Road Detection in Massive LIDAR Data

December 4, 2016

1 Objective:

From a given LIDAR data set for a geographic region, extracting the road network present in that region.

2 Goal:

The basic goal for our project is supervised machine learning. Initially we would be given a set of LIDAR data and the corresponding road network map for a particular region. We would try to identify certain points from the road netork map and locate the same in the LIDAR data set. After locating these points in the LIDAR data set we would analyse the data and try to extract various attributes related to the points located. We would also analyse LIDAR data where road is absent. Once the analysis is done we would train the system with various attributes of the LIDAR data for both absence and presence of road. Finally, using this learning we would predict the road network for a different region from the LIDAR data set of that particular geographic region.

3 Requisites:

- Study about LIDAR.

- Learn about LAS files, learn libLAS, compile and install libLAS.

- Study about Shapefiles as the map data would be shapefiles.

- Extract data from a LAS file and try to present it as an image file

- Learn and implement Supervised machine learning using SVM model.

4 Issues present from the past work done:

1. LAS tools was used which is not open source, hence need to use libLAS.

2. Previously open street map was used in which the data provided was only a vector which was the mid points only, hence the width of the road could not be determined. To overcome this, we will be using Shape files.

5 Timeline and Plan:

Task	Timeline
Extract data from LAS files.	15 Sep to 15 Oct
Finish the initial stage of learning and required installations. Study about Support Vector Machines (SVM).	01 Oct to 15 Oct
Supervised machine learning and training phase. Studying shape-files and understanding them.	15 Oct to 31 Oct
Create a program that would take a LAS file as an input and locate the road network present in that particular region.The output can be in the form of an image file showing the road network.	01 Nov to 15 Dec

6 Progress made so far:

We have successfully completed the first three tasks enlisted in the above section. We are in the final leg of the project where we are trying to draw the road network from a given LAS file. Following is the summary of the work done so far.

Initially we focussed on the processing of the LAS files using the python libLAS library. The header of the LAS file contains information such as the number of data points consisted in the LAS file, the maximum and minimum value of the x and y coordinates, the data format ID, so on and so forth. The attributes that each point in the LAS file will have depends on the data format ID. All the LIDAR data points have colour attribute when the data format ID is two and above. We created a python module which takes LAS files as input and after prossecing it generates an image file. It takes the rbg values of each LIDAR data point and colours the corresponding pixels of the image with the same colour as in the LIDAR data point. There is also a provision in our module to look into a particular rectangular region of the LIDAR data by providing the top and bottom coordinates of the diagonal, and the module then generates the image for that particular region. Thus we finally managed to draw each data point from the LIDAR file (LAS file) with their corresponding rbg values and generated an image file which solved our purpose of visualization of the data provided.

After processing the LAS files, we studied shape files. Shape files essentially are a popular geospatial vector data format for geographic information system (GIS) software. It is developed and regulated by Esri (Environmental Systems Research Institute) as an open specification for data interoperability among Esri and other GIS software products. We know that all points in the LIDAR data are stored in an UTM coordinate system, where x, y coordinates represent a particular point on the Earth. We installed pyproj library for converting the x, y coordinates to their corresponding longitute and latitude and vice versa. We began our study on the shape file for the Leon County. We worked on loading the shape file i.e. the complete geometry of the region for Leon County. After loading the shape file we tried to extract the part of the shape file corresponding to a particular LIDAR tile. We know that shapes of the shape files are polylines so we began drawing these polylines on the LIDAR image to find out the roads. After drawing the lines we processed the nearby LIDAR points and drew those to the LIDAR image. In this process we tried to extract the most of the LIDAR data points which would comprise the road.

We compared the shape file and LIDAR data points and tried to predict the roads with maximum accuracy. We labelled the points matching our prediction for the road as 1 and the points

not matching as 0. Furthermore, we have added the feature of getting the average height of the lidar data points by considering all the neighboring points. We have also created an estimator module which uses SVM (Support Vector Machine) learning model for prediction. It first trains the system with the previously computed training data. The training data contains rgb values along with average height.Then it takes a LAS file as input for prediction of the road network. Thus we succeeded in creating a training data set in the process.
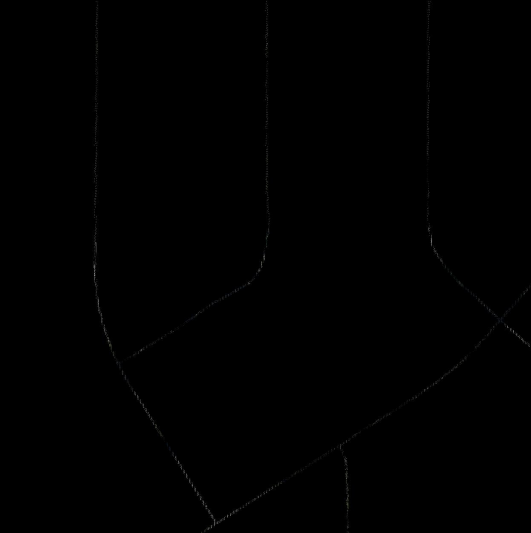
Following are the images depicting the extracted LIDAR data points from a shape file corresponding to a road network in the Leon County. Also the image generated after drawing the, corresponding extracted, road network on the LIDAR image is shown.
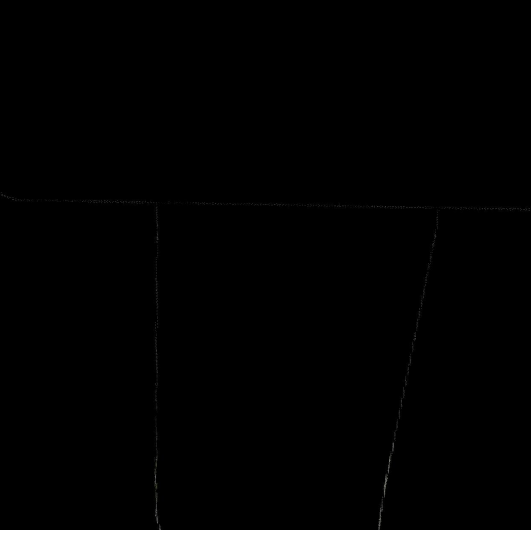
(a) LIDAR data points corresponding to a road extracted from a shape file

(b) Road drawn on LIDAR image

(c) LIDAR data points corresponding to a road extracted from a shape file

(d) Road drawn on LIDAR image

Figure 1: Road network prediction

7 Challenges and Setbacks:

The machine learning model currently used is not being able to predict the roads with complete accuracy. There may be various reasons for it.

- First of all we are using only default gamma and coefficient values of the SVM model.

- There are shadows on the road which dilute the rgb values.

- We may need to include more features because the current features may not be sufficient.

- Also the minimum distance between the polylines and the nearest data points drawn are dependent on the width of the road, which varies. We need to come up with an average width which would make the training dataset more compact.

- We have not been able to test the system properly because mosty it takes a very long time to train and predict. Our machines are not powerful enough to handle such huge amount of workload and crash at times, thus hindering the test procedure.

8 Final Outcome:

As planned and stated initially in our goals and plans we created a program that takes a LAS file of a particular region as an input and locate the road network present in that region. Also we present the output as an image file which shows the road network present in that region.