10-2005

Multi-Source Image Classification

Hillary Tribby

James Kroll
Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University, jkroll@sfasu.edu

Daniel Unger
Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University, unger@sfasu.edu

I-Kuai Hung
Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University, hungi@sfasu.edu

Hans Michael Williams
Arthur Temple College of Forestry and Agriculture Division of Stephen F. Austin State University, hwilliams@sfasu.edu

MULTI-SOURCE IMAGE CLASSIFICATION

Hillary Tribby, James Kroll, [1]Daniel Unger, I-Kuai Hung, Hans Williams

[1]Corresponding Author: Daniel Unger (unger@sfasu.edu)

Arthur Temple College of Forestry and Agriculture

Stephen F. Austin State University

Nacogdoches, Texas 75962

ABSTRACT

Since multi-source image classifications have the ability to exceed single source processes, such as traditional unsupervised classification methods, this paper will present the integration of four types of data: Lidar, elevation, multispectral and thermal. Using multi-source data and maximum likelihood classification methodology, as well as all possible permutations of data types, this paper will discuss ways to increase accuracy assessments of forested areas in east Texas and find the best combination of data sources.

Key Words: Land Cover, Accuracy, Lidar, Terrain, Canopy Elevation

INTRODUCTION

Land cover classification is a method to demarcate the landscape for the purpose of determining the dominant cover and monitoring the changes in land cover over time. Numerous schemes have been developed for use with land cover and land use classifications, but early in satellite remote sensing Anderson and his colleagues (1976) developed and published a rubric by which to define dominant cover.

Although important information can be extracted from a single satellite image, the inclusion of multiple sources and types of data may be best for analysis. This project assessed the utility of combining terrain and canopy elevation values derived from lidar (light detection and ranging) with traditional multispectral and thermal satellite imagery to increase the accuracy of land cover classifications.

OBJECTIVES

Since remote sensing technologies are constantly changing incorporating data into any given remote sensing project has become more complex. This project evaluated and recommends which data sources should be integrated into image classifications in order to produce the most accurate land cover map. These results will aid decision-makers in determining what data will be included and provide the most accurate information possible. Our primary objective was to evaluate the accuracy of single and multi-source image land cover classifications based on the hypotheses that: (1) the addition of lidar derived components will

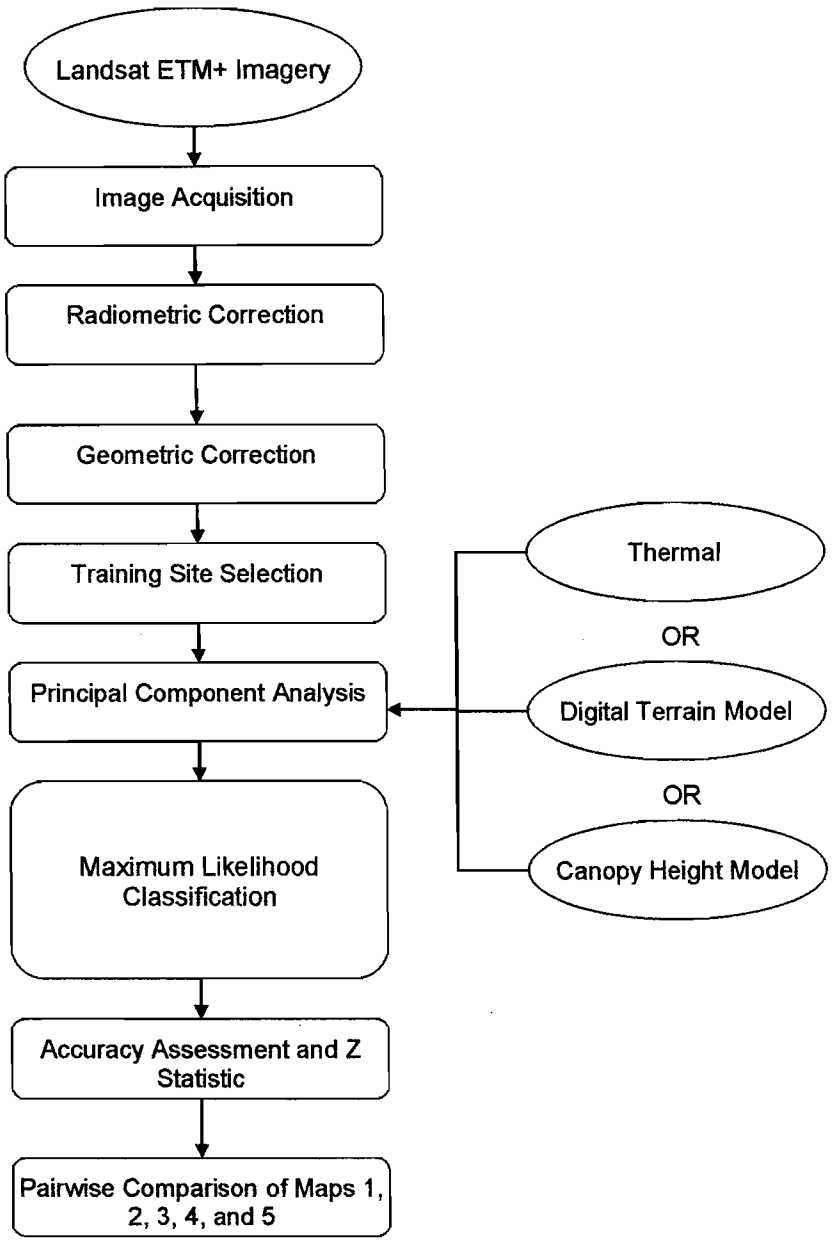
increase land cover map accuracy; and, (2) the addition of thermal data will increase land cover map accuracy.

METHODS

Using the inherent maximum likelihood (ML) algorithm in ENVI® (Environment for Visualizing Images), the Principal Components of a Landsat ETM+ scene were classified according to the Anderson Land Use Land Cover System (Table 1; Figure 1). Since lidar data were collected as vector data, it must be transformed into a continuous format. Moving windows that capture the highest elevation point for a given area were used to generate a surface elevation

Table 1. Eigenvalues associated with each of the Landsat ETM+ Principal Component bands.

Band	Eigenvalue	Percent Variability
1	877.5889	89.78
2	71.7970	7.34
3	20.6104	2.11
4	4.2251	0.43
5	1.8339	0.19
6	1.4560	0.15



Including multiple sources of data for integration into a classification requires data normalization and stacking.

Figure 1. Schematic of image classification methodology and accuracy assessment.

model (Popescu, Wynne & Nelson 2002). The digital terrain model was created using the second lidar return (Woolard & Colby, 2002; Lee & Shan, 2003; Popescu, 2004). The second return represented the laser signal hits that penetrated the canopy and returned from the ground. Using the Geostatistical Analyst extension in ArcGIS 9.0®, a digital terrain model was generated using Radial Basis Function. Numerous interpolation techniques were used and the RBF-Multiquadric had the lowest Root Mean Square Error. The terrain model was subtracted from the surface model to generate the Canopy Height Model.

The $\hat{K}$ statistic is a measure of agreement between the reference data and the classified land cover map and is calculated using the variables in the error matrix:

$$\hat{K} = \frac{n\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} x_{i+}x_{+i}}{n^2 - \sum_{i=1}^{k} x_{i+}x_{+i}}$$

Where k is the number of rows, x_{ii} is the number of observations in row i and column i, x_{i+} and x_{+i} are the marginal totals of row i and column i, and n is the number of observations (Congalton, 1991). To determine the Z statistic the variance must be calculated using the Delta method:

$$\hat{\text{var}}(\hat{K}) = \frac{1}{n}\left\{\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4}\right\}$$

where

$$\theta_1 = \frac{1}{n}\sum_{i=1}^{k} x_{ii}$$

$$\theta_2 = \frac{1}{n^2}\sum_{i=1}^{k} x_{i+}x_{+i}$$

$$\theta_3 = \frac{1}{n^2}\sum_{i=1}^{k} x_{ii}\left(x_{i+} + x_{+i}\right)$$

and

$$\theta_4 = \frac{1}{n^3}\sum_{i=1}^{k}\sum_{j=1}^{k} x_{ij}\left(x_{j+} + x_{+i}\right)^2.$$

Once variance has been calculated the Z statistic can be calculated:

$$Z = \frac{\hat{K}_1}{\sqrt{\hat{\text{var}}\left(\hat{K}_1\right)}}$$

The Z statistic informs the researcher whether or not the agreement between the reference data and the land cover map was greater than random chance. Since the Z statistic is standardized and normally distributed, with the null hypothesis that $\hat{K}$ is equal to zero and the alternative hypothesis that $\hat{K}$ is not equal to zero, if $Z \geq Z_{\alpha/2}$ then the null hypothesis is rejected. Z scores

with a value greater than one standard deviation from the mean or 1.960 is deemed significant. The Z statistic was also used to determine significant differences between two map accuracies:

$$Z = \frac{\left|\hat{K}_1 - \hat{K}_2\right|}{\sqrt{\hat{\text{var}}\left(\hat{K}_1\right) + \hat{\text{var}}\left(\hat{K}_2\right)}}$$

By calculating the $\hat{K}$, variance, and Z statistics for two error matrices, significant differences can be identified (Congalton & Green, 1999).

RESULTS

Of the five land cover maps, all map Z scores were deemed significant, indicating that the classifications were better than random chance (Table 2). Differences between the maps were not deemed to be significantly different, indicating that the addition of multiple sources of data did not effect the supervised classification of Landsat ETM+ data (Table 3).

Although the incorporation of multiple sources did not significantly effect the accuracy assessment, the inclusion of multiple data sources did produced the highest kappa statistic. Future studies should assess the accuracy of higher spatial resolution image classifications that include similar data.

Table 2. Land cover map results showing significant Z scores.

Land Cover Map	Overall Accuracy %	Kappa Coefficient	Variance	Z Score
Principal Component Analysis (PCA)	65.63	0.5816	0.001292	16.184
PCA and Canopy Height Model (CHM)	64.23	0.5664	0.003632	9.400
PCA and Digital Terrain Model (DTM)	69.30	0.6283	0.003112	11.262
PCA and Thermal Data	67.61	0.6050	0.001074	18.456
PCA, CHM, DTM, and Thermal	70.14	0.6384	0.004320	9.713

Table 3. Kappa analysis results describing the pairwise comparisons of error matrices.

Map_1	Map_2	Z Score
PCA	PCA and CHM	0.2167
PCA	PCA and DTM	0.7025
PCA	PCA and Thermal	0.4793
PCA	PCA, CHM, DTM, and Thermal	0.7577
PCA and CHM	PCA and DTM	0.7529
PCA and CHM	PCA and Thermal	0.5615
PCA and CHM	PCA, CHM, DTM, and Thermal	0.8070
PCA and DTM	PCA and Thermal	0.3602
PCA and DTM	PCA, CHM, DTM, and Thermal	0.1176
PCA and Thermal	PCA, CHM, DTM, and Thermal	0.4554

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T. & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data (Geological survey professional paper 964). Washington, DC: U.S. Government Printing Office.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Rem. Sens. Env., 37*, 35-46.

Congalton, R.G. & Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton, FL: Lewis Pubs.

Lee, D. S. & Shan, J. (2003). Combining lidar elevation data and IKONOS multispectral imagery for coastal classification mapping. *Marine Geodesy, 26,* 117-127.

Popescu, S. C., Wynne, R. H. & Nelson, R. F. (2002). Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Comp. & Electron. Ag., 37*, 71-95.

Popescu, S. C. (2004). Seeing the trees in the forest: Using small-footprint airborne lidar measurements. PhD. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA. 155 p.

Woolard, J. W. & Colby, J. D. (2002). Spatial characterization, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina. *Geomorphology, 48,* 269-287.