FOREST/NONFOREST CLASSIFICATION OF LANDSAT TM DATA FOR ANNUAL INVENTORY PHASE ONE STRATIFICATION[1]

Jim Rack[2]

Abstract—Launch of Landsat 7 creates the opportunity to use relatively inexpensive and regularly acquired land cover data as an alternative to high altitude aerial photography. Creating a forest/nonforest mask from satellite imagery may offer a cost-effective alternative to interpretation of aerial photography for Phase One stratification of annual inventory plots. This paper describes the procedures: they include image rectification, registration, and spatial filtering to allow accurate co-location with field plots and attempt to compensate for minor plot location errors. Identification of clouds and their removal from further analysis is outlined. Image alarms are described as a coarse filter for arriving at a forest/nonforest mask, with unsupervised classification as the fine filter. Accuracy assessment results for single-date, dual-date, filtered and unfiltered combinations are reported, as well as cost estimates.

INTRODUCTION

Phase One stratification for Forest Inventory and Analysis (FIA) survey purposes has traditionally been accomplished through detailed photo interpretation of the most current high altitude photography. Conversion from periodic to annual inventories necessitates acquisition of current imagery more often than high altitude photo programs can provide at traditional funding levels. Satellite image analysis offers a cost effective alternative.

One advantage of satellite imagery is the ability to machine-process large areas in a relatively short time. Another is that imagery is acquired on a regular cycle. Costs of imagery and analyses are lower. Disadvantages relate to accuracy: resolution is not as fine as aerial photography, and the human ability to interpret context, shape, and texture are lost.

The following methodologies were developed and tested in an attempt to apply well-established, simple techniques that could be quickly and easily implemented across a large program such as FIA.

IMAGE ACQUISITION AND PREPROCESSING

Midsummer imagery may offer the best opportunity for discriminating forest from nonforest if only a single date is used in the analysis. Addition of imagery from another season may improve accuracies by incorporating phenological differences.

Although the choice of sensors is expanding, driven mostly by a desire for increased spatial resolution, Landsat Thematic Mapper (TM) remains one of the better choices when considering classification needs of a forest target. Its spatial resolution is somewhat smaller than the size of an FIA plot, yet not so small as to overwhelm storage and processing capacities when dealing with large land areas. Its spectral resolution is greater than most other commercial systems and offers better classification potential.

Two dates of imagery were used in this study. The first was a July 24, 1999 Landsat 5 scene from Path 27, Row 26 shifted 70 percent south. The second was an October 13, 1999 leaf-off Landsat 7 scene from Path 26, Row 27. Both were rectified to the Minnesota Department of Natural Resources (DNR) standard of extended zone 15 UTM projection, NAD83 datum, using the MN Department of Transportation ARC/INFO roads coverage for ground control points.

The October scene needed some cloud removal. A Normalized Difference Cloud Index, (TM5 - TM6) / (TM5 + TM6) was calculated and added to the image as an additional band. An unsupervised classification of 70 classes was then created and cloud and cloud-shadow classes identified and masked out. Another unsupervised classification of 150 classes was run on the cloud-free images to identify obvious water and remove it from further analysis.

The remaining unmasked imagery contained forest and nonforest land pixels that can be roughly separated using an image alarm available in many types of image analysis software. The analyst roams the imagery digitizing a variety of coniferous and deciduous stands, trying to include the range of variability for each. I selected 30 of each and merged the 30 separate polygons into one coniferous and one deciduous signature. The image alarm allows the analyst to edit parallelepiped limits for bands 3, 4, and 5. The analyst interactively edits until the pixels alarmed reasonably represent the labeled class. If two dates of imagery are used, there are six bands to edit. The pixels identified with the alarm are separated into a preliminary “forest” area of interest (AOI) that will be further classified. The remaining pixels are identified as a preliminary “nonforest” AOI.

CLASSIFICATION REFINEMENT

Each of the AOIs are further classified using a 35-class unsupervised technique to build signatures, followed by a maximum likelihood supervised classification of the pixels.

[1] Paper presented at the Second Annual Forest Inventory and Anaylsis (FIA) Symposium, Salt Lake City, UT, October 17–18, 2000.

[2] Image Analyst, MN Department of Natural Resources, Division of Forestry—Resource Assessment, 413 SE 13th Street, Grand Rapids, MN 55744.

This classification will do a better job than the alarm of distinguishing fine differences between forest and nonforest. These 35 classes are labeled using 3 or more sets of the best available aerial photography. Photos should be well distributed across the image.

Some signatures from this classification will be confused and require further refinement. Each is identified as an AOI, and another unsupervised classification is performed using some number of subclasses, which are then relabeled using the photographs. Eventually, most ambiguity should be removed, and a model can be written to recode all pixels to either forest or nonforest classes.

Two spatial concerns related to FIA plots need to be addressed. One is accuracy of plot locations. A small test of plot location accuracy conducted in our office on 1990 era plots revealed a root mean square error of approximately 50 meters, even after plot locations had been extensively edited. The second refers to the difference in area between a pixel and an FIA plot cluster. One pixel is a square 30 meters on a side. It would take approximately 9 pixels in a 3 by 3 (3x3) matrix to cover the same area represented by the new 4 subplot cluster or the old 10 subplot cluster. To compensate somewhat for the inaccurate locations of the FIA plots and the difference in size between a pixel and a plot, a 3x3 majority filter was used to assign the majority makeup of a 9-pixel area to the center pixel. This 3x3 filter matrix passes over each pixel in the classification and outputs a new filtered classification. Results are reported for both the single-pixel classifications and the filtered classifications.

ADDITIONAL STEPS FOR TWO-DATE CLASSIFICATIONS

Using two dates of imagery requires some additional work in preprocessing. Registration between the two images must be checked: if pixel locations are not coincident, the images will require additional registration work. An overlap area must be identified so that analyses are restricted to pixels that contain data from all bands of both input images. Layers from the two images will have to be “stacked” to create one image for analysis, containing as many as 12 bands of data.

RESULTS AND DISCUSSION

Classifications were compared to 106 actual FIA plots field measured in 1998 and 1999. Four classification combinations were compared to various combinations of FIA ground land use (GLU) and history. Histories of “clearcut”, “natural significant disturbance”, and “man-caused significant disturbance” on forested plots were considered to be in a nonforested state for accuracy assessment purposes. Plots were also checked visually on the imagery to see if current condition matched attributes in the database: for example, if a forested plot was clearcut after the field visit and before the image date. An additional 44 nonforested plots were added to compensate for the small number of field-visited nonforest plots; these were identified by “PI_LAND_USE” and “GLU” codes of nonforest from annual inventory plots selected in 1994-1996. Only the overlap area of the two images was included in the accuracy assessment.

The single-date unfiltered classification accuracy assessment matrix, table 1, shows a certain bias on the part of the analyst towards an aggressive classification of “forest”. This bias offers the advantage of insuring that all or most of the actual forested plots are selected for field visits, but has a negative effect on estimates of forest area.

Results (table 2) from filtering the single-date classification indicate a slight improvement in overall accuracy from 85 percent to 88 percent. The filtering caused six plots classified as “forest” to change to a “nonforest” classification and one plot classified as “nonforest” to change to “forest”. Visual inspection of the plot locations changed by filtering confirmed the neighborhood of pixels to be a generally better representation of conditions at the plot than the single pixel at “plot center”.

Using a second image in the classification also improved the classification accuracy, as shown in table 3, from 85 percent to 88 percent. However, filtering this classification slightly reduced the accuracy, from 88 percent to 87 percent (table 4). In this case, filtering caused the classification of five plots to change from forested to nonforested and nine plots from nonforested to forested.

Visual inspection of the errors from these matrices indicates that plot location is a chief contributor. Plots near borders of

Table 1—Single-date classification, 7/24/99, unfiltered

	FIA forest	FIA nonforest	Total	User's accuracy
				Percent
Classified forest	81	23	104	78
Classified nonforest	0	46	46	100
Total	81	69	150	
				Overall acccuracy
Producer's accuracy (%)	100	67		85

Table 2—Single-date classification, 7/24/99, unfiltered with 3x3 majority filter

	FIA forest	FIA nonforest	Total	User's accuracy
				Percent
Classified forest	81	18	99	82
Classified nonforest	0	51	51	100
Total	81	69	150	
				Overall acccuracy
Producer's accuracy (%)	100	74		85

Table 3—Two-date classification, 7/24/99, unfiltered and 10/13/99, unfiltered

	FIA forest	FIA nonforest	Total	User's accuracy
				Percent
Classified forest	64	13	77	83
Classified nonforest	3	56	59	95
Total	67	69	136	
				Overall acccuracy
Producer's accuracy (%)	96	81		88

Table 4—Two-date classification, 7/24/99, unfiltered and 10/13/99, filtered with 3x3 majority filter

	FIA forest	FIA nonforest	Total	User's accuracy
				Percent
Classified forest	65	16	81	80
Classified nonforest	2	53	55	96
Total	67	69	136	
				Overall acccuracy
Producer's accuracy (%)	97	77		87

forest/nonforest conditions have a higher probability of being labeled incorrectly, especially if plot locations are imprecise. As plot locations are updated with high-accuracy Global Positioning Systems during field visits, classification accuracies should also improve. Other conditions that contributed to errors were plots labeled “marsh without trees” and “right of way,” which tended to be misclassified as forest.

The general conclusion one may draw from this small test is that simple image processing techniques of satellite imagery can offer almost 90 percent accuracies of forest/nonforest discrimination. Whether this is adequate for Phase One stratification of FIA, or repeatable on other landscapes, is yet to be determined.

COSTS

Image costs will vary depending on the source of the data. Landsat 7 scenes carry about a $600-$800 price dependent on the level of processing. The image analyses will take between 7 and 10 days per scene center depending on whether it is single date or dual date and the presence of clouds. Twenty scenes of Landsat 7 for Minnesota would cost about $13,000 and image analysis would add $40,000-$50,000. Round numbers would show a cost of about $53,000 for the 53 million acres in Minnesota or $0.001/acre ($0.64/square mile). These are estimates only and not to be considered universal for all users. Individual circumstances could alter these numbers substantially.