Supervised Parametric Classification of Aerial LiDAR Data

Amin P. Charaniya, Roberto Manduchi, and Suresh K. Lodha
University of California, Santa Cruz
{amin,manduchi,lodha}@soe.ucsc.edu

Abstract

In this work, we classify 3D aerial LiDAR height data into roads, grass, buildings, and trees using a supervised parametric classification algorithm. Since the terrain is highly undulating, we subtract the terrain elevations using digital elevation models (DEMs, easily available from the United States Geological Survey (USGS)) to obtain the height of objects from a flat level. In addition to this height information, we use height texture (variation in height), intensity (amplitude of lidar response), and multiple (two) returns from lidar to classify the data. Furthermore, we have used luminance (measured in the visible spectrum) from aerial imagery as the fifth feature for classification. We have used mixture of Gaussian models for modeling the training data. Model parameters and the posterior probabilities are estimated using Expectation-Maximization (EM) algorithm. We have experimented with different number of components per model and found that four components per model yield satisfactory results. We have tested the results using leave-one-out as well as random $\frac{n}{2}$ test. Classification results are in the range of 66% – 84% depending upon the combination of features used that compares very favorably with. train-all-test-all results of 85%. Further improvement is achieved using spatial coherence.

1. Introduction

Traditionally, 2D imaging techniques have been most popular in computer vision and image processing. In last few years, we have seen emergence of 3D range sensors. Mobile ground-based range sensors for large scale data collection is being increasingly used in several applications. In this work, we have used Airborne Laser Scanning also referred to as Aerial LiDAR (Light Detection and Ranging). It has emerged as a very popular technique for acquiring terrain elevation data. Short data acquisition and processing times, relatively high accuracy and point density, and low cost have caused LiDAR to be preferred over traditional aerial photogrammetric techniques. An hour of data collection can result in over 10 million points with point densities in the range of 1m to 0.25m. An entire city can be scanned in a matter of few hours. The resulting cloud of 3D points consists of a mixture of terrain, vegetation, building roofs, vehicles and other natural and man-made objects. Although laser range scanning technology has been in existence for more than 20 years, the development of supporting systems such as highly accurate GPS (Global Positioning System) orientation sensors have become available or affordable only in the last few years. Due to these recent developments, LiDAR data can be geo-spatially registered much more accurately which in turn helps to produce highly accurate and high-resolution digital surface models (DSMs).

An important task with Aerial LiDAR data is to classify it into meaningful categories. The raw LiDAR point cloud consists of a mixture of terrain, vegetation, buildings and other natural and man-made structures. Different types of objects require different methods for modeling, analyses and visualization. Therefore before applying any algorithms on the raw dataset, it needs to be classified into disjoint classes representing ground objects such as roads, soil, green and dry grass, concrete pathways and non- ground objects such as building roofs, trees and vehicles. In this work we have classified the LiDAR dataset into four disjoint classes - trees, grass, roads and building roofs. In order to accomplish this task we make use of the fact that each class exhibits homogeneity or patterns in a certain feature space. The objective is then to identify the correct features that can be used for discrimination in presence of outliers and random noise.

2. Background and Previous Work

2.1. Overview of LiDAR data

A typical LiDAR system consists of a laser range finder, differential GPS, inertial navigation sensors, a computer and some storage media and optionally other sensors such as Digital Cameras and multi-spectral cameras. Typically pulse lasers are used with wavelengths in the range of 1040-1060 nm. Some systems also use continuous-wave lasers. The system usually provides a number of variable parameters including the scan angle, pulse rate, beam divergence, maximum number of returns per pulse and scanning pattern. The data is usually acquired as a set of overlapping strips, each consisting of multiple scan lines. Each scan line con-

sists of a number of echoes. Generally, it is a requirement that no pulse be transmitted until the echo of the previous pulse is received. Most LiDAR systems can report multiple returns reflected from the surface. The data is generated at thousands of points per second and an hour of data collection can result in over 10 million points. Once the DGPS positions are determined, the scanner position and sensor orientation are used to compute the position of the laser spot on the ground.

LiDAR dataset consists of irregularly-spaced 2.5-D points where the elevation z has a unique value as a function of x and y. Each data point is composed of 3D position, a unique timestamp, and received signal intensity I. The intensity of the reflected light depends on the surface characteristics, the wavelength of light used and the angle of incidence. In contrast to intensity I, reflectance R refers to $\frac{I}{E_T}$, where E_T is transmitted signal intensity. For infra-red lasers with wavelength in the range of about 1000 nm, grass has reflectance of about 50%, asphalt roads reflect about 10-20%, trees (coniferous 30% and deciduous 60%) and concrete structures reflect approximately 25% of the light [8]. Most of the newer LiDAR scanners can usually record more than one return signals for a single transmitted pulse. In our case, we had multiple (two) returns that we discuss later.

2.2. Previous Work

Previous work on aerial lidar data classification can be broadly put into two categories: (i) classification of aerial lidar data into terrain and non-terrain points, and (ii) classification of aerial lidar data into features such as trees, buildings, etc.

We first describe the previous classification work into terrain and non-terrain points. This research has been motivated with the objective of generating digital terrain models. Kraus and Pfeifer [11] have used an iterative linear prediction scheme for removing vegetation points in forested areas. Vosselman et. al. [18] have used gradient-based techniques to separate building points from terrain points. Zhang et. al. [20] have utilized an iterative technique using progressive morphological filters of varying sizes for estimating suitable elevation thresholds in a local region, and thereby separating terrain points from non-terrain points. We also obtained aerial LiDAR data classified into terrain and non-terrain points provided to us by the data collection company using some undisclosed algorithm. However, we did not use this classified data. Our objective in this work iss to perform classification of original lidar data into four categories – trees, grass, roads, and buildings.

We now describe the previous efforts of classification of lidar data into features. Axelsson [1] has presented algorithms for filtering and classification of data points into terrain, buildings and electrical power lines using aerial LiDAR data, intensity returend by the LiDAR, and multiple returns. The method uses curvature based minimum description length criterion for classification. They have presented results of processing about 100,000 points with approximate point density of 8 points/m^2 visually. There is no discussion of the quality of results obtained. Maas [12] has used height texture for segmentation of lidar data. Filin [5] has proposed a surface clustering technique for identifying regions in LiDAR data that exhibit homogeneity in a certain feature space consisting of position, tangent plane and relative height difference attributes for every point. The surfaces are categorized as high vegetation (that exhibit rapid variations in slopes and height jumps), low vegetation, smooth surfaces and planar surfaces. Song et. al. [17] have focused on assessing separation of different materials such as trees, grass, asphalt (roads), and roofs based on intensity data that has been interpolated using three different techniques – inverse distance weighting, median filtering and Kriging. They observe that different interpolation techniques can enhance or suppress separability. Hebert et al. [7] have presented an outline of some classification approaches as well.

It appears that most of the previous work in classification of aerial LiDAR data has concentrated on unsupervised clustering on a smaller number of classes often resulting in coarse classification. In this work we have used supervised parametric classification with four classes. We use mixture of Gaussian models and train our data using Expectation-Maximization (EM) algorithm. Many approaches use mixture models [14, 10, 4, 3] for parametric classification. Recently Macedo et. al. [9] have also used ground-based laser for discriminating between grass and rocks (and other non-penetrable objects). In addition to LiDAR data, we decided to use aerial imagery as well because it has been suggested that using both geometry and imagery data can improve the results of classification [2]. Similarly, fusing separate color-based and texture-based classifications can also result in better classification [13, 15].

Automatic terrain classification has been used for autonomous terrain navigation (for example in exploration of Mars) [2] and for building 3D urban models [19, 6].

3. Data Classification

3.1 Data Collection and Preparation

LiDAR dataset for University of California at Santa Cruz and Santa Cruz City was acquired in October 2001 by Airborne1 Inc. The data was collected for approximately 8 square miles of target region. In order to obtain DGPS position for the scanner, reference GPS stations were set up at two National Geodetic Survey (NGS) ground control points lying within 10 miles of the target area. A 1064 nm laser at a pulse rate of 25 KHz was used for data collection. The raw data consists of about 36 million points with an average

point spacing of 0.26 meters. We resampled this irregular LiDAR point cloud on a regular grid with a grid size of 0.5m using nearest neighbor interpolation.

Since the terrain is highly undulating, we wanted to subtract terrain elevations from the aerial LiDAR data to work with the height from a flat level. To this purpose, we acquired DEM (digital elevation models). Digital Elevation Models at various resolutions can be obtained from USGS for the entire United States. We have acquired 10-meter DEMs for the San Francisco Bay Area. Due to the lower resolution, these DEMs have relatively low accuracy. We upsampled this DEM also on a grid size of 0.5m using bilinear interpolation to match with the aerial LiDAR grid.

In addition, we have used high resolution (.5ft/pixel) ortho-rectified gray-scale aerial imagery. Similar to the aerial LiDAR, aerial imagery is geo-referenced using NAD83 State Plane Coordinate System, California Zone III. The aerial imagery is downsampled to 0.5m/pixel to match with the aerial LiDAR as well.

3.2 Supervised Classification

Traditionally, there have been two main approaches to classification [3] - *supervised classification* and *unsupervised classification* (usually referred to as segmentation or clustering). In supervised classification we have a set of data samples that have class labels associated with them. This set is called the training dataset and is used to estimate the parameters of the classifier. The classifier is then tested on an unknown datasets referred to as the test dataset. An important underlying assumption is that the test dataset is similar in terms of distribution of features to the training dataset (i.e. the classifier must have observed similar features in the training in order to perform a good classification).

Here we consider the problem of assigning a class label to a d dimensional data sample **x** where d is the number of features in the feature vector **x**. Assuming that there are C classes, the posterior probability of a data sample **x** belonging to a particular class i can be computed using Bayes rule as:

$$P(i|x) = \frac{p(x|i)P(i)}{p(x)} \tag{1}$$

where $p(x) = \sum p(x|i)P(i)$, $P(i)$ is the prior probability of class i.

Assuming that we have no prior information about $P(i)$, it is usually safe to assume that $P(i)$'s for all the classes are equal $(1/C)$. Therefore, in order to determine the posterior probability $P(i|x)$ we need to determine the class-conditional densities $p(x|i)$. Finally, the data sample **x** is assigned to the class i for which $P(i|x)$ is maximum.

Mixture models are often used for modeling the class-conditional densities $p(x|i)$. A mixture model consists of linear combinations of M basis functions where M is treated as one of the parameters of the model. For example a Gaussian mixture can be expressed as:

$$p(x|i) = \sum_{j=1}^{M} P_i(j) G(x|\mu_j, \Sigma_j) \tag{2}$$

The model parameters (μ_j, Σ_j) for the Gaussian models and the mixing parameters $P_i(j)$ are estimated iteratively using Expectation Maximization (EM) algorithm [3, 4] on training samples.

3.3 Classes and Training

We classified the dataset into 4 classes:

- Trees (includes coniferous and deciduous trees)
- Grass (includes green and dry grass)
- Roads (asphalt roads, concrete pathways and soil)
- Roofs

Datasets for ten different regions of the UCSC Campus were created and manually labeled for training and validation. The size of these data sets vary from 100,000 points to 150,000 points. Presence of different classes – trees, grass, roads, and roofs – vary within these data sets. Roughly 25-30% of these data sets were trained to cover these 4 classes adequately.

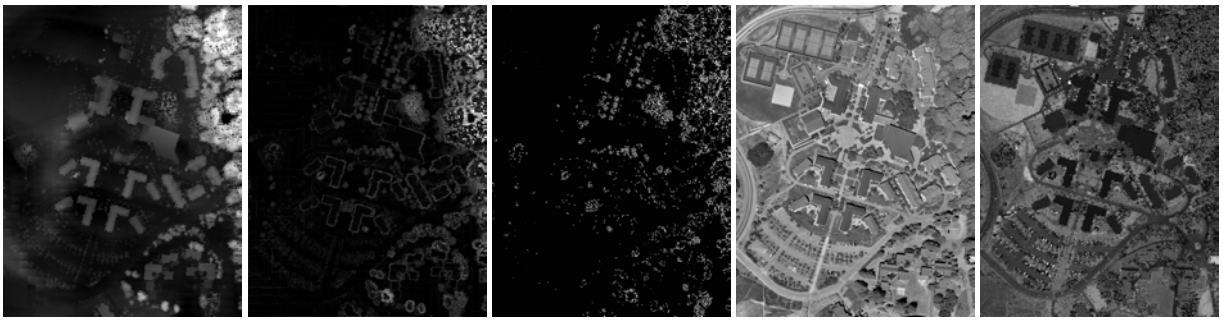
3.4 Features

We identified five features to be used for data classification purposes: normalized height, height variation, multiple returns, luminance, and intensity.

- *Normalized Height (H)*: The LiDAR data is normalized by subtracting the USGS DEM elevation from the LiDAR height values on a .5m grid.
- *Height Variation (hvar)*: Local height variation is usually computed using a small window around a data sample and is one of the most commonly used texture feature [9]. There are several possibilities such as standard deviation, absolute deviation from the mean, and the difference between the maximum and minimum height values. After some experimentation, we settled on the third measure listed above – difference between the maximum and minimum height values within a window. Here we have used a window size of 3*3 pixels ($2.25m^2$). It is expected that there is significant height variation in areas of high vegetation where some laser pulses penetrate the canopy while others return from the top. This is indeed apparent from local

(a) School of Engineering (b) Crown College (c) East Field House (d) Physical Plant

Figure 1: Sample datasets used



(a) Height (H) (b) Height Variation (hvar) (c) Multiple Returns (diff) (d) Luminance (L) (e) Intensity (I)

Figure 2: Five features used in data classification for one of the ten training data sets

height histograms. One of disadvantages of using this feature is that building roof edges can sometimes get misclassified as trees due to large height variation.

- *Multiple Returns (diff)*: Most of the newer LiDAR scanners can usually record more than one return signals for a single transmitted pulse. If the transmitted laser signal hits a hard surface such as terrain or the middle of a building roof, there is only one return. However, if the laser pulse hits the leaves or branches of trees or even the boundaries of roofs, there are at least two recorded returns, one from the top of the tree or roof and the other from the ground. Thus, LiDAR point cloud can be considered as a set of functions:

  $LiDAR = \{z_{first}, i_{first}, z_{last}, i_{last}\}$, where z is the height function and i is the intensity function.

  We have obtained both, the first and the last return datasets and have used the *diff = height difference between the first and last returns* as one of the features. The first and the last returns are associated using the timestamps. z_{first}, i_{first} exist for a subset of values of z_{last}, i_{last}. For the values of z_{last}, i_{last} for which we do not have the corresponding first returns, we assume that both returns are the same and hence there will be zero height difference. As with height variation, this feature can also be effectively used to identify high vegetation areas.

  One of the anomalies we observed was that sometimes, the first return height is less than the last return height. One possible reason for this could be the presence of noise, although it needs further investigation.

- *Luminance (L)*: Luminance corresponds to the response of the terrain and non-terrain surfaces to visible light. This is obtained from the gray-scale aerial image.

- *Intensity (I)*: Along with the height values, aerial LiDAR data usually contains the amplitude of the response reflected back from the terrain to the laser scanner. We refer to it as intensity. Since the laser scanner uses light in the near infra-red spectrum, we expect that the intensity provides information that is complementary to luminance (which is measured in the visible spectrum).

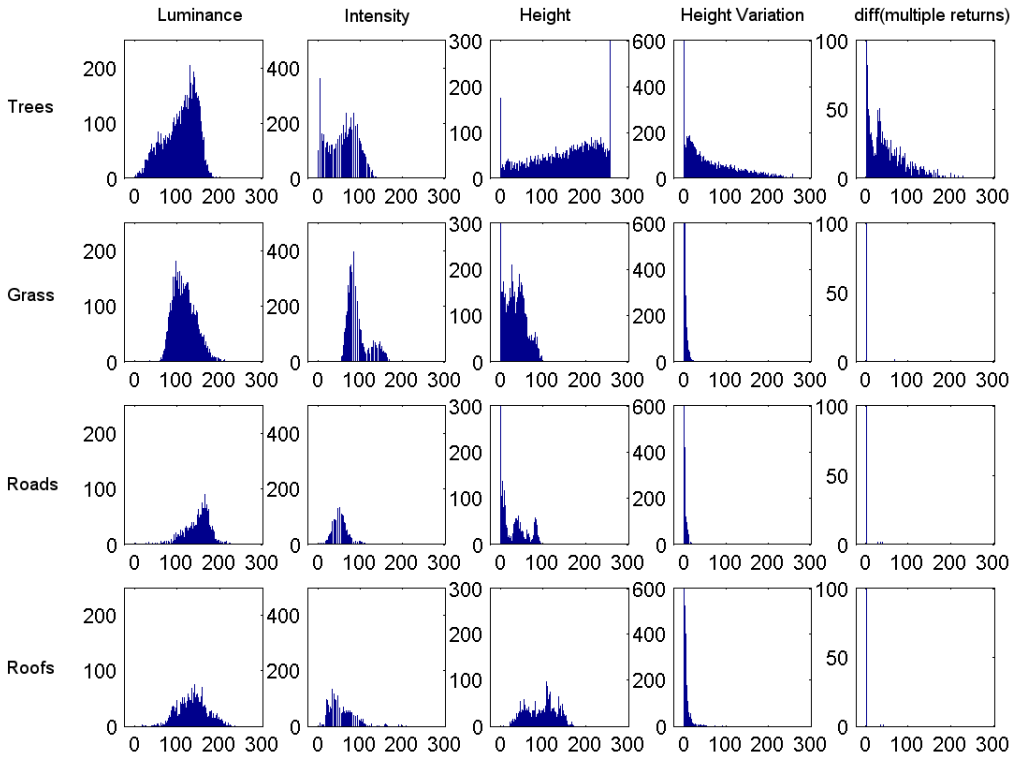


Figure 3: Marginal histograms for five features/four classes: x-axis represents the values of the features normalized between 0 to 255 and y-axis represents the number of points. Actual values of luminance, intensity, height, height variation, and diff vary from 0-255, 0-20, 0-50 meters, 0-50 meters, and 0-50 meters respectively.

Figure 2 shows each of the above-mentioned features for the College 8 area of UCSC Campus.

4. Results

Marginal Histograms: Figure 3 shows class/feature histograms for the training data. It should be noted that these are marginal histograms and and therefore do not show inter-feature correlation, which is exploited in our Mixture of Gaussian models. However, looking at the marginal histograms gives us some sense of relative complexity within the features.

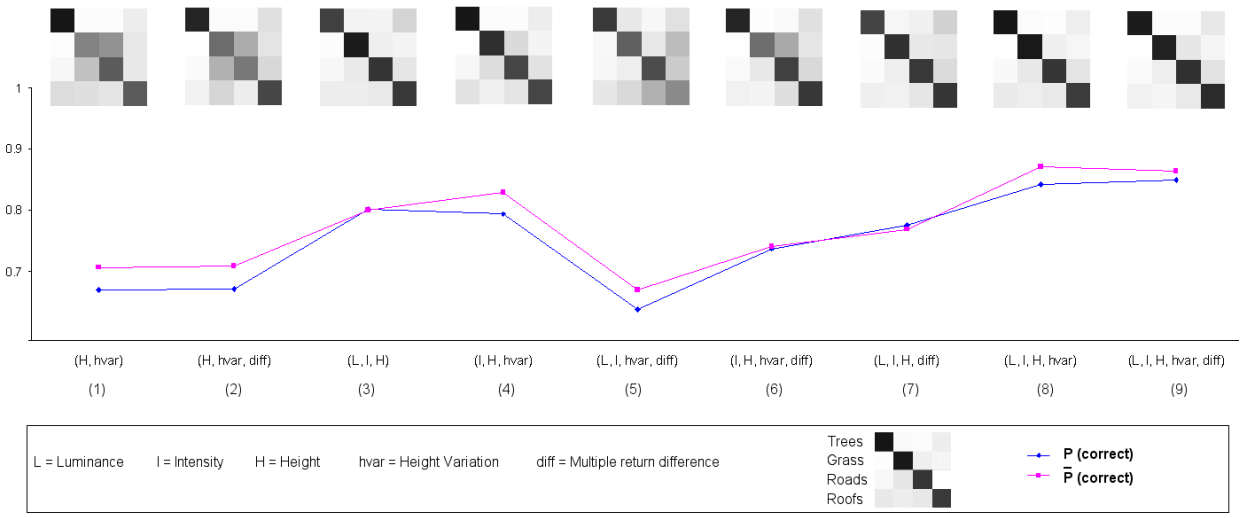
Number of components/mixture: Automatically determining the number of components for every mixture (equation 2) from the training data is not a trivial problem. Several well-known methods exist for estimating the number of components [16]. However, in our experience, such methods are not satisfactory. Therefore, we chose to decide the number of modes empirically. We experimented using 2, 4, 5 and 6 components per mixture. We noticed that the results improved significantly between 2 and 4 components. Adding more components did not improve the results very much. Moreover, by adding more components we increase the computational complexity as well as run the risk of overfitting the data. Therefore, we have used four components for each mixture model.

Leave-One-Out-Test: The model parameters and the posterior distributions are estimated using Expectation Maximization algorithm [3, 4].

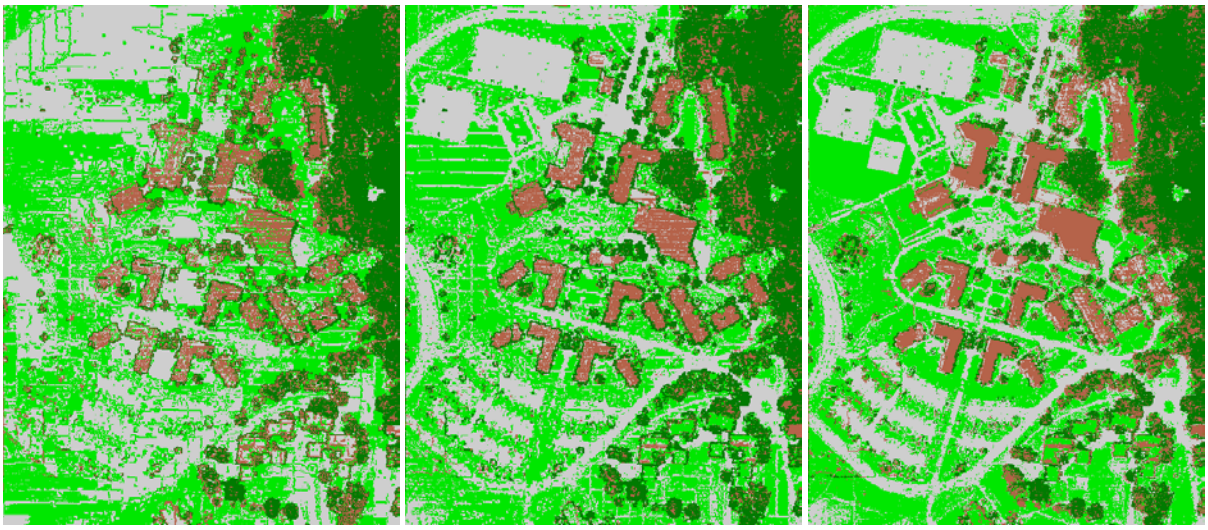
Figure 4 summarizes results of leave-one-out test. We performed this test using several combinations of the above-mentioned features. The figure shows a graph of $P(correct)$ and $\overline{P}(correct)$ for each these combinations along with their confusion matrices where,

$$P(correct) = \frac{1}{C}\sum_{i=1}^{C}\frac{n_{i'=i}}{n_i} \tag{3}$$

i' is the class assigned to a pixel for which the true class is i, and n_i is the total number of pixels assigned to class i.



(a) Classification results with confusion matrices



(b) Using just height and height texture

(c) Just LiDAR (no aerial image)

(d) All features used

Figure 4: Classification results in increasing order of (number of features, more accurate results): height H is effective in overall classification; height variation *hvar* is effective in tree classification; L and I together are effective in grass vs. road classification.

C is the total number of classes ($C = 4$ in our case).

$$\overline{P}(correct) = \frac{1}{N} \sum_{i=1}^{C} n_{i'=i} \quad (4)$$

where N is the total number of labeled pixels. This is the normalized probability that is the average of $P(correct)$ for each class weighted by the number of pixels assigned to it. The confusion matrices visualize the results for each class. The rows of the matrix show the true classes and the columns indicate the classes assigned by the classifier. In case of perfect classification the diagonal elements are all one (black) and the other elements are all zeros (white).

Random $\frac{n}{2}$ and Train-all-Test-All tests: In most cases, results of random n/2 test (randomly choosing half the data for training and other half for testing) closely follow the leave-one-out test. Therefore, for the sake of brevity we have chosen not to discuss these results here. Train-all-Test-all results were marginally better than the leave-one-out results in some cases, such as 85%, where all the five features were used. Again, for lack of space, we leave out those results.

Observations: We observe that using more features produces better results sometimes, but not always. However, some of the combinations seem to be better than the others. Here we briefly discuss some of the combinations.

1. *H, hvar:* Using just the height and height texture from the LiDAR data we find that detecting trees is quite effective (about 91%). It is also evident that there is a lot of confusion between grass and roads. Both these classes have similar height and height variation.

2. *H, hvar, diff:* Adding multiple return difference does not improve the results of (1) significantly.

3. *L, I, H:* By using luminance and intensity we have improved the overall results. However, due to the omission of height variation the classification of trees is worse than (1) and (2).

4. *I, H, hvar:* Assuming that we do not have aerial imagery available, we use only the intensity, height and height variation. In this case tree classification has improved. However, the overall results are slightly worse. This indicates how well we can do when no other supporting data is available.

5. *L, I, hvar, diff:* Excluding the height information results in the worst classification among the combinations that we have tried. Therefore, the height feature plays a very important role in classification.

6. *I, H, hvar, diff:* It is surprising to see that adding multiple return difference to (4) worsens the results. This is primarily due to misclassification of grass patches.

7. *L, I, H, diff:* Similarly adding multiple return difference to (3) lowers the results. Here too we observe the same effect (as in (6)). Including multiple return difference improves classification of roads and buildings typically by 5 to 6%.

8. *L, I, H, hvar:* Adding height variation feature to (3) dramatically improves the results. This is primarily because of improved classification of high vegetation areas.

9. *L, I, H, hvar, diff:* Finally, adding mutliple returns, improves overall results only marginally.

We can briefly summarize few important observations:

- Height feature is an important classifier for terrain.
- Height variation plays an important role in classification of high vegetation areas.
- Light features (luminance, intensity) are useful for separating low vegetation (grass) and roads.
- Adding multiple return difference improves classification of roads and buildings by only 5-6% and decreases the accuracy in other cases.

Spatial Coherence: The classification done so far is point-based. Each individual point is classified according to its position in the feature space. However, most classes including trees, grass, roofs etc. span across hundreds of data points that are close to each other in position space. Therefore it makes sense to exploit this spatial coherence in classification. The probability of a data sample belonging to a particular class is affected significantly by that of its neighbors. Enforcing spatial coherence constraints can be done as a post-process to classification and can be carried out in a number of different ways. One of the simplest ways would be to use a *max voting* filter where the data sample is assigned a class that occurs most frequently in its neighborhood. Here we have used a window size of 3 by 3 pixels ($2.25m^2$). Figure 5 shows the results with and without enforcing spatial coherence constraints. It can be seen that the results are, on an average 3 – 4% better with enforcing spatial coherence constraints.

5. Conclusions and Future Directions

We have presented the results of supervised classification of aerial LiDAR data using mixture of Gaussian models. Using this method we have been able to effectively classify the

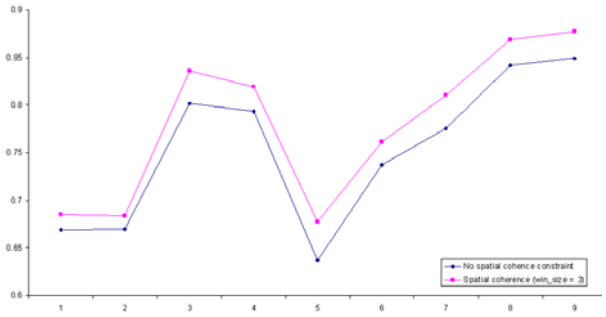


Figure 5: Results with and without enforcing spatial coherence constraints

dataset. More importantly, our results have identified what features may be appropriate for certain classes. We plan to investigate classification by fusing results of multiple classifiers. We also hope to improve the classification results further by identifying noise and outliers in the dataset before classification.

Acknowledgements

We would like to thank Airborne1 Corporation for helping us acquire the LiDAR data. This research is partially supported by the Multi-disciplinary Research Initiative (MURI) grant by U.S. Army Research Office under Agreement Number DAAD19-00-1-0352 and the NSF grant ACI-0222900.

References

[1] Peter Axelsson. Processing of laser scanner data -algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3):138–147, 1999.

[2] P. Bellutta, R. Manduchi, L. Matthies, K. Owens, and A. Rankin. Terrain perception for demo iii. In *IEEE Intelligent Vehicles Symposium 2000*, 2000.

[3] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[4] R.O. Duda, P.E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, New york, 2001.

[5] Sagi Filin. Surface clustering from airborne laser scanning data sagi filin. In *ISPRS Commission III, Symposium 2002 September 9 - 13, 2002, Graz, Austria*, pages A–119 ff (6 pages), 2002.

[6] Christian Frueh and Avideh Zakhor. Constructing 3D city models by merging ground-based and airborne views. In *IEEE Conference on Computer Vision and Pattern Recognition 2003, June 2003*, 2003.

[7] Martial Hebert, Nicolas Vandapel, Stefan Keller, and Raghavendra Rao Donamukkala. Evaluation and comparison of terrain classification techniques from ladar data for autonomous navigation. In *23rd Army Science Conference*, December 2002.

[8] Simon J. Hook. Aster spectral library, 2002. http://speclib.jpl.nasa.gov, last modified, Sept. 24, 2002.

[9] L. Matthies J. Macedo, R. Manduchi. Ladar-based discrimination of grass from obstacles for autonomous navigation. In *Seventh International Symposium on Experimental Robotics (ISER'00)*, 2000.

[10] M.J. Jones and J.M. Rehg. Statistical color models with application to skin detection. In *Cambridge Research Laboratory Technical Report CRL 98/11*, 1998.

[11] K. Kraus and N. Pfeifer. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53:193–203, 1998.

[12] Hans-Gerd Maas. The potential of height texture measures for the segmentation of airborne laserscanner data. *Fourth International Airborne Remote Sensing Conference and Exhibition*, 21st Canadian Symposium on Remote Sensing:Ottawa, Ontario, Canada, 1999.

[13] Roberto Manduchi. Bayesian fusion of color and texture segmentations. In *Seventh International Conference on Computer Vision, Sept 1999*, 1999.

[14] H. Riad and R. Mohr. Gaussian mixture densities for indexing of localized objects in video sequences. In *INRIA Technical Report RR-3905*, 2000.

[15] X. Shi and R. Manduchi. A study on bayes feature fusion for image classification. In *IEEE Workshop on Statistical Analysis in Computer Vision, Madison, Wisconsin, June 2003*, 2003.

[16] Padhraic Smyth. Clustering using monte carlo cross-validation. In *Knowledge Discovery and Data Mining*, pages 126–133, 1996.

[17] Jeong Heon Song, Soo Hee Han, Ki Yun Yu, and Yong Il Kim. A study on using lidar intensity data for land cover classification. In *ISPRS Commission III, Symposium 2002 September 9 - 13, 2002, Graz, Austria*, 2002.

[18] George Vosselman. Slope based filtering of laser altimetry data. *International Archives of Photogrammetry and Remote Sensing*, XXXIII, Amsterdam, 2000, 2000.

[19] Suya You, Jinhui Hu, Ulrich Neumann, and Pamela Fox. Urban site modeling from lidar. In *Second International Workshop on Computer Graphics and Geometric Modeling*, 2003.

[20] Keqi Zhang, Shu-Ching Cheng, Dean Whitman, Mei-Ling Shyu, Jianhua Yan, and Chengcui Zhang. A progressive morphological filter for removing non-ground measurements from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41, No. 4, April 2003:872–882, 2003.