LAND COVER CLASSIFICATION IN BEIJING AREA BASED ON CBERS-02B CCD DATA

Wenwen Wang [ab, *], Yonghong Zhang [a]

[a] Chinese Academy of Surveying and Mapping,No.16 Beitaiping Road,Haidian District,Beijing,P.R.China,wenwen.www@gmail.com

[b] Liaoning Technology University, No.47 Zhonghua Road, Fuxin,Liaoning

KEY WORDS: CBERS-02B, CCD, Land Cover Classification, DTC, Accuracy Assessment

ABSTRACT:

In this paper, we studied the potential of land cover classification using multi-spectral CBERS-02B CCD data. CEBERS-02B is the third of the series earth resource satellite of China. It was launched on Sept. 19th 2007. CCD is one of the payloads, and operates with a standard 5 bands setting through the visible to near infrared region of the electromagnetic spectrum. Compared with the main counterpart, Landsat TM/ETM+, the CBERS-02B CCD has the advantage of higher spatial resolution at 19.5m. This paper evaluated the ability of land cover classification with CBERS-02B CCD data through taking Beijing as the test area.

CCD level2 data were used for this research. Before the classification, the data preprocessing was implemented to improve the geometrical accuracy of CCD. Classification training samples were selected from reference to vegetation atlas and land use map of China, high resolution satellite image and field collecting points. After analyzing the characteristics of samples of each land cover type, this research choose the decision tree classifier to accomplish this assignment. The decision rules were built with the spectrum bands of CCD and the ancillary variables. Finally the author focused on accuracy assessment using confusion matrix and kappa coefficient. Results form the assessment showed class accuracies ranged between 55.18 to 100%, and the overall accuracy was 86.21%.

1. INTRODUCTION

Land cover classification is always a hotspot in remote sensing applications. Effective tools that support the inventory of land cover changes are of paramount importance to analyze environmental and natural impacts on the biosphere. Knowledge about land use and land cover has become increasingly important as the nation plans to overcome the problems of haphazard, uncontrolled development, deteriorating environmental quality, loss of prime agricultural lands, destruction of important wetlands, and loss of fish and wildlife habitat(James *et al*.,2001).

The needs of land cover data are very broad. CBERS-02B CCD with its relatively higher resolution (19.5m) can provide a suited datasets for land cover mapping. This research was to perform a preliminary work that estimated the feasibility of using CBERS-02B CCD for land cover classification from the classification accuracy. Firstly, preprocessing of CCD data was necessary. It included the band-registration and geometric correction. Because the band-to-band misregistration would decrease the definition and classification accuracy of CCD imagery, and the geometric correction was done to improve the consistency of multi-temporal data of CCD and other ancillary data layers. Secondly, the sample selection was a crucial part. It determined the analysis of sample characters and classification result. This research referenced a lot of existing datum such as vegetation atlas and land use map of China and some field collecting points to help the sample selection. Thirdly, after analyzing the samples, the author built the rules for the decision tree classifier. The variables of the decision rules included multi-temporal spectrum bands of CCD, the NDVI index and DEM layer. DEM as an ancillary data played an important role in separating the classes with confusions on spectrum characters and NDVI indexes. Finally, after the classification, the confusion matrix and kappa coefficient were applied in validation part. The reference data was collected equally cover the whole region. The classification accuracy gave an impersonal estimation on the potential of CBERS-02B CCD data for land cover mapping.

2. MATERIAL AND METHOD

2.1 CCD Data And Preprocessing

Full resolution CCD level 2 data of Beijing was obtained. Specifications of the spectral bands are given in Table1. The fifth band is panchromatic, so only the first 4 bands were used. This paper selected CCD data of two acquisitioning dates corresponding to peak and late vegetation green-up. The scene data were August 22nd (T1) and December 30th (T2) in 2008. The strategy developed for imagery selection was designed to meet the requirements of multi-temporal data, which were collected to assist the separability between classes.

The level 2 data is a product after radiometric calibration and systematically geometric correction. But it has the problems of band-to-band misregistration and large geolocation error. For the first 4 bands, the registration error is about 1 pixel between band 1 and others. For the classification was toward to single pixels, the misregistration would decrease the classification accuracy. The large geolocation error also blocked the application of CCD data. To deal with these problems, the third

* Corresponding author. Tel. +86-10-88217728

band (red band) was selected as the reference, and then band 1 was co-registered to it. Finally, all co-registered bands were geometrically rectified with 2-order polynomial model to a Landsat TM ortho-rectified image. The preprocessing improved the consistence of the multi-temporal data and the ancillary datasets.

band number	spectral range /nm
1	0.45-0.52
2	0.52-0.59
3	0.63-0.69
4	0.77-0.89
5	0.51-0.73

Table 1. CBERS-02B CCD band setting

2.2 Ancillary Data

Except for spectrum information of satellite data, successful land cover mapping typically needs ancillary data for improvement. The type of ancillary data available and the method used to classify them both play a large role in the successful of the classification. SRTM DEM was the primary ancillary data layer. It was used to separate the classes with confusions on spectrum characters and NDVI indexes. Additional ancillary data, such as vegetation atlas and land use map of China, high resolution satellite image and field collecting points were applied in training sample selection and classification accuracy assessment.

2.3 Legend

The classification legend used in this paper agreed the Land Cover Classification System (LCCS) (Collin *et al*., 2004) and the situation of Beijing area. It included 8 classes: Evergreen Needleleaf Forest, Deciduous Broadleaf Forest, Shrub, Crop, Grass, Urban, Water and Shadow.

2.4 Method

Decision tree classification (DTC) is the chosen method. Decision tree classifier is one of the possible approaches to multistage decision making. The basic idea involved in any multistage approach is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained this way would resemble the intended desired solution (Rasoul *et al*., 1991). Advantages it offers include: (1) it is non-parametric and therefore independent of the distribution of class signature, (2) it can handle both continuous and nominal data, (3) it generates interpretable classification rules, and (4) it is fast to train and often as accurate as, or even slightly more accurate than many other classifiers (Homer *et al*., 2004).

DTC is a supervised classification method that relies on large amounts of training data. The training data namely sample selection is a crucial part of paper which directly influenced the result. The samples were initially collected to improve classification efficiency through referencing a variety of sources including high-resolution QuickBird imagery, existing land use maps and field collected points. Then statistics of pure samples were computed. After analyzing the characteristics of samples of different land covers, DTC rules were founded using multi-temporal CCD datasets combined with NDVI and ancillary variables.

2.5 Classification Features and DTC Rules

The classification features are variables to constitute decision rules. Appropriate and efficient features play a crucial part in decision tree classification. Features chosen in this paper included the 1-4 spectrum bands of CCD of two acquisition dates, NDVI and DEM.

The Normalized Difference Vegetation Index (NDVI) is a simple numerical indicator that can be used to analyze remote sensing measurements. NDVI is calculated from the visible and near-infrared light reflected by vegetation. Healthy vegetation absorbs most of the visible light that hits it, and reflects a large portion of the near-infrared light. Written mathematically, the formula is:

$$NDVI = (NIR - VIS)/(NIR + VIS) \quad (1)$$

Calculations of NDVI for a given pixel always result in a number that ranges from minus one (-1) to plus one (+1); Negative values of NDVI correspond to water. Values close to zero correspond to barren areas of rock, sand, or snow. Moderate values represent shrub and grassland, while high values indicate temperate and tropical rainforests.

Legend	DTC Rules
Evergreen Needleleaf Forest	$60<B_4(T1)<125$, $DEM\le240$, $0.27<NDVI(T1)<0.4$
Deciduous Broadleaf Forest	$B_4(T1)>60$, $DEM\le240$, $NDVI(T1)\ge0.4$, $B_4(T2)<45$
	$B_4(T1)>115$, $DEM>240$, $NDVI(T1)\ge0.5$
Shrub	$B_4(T1)>115$, $DEM>240$, $0.27< NDVI(T1)<0.5$
Crop	$B_4(T1)>60$, $DEM\le240$, $NDVI(T1)\ge0.4$, $B_4(T2)\ge45$
Grass	$B_4(T1)>60$, $DEM\le240$, $0.27<NDVI(T1)<0.4$, $B_4(T1)\ge125$
Urban	$B_4(T1)>60$, $DEM\le240$, $NDVI(T1)\le0.27$, $B_3(T1)\ge95$
	$B_4(T1)>60$, $DEM\le240$, $NDVI(T1)\le0.27$, $65\le B_3(T1)<95$
	$B_4(T1)>60$, $DEM\le240$, $NDVI(T1)\le0.27$, $B_3(T1)<65$
Water	$B_4(T1)\le60$, $B_3(T1)\le59$
Shadow	Building Shadow: $B_4(T1)\le60$, $B_3(T1)>59$
	Mountain Shadow: $B_4(T1)>60$, $DEM>240$, $B_4(T1)\le115$

Table 2. Decision Rules of Decision Tree Classifier
(B_i represent the DN value of band i.)

Due to the slightly difference between the classes in the first 3 bands of CCD, NDVI was a variable of consequence. It was selected to extract the vegetation and separate them assisted by band 4 and diverse representations of vegetation in the two acquisition dates.

DEM data layer was a main variable of decision rules. The elevation and temperature caused some spectrum discrepancy on the same class and increase the confusions on different classes. For the CCD data, there were some overlaps of spectrum band and NDVI values between crop, grass and deciduous broadleaf forest. Therefore, DEM information was used in this paper to separate the forest on the mountain firstly, which improved the classification accuracy of these classes. The DTC rules are described in detail in table 2.

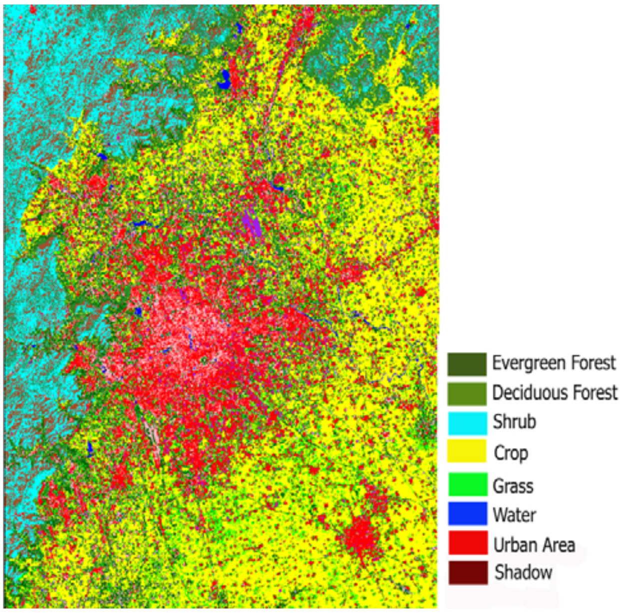

Figure 1. Land cover classification results of Beijing

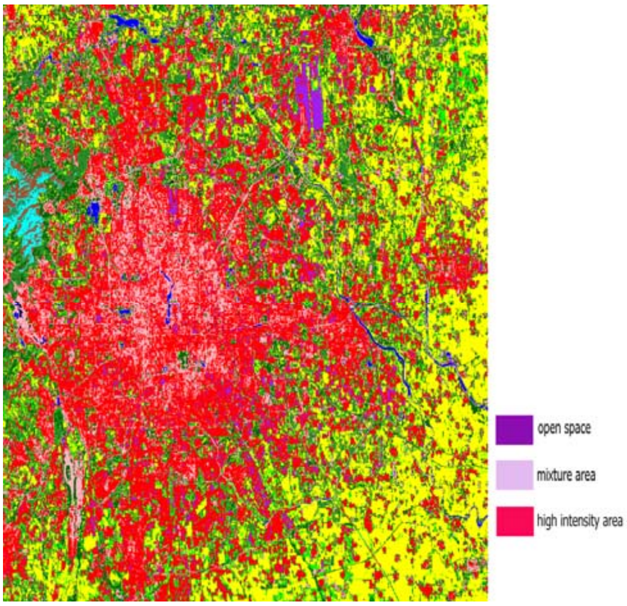

Figure 2. Urban area classification results of Beijing

3. CLASSIFICATION RESULTS AND ACCURACY ASSESSMENT

Decision tree classification offers an efficient way to classify the multi-layers included spectrum bands, NDVI index and DEM information. For this paper, 8 classes of land cover were mapped (Figure 1), and the author separate the urban area as three subclasses to provide a view of the construct form of the city (Figure 2). The subclasses were the following: （1）open space represented parks, air-depot and squares; （2）areas with a mixture of constructed materials and vegetation; （3）high intensity areas almost covered with row houses or commercial / industrial centre.

The classification accuracy was assessment using confusion matrix and kappa coefficient (Table3). The reference data which equally cover the whole area was selected to validate the classification result. The vegetation atlas and land use map of China, high resolution satellite images and GoogleEarth tool were applied in the reference data selected processing. The assessment result showed that class accuracies ranged between 55.18 to 100%, and the overall accuracy was 86.21%. The producer and user accuracy of deciduous broadleaf forest were lower. There were some intermixing between deciduous broadleaf forest and shrub, and also between crop and grass.

4. CONCLUSION

This study focused on the land cover classification using CBERS-02B CCD data. Based on land cover results of Beijing area described in this paper, it was concluded the CBERS-02B CCD data used as the data resource gained 86.21% overall accuracy applying decision tree classifier. Form the accuracy of every class, we elicited that there were some intermixing between deciduous broadleaf forest and shrub, and also between crop and grass, which because these classes had superposition on values of spectrum bands and the NDVI index, and also there were fewer bands available and slightly difference between the classes in the first 3 bands of CCD.

To sum up, the classification accuracy satisfied the precision requirements of land detection, and the decision rules built in this paper were efficient. Based on these initial results, we believe the CCD provided a dataset for land cover classification. And it has the potential to foster larger region and further exploration.

REFERENCES

Collin Homer, Chengquan Huang, Limin Yang, Bruce Wylie, and Michael Coan, 2004. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering & Remote Sensing*, 70(7), pp. 829–840.

C. M. D. Pinho,F. C. Silva,L. M. C. Fonseca, A. M. V. Monteiro, 2008. Intra-urban land cover classification from high-resolution images using the C4.5 algorithm. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Vol. XXXVII. Part B7. Beijing.

Giles M. Foody, 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, (80), pp. 185– 201.

James R. Anderson, Ernest E. Hardy, John T. Roach, and Richard E. Wttmer, 2001. A land use and land cover classification system for use with remote sensor data. *Geological Survey*.

M. Herold. P. Mayaux, C.E. Woodcock, A. Baccini, C. Schmullius, 2008. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing

Classification Result	Refrence data									
	Evergreen Needleleaf forest	Deciduous Broadleaf Forest	Shrub	Crop	Grass	Water	Urban Area	Shadow	Total	User Acc.
Evergreen Needleleaf forest	471	0	0	15	0	0	0	0	486	91.81
Deciduous Broadleaf Forest	0	271	172	0	75	0	0	0	518	52.32
Shrub	0	125	454	0	0	0	0	0	579	78.41
Crop	0	0	0	608	150	0	0	0	758	96.91
Grass	0	0	0	9	277	0	0	0	286	96.85
Water	0	0	0	0	0	439	0	0	439	100
Urban Area	42	0	0	0	0	0	833	0	875	95.2
Shadow	0	0	0	0	0	0	0	324	324	100
Total	513	396	626	632	502	439	833	324	4265	
Prod. Acc.	91.81	68.43	96.20	100	55.18	100	91.81	100		

Overall Accuracy = 3677/4265 =86.21%

Kappa Coefficient = 0.8405

Table 3. Confusion matrix of CCD classification

1 km datasets. *Remote Sensing of Environmen*, (112), pp.2538–2556.

S. Rasoul, Safavian and David Landgrebe, 1991 A survey of decision tree classifier methodology. *IEEE Trans. Systems, Man, & Cybernetics*, 21(3), pp. 660-674.

ACKNOWLEDGEMENTS

This work has been supported by the National Key Basic Research and Development Program, China, under project number 2006CB701303.