THE LAND COVER MAPPING IN NORTH KOREA USING MODIS IMAGE; THE CLASSIFICATION ACCURACY ENHANCEMENT FOR INACCESSIBLE AREA USING GOOGLE EARTH

Su-young Cha*, Chong-hwa Park**

Graduate School of Environmental Studies, Seoul National University, Korea, younga74@gmail.com
Graduate School of Environmental Studies, Seoul National University, Korea, rsgis@snu.ac.kr

ABSTRACT: A major obstacle to classify and validate Land Cover maps is the high cost of generating reference data or multiple thematic maps for subsequent comparative analysis. In case of inaccessible area such as North Korea, the high resolution satellite imagery may be used as in situ data so as to overcome the lack of reliable reference data. The objective of this paper is to investigate the possibility of utilizing QuickBird (0.6m) of North Korea obtained from Google Earth data provided thru internet. Monthly NDVI images of nine months from the summer of 2004 were classified into L=54 cluster using ISODATA algorithm, and these L clusters were assigned to 7 classes; coniferous forest, deciduous forest, mixed forest, paddy field, dry field, water and built-up area. The overall accuracy and Kappa index were 85.98% and 0.82, respectively, which represents about 10% point increase of classification accuracy than our previous study based on GCP point data around North Korea. Thus we can conclude that Google Earth may be used to substitute the traditional in situ data collection on the site where the accessibility is severely limited.

KEY WORDS: Land cover classification, Classification accuracy assessment, Google Earth, QuickBird, MODIS, North Korea

1. INTRODUCTION

Land cover (LC) is one of key factors for estimating net primary production, anthropogenic impact on global environment, and so on (Matsuoka, 2001). The pace of land cover change of the world, including North Korea, is very fast. Especially, the large-scale deterioration of agricultural fields and forests has been severe problems after the collapse of communism in the East European bloc countries and several flooding incidents of 1990s. Much of the steep slopes of the country have been destroyed by slash-and-burn practices by local residents (Palm, 2005). The status of land degradation is essential for the forest restoration of the country.

From the 1990s, several scientific researches have been carried out for the land cover mapping of North Korea using remote sensing. Lee (1994) tried to classify primary vegetation cover types all over the Korean Peninsula using the multi-temporal Advanced Very High Resolution Radiometer (AVHRR) imagery. Though AVHRR data have the advantage of a high temporal resolution, the coarse spatial resolution of 1km is a problem. Lee et al., (1998) used Landsat images to survey forest classification and growing stock of North Korea. Lee et al., (1999) used Landsat data covering three different time periods to compare the content and characteristic of forest cover changes for the province of Pyongyang and Heasan in North Korea. However, the land cover mapping with a single temporal image can not take into account the characteristic of phenology in forests and agricultural areas. Multi-temporal MODIS imagery can help to enhance the classification accuracy or make a reasonable qualified land cover map compared with a single date image. Kim et al., (2007) investigated the optimal land cover classification algorithm for the monitoring North Korea with MODIS multi-temporal data. They showed that ISODATA and SMA algorithm resulted in a higher classification accuracy of forest and agricultural categories, but the overall classification accuracy of SOM algorithm (73.57%) resulted in higher compared ISODATA (69.03%) and SOM (64.28%).

To improve classification accuracy it is important to choose a reference data. Historically many studies have been used: ground visit, aerial photography, airborne video data and satellite imagery. Especially an aerial photography has long been served in the creation of LC maps, as a mapping base and more recently as a source of higher-resolution reference data, for the comparison with maps produced by the classification of satellite imagery. But, as in many remote areas in developing countries, data sources for producing and assessing accuracy of LC maps for study area are limited (Sydenstricker-Neto, J., et al., 2004). In case of North Korea, historical aerial photographs are not available and the ground visit with GPS machine has not allowed because of their dictatorial government. Due to the limited reference data of North Korea, studies (Lee, 1994; Lee et al., 1998; Lee et al., 1999) used the statistical data which was published by various resources, but they are not reliable. Kim et al., (2007) used the ground truth data with GPS machine around North Korea, which were collected in the border of China and South Korea. However, there area limits to investigate the inside of North Korea in a regional scale, too. Therefore, Google Earth can be a good alternative when we classify an inaccessible area.

2. METHODS

2.1 Google Earth

Since Google Earth launched in 2005, many people use Google Earth to explore the world around them. It maps the earth by the superimposition of images obtained from satellite imagery, aerial photography and GIS 3D globe (http://en.wikipedia.org/). Most land is covered in at least 15 meters of resolution and some really high resolution insets in cities. New high resolution and Large Digital Globe, QuickBird (0.6m), update includes areas in cites plus some interesting region (http://earth.google.com).

The study area is located in the East Asia, northern half of the Korean peninsula bordering the Korea Bay and the East Sea. It covers 122,762 km^2 between 38~44 north latitude, and 124~131 east longitude. In Google Earth, currently, 147 high resolution satellite images from Quickbird for the period of the year 2002 to 2007 are available over North Korea, which covers about 30% of the land area.

For more professional uses, Google Earth Pro allows us to import the GIS files such as .shp and .tab through the GIS Data Importing Module. To evaluate whether Google Earth data can be used as a reference data, the test of location error was conducted in five points in South Korea with the GPS equipment and we found that the location error of Google Earth is less than 1 m in the surface of the earth. The internal coordinates system of Google Earth is geographic coordinates on the World Geodetic System of 1984 (WGS84) datum.

2.2 MODIS imagery DB

The Moderate Resolution Imaging Spectroradiometer (MODIS) offers an opportunity for detailed, large area Land use/Land cover characterization by providing global coverage with high temporal resolution (1~2 days) and intermediate spatial resolution (250m). It includes a time series of visible red (620-670 nm) and near infrared (841-876 nm) surface reflectance (Justice & Townshend, 2002; Wardlow, et al., 2007).

Monthly MODIS data from June of 2004 to May of 2005 were generated by applying monthly composite of the Normalized Difference Vegetation Index (NDVI) using 65 MODIS 1B images with 250m resolution acquired over from Japan MODIS Data Service Centre (http://webmodis.iis.u-tokyo.ac.jp). The NDVI is a normalized difference measure comparing the near infrared and visible red bands defined by the formula.

$$NDVI = (\rho_{NIR} - \rho_{red} / \rho_{NIR} + \rho_{red})$$

But monthly data might cause severe classification errors were excluded from the DB. Monthly data from December of 2004 to February of 2005 were excluded because of snow cover, and monthly data of July of 2005 were excluded because of the remaining cloud cover. Monthly NDVI MVC composites were classified with Iterative Self-Organizing Data Analysis (ISODATA) algorithm (100 Iteration), a widely used variant on the K-means method for unsupervised clustering. Geographic correction parameters were set UTM projection (Zone 52N) and WGS 84 Datum (Fig.1).

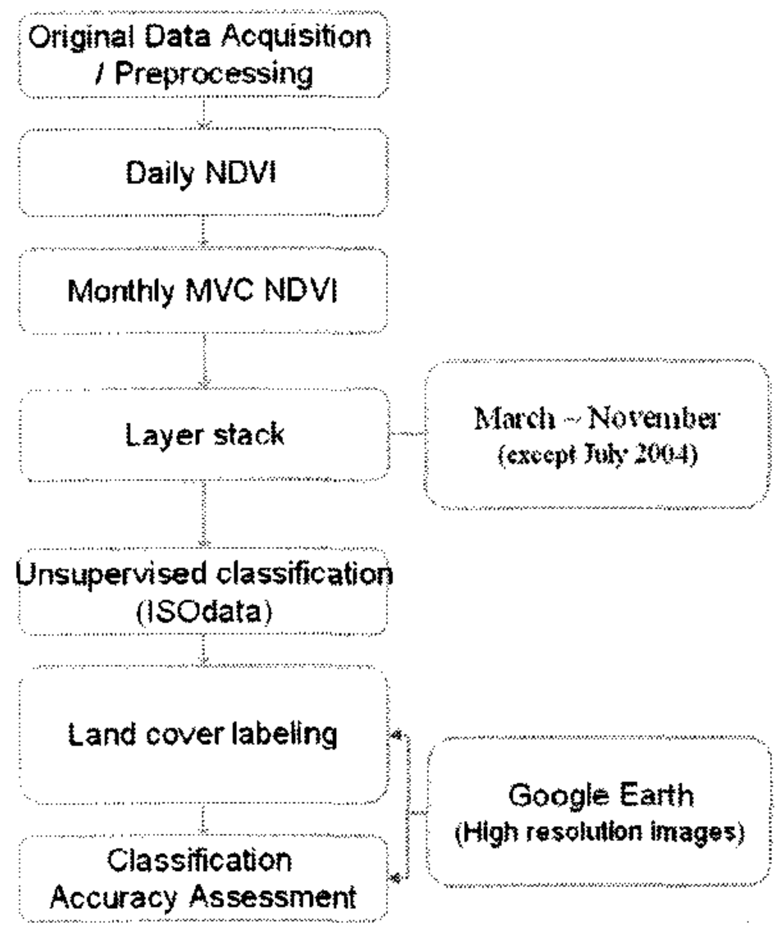

Figure1. Flow diagram of data processing

2.3 Classification scheme

A classification scheme has two critical components: (1) a set of labels and (2) a set of rules or definition such as a dichotomous key for assigning labels (Green, 2004). All study partners worked together to develop and finalize the classification scheme with the necessary labels and rules. And, seventy-three training pixels were identified from the high resolution imagery ($\leq$ 0.6m) provided by Google Earth through visual interpretation referring high-resolution satellite imagery interpretation key (Lillesand, 2004). Especially, we mainly selected sample points from winter season images. They can make easier to distinguish between coniferous and deciduous because of the characteristic of leafless canopies.

We utilized a seven-category LC classification scheme. The classes included coniferous forest, deciduous forest, mixed forest, paddy field, dry field, water and built-up area. The water class is all areas of open water, generally with less than 25% vegetation. The built-up class characterized by the high percentage of construction materials. The coniferous class contained at least 75% conifer and more than 25% deciduous. The deciduous class had at least 75% deciduous and no more than 25% coniferous. All other forest types were labelled as a mixed wood. To discriminate the crop land, if they are in water during its cultivation period, they were labelled as a paddy field. If not, the rest were labelled as a dry field. The bare land and grassland were not categorized. Because a bare land is difficult to distinguish from truly bare land areas at the MODIS 250m pixel size, it was included in a built-up class. And we could not acquire enough the training data detecting the difference between grassland and deciduous forest, thus grassland was included in the deciduous class.

2.4 Sampling procedure

A generally accepted rule of thumb is to use a minimum of 50 samples for each LC category in the error matrix. If the area is especially large or the classification has a large number of LC categories (i.e., more than 12 categories), the minimum number of samples should be increased to 75 to 100 samples per category.

In this study sample points were selected using a stratified random sampling design, stratified by LC area for each of the accuracy assessments. Research by Conglaton (1988) indicates that random and stratified random samplings are the optimal sampling designs for accuracy assessment. Interpretations were based on the majority of a 3 x 3 pixel window (Congalton and Green, 1999; Khorram, 2004). We selected the homogeneous patches through 3 x 3 pixel blocks window. For rare classes (e.g., water), pixel sample points were chosen with at least six pixels in the window belonging to the same class.

First we selected the 99 QuickBird images in the year 2004~2005 among the whole 150 QuickBird images in the year 2002~2007 which covers North Korea in Google Earth presently, because MODIS data used in this study are the year 2004~2005. Second primary sample unit (PSU) were 3 x 3 km quadrangle area. Sample units are the portion of the landscape that will be sampled for the accuracy assessment. 15~35 PSU per class were selected across all the quadrangles. Third random pixel sampling was performed. Finally, 215 PSU for seven land cover classes were selected and a total of 1049 reference samples were collected. A stratified random sampling design was incorporated with samples apportioned by LC area, using a minimum sample size of n = 50 for rare classes. After evaluation of selected sample points in each reference data set, an error matrix was constructed. It compares map class labels with reference data labels for each LC classification. Overall map accuracy and class-specific user and producer

3. RESULTS

The results of this study can be summarized as follows. The monthly NDVI images were classified into L = 54 cluster using ISODATA algorithm and these L clusters were assigned to K = 7 classes, which were labelled into the classes of forest (coniferous, deciduous, mixed), crop land (paddy, dry), water and built-up. Figure 2 illustrates the land cover classification with MODIS of North Korea. Overall accuracy is 85.98% and Kappa index is 0.82. Accuracy for specific classes ranged from 48% and 97%, achieving $\geq$ 96% for the deciduous forest. The producer's accuracy of the forest and crop land is quite good. This is because the MODIS NDVI data for the classification process is sensitive to vegetation. And the sample size of forest area and crop land is larger than built-up and water. It makes easier to collect the reference data of a good quality. The water class was particularly difficult to classify and attained accuracies below 50%. The sample size for this particular cover type was relatively small, which may have contributed to this poor outcome. It is because that Water and Built-up, although spectrally distinct, were easily biased along edge pixels.

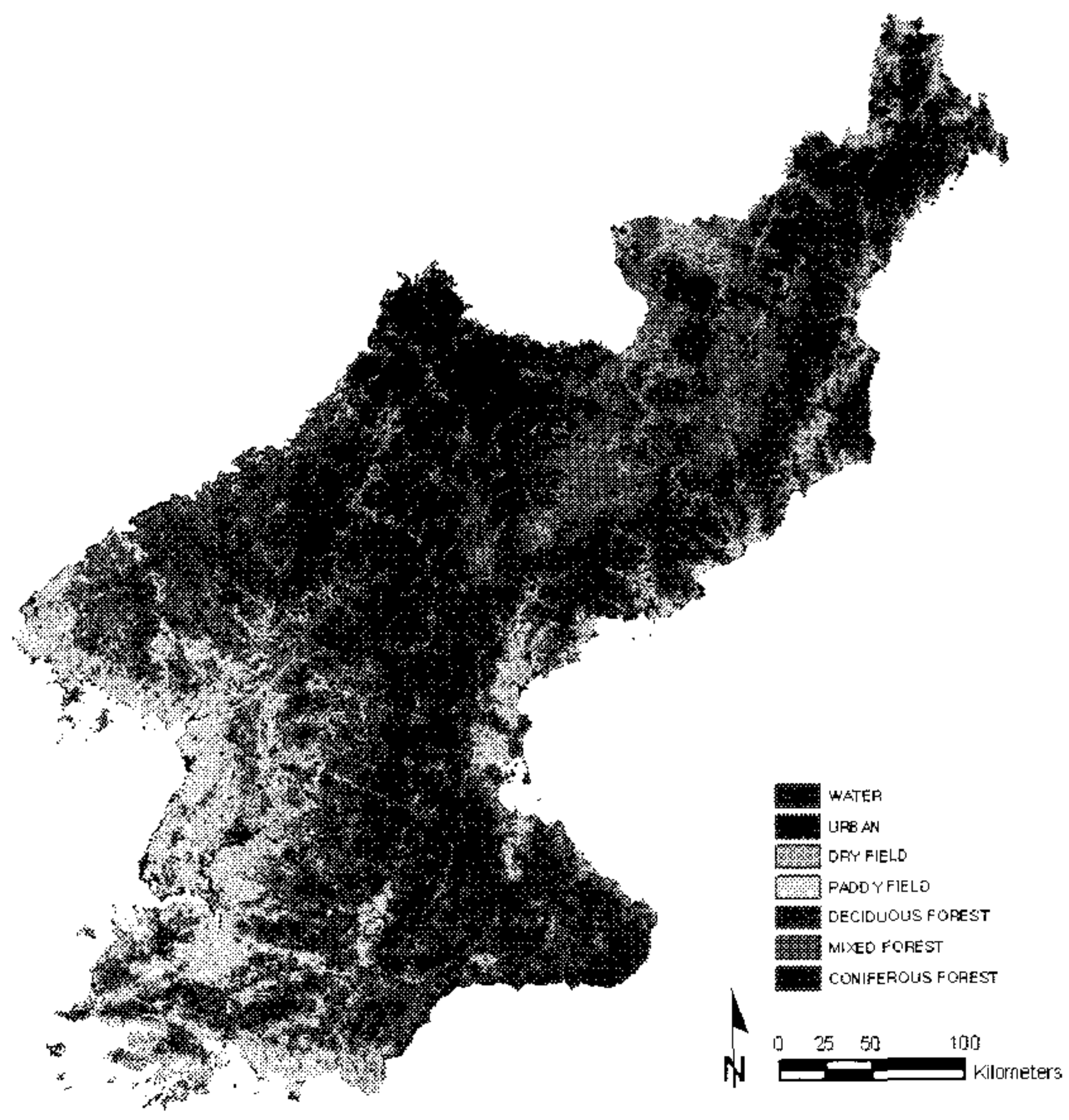

Figure2. Land cover classification in 2004~2005

4. CONCLUSIONS

The results discussed in this chapter indicate that Google Earth can be used successfully to perform classification accuracy assessment on LC maps derived from multi-temporal MODIS. In addition, this study shows that our result (85%) was achieved considerable enhancement compared with our previous study (76%), which was similar to the level of classification scheme and data that used by our study. It represents about 10% point increase of classification accuracy. Therefore, by acquiring reference data through Google Earth, we could reduce the classification error.

Up to date, even though the high spatial resolution satellite imagery ($\leq$ 1m) has a superior ability in understanding natural and built-up environment through visual interpretation, it has barriers due to its expensive cost and complex pre-processing methods. Google Earth makes easier the cost-effective collection of a statistically meaningful number of sample points. Besides, Land cover map can be obtained more exactly even in inaccessible areas with limited reference data. Furthermore, this study can be an efficient case study mapping Global land cover product.

ACKNOWLEDGEMENTS

"This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government(MOST) (Cooperative Research for Restoration of Degraded Ecosystems in Northeast Asian Regions)."

REFERENCES

Congalton, R.G., 1988. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data, *Photogram. Eng. Remote Sens.* 54: 587-592.

Green, K., and Congalton, R. G., 2004. An Error Matrix Approach to Fuzzy Accuracy Assessment: The NIMA Geocover Project, Remote Sensing and GIS Accuracy Assessment, CRC PRESS: 163-172.

Khorram, S., Knight, J. F., and Cakir, H. I., 2004. Thematic Accuracy Assessment of Regional Scale Land-Cover Data, Remote Sensing and GIS Accuracy Assessment, CRC PRESS: 91-103.

Kim, D. H., Jeong, S. G., and Park, J. H., 2007. Comparison of Three Land Cover Classification Algorithms – ISODATA, SMA and SOM- for the Monitoring of North Korea with MODIS Multi-temporal Data. *Korean Journal of Remote Sensing,* 23(3): 1-8.

Lillesand, T. M. and Kiefer, R. W. and Chipman, J. W., 2004. *Remote Sensing and Image Interpretation(5th),* John Wiley & Sons, Inc.

Lee, K. S., 1994. Vegetation Cover Type Mapping over the Korean Peninsula Using Multitemporal AVHRR Data, *Jour. Korean For. Soc.,* 83(4): 441-449.

Lee, K. S., Joung, M. R., and Yoon, J. S., 1999. Content and Characteristics of Forest Cover Changes in North Korea, *Jour. Korean For. Soc.,* 88(3): 352-363.

Lee, S. H., Choung, S. H., and Song, J. H., 1998. Forest Resources Inventory of North Korea using Satellite Remote Sensing Data, *RI. J. For. Sci.,* 58: 1-13.

Matsuoka, M., et. al., 2001. Feasibility study of land cover classification using bidirectional reflectance distribution function model, *IGARSS 2001, IEEE 2001 International,* 5: 2224 – 2226.

Palm et al., 2005. In: C.A. Palm, S.A. Vosti, P.A. Sanchez and P.J. Ericksen, Editors, *Slash-and-Burn Agriculture: The Search for Alternatives,* Columbia University Press.

Sydenstricker-Neto, J., Parmenter, A. W., and DeGloria, S. D., 2004. Participatory Reference Data Collection Methods for Accuracy Assessment of Land-Cover Change Maps, Remote Sensing and GIS Accuracy Assessment, CRC PRESS: 75-90.

Wardlow, B. D., Egbert, S. L., and Kastens, J. H., 2007. Analysis of time-series MODIS 250m vegetation index data for crop classification in the U.S. Central Great Plains, *Remote Sensing of Environment,* 108: 290-310.