Research on the Application of Decision Tree in the Classification of Shandong Peninsula Land Use and Land Cover Change

Dehong Duan
Department of science and technology
Shandong University of Science and Technology
Qingdao, China
duandh@foxmail.com

Lin Sun
Geomatics College
Shandong University of Science and Technology
Qingdao, China
Sunlin6@126.com

Genxia Wang
College of mechanical and electrical engineering
Qingdao Binhai College
Qingdao, China
wgxia@foxmail.com

Zhengzhong Xu
Department of science and technology
Shandong University of Science and Technology
Qingdao, China
89152721@qq.com

***Abstract*—Land use type is the main content of land use and land cover studies, and the Land Dynamic Monitoring has been focusing on getting information of land cover on the global and regional scale quickly and accurately. This paper takes Shandong Peninsula as an example, MODIS remote sensing as data source, and selects MODIS09Q1 products of the first period in 8-day. Such group index number as NDVI, ENI, NDWI, NDMI will be illustrated to analyze the spectral relation between different land use types and construct new wave bands, finally the application of the decision tree in Shandong Peninsula land use and vegetation cover status will be researched. The results show that the adoption of MODIS data in macroscopical land use classification can achieve high classification accuracy. Such land types as the cultivated land and urban residential sites, which distribute in complete range have a higher accuracy, while such scattered land types as rural residential land and grassland have a lower accuracy.**

***Index Terms*—MODIS, land use classification, decision trees, Shandong Peninsula.**

I. INTRODUCTION

Land use or vegetation cover information is an important index to evaluate regional ecological environment, which occupies an important position in the study of regional sustainable development. The development of satellite remote sensing technology provides convenience for extraction of information in land use changes in a long time. Satellite remote sensing technology offers a wide view and continuous coverage areas, which enables to conduct dynamic monitoring. In 1960s, especially since in the middle of 90s, with the progress of satellite and aerospace technology and the rapid development of computer technology, the use of remote sensing technology can help people obtain huge amounts of information in a short period, which provides favorable conditions to obtain the data of changes in the land use classification or land vegetation cover. At present, in the research of land use classification or land vegetation cover changes, remote sensing technology has become one of the important tools.

In 1998, the MODIS sensor with medium resolution had been installed to EOSAM (the morning orbit) and PM (afternoon orbit) satellite series, and in 2000, American LPDAAC released data from MODIS sensor. MODIS still adopts methods of traditional imaging radiometer, operation period of 1-2 days, daily or every two days for global observation data at a time. Ground spatial resolution is 250m (2 bands), 500m (5 bands) and 1000m (29 bands), scanning width of 2330km. with high time resolution, wide coverage, timely acquisition of information, low cost,etc. MODIS remote sensing data has become the preferred data of a large scale of land use or cover vegetation monitoring as well as resources and environmental monitoring.

The late 1990s, emerging data mining technology makes the decision tree become recognized as the strong technology to construct the decision system, and also brings new ideas and methods to the classification of remote sensing images. Because of simple calculation, reused rules, artificial interference, and high classification accuracy, the decision tree has become the hot research point in the field of present remote sensing images classification. With Shandong Peninsula as the study area, with multi-temporal advantages of MODIS data and combination with advantages of the multi-spectral information, this paper builds decision tree classifier, to achieve automatic classification of regional land use in time.

II. THE SELECTION OF THE STUDY AREA'S DATA SOURCE

Shandong peninsula is located in the eastern part of Shandong Province, including the whole area of Qingdao, Yantai, Weihai and most parts or parts of Weifang, Rizhao and

Dongying. It is between the Yellow Sea and Bohai. The Peninsula belongs to warm temperate humid monsoon climate, with natural vegetation-the warm temperate deciduous broad-leaved forest, among which the oak is the main tree species, such as quercus serrata and serrata, and quercus serrata is the most. The zonal soil is the typical brown forest soil.

Based on MODIS remote sensing data, this study selects 8 days of MODIS09A1 from June to August in 2006 in Shandong Peninsula and synthetic MODIS13Q1 from MODIS09Q1 throughout the year, with the help of MODIS product processing software-MRT（MODISTOOL） to deal with MODIS data product, extract band information, converse projection type, unify the resolution to 250m. Apply the ENVI software for stitching, image cropping, data format conversion, wave superposition. At the same time, this paper will collect status data set of Shandong 1:500 thousand land use, and analysis report of 2006 annual land use as auxiliary data.

III. Information Extraction from Land Use / Land Cover Based on Decision Tree

A. Analysis of Typical Spectral Characteristics

MODIS provides 7 common optical reflectance bands, and because the band is narrower, the atmospheric effect weakens, data quality is better.

TABLE I. The wave length sand spatial resolution of MODIS first seven bands

Band	wavelength Range (μm)	Function
band1	0.620-0.670	Identification of vegetation types and distribution
band2	0.841-0.876	
band3	0.459-0.479	Measurements of aerosol content
band4	0.545-0.565	Identification of vegetation distribution and growth
band5	1.230-1.250	Identification of desertification and salinization
band6	1.628-1.652	
band7	2.105-2.155	Vegetation growth, soil material, soil degradation

The spectral difference between objects is the analysis theory to interpretate remote sensing images, and the computer automatic classification is based on a comprehensive grasp of spectral characteristics of the classification of objects in each band. With reference to IGBP classification system, according to the image visual preliminary interpretation, combined with the actual situation of land use in Shandong Peninsula, the terrain is divided into city residential, rural residential, water, farmland, grassland, woodland and others. In view of different object types, at least 10 sampling points are selected, and the maximum, minimum and mean statistics are calculated. The mean spectrum curve of the land use types as shown in figure 1, according to the sampling points.

B. Information Extraction of Objects Features

Based on the spectral characteristics of various objects in remote sensing images, in land use and land cover classification, the pixel is divided into various types of land cover types. In the use of decision tree classification model, such as NDVI, EVI, DN, spectral values and other variables are used to set the classification rules, therefore, to select reasonable feature variables and corresponding reasonable appropriate threshold is a key application of decision tree classification method. This paper selects first 7 bands of MODIS data as basic data, the first 7 bands are expressed as b_1 to b_7, and NDVI, EVI, NDWI, NDMI and the band operation index are selected to determine the appropriate threshold .

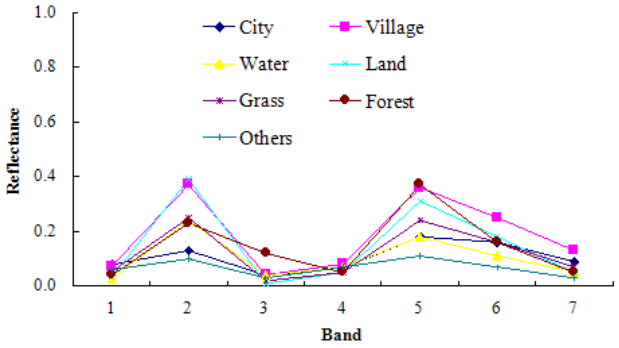

Fig. 1. Example of a figure caption.

- Extracting lake by NDVI index. After analysis, the threshold in the use type of lakes and other objects is 0.12, namely NDVI<0.12 for water.
- According to the spectral reflectance curve, extraction of city residential by the band operation b_7/b_6, the threshold is 0.58, namely $b_7/b_6>0.58$ for city residents.
- $b_2/b_1>6.1$ place for vegetation on the basis of crop characteristics in Shandong Peninsula, according to the June and August NDVI difference value, cultivated land and woodland can be distinguished.
- $b_1>0.15$ for rural residential land.

Based on the above analysis, decision tree model is shown in figure 2.

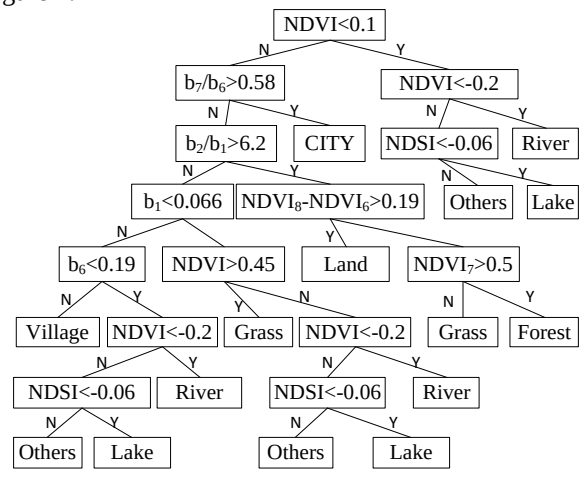

Fig. 2. The model of decision tree.

IV. Classification Results and Accuracy Evaluation

A. Classification results

According to the decision tree classification model to classify the land of the Shandong Peninsula, the land use

classification results after running the classification model, as shown in figure 3.

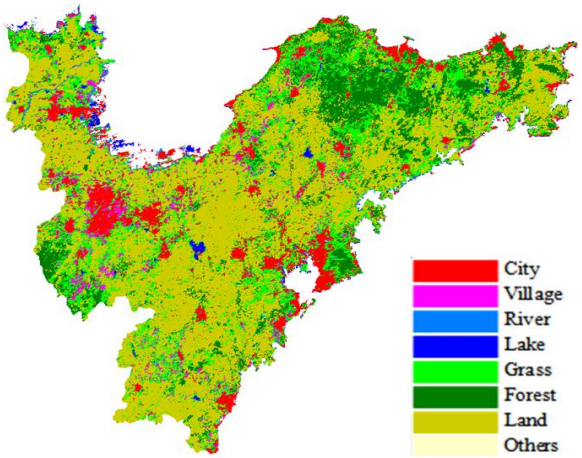

Fig. 3. Land use classification map of Shandong Province in 2006

B. Accuracy Assessment

The accuracy assessment is an integral part of remote sensing classification, therefore, the classification of the land of the Shandong Peninsula, classification accuracy of the classification results in evaluation of the region's high-resolution MODIS images.

Stratified sampling method, in order to ensure the accuracy of evaluation, it is from one can be divided into different sub-population overall, according to the provisions of the proportion of randomly selected samples from different layers.

The advantage of this method is that the representativeness of the sample, the sampling error is relatively small.

The disadvantage is even more complicated sampling procedures than simple random sampling.

In the end, a total of eight types of feature select 3572 sampling points, and create confusion matrix around this calculation accuracy of the producer, the overall accuracy and kappa coefficient, the results as shown in Table 2.

TABLE II. CONFUSION MATRIX OF SAMPLES AND THE LAND USE CLSSIFICAN RESULT

Class	city	village	land	grass	forest	river	lake	others	Total	Producer's Accuracy
city	779	26	0	45	0	0	44	57	951	81.91%
village	35	69	0	6	0	19	0	1	130	53.08%
land	0	8	196	66	2	2	9	0	283	69.26%
grass	8	11	0	116	11	1	128	2	277	41.88%
forest	0	0	3	7	490	0	0	0	500	98.00%
river	2	0	0	0	0	55	6	15	78	70.51%
lake	0	0	0	0	0	0	847	5	852	99.41%
others	0	0	0	0	0	0	1	500	501	99.80%
Total	824	114	199	240	503	77	1035	580	3572	
Overall Accuracy=85.44%						Kappa Coefficient=0.8215				

As can be seen from Table 2, limited by the spatial resolution of the MODIS data, the accuracy difference between the material around the large urban settlements, woodland, lakes and other types of classification accuracy is higher, while the rural settlements, grassland classification accuracy is poor, For the broken and scattered villages, the grass, the classification accuracy is not high, but the overall accuracy is 85.44%, Kappa coefficient was 0.8215.

V. CONCLUSIONS

In this study, based on the MODIS remote sensing data, using the decision tree classification method using the classification experiment is done on the Shandong peninsula of land, use the characteristic of multispectral MODIS data and the spectral differences among, by selecting the NDVI index and the band, and select the appropriate threshold, constructs the decision tree classification, and achieved good classification results, especially for city residents, woodland, large lake, the classification accuracy is higher, Laid the foundation for the next step on the regional scale and large-scale land use.

However, due to the limit of the spatial resolution of MODIS data, for broken land classification accuracy is lower, by using multiple remote sensing data to compensate for the lack of spatial resolution.

ACKNOWLEDGMENT

This work was financially supported by "Promotive research fund for young and middle-aged scientisits of Shandong Province"(BS2011HZ010), "the Project of Shandong Province Higher Educational Humanities and Social Science Program" (J12LH02), "Shandong province soft science research project" (2012RKB01356), "the Key Laboratory of Surveying and Mapping Technology on Island and Reef, State Bureau of Surveying and Mapping, China"(2012B06).

REFERENCES

[1] Price, John C. "Comparing MODIS and ETM+ data for regional and global land classification." Remote Sensing of Environment. Vol. 86, pp. 491–499, August 2003.

[2] M.C Hansen, R.S DeFries, J.R.G Townshend, L Marufu, R Sohlberg. "Development of a MODIS tree cover validation data set for Western Province, Zambia." Remote Sensing of Environment. Vol. 83, pp. 320-335, November 2002.

[3] L Yuan, L Nianfeng. "Macroscopic assessment of land degradation on the Songliao plain using MODIS data." Geography and Geo-information Science. Vol. 20, pp. 22-25, July 2004.

[4] X Zhang, R Sun, B Zhang, Q Tong. "Land cover classification of the North China Plain using MODIS_EVI time series." ISPRS Journal of Photogrammetry and Remote Sensing. Vol. 63, pp. 476-484, July 2008.

[5] Y Liu, Z Niu, C Wang. Research and application of the decision tree classification using MODIS data." Journal of Remote Sensing. Vol. 9, pp. 405-405, April 2005.

[6] BD Wardlow, SL Egbert. "Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains." Remote Sensing of Environment. Vol. 112, pp. 1096-1116, March 2008.

[7] RS Lunetta, JF Knight, J Ediriwickrema, JG Lyon, LD Worthy. "Land-cover change detection using multi-temporal MODIS NDVI data." Remote sensing of environment. Vol. 105, pp. 142–154, November 2006.