Assessment of Forest Cover Extraction Methods in PA

Project Report, 2017

Yasser Ayad, Ph.D., GISP
yayad@clarion.edu

Project funded by PennsylvaniaView, a full member of AmericaView

Clarion University of Pennsylvania

Assessment of Forest Cover Extraction Methods in PA

Yasser Ayad
Professor, GIS
yayad@clarion.edu

Abstract

With the introduction of the state-wide LiDAR data in PA in the early to mid 2000 many methods have been developed to extract land cover information using point cloud LAS files in conjunction with other high resolution remotely sensed data. This study suggests the comparison of different methodologies to extract forest coverage using LiDAR data as well as the NAIP multispectral images. The results will be evaluated using accuracy assessment techniques in order to extract the best available method as well as the optimum spatial resolution for forest coverage extraction using the freely available remotely sensed datasets for Pennsylvania.

Keywords: Land Classification, Forest Extraction, Tree Cover, Accuracy Assessment, GIS, LiDAR, NAIP, Multispectral

Contents

List of Figures

List of Tables

Introduction

The extraction of high-resolution forest cover from remotely sensed data has been a challenge. The smaller the cell size the more variations of land cover and spectral information are present which adds to the complexity of automated processes. This study introduces an assessment of two high-resolution dataset classifications. First by using LiDAR point cloud data to extract forest cover, which will be based on the class and return values of each point. And, second, using traditional multispectral image classification from the National Agricultural Imagery Program (NAIP). With the multispectral image classification two methods were applied, first a straightforward unsupervised classification using the available 4 bands, and second using Principal Component Analysis to isolate variations in those bands, then applying unsupervised classification. Furthermore, the assessment include testing the suggested variations of classification methods at different spatial resolutions.

Data

Clarion County in PA was selected as a study area for this project (Figure 1). The corresponding data used in this study are summarized in (*Table 1*).

Table 1. Collected datasets for the current project

Dataset	Source	Notes
LiDAR data	PAMAP (2004)	LiDAR point cloud data stored as LAS files
NAIP-Infrared	USDA (2010)	Multispectral (IR, R, G, & B) tiles. Available for download from PASDA

75 NAIP multispectral tiles as well as 202 LiDAR point cloud (LAS) tiles were collected to cover Clarion County. They were both stored as mosaic datasets for further analysis.

Methods

Two separate methodologies were applied on the available datasets. LiDAR data processing involved the identification and extraction of classes and returns pertaining to forest and tree covers, while NAIP was manipulated with common image classification techniques. Both datasets were classified at 8ft (2.4m), 12ft (4m) and 15ft (5m) cell sizes. In the following sections, each methodology is described in more details.

LiDAR Data Processing

Observing the different returns from the LAS dataset revealed that both classes 1 and 12 contain returns that represent forest/tree coverage. With a closer look at each, all returns from class 12 could safely be representative of forest/tree coverage. Returns from class 1, on the other hand, did not include only trees. Some structures and rooftops were mainly included in the first returns. The first and last of many, as well as returns 2, 3 and 4 were the only ones that included other form of larger vegetation forms such as trees and shrubs. Figure 2 shows both classes, 1 (red) and 12 (green), this specific sample includes all returns from both classes. Figure 3 shows the designated classes and returns in an area where structures and building rooftops existed, they show in red and they are mixed with other returns from class 1 that are designated as mid to high vegetation as well. Figure 4, on the other hand, shows both classes, but with only the returns from class 1 that represent medium to high vegetation. Rooftops and other structures were omitted.

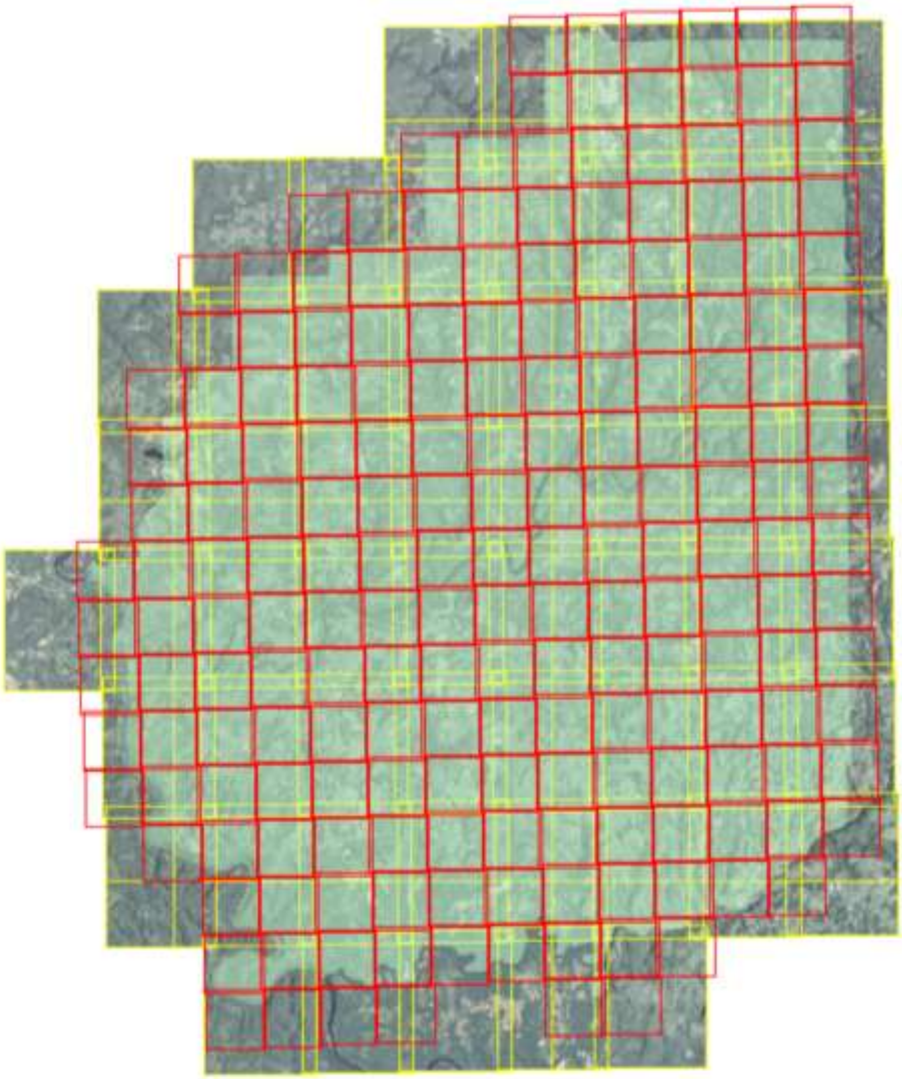

Figure 1. Clarion County boundaries (shown in transparent green). LiDAR point cloud tiles are shown in red, and the NAIP multispectral aerial photo tiles are shown in yellow.

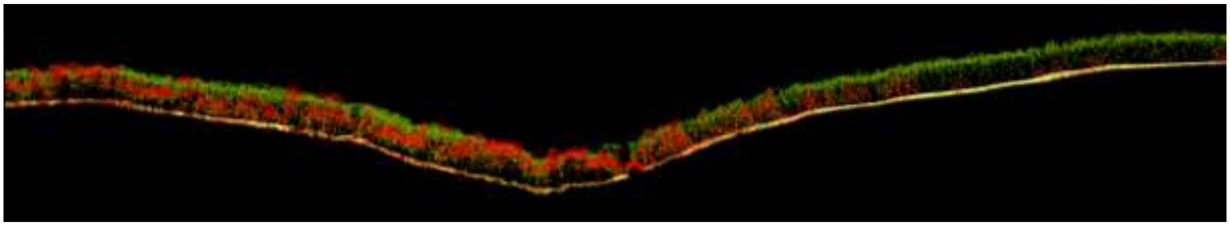

Figure 2. Sample LAS profile for a forested area. Green points are class 12 (overlap/reserved), while red ones are class 1 (unassigned)

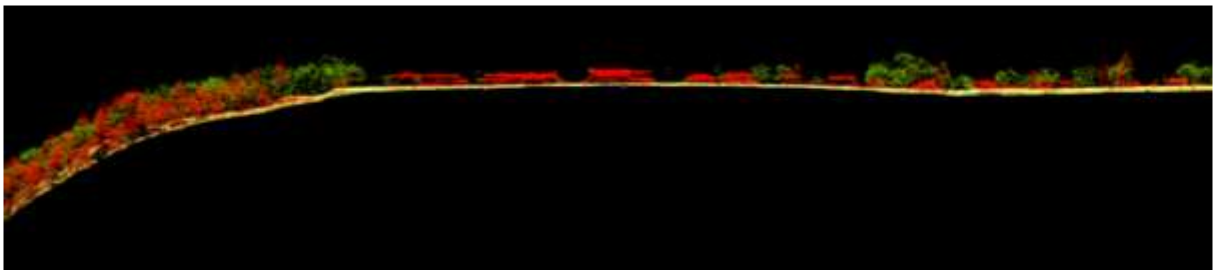

Figure 3. Sample LAS profile depicting class 12 (green) as well as mixed class 1 (red). Building rooftops as well as some mid to high vegetation covers are both shown in red.

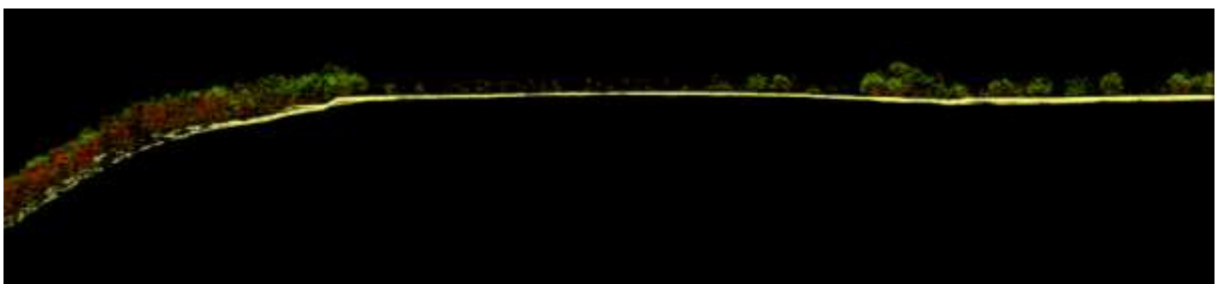

Figure 4. Sample LAS Profile depicting class 12 (green) as well as the selected returns for class 1 (red) that represent vegetation. Building rooftops are omitted.

Those classes and their corresponding returns that mostly represent forest/tree covers were rasterized and classified as forest (1) using the forest extraction model (Figure 5).

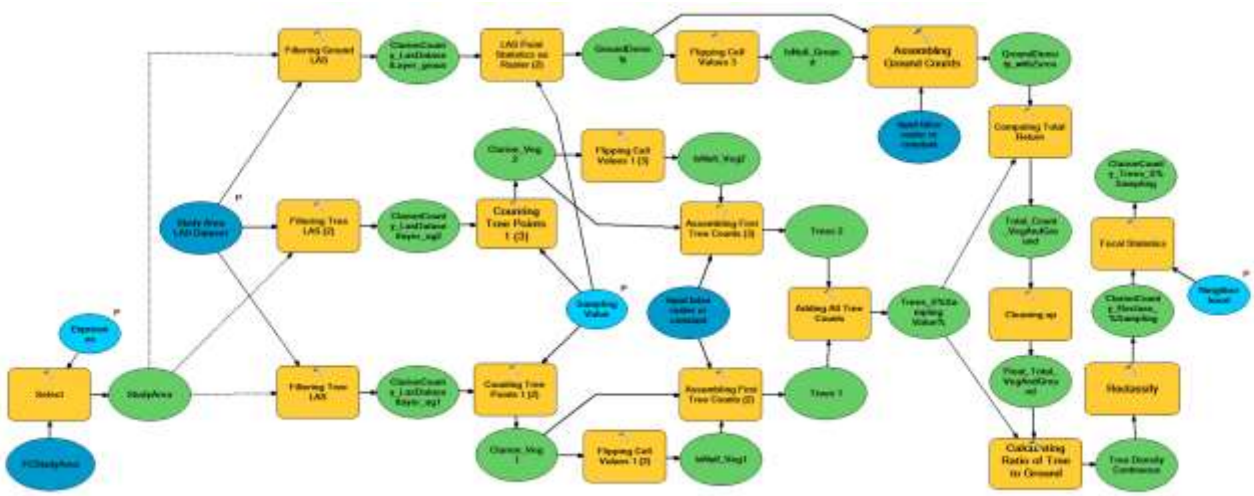

Figure 5. LiDAR Forest Extraction Model

The model filters the LAS dataset for the ground as well as the designated forest/tree cloud points using the previously mentioned class and return numbers. It then calculates the percent tree cover within the designated cell size (8, 12, and 15 ft), and uses the maximum tree density for forested areas identification. A final smoothening of the classified dataset is carried out in order to remove additional noise from the final result.

NAIP Data Processing

The initial objective of this study is to produce an automated classification method. And since the NAIP multispectral photos were available, ISO clustering unsupervised classification was employed for each of the LiDAR comparable cell sizes (8ft (2.4m), 12ft (4m), and 15ft (5m)). This was carried out in two ways: A direct, straightforward ISO unsupervised classification of the multiband aerial photo, and an ISO unsupervised classification of the derived principal components of the same set of bands.

Direct Unsupervised Classification

ISO clustering technique was applied to the 4-band NAIP aerial photographs for each of the designated cell sizes (Figure 6). The result was very rough and, by visually assessing it, it was obvious that the committed misclassified pixels were significant all over the study area. Therefore, other techniques were explored, including working with Vegetation Indices (VI) as well as performing a Principal Component Analysis (PCA) on the available bands. The VI methodology did not yield a visually different separation than the ISO clustering technique, while the PCA method revealed good extraction of forest versus non-forested areas.

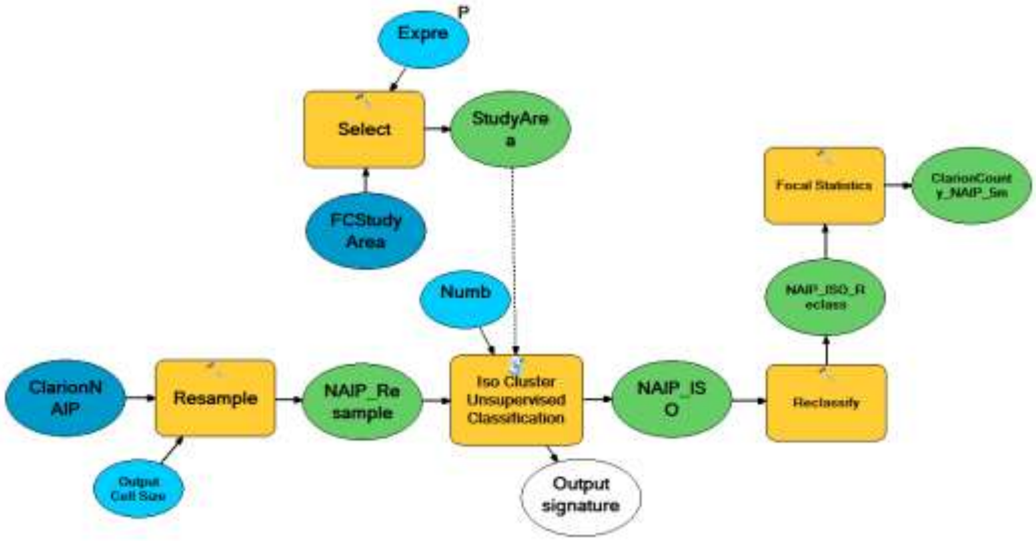

Figure 6. ISO Clustering Model used for the NAIP data classification

Unsupervised Classification of Derived Principal Components

The PCA methodology calculated 4 outputs that were then used as inputs for an unsupervised ISO clustering method (Figure 7). The significant separation of forest versus non-forested areas was obvious, and revealed a better overall classification than the ISO clustering applied to the raw multispectral photos.

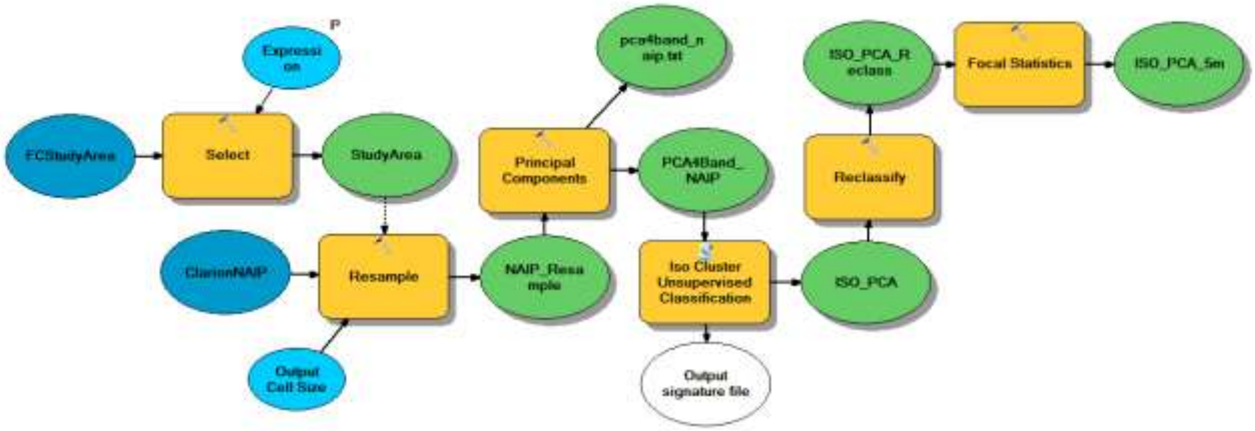

Figure 7. Principal Component Analysis and ISO clustering Model used for the NAIP data classification

Accuracy Assessment

One hundred (100) random points were generated to cover the study area. The ground truth values for each of those points were extracted from the latest NAIP aerial photographs (notice the 6 years time difference between the LiDAR (2004) and the NAIP imagery (2010)). The classified values from the resulting 9 classified images; three cell sizes (8, 12 and 15 ft) and three classification methods (LiDAR, ISO NAIP, and PCA ISO NAIP) were then cross tabulated against the extracted ground truth. The resulting transition matrices revealed that the LiDAR classification showed an overall higher accuracy than the NAIP classifications (Figure 8). It showed an accuracy improvement with larger cell size (15ft) at 89%, while the ISO classification showed a steady and clear decrease in accuracy with increasing cell sizes (79% at 8ft, 75% at 12ft and 74% at 15 ft), while the PCA ISO classification accuracy was almost unaffected by the change in cell size.

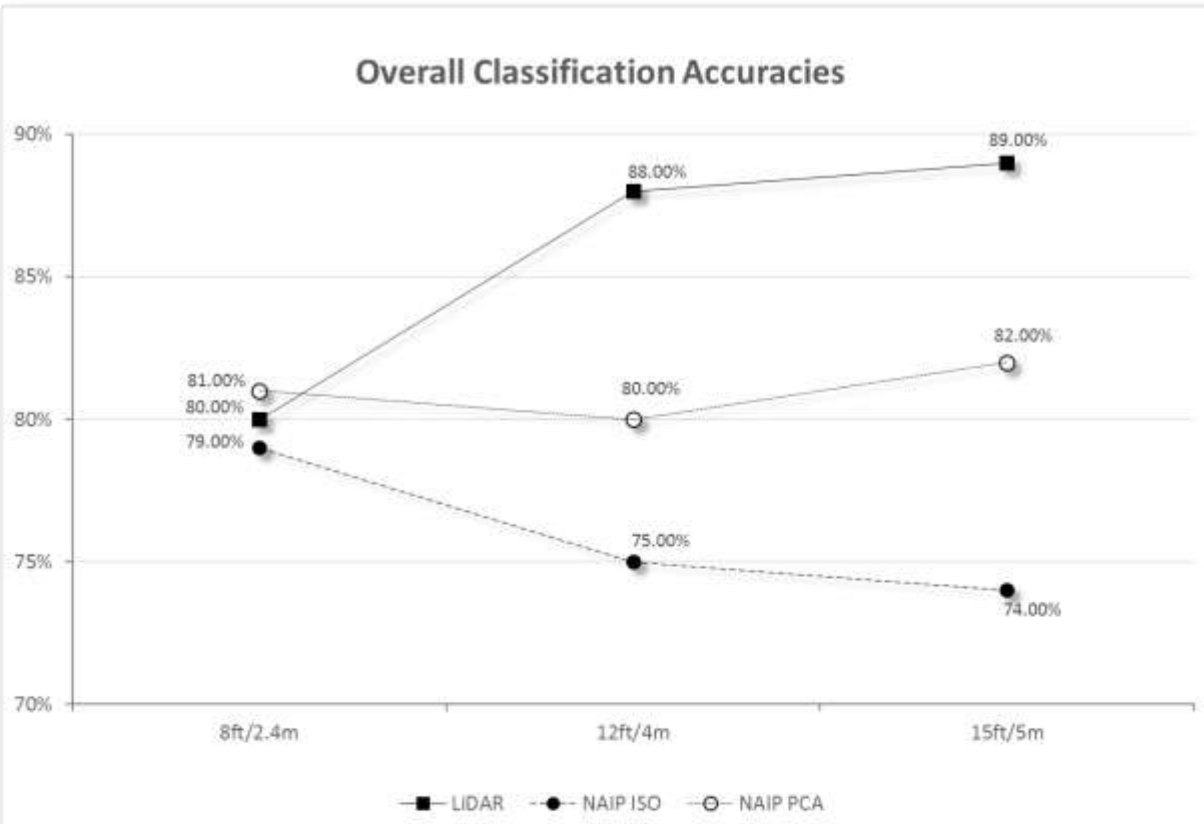

Figure 8. Overall classification accuracy assessment

Results and Discussion

Nine forest cover maps were created (Figure 9, Figure 10, and Figure 11). The differences between the different results might be visually similar but some variations prevail. It is important to note that by looking closely at the results, it becomes clear that the ISO classification of the multispectral NAIP aerial photographs tend to commit more non-forested areas (open fields, grassland, and other vegetation-covered surfaces) into the designated forest class. Which significantly increased the capability, and accuracy percentage, of the classified forested cells, but, on the other hand, introduced a significantly low accuracy for the non-forested/no tree cover areas.

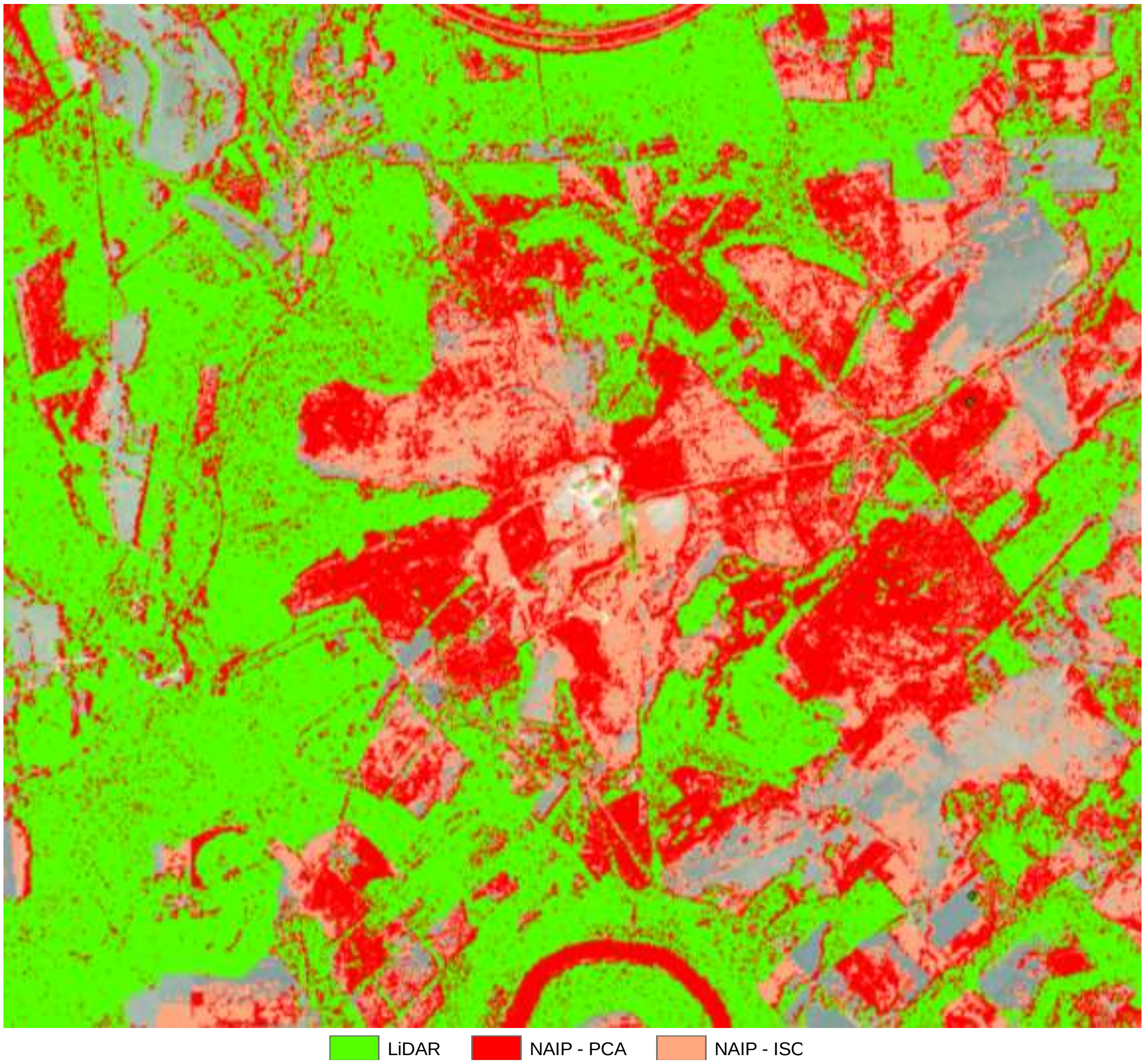

Figure 9.Identified forested/tree covered areas at 8ft/2.4m cell size

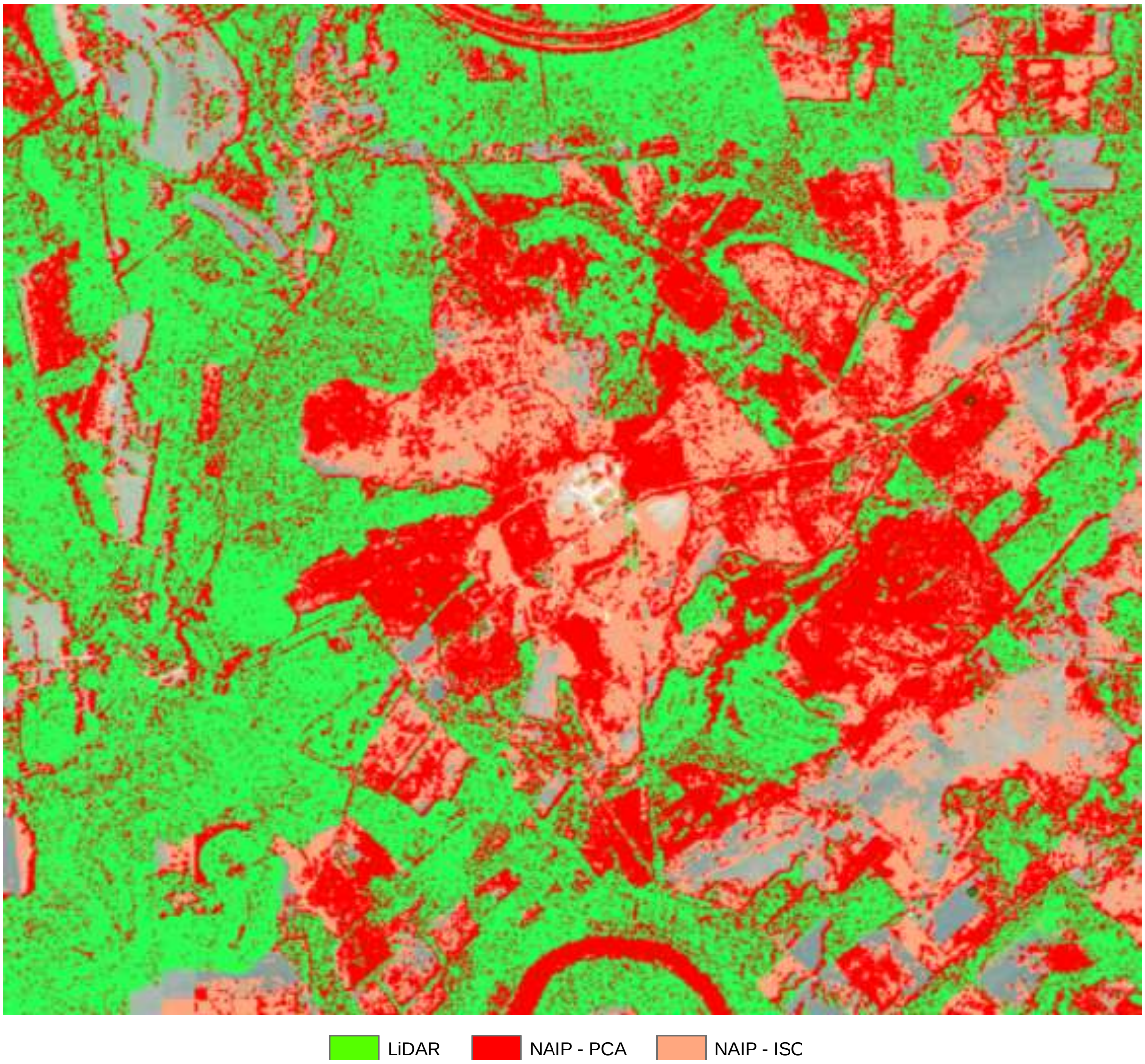

Figure 10. Identified forested/tree covered areas at 12ft/4m cell size

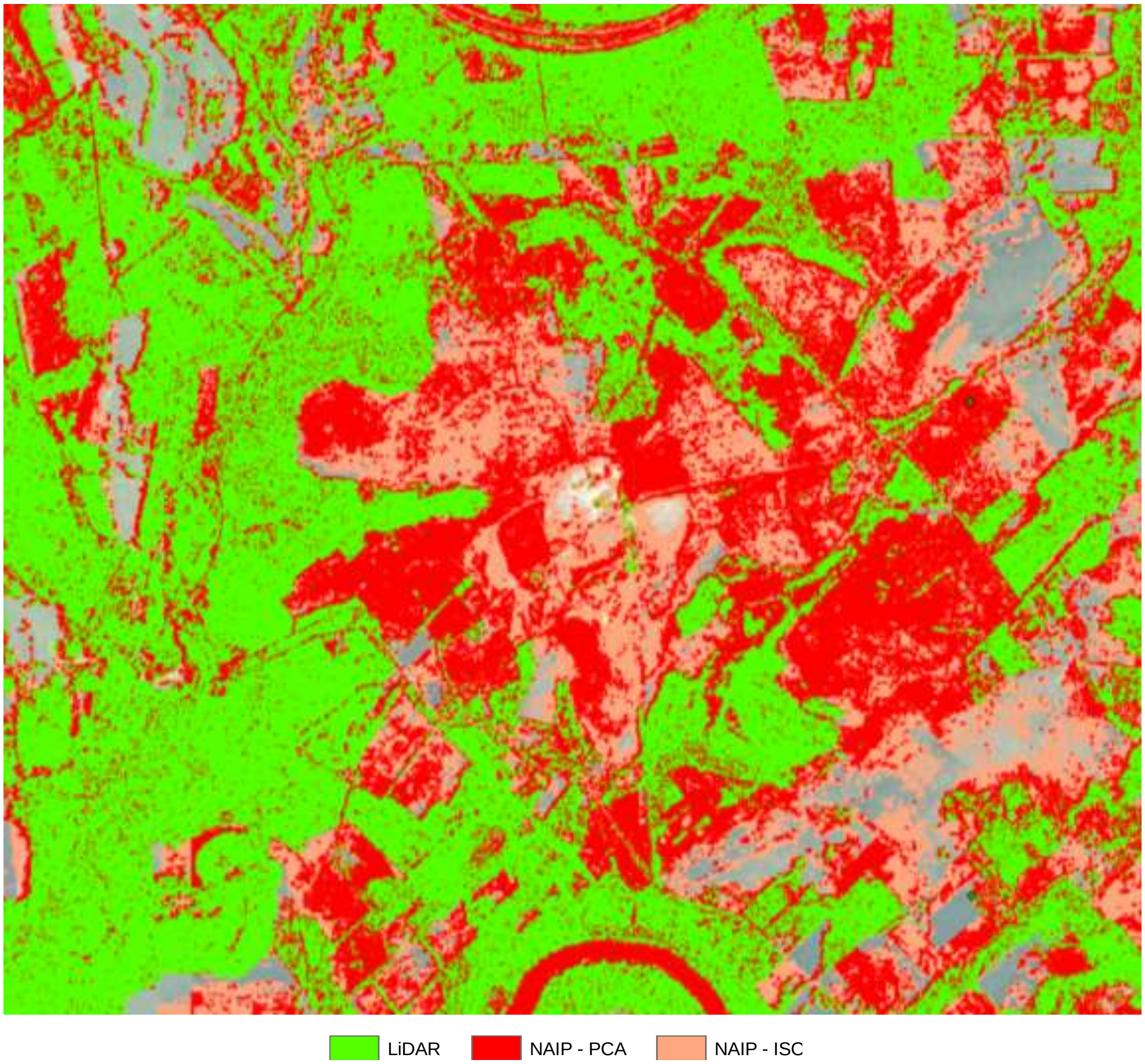

Figure 11. Identified forested/tree covered areas at 15ft/5m cell size

By taking a closer look at the resulting transition matrices (Table 2), although the overall accuracy of the LiDAR data (A, D, and G) is improved with increased cell size, but the committed non-forested accuracy decreased from 93.94% to 87.88% in favor of a forest classification improvement from 73.13% at 8ft, 85.07% at 12ft, to 89.55% at 15ft.

It is also noted that although the overall accuracy of both methods of the NAIP image classification is relatively acceptable (between 74% and 82%), but the committed non-forested cell classification accuracy is pretty low (21.21% to 60.61%) (B, C, E, F, H, and I). This suggests that the multispectral nature of the NAIP imagery can introduce some ambiguity in the identification of forest coverage and significantly commits other classes (mostly grassland and open fields).

Table 2. Transition matrices of the resulting 9 forest cover maps; three classification methods were tested at each of the given cell sizes (8ft/2.4m, 12ft/4m, and 15ft/5m).

8ft/2.4m — A

Ground Truth \ LiDAR (8ft)	0	1	Total	
0	31	2	33	93.94%
1	18	49	67	73.13%
Total	49	51	100	↑
	63.27%	96.08%	←	Correctly Classified
Overall Accuracy	80.00%			

12ft/4m — D

Ground Truth \ LiDAR (12ft)	0	1	Total	
0	31	2	33	93.94%
1	10	57	67	85.07%
Total	41	59	100	↑
	75.61%	96.61%	←	Correctly Classified
Overall Accuracy	88.00%			

15ft/5m — G

Ground Truth \ LiDAR (15ft)	0	1	Total	
0	29	4	33	87.88%
1	7	60	67	89.55%
Total	36	64	100	↑
	80.56%	93.75%	←	Correctly Classified
Overall Accuracy	89.00%			

8ft/2.4m — B

Ground Truth \ NAIP (2.4m)	0	1	Total	
0	12	21	33	36.36%
1	0	67	67	100.00%
Total	12	88	100	↑
	100.00%	76.14%	←	Correctly Classified
Overall Accuracy	79.00%			

12ft/4m — E

Ground Truth \ NAIP (4m)	0	1	Total	
0	8	25	33	24.24%
1	0	67	67	100.00%
Total	8	92	100	↑
	100.00%	72.83%	←	Correctly Classified
Overall Accuracy	75.00%			

15ft/5m — H

Ground Truth \ NAIP (5m)	0	1	Total	
0	7	26	33	21.21%
1	0	67	67	100.00%
Total	7	93	100	↑
	100.00%	72.04%	←	Correctly Classified
Overall Accuracy	74.00%			

8ft/2.4m — C

Ground Truth \ NAIP PCA (2.4m)	0	1	Total	
0	20	13	33	60.61%
1	6	61	67	91.04%
Total	26	74	100	↑
	76.92%	82.43%	←	Correctly Classified
Overall Accuracy	81.00%			

12ft/4m — F

Ground Truth \ NAIP PCA (4m)	0	1	Total	
0	18	15	33	54.55%
1	5	62	67	92.54%
Total	23	77	100	↑
	78.26%	80.52%	←	Correctly Classified
Overall Accuracy	80.00%			

15ft/5m — I

Ground Truth \ NAIP PCA (5m)	0	1	Total	
0	18	15	33	54.55%
1	3	64	67	95.52%
Total	21	79	100	↑
	85.71%	81.01%	←	Correctly Classified
Overall Accuracy	82.00%			

Conclusions

With the high-resolution nature of the available spatial data for Pennsylvania, the goal of reaching better land cover classifications in general and more accurate forest cover identification in specific is possible. However, the spectral variability, shadows and other interfering factors might cause some challenges in achieving the best results. In this study, two different datasets were used in order to test the accuracy of the extraction of forest coverage using different classification methods at a variety of high-resolution cell sizes.

The LiDAR point cloud dataset showed the most promising result overall. The nature of the data and its resolution enabled the extraction of features based on the signal return from an active light sensor. Which enables the identification of features based on not only their elevation, but also the intensity of the return signal. Different cover classes as well as return levels can be used to determine the cover types at a relative high resolution (about 6-10ft cell size area at most).

On the other hand, the National Agricultural Imagery Program multispectral (IR, R, G, and B) aerial photos are provided on a regular basis from the US Department of Agriculture for most of the US on a regular basis. They are provided at 1m cell size (about 3ft).

In this study, unsupervised classification, as well as a hybrid Principal Component Analysis coupled with an unsupervised classification technique were applied to the NAIP imagery. The Principal Component Analysis has significantly improved the identification of forest cover, and it is suggested that, for future work, it could be improved upon by adding more classes to the ISO clustering, taking advantage of the PCA separation, and exploring the variations of resulting classification. Nevertheless, even with the current improvement, the LiDAR LAS data manipulation provided an overall higher accuracy for both omitted as well as committed cell classification.