Pine-Hardwood classification with LiDAR-derived canopy data and multispectral image for mixed forest in Huntsville, TX

Xiangmin Sun, Ryan Sheridan, Sorin C. Popescu

Ecosystem Science and Management Department, Texas A&M University, College Station, Texas, USA.

Abstract: Remote sensing has innovated and reshaped many traditional forestry investigation methods. Light detection and ranging (LiDAR) and high resolution imagery has been combined and utilized to extract many forest inventory data. Height bins (voxel) method to extract height points information within different height range from LiDAR point cloud is a promising tool for tree species identification. We utilized canopy density for every cell within different height bins and general height statistic parameters from LiDAR and National Agricultural Imagery Program (NAIP) imagery to differentiate pine and deciduous in a mixed forest in southeastern U.S.. The Principal Component Analysis (PCA) method was utilized to find the correlation among these bands, and then decision tree method was applied to do classification based on PCA bands. The PCA result shows no significant correlation between these bands, and the first Principal Component (PC) and first 5 PCs accounted for about 50% and 80% of the total variance, so the classification accuracy based on these PCs is relatively satisfactory. The differentiation between pine and deciduous, however, was weakened by sparse distribution of deciduous and loose correlation between the original bands, but this less correlation could prompt more advanced choice of LiDAR and spectral information to differentiate pine and deciduous based on further physical investigation. So there is merit in including variables within height-bins for tree species classification.

1. Introduction:

The obtaining of reliable and precise forest inventory information has been always fundamental and crucial for forest resources assessment and management, but was historically rather expensive, labor consuming and lasted very long time due to aerial photo interpretation, field investigation, and scaling problems from forest plot result to landscape. Many traditional studies on tree classification are carried out by human visual interpretation using aerial photos. The use of remote sensing could not only provide several traditional attributes of forest inventory, but also provide information that not currently part of an existing forest inventory, such as Leaf Area Index (LAI). The core of current remote sensing application in forest inventory is to apply an empirical classification or estimation technique, including establishing an empirical relationship between spectral measurements and field estimates of species or canopy density[1].

Tree species identification constitutes a bottleneck in remote sensing-based forest inventory, due to overlap and bidirectional reflectance during differentiating features [2]. The introduction of airborne LiDAR was a breakthrough, since it permits accurate three-dimensional (3D) probing of the vegetation and terrain, which cannot be done reliably in passive RS data [2]. As such, different species of trees will reflect LiDAR pulses differently, so the individual crown shapes allows for identification of predominant tree species. LiDAR return intensity, which has not been as heavily explored for its usefulness in forestry due to the difficulty in radiometric calibration, could merit feature extraction on imagery[3]. The use of LiDAR height and intensity data to classify land cover via an object-oriented approach could result an accurate land cover maps with accuracies ranging from 94% to 98% [4]. Examination on the distribution of the laser points within the contour of a single tree could distinguish different species of tree. Figure1 shows sample point distributions from three different tree species, so by analyzing the number of variables within the point distributions can identify the tree individual tree species.[5] Species tree height and NDVI have shown to be the two most important LiDAR derived parameters for classification of savanna tree species[6]. A proposed method was to capture the geometrical difference between deciduous and coniferous tree by looking the geometrical properties of the crown shape (spherical, conical or cylindrical) based on LiDAR point cloud, and derived the internal structures (bole and branches) of the LiDAR tree further [7].

LiDAR and high resolution aerial photography have been investigated as means to extract forest data, such as biomass, timber volume, and stand dynamics. Data fusion with multispectral optical data and local filtering with variable window size was used to estimate tree height, and the integration with co-registered multi- and hyper-spectral digital imagery makes LiDAR a realistic precision forestry alternative to traditional measurement in forest inventory, and the fusion of LiDAR and image could bring dramatic gains in characterizing the 3D structure of the forest canopy[8]. A novel calibration technique for estimating individual tree heights with a more reliable prediction based on shape characteristics of a marginal height distribution of the whole first-return point cloud of each tree showed a reduction of the RMSE of the tree heights of about 20%, so improved the species classification accuracy markedly under leaf-off and leaf-on conditions[9]. A parcel-based unsupervised classification approach was employed with the first two Principal Components from 12 selected wavebands of HyMap data and a Digital Canopy Height Model extracted from LiDAR data, and the resultant thematic classes contain information

on species composition and structure, thus an ecologically meaningful thematic product for complex woodland environment[10]. A voxel-based LiDAR method for estimating crown base height for deciduous and pine trees was carried out in southeastern United States, which illustrates that height bins, generating images of the vertical structure of forest vegetation, are an innovative LiDAR-derived product that has the potential to become a standardized imagery product for LiDAR applications in ecosystem studies [11]. The intensity and the pulse width derived from small-footprint full waveform are used to detect coniferous and deciduous trees by an unsupervised classification, and an overall accuracy of 85% in a leaf-on situation and 96% in a leaf-off situation.[12]

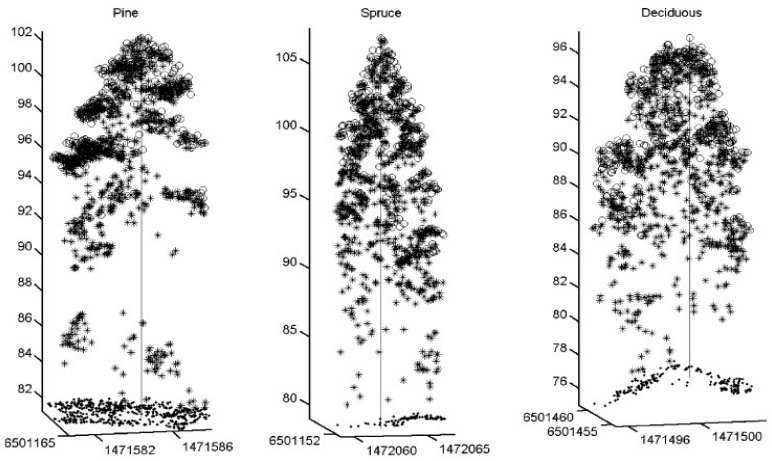


Figure 1: sample LiDAR point distributions for different tree species [5]

The objective of this study was to differentiate and classify the pine and deciduous with combined multispectral images and LiDAR-derived voxel-based images of canopy structure, including canopy density and height statistical indexes in a mixed pine-deciduous forest in southeastern United States. We hypothesized that the canopy density information and height statistical parameters available in LiDAR-derived data, combined with vegetation spectral and textual information, could be utilized to differentiate pine and deciduous part of forest. Of interest was four specific aspects listed below:

- To examine LiDAR-derived canopy parameters for pine and deciduous in Huntsville
- To develop a frame to fuse LiDAR data with multispectral imagery for tree classification
- To develop classification parameters and method for mixed pine-hardwood forest
- To examine the capabilities of LiDAR data in combination with multispectral imagery to distinguish pines from hardwoods

2. Material and Methods

2.1 study area

This study area is located near Huntsville, East Texas, Southern United States(Fig.2), with a rectangular region defined by 95°24'57"W-30°39'36"N and 95°21'33"W-30°44'12"N. According to the Vegetation Types of Texas (1984), the vegetation type of the study area is Pine-Hardwood Forest.[13] The study area consists of Loblolly pine (Pinus taeda L.) in various developmental stages, including young, mature, and withered or dead Loblolly pine stands (red tone in fig.2) in the Sam Houston National Forest, and upland and bottomland hardwoods are comprised by Water Oak (Quercus nigra L.), Southern Red Oak (Quercus falcata Michx.), White Oak (Quercus alba L.), Sweetgum (Liquidambar styraciflua L.) and Winged Elm (Ulmus alata Michx.). Loblolly is widely planted in this region due to its economic value, while deciduous trees are also widely distributed, especially in several valleys. The topographic characteristic of the study area is gentle slopes with elevation fluctuating between 62 to 105 m.

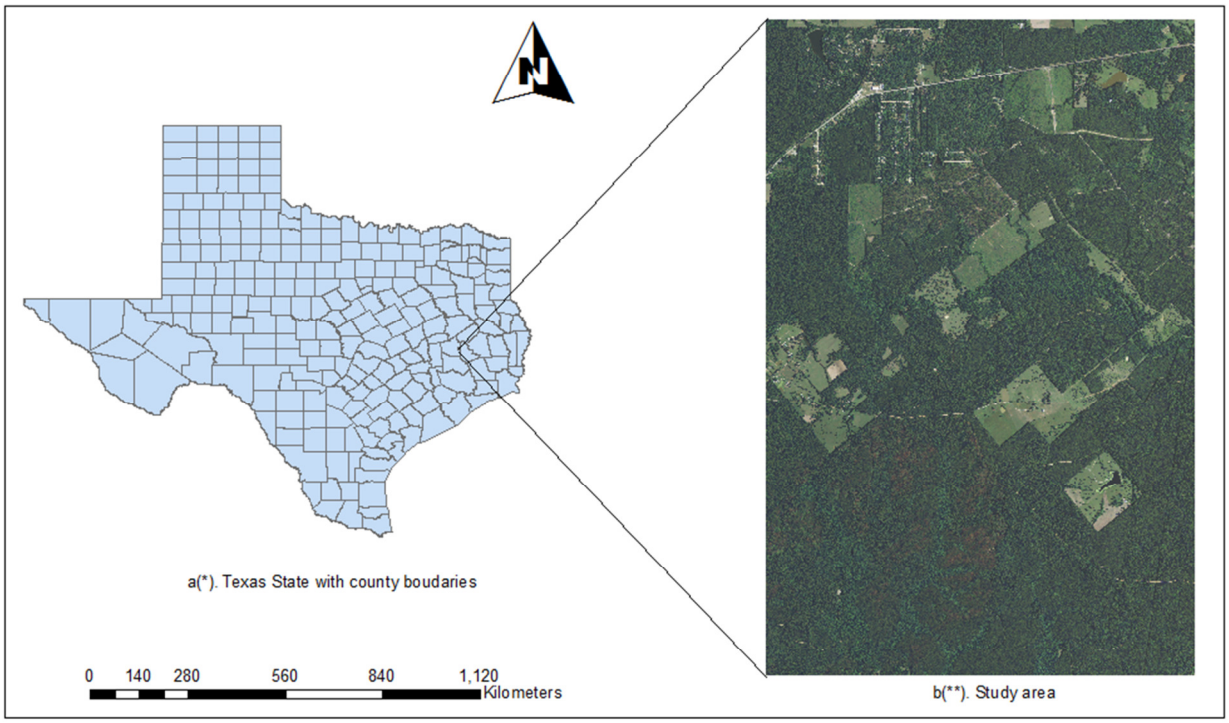


Fig.2 Study area located in Huntsville, TX. shown as a true color composite of national agricultural imagery program (NAIP)

(*)StratMap County Boundaries from TNRIS

(**) the study area NAIP image shown in true color composite (Red for band 1, green for band 2, and blue for band 3)

2.2 Data

2.2.1 Airborne LiDAR

Discrete return LiDAR data were acquired in early November 2010 by Terrapoint Inc. (acquired by GeoDigital Inc. in June, 2011), with 506,502,099 point records, and the number of points by return was

374 204 482, 112 766 004, 18 324 897, 1 177 768, and 28 692 from first return to the 6th return, respectively. The projection system is GT-Model-Type-Geo-Key UTM15-Northern hemisphere, and the unit system is linear meter. The scanning angle range for this LiDAR Dataset is ±29°. The point file was provided without classification.

2.2.2 Multispectral imagery

This dataset included four National Agriculture Imagery Program (NAIP) orthophotos (Fig.1) from the Digital Ortho Quadrangle (DOQQ) Datasets of Texas Natural Resources Information System (TNRIS), which were acquired in May of 2010, the same year with LiDAR acquisition year. These NAIP images have a 1 m spatial resolution with four spectral bans, namely, Red (604-664 nm), green (533-587 nm), blue (420-492 nm) and Near-infrared(NIR) band (833-920 nm).

2.3 Data processing

The processing of LiDAR and NAIP images consist of several steps. For initial LiDAR data, classification on canopy and ground was made and the Canopy Height Model (CHM) was derived, and then to extract canopy density and height distribution information for different height bins form CHM with 2 m resolution. The Normalized Difference Vegetation Index (NDVI) and Texture information from NAIP images were also extracted by band math tool of ENVI (Version: 5.0.3, Exelis Visual Information Solutions, Inc.), and then resized to same 2 m to LiDAR-derived raster file. And then stacked all these bands derived from LiDAR and NAIP together with to obtain uncorrelated principal components (PCs) and the different weights the original bands carries in the overall variability though Principle Component Analysis (PCA), and lastly decision tree (DT) classification method and support vector machine (SVM) were applied to do classification, with assist of further Principal Component Analysis (PCA) for representative pine-hardwood sub-region to extract particular information to differentiate pine and deciduous. The overall flow chart is shown in Fig.3.

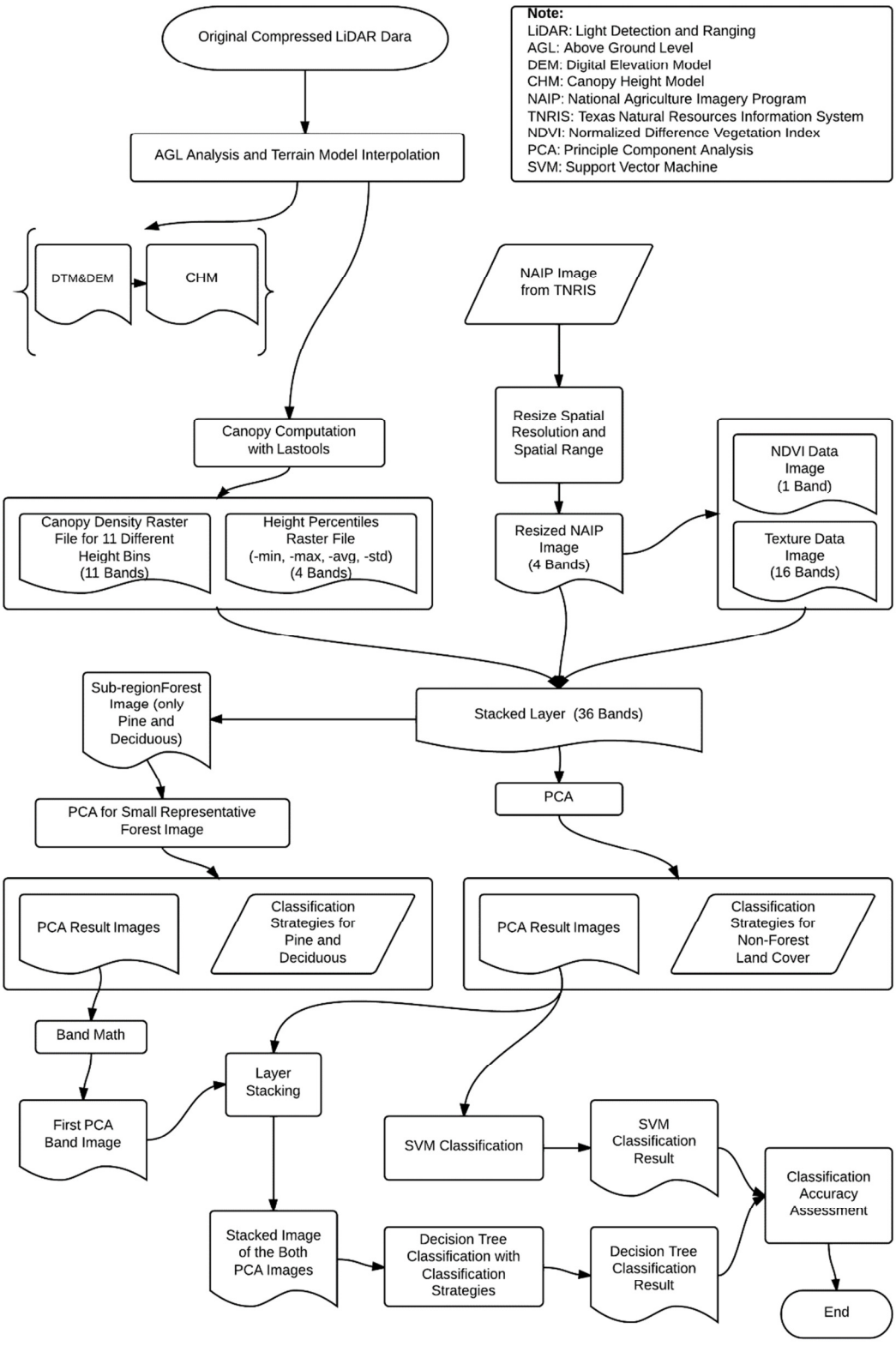


Fig.3 Flow chart of the overall pine-deciduous tree classification

2.3.1 Canopy Height Model (CHM) mining from LiDAR data

A digital terrain model (DTM) was generated by inferring the last return (i.e. 6th) LiDAR point elevation from vegetation and returns from bare soil with 5 m spatial resolution using the AGL analysis by QTM (V7.1.5, Applied Imagery Inc.), and then smoothed further by noise removal. As such, the CHM was generated by subtracting the resampled DEM from original point cloud to generate AGL (above ground level) at a resolution of 5 m via subtract tool of Quick Terrain Modeler. The resultant AGL point cloud, which symbolizes a three-dimensional point clouds characterizing above ground height, especially trees, across the landscape within the study area, is available for following canopy analysis.

2.3.2 Canopy structure information derived from AGL point cloud

Forest canopy density and height are two essential variables in a number of environmental applications, including the estimation of biomass[14], canopy fuel[15], and biodiversity[16]. Accordingly, two main aspects of canopy structure information were obtained from AGL, which were canopy density within each cell of different height bins and statistical index of the canopy height distribution. Relative density raster files were computed by 'Lascanopy' tool of Lastools (rapidlasso GmbH Inc.) which divided height count in every cell of each height bin by the total number of points and scaled to a percentage. 2 m was chose for the 'step' or spatial resolution with consideration of the point density for every gridded unit area. 10% buffer boundary was applied during this process to confirm analysis accuracy. According to the distribution of height points and previous studies the height bins consisted of 11 different intervals, as shown in Tab.1 and Fig.4.

Tab. 1 Height-bins information

height-bins	Elevation range (m)
Bin 1	0.0-0.5
Bin 2	0.5-1.0
Bin 3	1.0-1.5
Bin 4	1.5-2.0
Bin 5	2.0-5.0
Bin 6	5.0-10.0
Bin 7	10.0-15.0
Bin 8	15.0-20.0
Bin 9	20.0-25.0
Bin 10	25.0-30.0
Bin 11	30.0-70.0

Meanwhile, several common statistical values, such as minimum, maximum, average, and standard deviation of all height above the ground ('-height_cut_off 0.0') were also computed by 'Lascanopy' tool of Lastools. Due to different characteristics of point height distribution near the ground surface of different trees (Fig.1), the height cutoff was rearranged from default breast height (1.37m) to ground surface 0.0 m, so all points were computed for these four height parameters. Meanwhile, since the height bin model is more easily interpreted than the height percentile model[17], so some other height

percentiles were not computed rather. The CHM is relatively much more important since it carries canopy height information, which is an important parameter for tree classification[18]. Those four output raster file are shown in Fig.5.

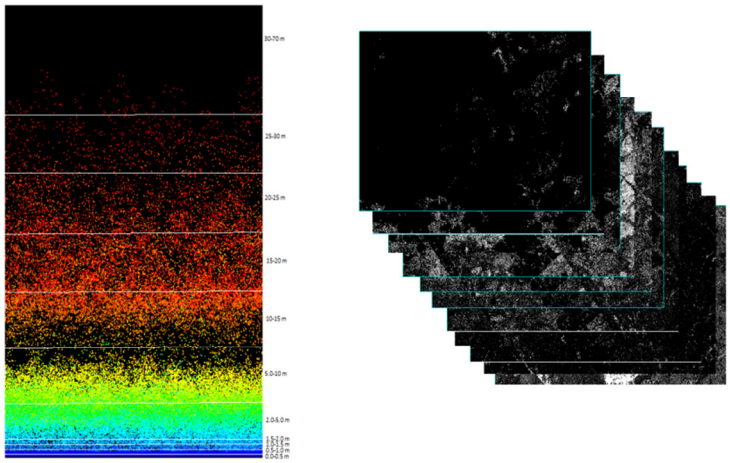


Fig. 4 height bins and its output density raster files

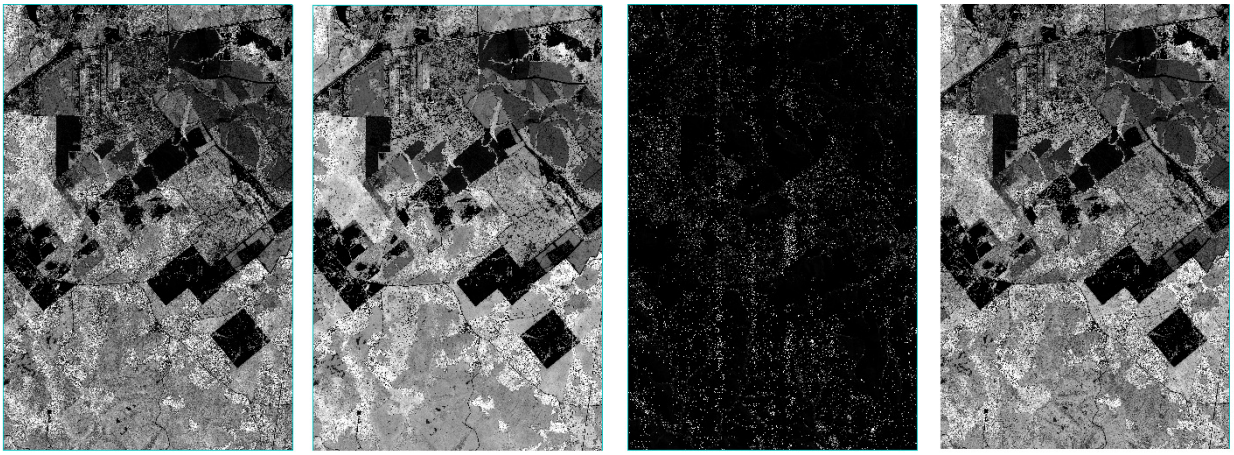


Fig.5 percentile images

2.3.3 NAIP image processing

Four NAIP images from TNRIS were mosaicked together geo-referentially in the beginning. An inspection of a few noticeable feature points were checked to confirm the well geographically registry to LiDAR derived raster files. These NAIP images were used to extract thematic information to distinguish different land cover types in this study, including forest (pines and deciduous), grassland, urban/bare soil, water, etc. A visual interpretation on these NAIP images, especially with a false color composite, showed relatively significant spectral difference among these different land covers.

The mosaicked NAIP images was resized with spatial resolution and extent according to raster files from AGL file mentioned above, and then used for computation of normalized difference vegetation index (NDVI) and texture based on its different four bands, namely red, green, blue, and NIR. NDVI carries vegetation information[19]. The image texture information is a quantification of the spatial variation of image tone that linked to changes in the spatial distribution of forest vegetation in both vertical and horizontal dimensions[20].

Co-occurrence-based texture filters was applied to the NAIP image to derive variation in brightness in those four bands, and four options were chosen in this study, which is variance, homogeneity, contrast, and entropy. As a result, 16 raster images were available after these four parameters[21] were computed for four bands.

- Variance: The local variance of the processing window. This value is based on the Greyscale Quantization Level that you specify.

$$f_4 = \sum_i \sum_j (i-u)^2 \, p(i,j)$$

- Homogeneity: ENVI computes homogeneity using the "inverse difference moment" equation. Values range from 0 to 1.0.

$$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} \, p(i,j)$$

- Contrast: ENVI computes contrast using the following equation:

$$f_2 = \sum_{n=0}^{N_g - 1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{\substack{j=1 \\ |i-j|=n}}^{N_g} p(i,j) \right\}$$

- Entropy: ENVI computes entropy using the following equation. Values range from 0 to the alog of the processing window size.

$$f_9 = -\sum_i \sum_j p(i,j) \log(p(i,j))$$

Where p(i,j) is the spatial co-occurrence matrix element.

2.3.4 Stacking layers derived from LiDAR and NAIP

With ENVI 5.0.3 (Exelis Visual Information Solutions, Inc.), a new multiband image with 2.0 meter spatial resolution was stacked based on 11 bands of canopy density image extracted from LiDAR derived CHM, 4 original bands, 1 NDVI band, and 16 bands of texture information resample and derived from NAIP, as is shown in Fig. 6, and will be subsequently referred as the LiDAR-NAIP stack. The first 11 bands were LiDAR derived height bins, and 12th-15th bands consisted of original NAIP bands, and then NDVI was added as 16th band, and 4 statistical bands were ranging from 17th to 20th band, and then 16 texture bands were added further to 36th band. We included all these bands to

convey as much information carried by LiDAR and NAIP image as possible to find which combination of bands implies most majority information of variety behind this stack image.

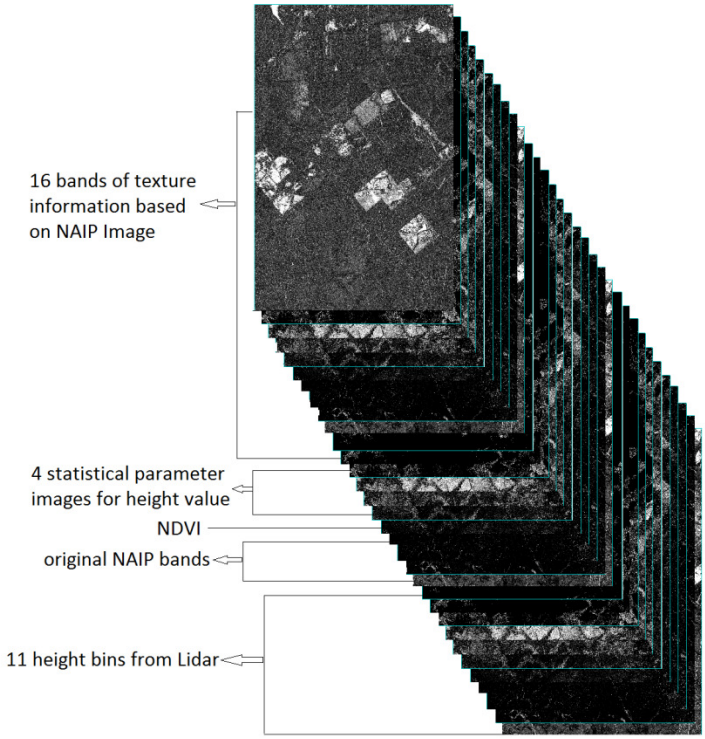
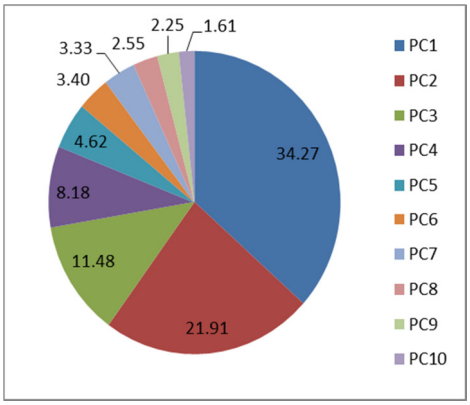


Fig.6 the LiDAR-NAIP stack image

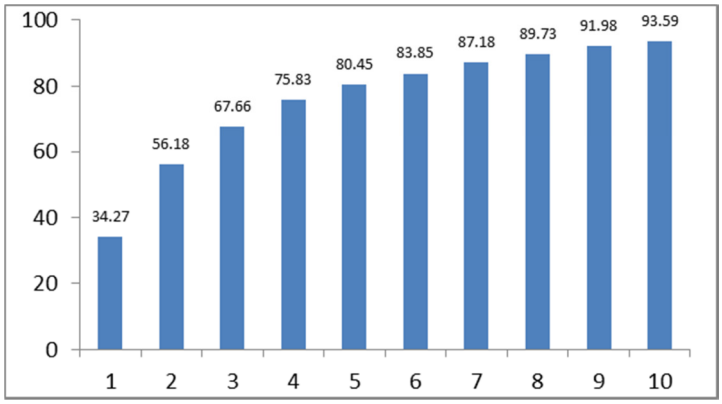
2.3.5 Principal Component Analysis (PCA)

A. first PCA:

PCA was applied to the LiDAR-NAIP stack image consisting of 36 bands, and then the first 10 of the output PCs is shown in Tab.2. The first principal component only accounts for only 34.27% of variance in the entire LiDAR and multispectral data, and the first 5 PCs comprise just approximately 80% of the total variance, and only the first 10 PCs can explain more than 90% variance, so it could be concluded that the correlation between LiDAR-derived canopy density and NAIP derived texture bands are much uncorrelated and their relation is not so strong,



(a) eigenvalue percent of the first 10 PCs



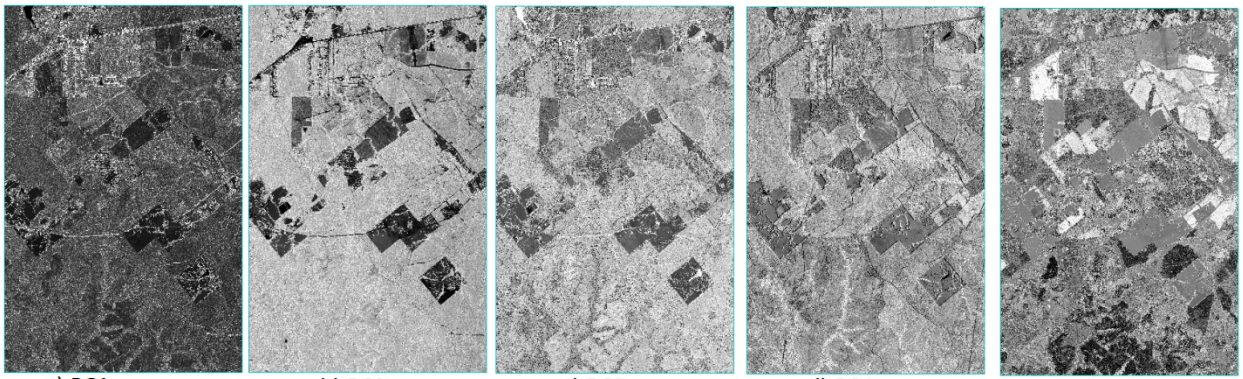
(b) cumulative percent of the first 10 PCs

Tab.2 the eigenvalue percent for first 10 PCs of the first PCA analysis

Upon further examination on the relationship between the PCs and original Bands, as shown in Tab.3, several significant uncorrelated bands could be detected by its correlation ratio with PCs. Band 4 (height-bin 1.5-2 m) is most strongly correlate with PC1 negatively, and generally the band 2,3,9, 14, 17, and 24 illustrated more strongly correlated with the first 3 PCs. And the first five PC images is shown in fig.7.

Tab. 3 the correlation between first 3 PCs and original bands

Eigenvector	PC 1	PC 2	PC 3	Note
Band 1	0.032897	0.00816	0.00287	
Band 2	-0.39901	-0.00567	-0.00236	height_bin 0.5_1m
Band 3	-0.3084	0.003389	-0.0072	height_bin 1_1.5m
Band 4	-0.6521	0.009803	-0.00422	height_bin 1.5_2m
Band 5	-0.09228	0.013354	0.012404	
Band 6	0.072979	-0.02188	-0.01957	
Band 7	-0.10273	0.024693	0.017078	
Band 8	-0.05961	0.01123	-0.00359	
Band 9	-0.33208	0.065159	-0.00529	height_bin 20_25m
Band 10	0.085635	-0.02956	-0.02894	
Band 11	-0.17239	0.159612	0.103063	
Band 12	-0.02399	0.023671	0.001208	
Band 13	-0.02581	0.0454	0.004732	
Band 14	0.138317	-0.79483	-0.19676	naip_blue
Band 15	0.014123	0.059887	-0.01011	
Band 16	0.056428	0.20749	-0.18558	
Band 17	-0.12519	-0.40217	0.598278	height_avg
Band 18	-0.06401	-0.14427	0.229266	
Band 19	-0.05951	-0.06155	-0.19935	
Band 20	0.107837	0.096266	0.356074	
Band 21	-0.01801	-0.00735	-0.46973	
Band 22	-0.03717	-0.04007	0.135989	
Band 23	0.006277	0.002547	0.004881	
Band 24	-0.21502	-0.2164	-0.22301	entropy_red



a) PC1 b) PC2 c) PC3 d) PC4 e) PC5

Fig.7 first five PC images

B. feature separateness within PCs

Region of Interest (ROIs) for every land cover were collected with reference to NAIP image (with False color composite) and Google Earth display across its entire range, as is shown in Fig.8. And 'stats' tool was run for every type of ROI to check its separateness for first 5 PCs, and the result (Tab.4) shows the mean and standard deviation (std) for digital number (DN)'s distribution of every land cover in each PCs.

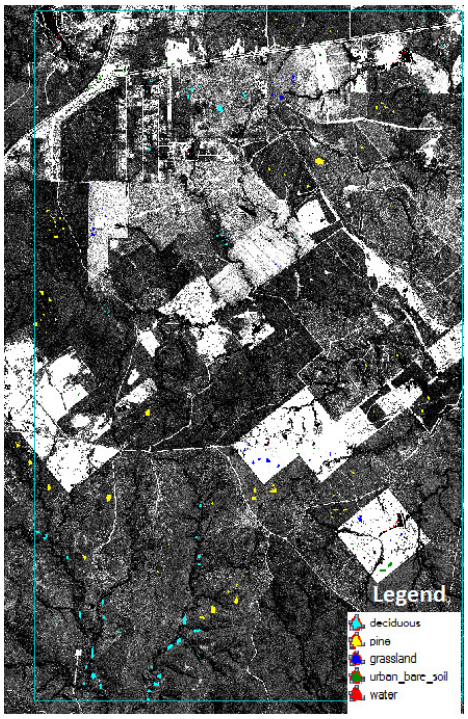


Fig.8 Region of Interest (ROI) is shown in Study Image

In this project, we roughly define the distribution range of DN for each Land cover as the mean ± std to include 68% of the DN values assuming that the DN distribution is normal. So from Tab.4, we can conclude that the first five bands cannot differentiate pine-hardwood very well, although other types of land cover could be distinguished relatively well via these PCs of the first PCA analysis, so a subset representative region of pine-hardwood forest was chosen for second PCA analysis.

Tab.4 Separateness check for every land cover in first 5 PCs of the first PCA

PCs	PC1		PC2		PC3		PC4		PC5	
	mean	std	mean	std	mean	std	mean	std	mean	std
water	-82.65	4.52	1.85	33.77	38.62	20.82	-51.08	28.78	8.29	6.94
urban/bare_soil	-14.75	115.69	-143.45	41.65	-33.36	66.30	-12.72	34.58	-10.31	10.22
pine	-19.18	34.79	15.10	22.29	13.01	19.16	3.55	21.36	-7.54	31.78
grassland	-32.75	45.73	-59.42	36.52	-36.26	15.47	-11.86	21.47	1.77	6.20
deciduous	12.53	54.33	16.94	35.49	-14.07	27.91	22.29	30.75	-0.53	21.27

Tab5. Separateness check for pine and deciduous in first 5 PCs of the 2nd PCA

PCs	Decideous		Pine	
	Mean	Stdev	Mean	Stdev
PC1	-23.20	59.58	28.81	29.76
PC2	11.93	39.28	-13.93	21.15
PC3	-3.58	23.41	19.49	19.21
PC4	5.58	20.74	4.95	14.69
PC5	3.81	21.97	-4.81	10.69

C. Second PCA for a subset representative image with Pine-Hardwood

The second PCA applied exclusively for a subset region with only two types of trees, pine and deciduous to check its separateness by the LiDAR- and NAIP- derived images. The subset region and its ROIs are shown in Fig.9, and its eigenvalues and its percentage of total variance are shown in Tab.6, and the first 5 PCs image are shown in Fig.10. The first PC still counts for only 42% of the total variance, and first 5 PCs just stand for approximately 80% of the total variance, which means no significant correlation exists among the canopy density and height statistical index and texture information derived from NAIP. Meanwhile, this 2nd PCA result are relatively similar to the 1st PCA.

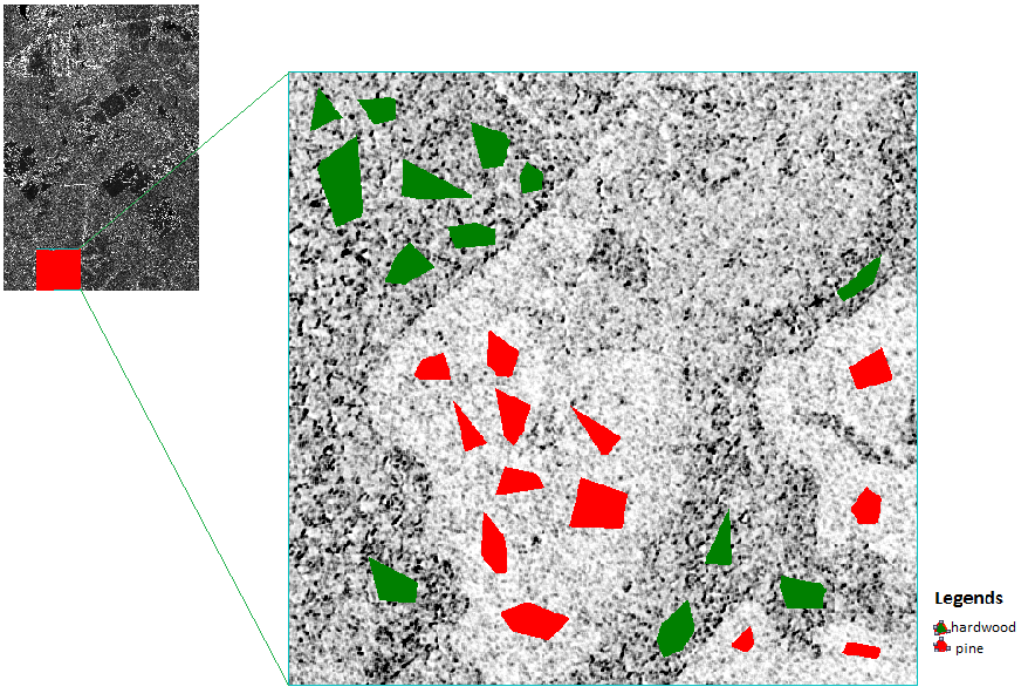


Fig.9 Subset representative image and its ROIs for 2nd PCA

Tab.6 Eigenvalues, and (cumulative) percentage of total variance

PCs	Eigenvalue	% of Total Variance	Cumulative
PC1	2893.110507	42.07491555	42.0749156
PC2	1194.30672	17.36897165	59.4438872
PC3	743.117434	10.80726201	70.2511492
PC4	430.000532	6.253558593	76.5047078
PC5	341.452213	4.965787858	81.4704957

a) PC1 2) PC2 3)PC3 4)PC4 5)PC5

Fig.10 First Five PC images for 2nd PCA

2.3.6 Stacking the first PCA bands with the PC1 band of the second PCA output

With combination of the two PCA analyses, the PC1 of the second PCA was stacked to the first PCA result to differentiate pine and hardwoods. This stacked layer will be referred subsequently as the composite PCA image with 37 bands since the 36 PCs from the 1st PCA result didn't removed. The 1st PC's linear relation with the original bands was extracted and computed with the same linear formula for the entire range with 'band math' tool, and then stacked to the first PCA resultant bands.

2.3.7 Classification_decision tree classification

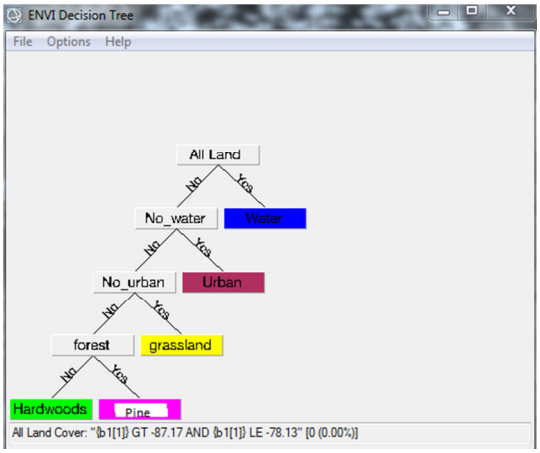


Fig.11 decision tree classification based on two PCA bands

As mentioned above, we assume the DN distribution for every ROI is same across the image, and then decision tree was used to classify different land cover based on its DN range defined as Mean± Std, so the Water range was -82.65±4.52 in PC1, Urban range was -143.45±41.65, and Grassland classified was -59.42±36.52. So relatively all these three land cover could separate by this range for different PCs. But the Pine and hardwood still could not differentiate completely by Mean± Std in the 1st PC of the 2nd PCA, so we divided them at the middle point between their Mean values in PC1, namely (-23.2+28.81)/2=2.8. The detail step of decision tree is shown in Fig.11. The thematic map from decision tree classification is shown in Fig.12. The classification accuracy was assessed by testing data collected based on visual interpretation on NAIP image and Google Earth display.

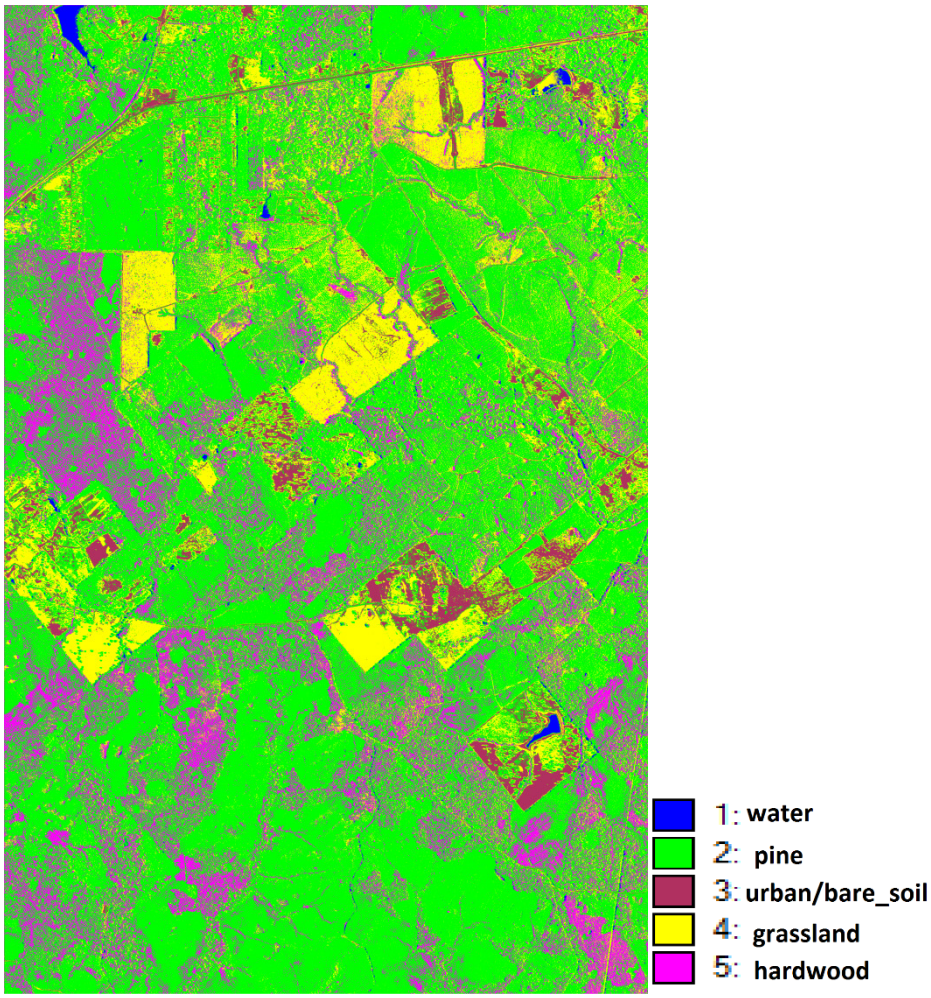


Fig. 12 Classification of decision tree

3. Results and discussion

3.1 PCA result

For the two PCA results, no strong correlation existed among these height bins, height statistic index, NAIP bands, and the texture information, because the first PC only accounts less than 50% of the total variance, and the height bins are more related to 1st PC for two PCA analysis, although a little difference existed for weights carried by height bins between two PCAs. The relation between PCs and original band is shown in Tab.7.

Tab.7 the relationship between PCs and original bands for 2nd PCA

Eigenvector	PC 1	PC 2	PC 3	note
Band 1	0.022607	0.006655	0.005433	
Band 2	0.027878	0.003202	0.000976	
Band 3	0.116155	0.006886	0.00647	height_bins_1_1.5m
Band 4	0.228929	0.00707	0.00508	height_bins_1.5_2m
Band 5	0.317313	0.011828	0.002739	height_bins_2_5m
Band 6	0.719801	0.025131	0.005325	height_bins_5_10m

The uncorrelated bands have different weights for PCs, the height-bins are much more strongly related to the 1st PC for two PCA result. Band 4, height-bin 1.5-2.0 m is the top significant band for 1st PCA result, while the third significant in 2nd PCA, so the influence from grassland and bare soil might been deteriorate the un-correlation between these bands, and the height-bins near the surface exhibited no big contribution to the PCs maybe due to its high homogeneity across the entire range, or due to point data of grassland or shrub. So the near ground surface height bin could not differentiate pine and deciduous very well. On the contrast, the height bins ranging from 0.5-10 m contributed major uncorrelation to these bands. Among NAIP bands, blue band was much more significant related with PC1 for 1st PCA result than other bands, the reason of which is unknown and needed to be further investigated. And among height statistic index, the average is much pronounced than the Maximum (CHM), so we may concluded that the maximum height of pine and deciduous are less different than the average height between them. For the texture bands, they may generally correlate with each other, and neither contributed to PCs significantly, except the limited contribution by entropy of red band.

3.2 Decision tree classification accuracy assessment

The classification based on the assumption of mean ± std range shows an overall accuracy of 71.4% (K_{hat}=0.60), the thematic map result is relatively satisfied due to the loose division assumption and small percentage of eigenvalue for these three PCs used for classification. The confusion matrix (%) for accuracy assessment is show in Tab.8, and the producer's and user's accuracy is listed in Tab.9.

Tab.8 confusion matrix for classification

Class	Ground Truth (%)				
	water	urban	grassland	pine	deciduous
water	73.12	0	0	0.19	0
urban	0.54	77.63	1.91	0	0
grassland	22.58	4.61	90.45	4.12	12.44
pine	3.76	17.11	3.18	85.96	69.78
deciduous	0	0.66	4.46	9.74	17.78
total	100	100	100	100	100

Tab.9 producer's and user's accuracy

class	Prod.Acc	User. Acc
water	73.12	99.27
urban	77.63	96.72
grassland	90.45	58.92
pine	85.96	70.18
deciduous	17.78	40

Based on visual inspection on the thematic map result in Fig.12, several observations could be made upon regarding the products. firstly, the separation between pine and deciduous isn't so satisfied enough, and 69.78% of deciduous reference points were classified as pine, so the producer's accuracy is only 17.78%, and user's accuracy is resultantly only 40% for deciduous. A significant misclassification for deciduous trees happened in an up north region with relatively sparse deciduous trees distribution, which was almost classified as pines. With dense deciduous trees distributed in lower valley region in southern part, the classification region was dramatically shrunk to narrow strips compared to reference data. This poor separation between deciduous and coniferous may be a result of intensity at different height bins, overall height distribution, and spatial texture heterogeneity. The producer's accuracy for pine is 85.96%, which is a very satisfactory number, and it means 85.96% of possibility that an "pine" on the ground was identified as a "pine" region on the classified image, so we can concluded that the pine's distribution range on the 1st PC of the second PCA is much convoluted than normal distribution, while the deciduous tree's counterpart distribution would be much loose due to its density range, so the misclassification is much more larger. Secondly, a problem associated with image classification at this spatial scale is shadowing of canopy, and we observed that tree shadows were often classified as water bodies, potentially overestimating the water range extent, or an independent class of shadow would be used in classification process. Thirdly, the classification between grassland and water was poor, might due to its small range of height points near ground surface, so their difference between different heights bins above 0.5 meters are very difficult to detect. And the height bins near the ground surface cannot differentiate the pine and deciduous very well, or even for all land covers in this project.

4. Conclusions

Although there was not significant increase in map accuracy when LiDAR-derived variables, such as canopy density for height bins and overall height statistic indexes, were considered and fusion with NAIP images and its texture information, there was merit in including these data to investigate their correlation between them via PCA. The canopy density ranging from 0.5 to 2.0 m accounts for most un-correlation between all these bands, which could be further used for pine-deciduous separation in mixed forest. The average index of canopy distribution is much more strongly related with PCs than CHM, which is slightly different with earlier findings[22]. The texture information and NDVI were not contributed much to the total variation, and could not differentiate pine and deciduous very well based on them. Since the first PCs or first 5 PCs accounts for a small part if the total variation, so PCA analysis isn't applied in this situation, so the tree decision tree classification based on loose assumption didn't provide a very satisfactory result for the classification result. The thematic result map for land cover failed to differentiate pine and deciduous very efficient due to the loose assumption and the wide range of variance for pine and deciduous. It should be noted that the result of this analysis only applies for this particular combination band of height bins, height statistic index, NAIP, and texture information from it. We simply compared initial outputs using these

variable data inputs, so it may still be possible to increase thematic map accuracy by implementing other classification methods rather than PCA, or even possible to improve decision tree approached based on PCA.

Future research should attempt thematic map creation from LiDAR intensity data and other derived information from high resolution imagery using stacked layers extracted from them. Meanwhile, the inclusion of LiDAR parameters and image, including height bins and texture information, should be explored for their usefulness individually and based on field survey before stacking them together to do analysis. The LiDAR data and image should be collected in same season, hopefully in leaf-off time to maximize the difference of canopy structure between pine and deciduous. Although this work support the importance of height bins in differentiate the pine and deciduous, high resolution imagery would still wanted to help increase thematic map accuracy. Other classification approaches should also be explored to examine their usefulness in differentiation of pine and deciduous.

References:

1. Franklin, S.E., et al., *Incorporating texture into classification of forest species composition from airborne multispectral images.* International Journal of Remote Sensing, 2000. **21**(1): p. 61-79.
2. Korpela, I., et al., *Tree Species Classification Using Airborne LiDAR - Effects of Stand and Tree Parameters, Downsizing of Training Set, Intensity Normalization, and Sensor Type.* Silva Fennica, 2010. **44**(2): p. 319-339.
3. Maxwell, A.E., A.C. Riley, and P. Kinder. *Comparison of LiDAR-derived data and high resolution true color imagery for extracting urban forest cover*. in *Proceedings of the 18th Central Hardwoods Forest Conference GTR-NRS-P*. 2012.
4. Brennan, R. and T.L. Webster, *Object-oriented land cover classification of lidar-derived surfaces.* Canadian Journal of Remote Sensing, 2006. **32**(2): p. 162-172.
5. Ahlberg, M.S., et al., *High-Resolution Environment Models to Support Rapid and Efficient Mission Planning and Training.*
6. Naidoo, L., et al., *Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment.* Isprs Journal of Photogrammetry and Remote Sensing, 2012. **69**: p. 167-179.
7. Ko, C., G. Sohn, and T.K. Remmel, *Tree genera classification with geometric features from high-density airborne LiDAR.* Canadian Journal of Remote Sensing, 2013. **39**: p. S73-S85.
8. Popescu, S.C. and R.H. Wynne, *Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height.* Photogrammetric Engineering and Remote Sensing, 2004. **70**(5): p. 589-604.
9. Brandtberg, T., *Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar.* Isprs Journal of Photogrammetry and Remote Sensing, 2007. **61**(5): p. 325-340.
10. Hill, R. and A. Thomson, *Mapping woodland species composition and structure using airborne spectral and LiDAR data.* International Journal of Remote Sensing, 2005. **26**(17): p. 3763-3779.
11. Popescu, S.C. and K. Zhao, *A voxel-based lidar method for estimating crown base height for deciduous and pine trees.* Remote Sensing of Environment, 2008. **112**(3): p. 767-781.
12. Reitberger, J., P. Krzystek, and U. Stilla, *Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees.* International journal of remote sensing, 2008. **29**(5): p. 1407-1431.
13. Texas. Parks and Wildlife Department., *The vegetation types of Texas*. 1984, Austin: The Department. 1 map.
14. Lefsky, M.A., et al., *Estimates of forest canopy height and aboveground biomass using ICESat.* Geophysical Research Letters, 2005. **32**(22).
15. Andersen, H.-E., R.J. McGaughey, and S.E. Reutebuch, *Estimating forest canopy fuel parameters using LIDAR data.* Remote sensing of Environment, 2005. **94**(4): p. 441-449.
16. Hubbell, S.P., et al., *Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest.* Science, 1999. **283**(5401): p. 554-557.
17. Ku, N.-W., et al., *Assessment of available rangeland woody plant biomass with a terrestrial LIDAR system.* Photogrammetric engineering and remote sensing, 2012. **78**(4): p. 349-361.
18. Pitkänen, J., et al., *Adaptive methods for individual tree detection on airborne laser based canopy height model.* International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2004. **36**(8): p. 187-191.
19. DeFries, R. and J. Townshend, *NDVI-derived land cover classifications at a global scale.* International Journal of Remote Sensing, 1994. **15**(17): p. 3567-3586.

20. Kuplich, T., P. Curran, and P. Atkinson, *Relating SAR image texture to the biomass of regenerating tropical forests.* International Journal of Remote Sensing, 2005. **26**(21): p. 4829-4854.
21. Inc., E. *Texture Filters*. 2014; Available from: http://www.exelisvis.com/docs/TextureFilters.html.
22. Koukoulas, S. and G.A. Blackburn, *Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data.* International Journal of Remote Sensing, 2005. **26**(3): p. 431-455.