

ICA BASED MODEL FOR EXTRACTING INFORMATION OF CONTAMINATIVE SOIL FROM HYPERION DATA

Xiaodong Wang, Xiangnan Liu*, Lele Su, Nan Jiang

School of Information Engineering, China University of Geosciences, Beijing, China, 100083 - liuxncugb@163.com

KEY WORDS: Hyper spectral, Entropy, Image understanding, Feature extraction, Environment, Modelling

ABSTRACT:

This article discusses the influence of soil contamination on plants, particularly on the influence of vegetation spectral characteristics. Based on the analysis of healthy and contaminated vegetation spectral characteristics, a Difference-NDVI is presented by subtracts NDVI from REDNDVI (Red Edge NDVI) as a contamination factor. The result of Difference-NDVI may contain some unidentified substances which have similar spectral curves in red and near-infrared waveband to vegetable spectral. Therefore, the contaminative information extraction is basically a three-step process. First, subtle information hiding in high-dimensional data is projected into low-dimensional feature space using Fast-ICA. Then vegetable information is extracted based on the independent component generated by Fast-ICA. At last, in order to exclude the interference of unidentified substance, the result of D-NDVI and ICA-based vegetable information are intersected. According to the experimental results obtained from Hyperion data, it's confirmed that the result is rational as anticipative.

1. INTRODUCTION

As the deterioration of environmental contamination, soil contamination has threatened ecological environment security and Sustainable development. To prevent the increasingly severe environmental contamination, a real-time, accurate effective inspecting data is necessary.

Hyperspectral remote sensing is one of the most important breakthrough that human had made in the field of earth-observation of the 20th century. Contrast with traditional multi-spectral remote sensing, hyperspectral is of high spectral resolution and successive wave band. The imaging spectrometer supplies hundreds of spectral information of wave band that can generate an integrated spectral curve for each pixel. Based on those spectral curves, hyperspectral imagery technology could distinguish the surface substance being of diagnostic spectral feature that traditional remote sensing could not. Because of the trait mentioned above, hyperspectral remote sensing technology has overwhelming dominance in the application of contamination monitoring, contamination feature extraction, and becoming one of the most important monitor means.

Unfortunately, soil contamination information is subtle and affected by many factors like atmosphere, underground water, pesticides and so on. And all those factors are interfered mutually all the time which make soil contamination invisible and very difficult to be distinguished directly from remote sensing imagery. But when soil contaminated, the internal structure, water content, chlorophyll content of vegetable grow on it will changed, so as to the spectral Reflection Characteristic. Several researchers (Jago et al., 1995; Buschmann et al., 1998) have theoretically investigated the relationship between contamination characteristic and hyperspectral spectral Reflection effect of vegetable that is contaminated. The ways that researchers have put forward mainly included in two sorts. One is based on the Inversion of vegetable Biochemical Parameters. Like multivariate regression

modelling which take the spectral data or its transformation form(spectral reflection' First-order Differential, second order Differential) as Independent Variable, take vegetable Biochemical Parameters (LAI, chlorophyll content) as Dependent Variable, build multivariate regression estimate modelling. Matson (1994) found that the first derivative spectra at a centre waveband of 1525-1564nm could describe the change of nitrogen content of canopy. Peterson (1988) pointed that infrared region absorption feature strongly affected by biochemical characteristics by study the relationship between AIS data and field measured chemical data. Spectral position Analysis is another inversion-based method which estimates the Biophysical and Biochemical Parameters according to special spectral position (red edge). The previous work on red edge has indicated that red edge position' change is closely related to content of chlorophyll of vegetable. While chlorophyll content reduce, red edge position will shift to shortwave which called "blue-shift" (Rock et al., 1993; Horler et al., 1983). The other is based on spectral spatial transformation which projects the hyperspectral data from spectral space into feature space that find the inherent interrelation of hyperspectral data and the hiding subtle information invisible in spectral space. Although a number of papers have been published in the general area of inverse model, little work has been carried out for the spectral spatial transformation. The main reason for the scarcity of research in this area mainly is that traditional Methods of spatial transformation (PCA, MNF) are designed to measure data second-order statistics, while contaminative information is subtle substance which is not sufficient to constitute reliable statistics and may not be captured by the second order statistics-based PCA or MNF in its components (Wang et al., 2005). So this paper proposes an Independent Component Analysis (ICA)-based model for soil contaminative information extraction which measure data statistical independency that exceeds second-order statistics (Wang et al., 2005). Therefore, ICA could detect the weak information hiding in hyperspectral imagery that traditional approach could not. Unfortunately, when project the information hiding in high-dimensional data

* Corresponding author.

into low-dimensional feature space, it is still difficult to extract contaminative information directly from Independent Component (IC) of ICA. But As mentioned above, if vegetable was contaminated, red edge position will shift to shortwave, combining with the characteristics of vegetable spectral, a contaminative factor D-NDVI is put forward which subtracts NDVI from RENDVI (red edge NDI).When vegetable grow healthily, the value of D-NDVI is big. When vegetable was contaminated, the value of D-NDVI will be smaller. Therefore, combining the D-NDVI and ICA implement end-member extraction, could detect contaminative information directly. A real Hyperion image experiment is performed to substantiate the utility of this model.

2. INDEPENDENT COMPONENT ANALYSIS (ICA)

ICA is a multivariate data analysis method that recovers the original independent component from the linear mixture of statistical independent sources. Stemming from Blind Source Separation (BSS), ICA has applied widely in several fields like speech analysis, medicine, and face detection. ICA based remote sensing data process has made great progress recently. Du (2005) used ICA to the application of hyperspectral imagery plant classification. Wang J applied ICA to the application of dimensionality reduction for Hyperspectral imagery (Wang et al., 2005).

ICA based on high-order statistics that detect connotative factor from observed signature. Suppose that there are N independent sources of vibration $S_i(t)$, $i = 1 \sim N$, A is a unknown $M \times N$ mixing matrix, and the observed signal $x_i(t)$, $i = 1 \sim M$. ICA view the mixing matrix X as the result of mixing A and source signal S(t).

$$X(t)=AS(t) \tag{1}$$

Therefore, the purpose of ICA is to calculate an unmixing matrix W that make the t-dimensional vector $Y(t) =WX(t) =WAS(t)$ be the optimum approximation of the original sources S(t), which means $WA=I$. ICA take two steps to demix the observed signal. Firstly, decorrelate the second-order correlation (whiting) by PCA. Then, take revolving matrix to make output component independent. To insure that the components of ICA is mutual independent, mutual information

Influence factor	Pigment	Cell structure	Water content
Wave band	400-750	750-1300	1300-2400

is introduced which measure the discrepancy of the product of the marginal density and joint distribution, if the components mutual independent, then the product of marginal density will equal the joint distribution.

3. MODEL AND METHODS

3.1 ICA based-on vegetable information extraction

ICA technology considers hyperspectral image as a mix of independent ground object by an unknown matrix, the purpose of ICA is to find how those independent objects mixed and then separates them. FastICA algorithm which is developed by Hyvarinen and Oja (1997) was used to separate the independent

component one by one. FastICA is an iterative algorithm based on Fixed-point theoretic. FastICA is used to process hyperspectral images widely for the fast computational speed. To generate ICs, an initial projection vector is required to initialization. A common method is to generate a vector randomly as the initial projection vector. As a result, the ICs generate earlier does not necessarily more important than the one generate later. For hyperspectral image the number of independent ground objects is always smaller than the number of bands. Therefore, the whiting is necessary to reduce the dimension and determine the number of independent ground objects P. ICA assumes that the hyperspectral image is a data cube $M \times N \times L$ where L is the number of the band and M, N is the size of each band image. P ICs could be generated by FastICA one by one sequentially where each of the ICs is produced by maximizing the negentropy and the number of p is decided by the process of whiting and much smaller than L. Based on the generated ICs, pixel purity index (PPI) was implemented to extract the vegetable endmember and the Spectral Angle Mapper (SAM) was used for signature identification.

ICA-based feature extraction based on high order statistics that could detect subtle information of hyperspectral data that traditional method MNF and PCA who based on second order statistics could not. Wang (2005) has demonstrated that the ICA based endmember extraction show advantage over MNF and PCA. The result of SAM is the vegetable information which includes the healthy and contaminative vegetable. Actually it is difficult to identify the contaminative information directly. Therefore, a contaminative factor is presented base on the spectral feature of healthy vegetable and contaminative vegetable.

3.2 Contaminative factor

Contaminative information is weak information which is always invisible in remote sensing imagery. So it is difficult to detect directly based on the spectral of soil. Most of the time, even the soil was contaminated, the soil spectral curve usually do not transformed. This may due to the fact that contaminative information is relatively small and not sufficient to constitute reliable statistics. But soil contamination will affect the internal structure, water content, chlorophyll content of vegetable. While Green vegetable' spectral is significantly related to absorption and reflection of light of the mesophyll cells, chlorophyll, content of water and biochemical composition (Table 1). Therefore, if the soil was contaminated, the spectral of vegetable grow on it will transform. Study the transform of vegetable spectral could detect the contamination.

Table 1 factor that influence the reflectivity of vegetable leaf

3.2.1 Spectral characteristics of healthy vegetable: Spectral of healthy vegetable has low reflectivity in blue and red light band at the central wavelength of 450nm and 650nm because of the selective absorption of chlorophyll in leaf. In green band, spectral reflectivity has a little reflection peak for the relatively high reflectivity. The spectral reflectivity increased rapidly at the waveband range from 680nm to750nm which is generally referred as "red edge" in hyperspectral study. Red edge is the most distinctive characteristics of vegetable spectral curve and the best important band to index the healthy condition of vegetable. In near-infrared region (700nm-1300nm) reflectivity is relatively high since the structure of leaf cells. There are

three absorption vales because of the intense absorption of water.

3.2.2 Spectral characteristics of contaminated vegetable: Vegetable reflectivity spectral is related to the biochemical components. If vegetable was contaminated, its content of water, chlorophyll and internal structure will change. The type and content of contamination have great influence on leaf pigment than internal structure and water content. The chlorophyll of vegetable leaf has intense absorption to visible band. In the near-infrared band vegetable has high reflectivity, spectral transmittance and extreme low absorptivity. Therefore, waveband range from 400nm to 900nm is very sensitive to the vegetable growth and vegetable difference. This region is very important for the vegetable contaminative information detection. Spectral difference between healthy vegetable and contaminated vegetable is relatively more obviously in this region. Although different contaminative factors may have different influence on vegetable, they all induce the change of content of leaf pigment. This may be due to the fact that the contaminative factors inhibit the enzyme activity and hinder the chlorophyll biosynthesis (Jiang et al., 1994). The change of pigment content induce the change of vegetable spectral characteristics that reflectivity increased in visible light and near-infrared region, red vale become vague and the green peak shift to long wave, The reflectivity reduce greatly in near-infrared region, the red edge shift to the shortwave (blue shift) (Hoque et al., 1992).

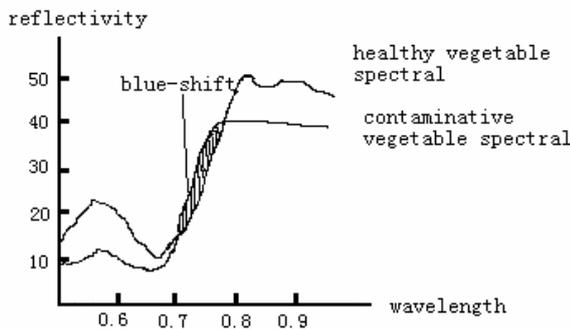


Figure 1. Healthy and contaminative vegetable spectral curve

3.3 Extracting contaminative factor

NDVI (Normalized Difference Vegetable Index) (Rouse et al., 1973) is the most widely used vegetable index which is expressed below as:

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \quad (2)$$

Where RRED is the reflectance of red light at a wavelength of 670nm, and RNIR is the reflectance of infrared light at 800 nm wavelength.

NDVI is widely used for estimating LAI or monitor Vegetation growing condition (Gong et al., 1995). Additionally it is also frequently as the health or stress index for vegetable in remote sensing studies (Morimoto et al., 1991). Several researches have indicated that the value of NDVI is related to vegetable

parameters of vegetable biochemical characteristics, vegetable fraction and photosynthesis.

RENDVI (Red Edge Normalized Difference Vegetable Index) is an improvement of NDVI which replace the red light band and infrared light band with the peak and vale of red edge (Sims et al., 2002). Boochs (1990) has pointed out that the red edge which obtain based on the hyperspectral spectral curve could supply enough information to detect the subtle difference of vegetable biochemical parameters.

RENDVI is expressed as follow:

$$RENDVI = \frac{R_{REP} - R_{REV}}{R_{REP} + R_{REV}} \quad (3)$$

Where RREP is the reflectance of red edge peak and RREV is the reflectance of red edge vale.

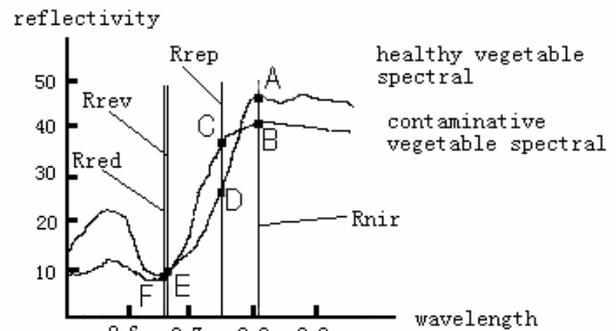


Figure 2. Spectral characteristics of healthy and contaminative vegetable

RRED is the position of red light at a wavelength of 670nm, RNIR is the position of infrared light at 800nm, RREP and RREV is the position of red edge peak and red edge vale which is calculated by the derivative operation (Miller, 1990). If the vegetable is healthy, the different value of NDVI (A-F/A+F) and RENDVI (D-E/D+E) would be bigger. Inversely, the different value of NDVI (B-F/B+F) and RENDVI(C-E/C+E) is smaller if the vegetable is contaminated (Figure 2).

Unfortunately, both NDVI and RENDVI could not detect contaminative information respectively. Nevertheless, observing the difference of healthy and contaminative vegetable spectral curve (Figure 3) and combining with the characteristics of NDVI and RENDVI, would find that the difference value of NDVI will be bigger if the vegetable is contaminated, while the difference value of RENDVI subtract from NDVI will be smaller if the vegetable is healthy. Based on this conclusion, this paper presents a contaminative factor Difference NDVI (D-NDVI) which is obtained by subtracting NDVI from RENDVI. The lower D-NDVI will be taking as the contaminative factor. One major issue for the contaminative factor is that the choice of threshold, it is difficult to select the threshold accurately without the support of field data. In the experiment next section, we plan to estimate the threshold by contrast the spectral curve. Another key issue arising in contaminative is that may have some substances not belong to vegetation but D-NDVI is small too. To address this issue, we decide to intersect contaminative factor and the vegetable distribution data obtained by the ICA.

3.4 ICA-Based Contaminative Information Extracting Model

- 1) Implement Fast-ICA to find p independent components, p is decided by the process of whitening.
- 2) Extract the vegetable information using PPI and SAM based on the result of Fast-ICA.
- 3) Perform NDVI and RENDVI, then calculate D-NDVI by subtracting NDVI from RENDVI and control the precision by choosing proper threshold.
- 4) To exclude the interference factor exist in D-NDVI, the image of D-NDVI which present the contaminative distribution obtained in step 3 was intersected with the image of vegetable information obtained in step 2.

In what follows we describe 2 major novelties derived from our proposed model.

- a) One of major novelties is the idea of contaminative factor that have not been explored before. The performance of this method is verified by the results of Hyperion data experiment in the following section.
- b) A second novelty is that ICA-based contaminative information extraction is a high order statistics-based method as opposed to commonly used MNF or PCA which are second-order statistics-based, MNF or PCA is difficult to capture contaminative information which is subtle substance and not sufficient to constitute reliable statistics.

4. EXPERIMENT AND RESULTS

In order to demonstrate the utility of ICA-based contaminative extraction model from Hyperion image, a Hyperion image data was used for experiments.

The Hyperion image scene was collected in 2004. It has size of 256×2905 pixels. This experiment select a representative 256×256 image (Figure 3). Within the scene there has several tidal flat on the left, a large of buildings on the right and a river running from right to left. The image was acquired by 242 spectral bands with spectral coverage from 355 to 2577nm and band 1-7, 58-78, 121-127, 167-178, 224-240 have been removed prior to the analysis due to water absorption and low SNR in those bands. As a result, a total of 176 bands were used for this experiment.



Figure 3. The Hyperion image scene

To implement ICA, the hyperspectral image cube $M \times N \times L$ should be transform to be a matrix S with L rows and $M \times N$

columns ($L \times MN$) according ICA theory. Actually each row of matrix S represents one particular band of hyperspectral image. After Fast-ICA performed, a matrix will be generated which include all independent component. To obtain those ICs, each row should be transform to a $M \times N$ matrix representing an independent component. Figure.4 shows the 21 ICs generated by Fast-ICA and each of the generated ICs represents a special object.

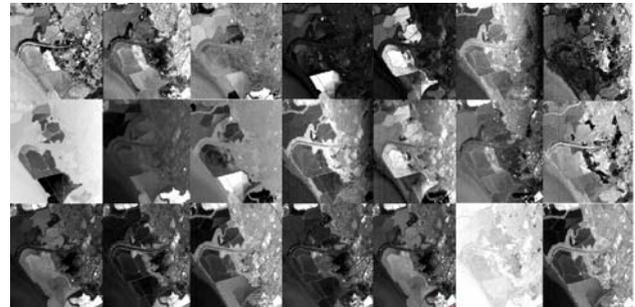


Figure 4. Twenty-one independent components produced by the Fast-ICA

Figure.5 show the distribution of vegetable performed by PPI and SAM based on the result of Fast-ICA. As mentioned above, it is still difficult to distinguish the contaminative vegetable from healthy vegetable base on the result of ICA-based vegetable information extraction.

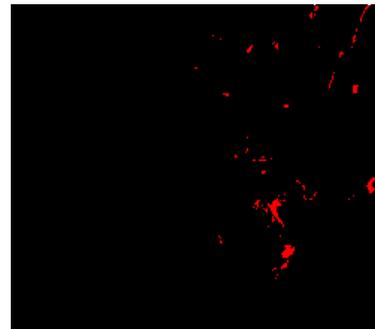


Figure 5. ICA-based vegetable information extraction

NDVI use the spectral difference of high reflectance of the near-infrared waveband and low reflectance of red light waveband of vegetable to estimate vegetable distribution or as the healthy index of vegetable. Figure.6 shows the NDVI and RENDVI of the hyperspectral image data which generally represent the distribution of vegetable.

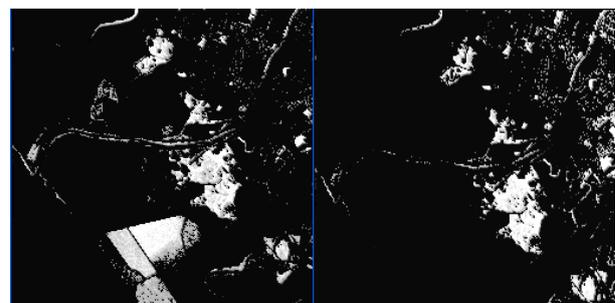


Figure 6. Comparison of NDVI and RENDVI represented vegetable distribution

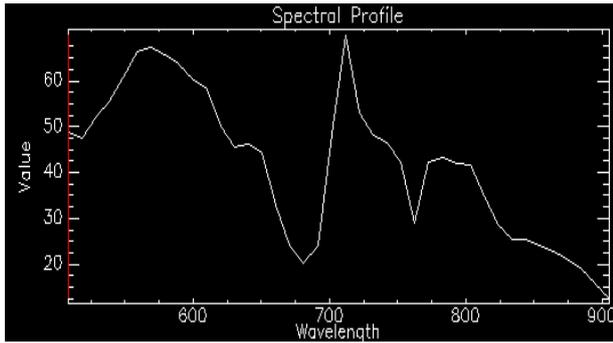


Figure 7. Spectral curve of unidentified substance having the similar spectral with vegetable in near-infrared and red light waveband

Interestingly, the value of NDVI and RENDVI of healthy vegetable is different from the value of contaminative vegetable. Therefore, the contaminated vegetable could be detected by the D-NDVI. One major issue for D-NDVI is that there are some substances showing in the result of D-NDVI do not belong to vegetation labeled “A” and “B” (Figure 9). This is due to the fact that the spectral curves of unidentified substance have the similar spectral with vegetable in near-infrared and red light waveband (Figure 7). In order to address this issue, this experiment combines the result of ICA-based vegetable information extraction that present the vegetable distribution (Figure 5) and the result of RENDVI-NDVI (Figure 8) that presents the contaminative information distribution to exclude the interference of non-vegetation.

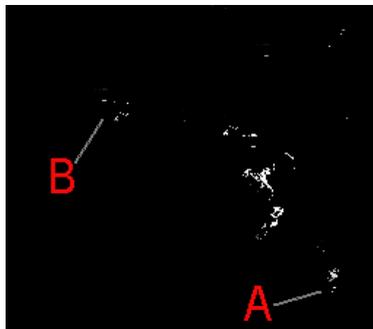


Figure 8. D-NDVI Where A and B present the unidentified substance having similar spectral curve with vegetable

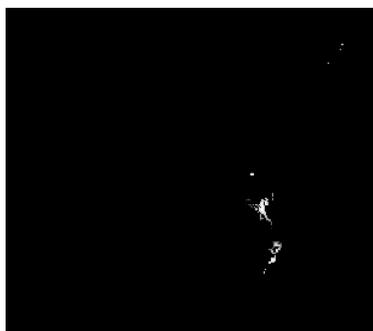


Figure 9. The result of contaminative information distribution

5. CONCLUSIONS

Vegetable spectral characteristics-based soil contaminative information extraction has found important application in the recent years. Several remarkable works has also been reported (Jago et al., 1995; Buschmann et al., 1998). But those research mainly base on the support of field data to detect contaminative information. They have the disadvantage of greater complexity and higher cost. Actually, we need to estimate contaminative information without field data. Interestingly, if vegetable was contaminated, the spectral curve will transform, the red edge will shift to blue light waveband. By comparing the spectral difference of healthy vegetable and contaminated vegetable and analyzing the vegetable contaminative mechanism, this paper proposed a new approach that introduce contaminative factor to estimate contamination distribution. Nevertheless, there are some substance having similar spectral characteristics in object waveband (red light and near-infrared waveband) interfere contaminative information extraction. To exclude that interference, this paper process the image data using Fast-ICA, then extract vegetable distribution information based on PPI and SAM technology. Finally, extract contaminative information by intersecting the result of D-NDVI with the result of ICA based-vegetable information extraction. A real Hyperion image experiment is conducted to validate the utility in real applications. However, one limitation of the results developed in this paper is the choice of threshold in step3. For the scarcity of field data, it is difficult to calculate the threshold exactly. In this experiment threshold is estimated by the spectral difference. Further research will be devoted to removing this limitation.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their thoughtful comments. This work is under the auspice of National High-tech R&D Program of China (863 program) (2007AA12Z174) and National Natural Science Foundation of China (40771155).

REFERENCES

- Boochs, F., Kupfer, G., 1990. Shape of the red edge as vitality indicator for plants [J]. *Int J Remote Sensing*, 11(10):1741-1753.
- Buschmann, C., Lichtenthaler, H.K., 1998. Principles and characterization of multi-colour fluorescence imaging of plants [J]. *Journal of Plant Physiology*, 152:297-314.
- Du, P, Zhao, HJ, Zhang, B, Zheng, LF, 2005. Independent component analysis for hyperspectral imagery plant classification. *Applications of neural networks and machine learning in image processing*, ix 5673: 71-81.
- Gong, P, Pu, RL, Miller, J.R., 1995. Coniferous forest leaf-area index estimation along the Oregon transects using compact airborne spectrographic imager data. *Photogramm Eng Remote Sens* 61 (9):1107-1117.
- Hoque, E., Huntzler, J.S., 1992. Spectral blue shift of red edge monitors damage class of beech trees [J]. *Remote sensing of environment*, (39):81-84.
- Horler, D. H. N, Barber J., 1983. The Red Edge of Plant Leaf Reflectance [M]. *Int J Remote Sens*, 4: 273-288.

Hyvarinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483 - 1492.

Jago, A.R., Cutler, M.E.J., Curran, P. J., 1999. Estimating canopy chlorophyll concentration from field and airborne spectra[J]. *Remote Sensing of Environment*, (68): 217-224.

Jiang, MY, Yang, WY, 1994. Damage of active oxygen degradation of chlorophyll of rice seedling under Osmotic stress [J]. *vegetable journal*, 36 (4):289-295.

Matson, P. A., et al., 1994. Seasonal changes in canopy chemistry across the Oregon transect: patterns and spectral measurement with remote sensing, *Ecol. Appl*, 4(2): 280~298

Miller, J.R., Hareew, W.U.J., 1990. Quantitative characterization of the vegetation red edge reflectance model [J]. *Int J Remote Sensing*, 11(10):1755.

Morimoto, Y., Takei, J., Kominami, H., Hanayama, H., Miwa, K., 1991. The vigor assessment of street trees, *Zelkova serrata*. *J Jpn Soc Reveg Technol* 17(1):9-15.

Peterson, D.L., et al., 1988. Remote sensing of forest canopy and leaf biochemical contents, *Remote Sens. Environ*, 24: 85~108

Rock, B.N., Hoshizaki, T., Miller, J.R., 1988. Comparison of in Situ and Airborne Spectral Measurements of the Blue Shift Associated with Forest Decline [J]. *Remote Sensing of Environment*, 24:109 21-27.

Sims, DA., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages *Remote Sensing of Environment*, 81: 337-354.

Wang, J, Chang, C-I, 2006. Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis. *IEEE transactions on geoscience and remote sensing*, vol. 44.

Wang, J, Chang, C-I, 2005. Dimensionality reduction by independent component analysis for hyperspectral image analysis, presented at the *IEEE Int. Geosci. Remote Sens. Symp.*, Seoul, Korea.