A HIERARCHICAL APPROACH TO HIGH-RESOLUTION HYPERSPECTRAL IMAGE CLASSIFICATION OF LITTLE MIAMI RIVER WATERSHED FOR ENVIRONMENTAL MODELING

Joon Heo, Michael Troyer, Jung Bin Lee, and Woo Sun Kim

*Yonsei University, School of Civil and Environmental Engineering
jheo@yonsei.ac.kr, ortolan@naver.com, kwsbulls-77@hanmail.net
** National Center for Environmental Assessment, U.S. Environmental Protection Agency
Troyer.Michael@epamail.epa.gov

ABSTRACT: Compact Airborne Spectrographic Imager (CASI) hyperspectral imagery was acquired over the Little Miami River Watershed (1756 square miles) in Ohio, U.S.A., which is one of the largest hyperspectral image acquisition. For the development of a 4m-resolution land cover dataset, a hierarchical approach was employed using two different classification algorithms: "Image Object Segmentation" for level-1 and "Spectral Angle Mapper" for level-2. This classification scheme was developed to overcome the spectral inseparability of urban and rural features and to deal with radiometric distortions due to cross-track illumination. The land cover class members were lentic, lotic, forest, corn, soybean, wheat, dry herbaceous, grass, urban barren, rural barren, urban/built, and unclassified. The final phase of processing was completed after an extensive Quality Assurance and Quality Control (QA/QC) phase. With respect to the eleven land cover class members, the overall accuracy with a total of 902 reference points was 83.9% at 4m resolution. The dataset is available for public research, and applications of this product will represent an improvement over more commonly utilized data of coarser spatial resolution such as National Land Cover Data (NLCD).

KEY WORDS: High-Resolution Hyperspectral Image; CASI; Land Use Classification; Non-Point Source Pollution; Water Quality Modeling; Image Object Segmentation, Spectral Angle Mapper

1. INTRODUCTION

1.1 Background

In 2001, an interdisciplinary group of scientists based at the U.S. Environmental Protection Agency (EPA) in Cincinnati, Ohio formed a collaborative effort to study the Little Miami River (LMR) and its watershed in southwestern Ohio, USA. The objective of the study was to improve the understanding of relationships between non-point sources of nutrients in watersheds, nutrient enrichment in rivers and streams, and the ecological responses to these stressors and was aimed at providing useful assessment approaches, information, and models to address the risks of excess nutrients in agricultural and urbanizing watersheds. Changes in water quality can indicate a change in some aspect of a terrestrial, riparian, or in-channel ecosystem. In spite of the advancement of environmental modeling, the applicability of the grid-based environmental models is limited due to insufficient information and data-intensive requirements. The ultimate goal of developing a LULC dataset of high spatial resolution is to help enhance the use of geographic and spatial analytic tools in risk assessments and to improve the scientific basis for environmental risk management decisions. From a technological point of view, an innovative hierarchical classification approach was devised, incorporating both object-based pattern recognition and spectral image processing techniques.

1.2 Project Site

The Little Miami River in southwestern Ohio, USA has a drainage area of 1756 square miles (1.15 million acres) and stretches in a southwesterly direction for 105.5 miles from its origin near South Charleston, Ohio to its confluence with the Ohio River east of Cincinnati, Ohio (Figure 1). It is one of the oldest river groups in the state, having become Ohio's first State and National Scenic River (Sanders, 2002).

2. METHOD

2.1 Pre-Processing

In addition to the need for geometric correction of remote sensing data in general, the acquisition of airborne optical image data is susceptible to a number of effects due to sun angle, target characteristics, atmospheric conditions, and others. Of these effects, acquisition geometry and total-scene radiance generally most directly affect the ability to produce high-quality mosaics and also to perform accurate multispectral or hyperspectral classifications. The following are those preprocessing issues that have to be considered before classification.

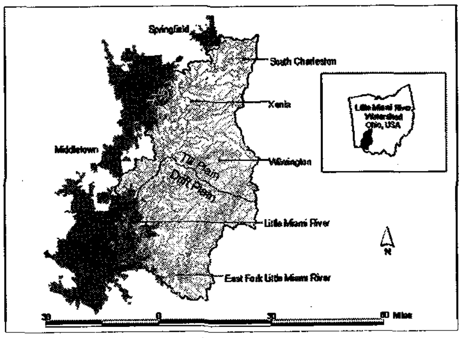

Figure1. Project Site Map of Little Miami River Watershed in Southwestern Ohio, USA

2.2 Main Part of Method

2.2.1 Cross-Track Illumination

The general method to correct cross-track illumination is to calculate the means for each sample of along-track pixels, to use a polynomial function to fit the average curves, and to obtain the correction factors by normalizing the fitted curves (Kennedy et al 1997; Research Systems Inc. 2003).

2.2.2 Flightline Radiance Variations

Flightline radiance variations are another issue occurring as a result of decreasing/increasing illumination over the elapsed time of the data acquisition period.

2.2.3 Image Classification

The proposed classification method employed a hierarchical approach using two different classification algorithms: "Image Object Segmentation" for level-1 and "Spectral Angle Mapper" for level-2. This classification scheme was developed to overcome the spectral inseparability of urban and rural features and to deal with radiometric distortions due to cross-track illumination.

In order to combine the effectiveness of both approaches, object image segmentation in eCognition (Definiens Imaging GmbH., 2003) was applied as a "Level 1" classification of water, urban, and rural features, which classification required consideration of large scale-factors as well as area-based parameters such as adjacency, texture and shape. Next, SAM in ENVI (Research Systems Inc., 2003) was applied as a "Level-2" classifier to distinguish the urban and rural areas into more specific classes (barren, built, grass, corn, soypbeans, wheat, etc.). Wire diagrams of the methodology used are shown in Figures 2 and 3. Throughout the process, image object segmentation created a spatially exclusive mask of urban and rural regions, and then each region was filled with the classification result from SAM. As a consequence, the final classification result remained a uni-scale product with a spatial resolution of 4m throughout the entire watershed.

Training samples were chosen along all flight lines and selected in order to capture intra-class variation. The same classification rules were applied to each flightline. After the first round of classification, the results were fine-tuned by adding training sets to accurately define inter-class boundaries.

The classification results for each individual flight line were assessed for accuracy and accepted if they did not show any overall discrepancies with respect to the aerial images.

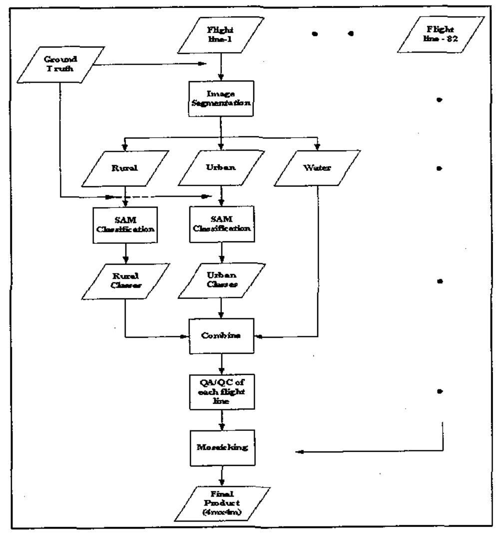

Figure 2. Flowchart of Classification Method

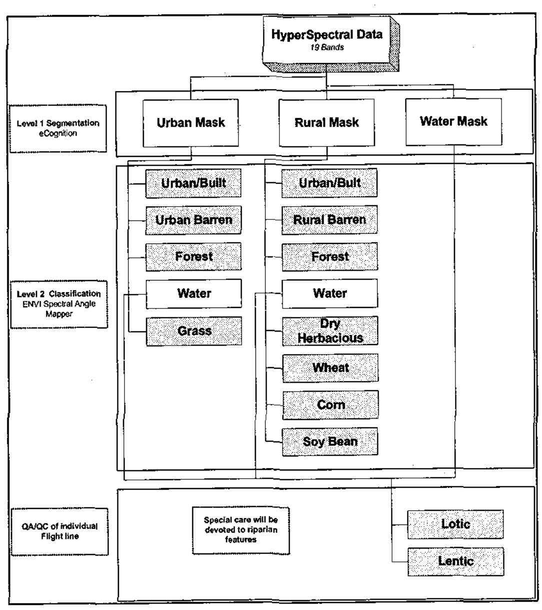

Figure 3. Schematic of Hierarchical Classification and Derived Classes (shadowed)

The classification results were also re-examined after joining the classified flight lines together in a mosaic of the watershed.

2.3 Post-Processing

Image post-processing and QA/QC following the Level 1 and 2 classifications included manual editing and map generalization to create a second LULC product. Manual editing was employed as a final QA/QC step for (1) differentiation of water bodies into lotic and lentic; (2) assignment of clouds, shadows, and haze to "unclassified" class. After the manual editing, a "clump-sieve-and-fill" technique was used to eliminate single pixels or groupings of pixels that were smaller than the minimum target mapping unit (e.g., random pixels of "Forest" denoting scattered trees in an otherwise homogeneous 40-acre plot of "Corn"). As a result, a second LULC mosaic product was produced, which eliminated the "salt and pepper" effect common in classifications of smaller pixel or "finer" spatial resolution imagery. For the smoothed product, the minimum mapping unit of the classification result was about 0.04 of an acre as represented by 10 pixel clusters (at 4m x 4m spatial resolution), or linear chains of (minimally) four contiguous pixels in any direction. The original, unsmoothed product remained at a 4m x 4m spatial resolution.

3. ACCURACY ASSESSMENT

The accuracy assessment was based on whether the majority of classed pixels in a 3x3 pixel window, centered on a ground truth site, agreed or not. Thus, if five or more pixels were classified as corn, and ground truth indicated corn, then the majority criterion was satisfied and "corn class" would be considered correct for that site.

Reference / Classified	Lentic	Lotic	Forest	Corn	Soy	Wheat	Dry Herb	Grass	Urban Barren	Rural Barren	Urban/Built	Total
Lentic	85	1	0	0	0	0	0	0	0	0	0	86
Lotic	3	17	0	0	0	0	0	0	0	0	0	20
Forest	0	2	95	0	0	0	3	4	0	2	0	106
Corn	0	0	1	106	3	0	1	0	0	0	0	111
Soy	0	0	0	3	107	2	5	1	0	0	0	118
Wheat	0	0	0	0	0	18	0	0	0	0	0	18
Dry Herb	2	0	3	9	12	8	82	16	0	8	0	140
Grass	1	0	1	0	0	0	8	75	8	0	0	93
Urban Barren	0	0	0	0	0	0	0	1	57	1	4	63
Rural Barren	6	1	0	2	0	10	1	3	0	20	1	44
Urban/Built	1	3	0	0	0	0	0	0	4	0	95	103
Total	98	24	100	120	122	38	100	100	69	31	100	902

Table 1. Classification Error Matrix

Class Names	Reference Total	Classified Total	Number Correct	Producer's Accuracy	User's Accuracy
Lentic	98	86	85	86.7%	98.8%
Lotic	24	20	17	70.8%	85.0%
Forest	100	106	95	95.0%	89.6%
Corn	120	111	106	88.3%	95.5%
Soy	122	118	107	87.7%	90.7%
Wheat	38	18	18	47.4%	100.0%
Dry Herb	100	140	82	82.0%	58.6%
Grass	100	93	75	75.0%	80.7%
Urban Barren	69	63	57	82.6%	90.5%
Rural Barren	31	44	20	64.5%	45.5%
Urban/Built	100	103	95	95.0%	92.2%

Overall Accuracy : 757(number correct) / 902(reference total) = 83.9%

Table 2. A Statistics Summary of the Accuracy Assessment

A standard error matrix was used in reporting the classification accuracies (Table 1). A total of 902 independent ground truth sites were used for the accuracy assessment, including primary data (i.e., data collected by EPA scientists in the field at the time of the overflights), and secondary data from 2002 and 2003 aerial images of the watershed. Table 2 presents a statistical summary of the accuracy assessment for the classification results using 902 reference points with respect to all eleven landcover classess. The overall accuracy of the product was 83.9%.

4. RESULT AND ANALYSIS

The overall classification accuracy was 83.9%, which is above the project's target of 80% accuracy. However, the producer's and user's accuracy of some classes fell short of the target of 80%. Nevertheless, the strength of this classification relative to other existing LULC datasets of this watershed, such as the National Land Cover Dataset or NLCD (Vogelmann et al., 2001), and the State of Ohio Land Cover (Ohio DNR, 1994), is the higher spatial resolution (e.g., 4m x 4m rather than the 30m x 30m pixel resolution of previous existing classifications of this watershed) for such a large area—4495 square kilometers, one of the world's largest hyperspectral image acquisitions. Figure 5 visualizes the difference between the classification results for 4m and 30m resolution, respectively. Development of the dataset was customized for water quality modeling with special care for linear strips of riparian vegetations.

Figure 4. Overview of classification of entire watershed at 4m resolution

A basic understanding of the physical (i.e., geological and anthropogenic) processes at work in the Little Miami River Watershed will help the user of this LULC dataset interpret some of the resulting land use patterns shown in Figure 4. These soils normally have better natural drainage and fertility than those of the southern half of the watershed (or "Drift Plain"). The southern half of the watershed has more deeply-leached, acidic, pre-Wisconsinan till and thin loess as well as very poorly-drained soils with fragipans (clays). The southern half of the watershed also exhibits relatively modest relief, but with dissected areas and somewhat more complex topography than the northern half (Omernik, 1987; Woods et al., 1998).

As such, the northern and southern parts of the watershed can be expected to have different types and proportions of certain land uses or land covers based on the differing soils and micro-climates found in these two distinct "ecoregions".

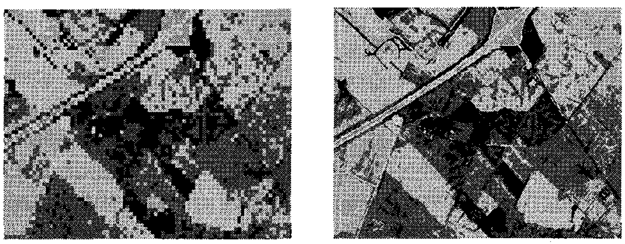

Figure 5. Comparison of 4m and 30m resolution classifications

Spatial patterns separating western and eastern portions of the watershed exist as well. Perhaps most notable is the western urban/exurban corridor stretching from Cincinnati (in the south) to Dayton and Xenia (in the north) and beyond, encompassing portions of Hamilton, Warren, and Montgomery Counties. These growing urban landscapes run parallel to and already straddle much of the main stem of the Little Miami River, which can be observed as a nearly contiguous linear band of riparian forest running upwards along the western part of the image. The eastern half of the watershed tends to be more agricultural in character, particularly in the north. But this characteristic appears to wane in the east-central part of the image near the city of Wilmington (a crossroads or pole for the primary economic sector in this region as well as a major air transportation hub), and in the south as well, particularly along the East Fork of the Little Miami River in Clermont County, where urban development and human population continues to rapidly grow.

5. CONCLUSION AND FUTURE RESEARCH

High-resolution, hyperspectral images were acquired and processed to produce a 4m x 4m land cover classification for the Little Miami River Watershed. A hierarchical combination of innovative approaches was used; subsequently, image object segmentation and spectral angle mapper (SAM) were applied, and the classification was successfully completed with the overall accuracy of 83.9%. The resulting classification product is an important dataset for a variety of environmental and geographic studies within the Little Miami River Watershed. Even given the predominance of the "dry herbaceous" class, the classification product remains meaningful for studying several urban and agricultural patterns or gradients as well as anthropogenic and natural processes within the watershed.

6. REFERENCES

Definiens Imaging GmbH. 2003. *eCognition 3.0.* Munich, Germany.

Kennedy, R., W. Cohen, G. Takao. 1997. Empirical methods to compensate for a view-angle-dependent brightness gradient in AVIRIS imagery. Remote Sensing of Environment, 62(3):277-291.

Ohio DNR (Ohio Department of Natural Resources). 1994. *State of Ohio Land Cover. Division of Real Estate and Land Management.* Columbus, OH.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Map (1:7,500,000). Annals of the Association of American Geographers. 77(1):118-125.

Research Systems Inc. 2003. *ENVI 3.6 (the Environment for Visualizing Images).* Boulder, CO.

Sanders, R.E., Ed. 2002. *A guide to Ohio streams.* Columbus, Ohio: Streams Committee, Ohio Chapter of the American Fisheries Society. In partnership with the Ohio Environmental Education Fund, Ohio Department of Natural Resources, and the Ohio Environmental Protection Agency.

Vogelmann, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie and N. Van Driel. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat thematic mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing,* 67:650-662.