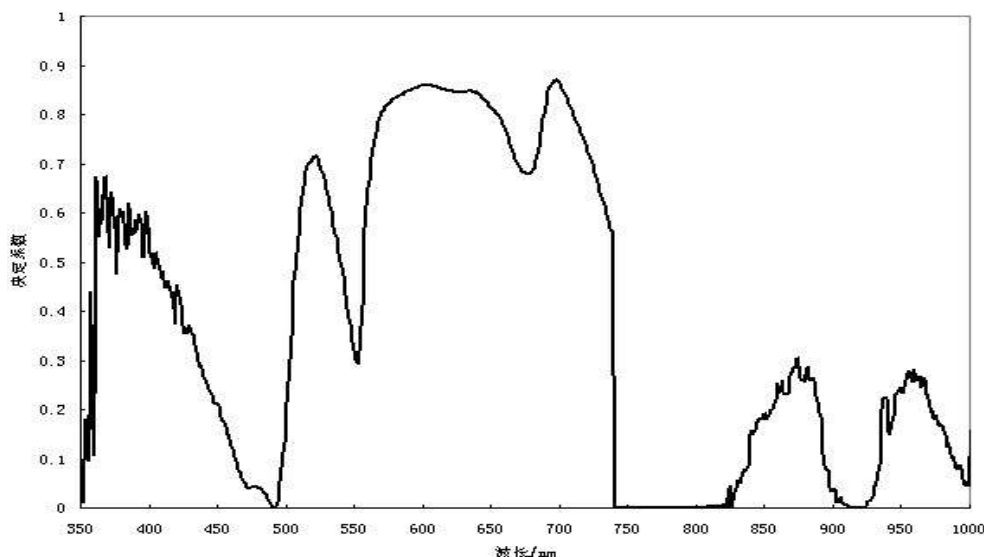
Fig. 4: Coefficient of determination for the maturity of tobacco leaf

Table 3: J-M distance of one-dimensional spectral feature

Spectral feature	J-M$_{1,2}$ distance	J-M$_{2,3}$ distance
C1	1.24	1.03
C2	1.24	0.88
C3	1.13	1.08
C4	0.93	0.97
C5	1.24	0.88
C9	0.64	0.72
C10	1.31	0.86

J-M distance of C6, C7 and C8 is seen in Table 2

Table 4: J-M distance of eigenvectors of 2-to-10-dimensional spectrum

Eigenvector	J-M$_{1,2}$ distance	J-M$_{2,3}$ distance
(C1, C3)	1.312	1.255
(C2, C5)	1.414	1.414
(C1, C2, C5)	1.414	1.414
(C1, C3, C6)	1.311	1.254
(C1, C2, C3, C5)	1.414	1.414
(C1, C3, C6, C9)	1.292	1.245
(C1, C2, C3, C4, C5)	1.414	1.414
(C1, C6, C7, C8, C10)	1.311	1.256
(C1, C2, C5, C8, C9, C10)	1.414	1.414
(C1, C3, C4, C7, C9, C10)	1.311	1.254
(C2, C4, C5, C6, C8, C9, C10)	1.414	1.414
(C1, C3, C4, C6, C7, C8, C10)	1.316	1.254
(C1, C2, C3, C4, C5, C6, C7, C8)	1.414	1.414
(C1, C3, C4, C6, C7, C8, C9, C10)	1.300	1.245
(C1, C2, C3, C5, C6, C7, C8, C9, C10)	1.414	1.414
(C1, C3, C4, C5, C6, C7, C8, C9, C10)	1.414	1.406
(C1, C2, C3, C4, C5, C6, C7, C8, C9, C10)	1.414	1.414

nm) and red light region (between the band 550 and 740 nm), which are, therefore, the sensitive bands of the maturity of tobacco leaves and the spectral features reflecting the maturity of tobacco leaves shall be constructed in the above-mentioned three sensitive bands. The reflectance factor at positions of 514, 629 and 650 nm, respectively selected in reference (Folin *et al.*, 2008) is located in the very last two bands.

The maturity process of tobacco leaves is a unidirectional and consistent process, changing step by step in the sequence of immaturity, initial maturity,

proper maturity and over maturity. Hence tobacco leaves among categories of adjacent maturity are easy to be confused and whereas tobacco leaves among categories of non-adjacent maturity are difficult to be confused. What's more, over-mature tobacco leaves and initial mature tobacco leaves show obvious difference in appearance; initial mature tobacco leaves and mature tobacco leaves are likely to be very similar in appearance. Therefore, the author measures the divisibility of spectral features mainly by means of calculating J-M distance among adjacent categories in the process of researching classification of the maturity, instead of taking in account J-M distance of non-adjacent categories.

Table 3 illustrates that the divisibility of spectrum containing none of the 10 spectral features constructed is excellent (Jinbao *et al.*, 2010), except that only two features, i.e., red edge position parameter (C1) and area ratio between absorption peak of blue-green light band and absorption peak of red light band (C3) share equal divisibility and other features cannot distinguish from categories of adjacent maturity effectively.

Using multiple one-dimensional spectral features as components to construct high-dimensional eigenvector is a widely used method in research on pattern recognition (Gonzales and Woods, 2003)

Calculate J-M distance of high-dimensional eigenvector formed by the 10 spectral features as shown in Table 4

Table 4 illustrates that Feature C2 and C5 are key components of constituting eigenvector, to be specific, J-M distance is determined to be 1.414 as long as C2 and C5 are contained in components of high-dimensional eigenvector; J-M distance will drop to different extents if only one of C2 and C5 is contained; J-M distance will drop more obviously if Feature C2 and C5 are not contained in the components; as far as classification of tobacco leaves of different maturity is

concerned, it may increase the divisibility to expand from one-dimensional features to high-dimensional eigenvector, it will not do that to expand high-dimensional eigenvector to higher-dimensional eigenvector and if Feature C2 and C5 are not contained in components of higher-dimensional eigenvector, the divisibility will reduce instead of increasing. Apart from that, the maximum value of J-M distance is less than 1.8, failing to reach good divisibility in standard sense as for the extracted 10 spectral features and the high-dimensional eigenvector formed by combination of the spectral features. Therefore, the effect is limited to optimize the classification feature of the maturity in the visible light bands by means of features optimization and by adding dimensionality of classification eigenvector.

CONCLUSION

It is found, through research on normalized absorption spectrum of the middle leaves of Yunyan 87 of different maturity, that spectral features correlated with absorption features between the band 550 and 750 nm can reflect the maturity of tobacco leaves very effectively; absorption feature in the blue-green light band between 350 and 550 nm is affected by the maturity as well which is manifested by the change of symmetry of two bulges on top of absorption peak; the absorption depth and the absorption peak area of absorption peak in blue-green light band and absorption peak in red light band will change along with the maturity of tobacco leaves. The 9 spectral parameters extracted from almost one hundred band information can be used as classification feature of the maturity of tobacco leaves besides the red edge position parameter. Nevertheless, the 10 one-dimensional features are not optimal to solve the problem of classifying the maturity of tobacco leaves under the condition of their single use. Therefore, the author sifts one-dimensional spectral features and high-dimensional spectral features formed by combining multiple one-dimensional spectral features by virtue of J-M distance method which achieves better inspection effect, by which the author finds that although the two-dimensional eigenvector formed by Feature C2 and C5 as well as higher-dimensional eigenvector containing the two above-mentioned spectral features can improve the divisibility and they are helpful for solving the problem of classifying the maturity of tobacco leaves, they are not best as to 2.0, the limiting value achievable in terms of J-M distance.

As hyper-spectral data can provide a large amount of information and the reflectance spectrum involves various transformations such as solving higher derivative and logarithm besides continuum removal, the author thinks that there are many spectral features left to be extracted, which remain to be verified together with the divisibility of high-dimensional eigenvector expanding based on these features.

REFERENCES

Curran, P.J., 1989. Remote sensing of foliar chemistry [J]. Remote Sens. Environ., 30: 271-278.

Degang, Z., Z. Yulin, X. Liu, L. Cheng-Lin and Z. Da-Fang, 2010. Land cover classification in china based on chosen bands of MODIS [J]. Remote Sens. Land Resour., No. (3): 108-133.

Folin, L., Z. Chunjiang, L. Liang-Yun, J. Exudate and J. Zhiwei, 2007. Quantitation of maturity off reshflue-curedtobaccoleaf [J]. Tobacco Sci. Technol., 234(1).

Folin, L., Z. Chunjiang, W. Ji-Hua and L. Liang-Yun, 2008. Spectral characteristic of flue-cured virginia tobacco leaves of different maturity grades and spectra discriminant classification [J]. J. Fujian Agric. Forestry Univ., 37(6).

Gonzales, R.C. and R.E. Woods, 2003. Digital Image Processing [M]. Publishing House of Electronics Industry, Beijing.

JianKang, L., G. QunZhao and X. JianBo, 2010. Study on Suitable Maturity Harvest of Tobacco Variety "Honghuadajinyuan" in Liangshan Area [J]. Southwest China J. Agric. Sci., 23(3): 656-659.

Jinbao, J., L. Yifan, G. Hai-Qiang, L. Yi-Qing and C. Yun-Hao, 2010. Research on spectrum characteristics of bean and identification under different disease stress [J]. Spectroscopy Spect. Anal., 32(10): 2775-2779.

Jing, X., W. Jihua, X. Song, X. Xu, B. Chen and W. Huang, 2010. Continuum removal method for cotton verticillium wilt severity monitoring with hyperspectral data [J]. CSAE, 26(1): 193-198.

Keyan, N., Y. Yanbin, S. Chang-Qing, L. Li, Z. Zhao, *et al.*, 2010. Canopy spectral characteristics and hazard rating inversion of polyphagotarsonemus latus on hot pepper [J]. Southwest China J. Agric. Sci., 25(5): 1689-1693.

Kokaly, R.F. and R.N. Clark, 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression [J]. Remote Sens. Environ., 67: 267-287.

Lichtenthaler, H.K. and C. Buschmann, 2001. Chlorophylls and Carotenoids: Measurement and Characterization by UV-VIS Spectroscopy. Current Protocols in Food Analytical Chemistry [M]. John Wiley and Sons, New York.

Miller, J.R. and E.W. Hare, 1990. Quantitative characterization of the vegetation red edge reflectance 1: An inverted-Gaussian reflectance model [J]. Remote Sens., 2(10): 1755-1773.

Niemann, K.O., 1995. Remote sensing of forest stand age using airborne spectrometer data. P. RS, 61(9): 1119-1127.

Qingxi, T. and B. Zhang, 2006. Hyperspectral Remote Sensing-Principles: Technology and Application [M]. Beijing Higher Education Press, Beijing.

Swain, P.H. and S.M. Davis, 1978. Remote Sensing: The Quantitative Approach [M]. Mc-Grow Hill Inc., New York.

Wenjiang, H., 2009. Remote Sensing Monitoring Mechanism of Crops Disease and its Applications [M]. China Agricultural Science and Technology Press, Beijing.

Xiangyang, L., L. Guoshun, S. Zhou, Y. Xie-Feng and Z. Chun-Hua, 2007. Predicting leaf maturity of flue-cured tobacco using red edge characteristics of laboratory spectrometry [J]. J. Remote Sens., 11(2).