

QUANTITATIVE ESTIMATION OF DESERTIFICATION DEGREE BASED ON PCA FUSION AND SVM USING CBERS-02B

HE Qisheng^a, Li Guoping^b, CAO Chunxiang^{a*}, ZHANG Hao^a, BAO Yunfei^a, CHANG Chaoyi^a, LI Xiaowen^c

^a State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China(hqsfei999@163.com, cao413@irsa.ac.cn, zhangh@irsa.ac.cn, byf_sheep@163.com, qiidee@vip.sina.com)

^b School of Earth and Space Sciences, Peking University, Beijing 100871, China - lgp@cnsa.gov.cn

^c State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and the Institute of Remote Sensing Applications of Chinese Academy of Sciences, Beijing 100875, China. lix@bnu.edu.cn

KEY WORDS: support vector machine (SVM); data fusion CBERS-02B soil desertification information, PCA

ABSTRACT:

Soil Desertification has severely threatened inner stability and sustainable development in arid areas, so extracting soil desertification information based on remote sensing and mastering its spatial distribution are of important practical significance. In this article, taking the Mu Us sand land as a case study, utilizing CBERS-02B including CCD data and HR data, the approach of effective remote-sensing information extraction for soil desertification in arid areas based on principal component fusion and SVM is discussed. Firstly, two dimensional discrete stationary wavelet transform combined average filter processing is presented to improve the image quality. Then principal component fusion and SVM classification were used to extract soil desertification information. Finally, the kernel type of radial basis function was selected and comparing the classification results in SVM classification with maximum likelihood classification and minimum distance classification qualitatively and quantitatively in terms of classification accuracy. The results suggest that the method is effective and the precision of this approach is very high, so it is an effective method for monitoring soil desertification changes utilizing CBERS-02B data in arid area.

1. INTRODUCTION

Desertification is the label for land degradation in arid, semi-arid, and dry sub-humid areas, collectively called dry-lands, and represents one of the most threatening environmental hazards due to the large amount of people and land at risk. Soil desertification has severely threatened inner stability and sustainable development in arid areas. Sandy desertification is one of the main forms of land degradation in China, especially in northern China (Wang T and Zhu Z D), which has been expanding since the 1950's and has exerted severe impact on regional socio-economic development and environmental security (Wang T, *et al*). Harsh physiographic conditions (sparse vegetation coverage, sandy soil and water deficiency), irrational land-use practices, and population augmentation are regarded as the forces triggering sandy desertification (Zhu Z D, Chen Y F). Therefore, sandy desertification assessment and monitoring are always concerning to researchers, the public, and policy-makers. Remote sensing data and techniques have been widely used to map soil desertification areas (Liu H J, *et al*, Wang X Q, *et al*, Guo J, *et al*, Liu W J, *et al*).

economic and social complex ecosystem. For a long time, due to lack of ecological protection awareness and blind pursuit of short profit, a vicious cycle of "population growth - lower standard of living - over-development - ecological degradation - low productivity - over-development" damaged the ecological environment. The location of the study area is shown in Fig.1.

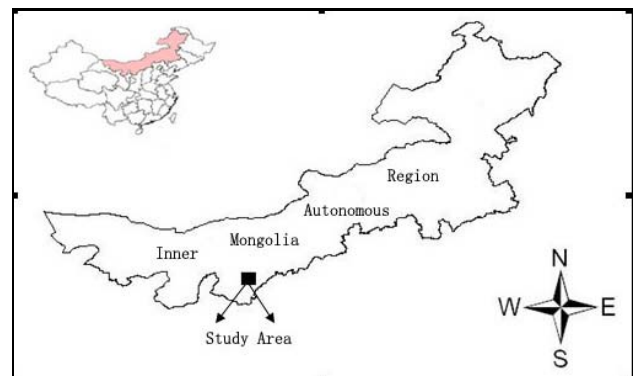


Figure.1 The location of study area

2. AREA DESCRIPTION AND DATA SET

2.1 Study area

Mu Us sand land is one of twelve large desert areas in the southeast of China's desert areas, at latitude 37°27.5' ~ 39°22.5'N, longitude 107°20' ~ 111°30'E. The climate of this area is located in arid, semi-arid transition zone, most of the area is in temperate semi-arid area, and has a fragile natural,

2.2 Data set

The CBERS-02B satellite was set up by China on September 19, 2007, with the highest resolution HR image by 2.36m and the middle resolution CCD image. The CBERS-02B HR data is taken almost simultaneously with multi-spectral CCD data

* Corresponding author.

using the different sensor system. The HR camera consists of three pieces of TDI CCD and the imaging modality is different from the traditional remote sensing optical cameras. The middle piece of image is first imaging, then the left and the right are time lagged imaging.

In this paper, CBERS02B CCD and HR data were selected to study the soil desertification extraction method. The data provided by the China Center for Resource Satellite Data and Applications were acquired 17October 2007, with the Path / Row: 9/56. The data charter of CBERS02B satellite is shown in Tab.1.

In this study, a representative region of desertification soils was selected. The coordinate of upper left was 39.5N and 108.48E and the coordinate of lower right was 39.37N and 108.7E in the ultimate selected image.

Payload	Band	Spectral range (μm)	Spatial resolution (m)
CCD Camera	B1	0.45~0.52	19.5
	B2	0.52~0.59	19.5
	B3	0.63~0.69	19.5
	B4	0.77~0.89	19.5
	B5	0.51~0.73	19.5
HR Camera	B6	0.5~0.8	2.36

Table.1 Data charter of CBERS02B satellite

3. METHOD

3.1 Pre-processing

From the system geometric correction products provided by China Center for Resources Satellite Data & Application, the imaging time inconsistency causes the marked band strips and the color aberration between different strips, which severely affects the quality of HR image. So firstly the wavelet transform combined filter processing is presented to remove strips and color aberration. In this paper, two dimensional discrete stationary wavelet transform combined average filter processing is presented to improve the image quality. First, average filter was used to smooth the image. Then a two dimensional discrete stationary wavelet transform method is introduced. The wavelet function of sym4 is selected at three levels, with the index of fixed form soft threshold and un-scaled white noise. Finally the geometric correction was made to HR and CCD data.

3.2 Fusion method

3.2.1 PCA fusion:

The principal component transform was made to the CCD data. The main purpose of principal component transform is that the useful information of original multi-band images is concentrated on the minimal number of the principal components of the new images and the principal components are uncorrelated. In other words, the information contained in all the principal components does not overlap with each other; thus the total volume of data is greatly reduced, redundant

information is eliminated, and effective information is highlighted. Then the physical meaning of each principal component transform was analyzed in extracting soil desertification information.

The CBERS-02B CCD1-4 bands were used to make principal component transform in this paper. The result showed that in the first principal component the band coefficients are proportional to the standard deviation, so the first principal component reflects the integrated information. In research area, the mainly distributed types are desertification soils in different levels including light desertification soils, middle desertification soils, and severe desertification soils. Different levels of desertification soils have different values. In the second principal component, the water body is emphasized with the values larger than others types.

PC	CBERS-02B CCD			
	B1	B2	B3	B4
PC1	0.3034	0.3370	0.6691	0.5888
PC2	0.4351	0.3007	0.3368	-0.7790
PC3	0.8264	-0.0790	-0.5174	0.2074
PC4	0.1888	-0.8887	0.4137	-0.0587

Table.2 Principal component eigenvector matrix of CBERS-02B CCD image in research area

PC	Eigenvalues	Percentage	Cumulative
PC1	282.19	85.12	85.12
PC2	38.68	11.67	96.79
PC3	8.03	2.42	99.21
PC4	2.62	0.79	100.00

Table.3 Principal component eigenvalue of CBERS-02B CCD image in research area

Then the HR data was contrast stretched to have approximately the same variance and average as the first principal component image. The stretched HR data was substituted for the first principal component image and the data was transformed back into original space. The HR data may be substituted for the first principal component image because the first principal component image normally contains all the information that is common to all the bands input to PCA, while spectral information unique to any of the input bands is mapped to the other principal components.

3.2.2 HIS and Brovey fusion:

The fusion methods of HIS and Brovey were used to fuse CBERS-02B CCD data and HR data to enhance the soil desertification information. As to the HIS transformation, the CCD data of B2,3 and 4 in RGB color space were transformed into three bands in HIS color space. Then the HR image was contrast stretched to have approximately the same variance and mean as the intensity (I) image. Then the stretched, high-spatial-resolution image is substituted for the intensity (I) image. Then the modified HIS dataset is transformed back into RGB color space using an inverse IHS transformation. The justification for replacing the intensity (I) component with the stretched HR image is that the two images have approximately

the same spectral characteristics. The Brovey fusion method is made by the normalized CCD image and HR image.

3.2.3 Comparison of fusion effect: Quantity estimation of fusion effect of different fusion method was undertaken. The results showed that PCA fusion has the best effect. Tab.4 shows that PCA fusion had the largest value of entropy and contained the most information at each band. The fusion image is shown in Fig.2.

Method		mean	stdev	entropy
HR image		92.5913	10.9536	5.4466
CBERS02B CCD4,3,2	R	77.1471	11.0298	5.3373
	G	77.2407	11.5468	5.5357
	B	51.6949	6.1329	4.6010
PCA	R	77.1453	10.6250	5.3733
	G	77.2395	11.7665	5.5697
	B	51.6945	6.2353	4.6325
HIS	R	49.1426	8.3631	4.0113
	G	50.2367	8.9634	4.2457
	B	33.6473	5.3267	3.8356
Brovey	R	34.1748	4.6414	4.1206
	G	34.2344	4.8211	4.2858
	B	22.7031	2.6507	3.4056

Table.4 quantity estimation of fusion effect of different fusion method

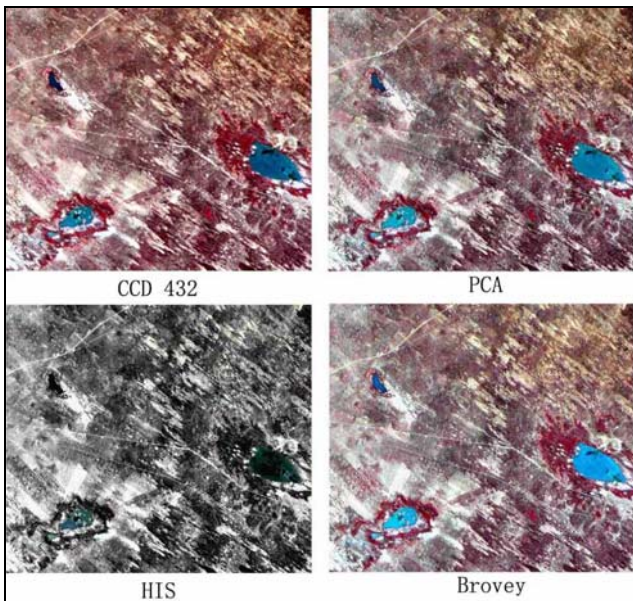


Figure.2 The results of different fusion methods

3.3 Classification

3.3.1 Classification method:

Support vector machines are a general class of learning architecture inspired from statistical learning theory that perform structural risk minimization on a nested set structure of separating Hyper-planes. Kernel based learning methods in general and the Support Vector Machines (SVM) in particular, have been introduced in the last years in learning theory for classification and regression tasks. SVM have been successfully applied to text categorization (Joachims, 1998) and face

recognition (Osuna et al.,1997). Recently, they have been successfully used for the classification of hyper-spectral remote sensing images (Bruzzone and Melgani, 2002) and automatic recognition of man-made objects in high resolution optical remote sensing images (Inglada, 2007).

SVM is an effective machine learning method based on structural risk minimization (SRM). The main idea is that by ensuring the minimization of empirical risk, the top margin of the confidence risk is minimized. SVM implements the following idea: it maps the input vector into a high-dimensional feature space and constructs the optimal separating hyper-plane in that space. The SVM learns a separating hyper-plane in order to maximize the boundary margin and to produce good generalization ability, which can avoid over-fitting (Lu B, Bengio Y, Keerthi S S, Zheng C H, Xu P).

The radial basis function (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

is selected as a SVM kernel function. When training the SVM classifier, two important parameters -- the penalty value C and kernel parameter γ -- should be predetermined using the training data. The two parameters are main factors affecting the classification accuracy of the SVM (Yang H C, et al).

3.3.2 Post classification processing: Since the information of remote sensing images is automatically extracted by computer for the individual pixel, the phenomenon that a large same kind distributes some scattered points with other kinds will emerge in the result images. These scattered kinds are often called "category noise". We used majority analysis to change spurious pixels within a large single class to that class. We entered a kernel size and the central pixel in the kernel was replaced with the class value that the majority of the pixels in the kernel have. In this article, the kernel size of 3x3 was selected to eliminate the impact of category noise and the results are shown in Fig.3.

3.3.3 Comparison analysis: In this paper the PCA fusion image was used to extract soil desertification information. The kernel type of radial basis function was selected and comparing the classification results of SVM classification with maximum likelihood classification and minimum distance classification qualitatively and quantitatively in terms of classification accuracy. The study results indicate that: image classification based on SVM can monitor the soil desertification more accurately than those of maximum likelihood classification and minimum distance classification, classification overall accuracy reaches 96.66%, while the classification overall accuracy of maximum likelihood classification and minimum distance classification reaches 90.54% and 86.43% respectively. The classification method based on SVM can be adapted to the fusion image classification and monitoring of soil desertification.

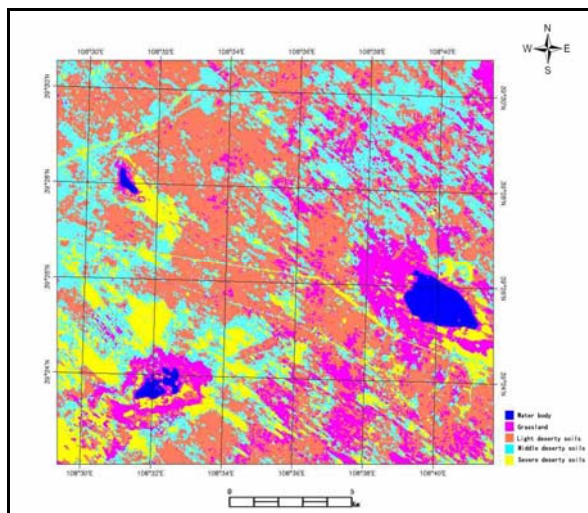


Figure. 3 The result of extraction of desertification soil information

4. CONCLUSION

In this article, the soil desertification information was successfully extracted by using PCA fusion and SVM method. Accuracy is high, so it is noted that this method in the extraction of soil desertification information has advantages.

As to CBERS02B data, the method of two dimensional discrete stationary wavelet transform combined average filter processing is valid to improve the image quality. It can remove some noise and color aberration.

The results also show that PCA fusion has the best effect in contrast with IHS and Brovey fusion method and the support vector machine (SVM) has the best effect in contrast with maximum likelihood classification and minimum distance classification. Although the method is used for CBERS-02B HR and CCD data, the method is also applicable in other data with lower and higher spatial resolution image.

ACKNOWLEDGEMENTS

This research has been supported by a grant from the Major State Basic Research Development Program of China (973 Program) (NO. 2007CB714400). We would like to acknowledge the support of the China Center for Resource Satellite Data and Applications in providing CBERS-02B CCD and HR data.

REFERENCES

Bengio Y. 2000. Gradient-based optimization of hyperparameters. *Neural Computation*, 12(8),pp.1889-1900.

Bruzzone, L., Melgani, F., 2002. Support vector machines for classification of hyperspectral remote-sensing images. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, vol. 1, pp. 506-508.

Chen Y F, Tang H P. 2005.Desertification in north China: background, anthropogenic impacts and failures in combating it. *Land Degradation and Development*, 16,pp.367-376.

Guo J, Wang T, Xue X, *et al.*2006. The Status and Distribution of Desertification in Maowusu Sandy Land. *Research of Soil and Water Conservation*,13(3),pp.198-199.

Inglada J.2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS Journal of Photogrammetry & Remote Sensing*,62,pp.236-248.

Joachims,T., 1998. Text categorization with support vector machines: learning with many relevant features. In: Nédellec, Claire,Rouveirol, Céline (Eds.), *Proceedings of the European Conference on Machine Learning*. Springer, pp. 137-142.

Keerthi S S. 2002. Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms . *IEEE Trans on Neural Networks*, 13(5):1225-1229.

Liu H J, Zhou C H, Cheng W M, *et al.*2008. Monitoring sandy desertification of Otindag Sandy Land based on multi-date remote sensing images. *Acta Ecologica Sinica*, 28(2), pp.627-635.

Liu W J, Chang Q R, Jia K L *et al.*2007. Study on land cover spectral character and remote sensing information extraction of desertification land. *Journal of Northwest SCI-Tech Univ of AGR and*, 35(9),pp.205-209.

Lu B, Wei X K, Bi D Y. 2005. Applications of support vector machines in classification. *Journal of Image & Graphics*, 10(8),pp.1029-1035.

Osuna, E., Freund, R., Girosi, F., 1997. Training support vector machines: an application to face detection. *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pp. 130-136.

Wang T, Zhu Z D. 2001. Studies on the sandy desertification in China.*Chinese Journal of Eco-agriculture*, 9(2),pp.7-12.

Wang T, Chen G T, Zhao H L, *et al.* 2006.Research progress on aeolian desertification process and controlling in north of China. *Journal of Desert Research*, 26(4),pp.507-516.

Wang X Q, Yang L A, Jiang Y, *et al.*2007. Dynamic Changes of Wetlands and Sandification in the South of Musu Sandy Land. *Journal of Northwest Forestry University*,22(4),pp. 73-77.

Xu P, Chan A K. 2003. An efficient algorithm on multi-class support vector machine model selection. In:*Proceedings of the International Joint Conference on Neural Networks*. Portland: IEEE,pp.3229-3232.

Yang H C, Zhang S B, Deng K Z, *et al.*2007. Research into a Feature Selection Method for Hyperspectral Imagery Using PS O and SVM. *Journal of China University of Mining & Technology*, 17(4),pp.473-478.

Zheng C H, Jiao L C, Zheng G W. ,2005. Genetic algorithm-based SVM for automatic target classification of remote sensing images. *Control and Decision*, 20(11),pp.1212-1216.

Zhu Z D, Chen G T.1994. *Sandy desertification in China*. Beijing: Science Press, pp. 20-33.