Application of Data Mining Techniques for the Development of 3D Laser Scan Data Management Program

Choi, Seung Pil [1], Shin, Moon Seung [2] *, Yang, In Tae [2] and Acharya, Tri Dev [2]

[1] *Department of Civil Engineering, Catholic Kwandong University, Gangneung, Gangwon, 25601, Korea.*
[2] *Department of Civil Engineering, Kangwon National University, Chuncheon, Gangwon, 24341, Korea.*
Corresponding Author: (Shin, Moon Seung)

Abstract

Remotely sensed '3D Laser Scan' technology is the most precise surveying technique to obtain three dimensional geospatial information. Due to its ease and precision, the application in various fields has been much preferred. But the data size are quite heavy and require special and expensive proprietary software for conversion into readable formats. In this study, we developed a python based LiDAR Read and Extract (LiRE) program with various python modules available. The program is useful to read, view, filter and export LAS format data in in various formats. Also, it uses unsupervised and supervised methods for classification of the data in different classes based on training data. The exported results can be easily used in other civil related programs.

Keywords: 3D Laser Scan, LiDAR, Data Mining, LiRE, Python.

INTRODUCTION

In recent days '3D Laser Scan' technology has extremely been useful in surveying and mapping, forestry, construction, field for 3D documentation of cultural heritage, reverse engineering etc. [1-3]. The technology uses pulsed laser light to measure the return time of reflected pulses and thus with known wavelength 3D points are made, without in contact with the actual object. The produced result are highly accurate dense cloud points. For various civil engineering software, these data are too heavy to handle and require expensive proprietary software for conversion into readable formats. And due to the inconvenience of the data handling, the application of '3D Laser Scan' data has been limited. Developing a program to read, visualize and export these type of data would be very useful in further expanding the applications where high precisions are necessary.

In Korea, researchers has been applied various techniques for the successful application of LiDAR data in civil engineering domain. Gi *et.al.* [4] used Local Indicators Of Spatial Association (LISA) statistics variation for the extraction of buildings automatically. Similarly, Kim *et.al.* [5] proposed new extraction algorithm for artificial structures in 3D cloud points based on RGB and Hough transformation. Lee *et. al.* [6] filtered Airborne Laser Scanning (ALS) points into ground and non-ground effectively through labelling and window based algorithm by utilizing 2D adjacency based on scan line. Choi *et. al.* [7] presented techniques for classifying terrestrial LiDAR data using factors and clustering analyses. Also, Kim and Kim [8] has used the tile-based topographical classification of airborne LiDAR data extraction of building outlines.

Most of the existing studies have developed data classification models and criteria that are limited to specific regions, requiring modifications to data classification models and standards to apply to other regions. In addition, these studies utilize computer programming tools such as MATLAB, C, C++, etc. to design and implement data classification models, and thus require a large number of time and development costs. Nowadays data mining are widely used to identify pattern information using learning data as a means of learning and using these pattern information to solve the classification of new input data that determines the properties of new input data. Typical classification models include Deterministic Tree, Rule Set, Support Vector Machine, Bayesian Network, and Neural Network Models. Depending on how one choose models and types of classes in dataset, the accuracy of the classification varies. Hence for highly precise results, finding out more sophisticated attributes are necessary as well. Waikato Environment for Knowledge Analysis (WEKA) [9] is commonly used opens source program to help users identify the model and classify characteristics of the aforementioned data mining, and assist users with easy search for classification models and classifications. Lee *et.al.* [10] developed a topographical classifier development support system cooperating with data mining tool WEKA to help users to construct accurate topographical classification systems. However, the program is dependent on data mining techniques from WEKA and uses images as well all attributes of ALS data.

In this study, a program based on python named LiDAR Read and Export (LiRE) was developed for 3D Laser Scan Data management. The program read and view LAS files then export them in comma separated or tab delineated file as well as direct input data. Moreover, the data can be filtered based on the area, height, and lower or higher to average value and visualize them in 2D and 3D point and surface plots. Finally, the data can be directly classified based on unsupervised or training data based

supervised data mining based classification techniques. Following sections describes the features of the program in more details along with data from an actual terrestrial LiDAR.

3D LASER SCAN DATA MANAGEMENT PROGRAM

The whole program was coded in Python and its open source modules. To minimize the effort in installing multiple modules Anaconda package was used for the complete scientific package.The GUI was based on wxPython and is shown in Figure 1. Table 1 shows the details of all the functions of the program. The program has five menus: File which opens and export LAS, comma separated or tab delineated files, View for 2D and 3D plots, Filter for filtering and generating new files, Data mining for classification and lastly about for information. The las file is read by LibLAS module and saved using NumPy and SciPy modules, whereas the visualization is done using matplotlib and mayavi modules. The unsupervised and supervised classification is done using scikit-learn and orange modules respectively.

Table 1. Section name and functions details.

Menu	Sub Menu	Function
File	New	Reset file selection
	Open	Open las file and training input
	Export	Export in las, txt, tab file
	Quit	
View	Feature Boundary Line	Visualize boundary, 2D and 3D plots
	RGB Scatter Plots	
	2D Scatter Plot	
	2D TIN Plot	
	2D Contour Plot	
	3D Scatter Plot	
	3D TIN Plot	
	3D Overview	
Filter	Area(XY Range)	Filter 3D points based on the area range, elevation
	Elevation	
	Intensity	
	LowFeature Points	
	HighFeature Points	
	Density	
Data Mining	Unsupervised KMeans	Using K-means clusters
	Classification Tree	Requires training input data in file menu
	K-Nearest Neighbours	
	Naive Bayesian	
	Support Vector Machine	

File menu

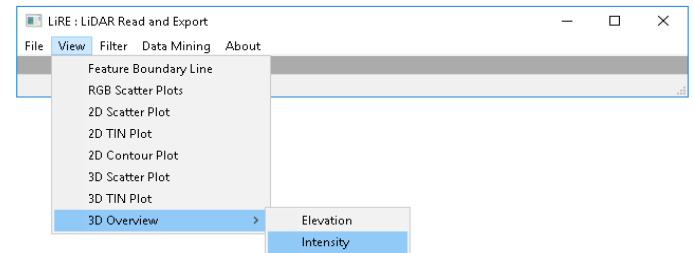Using different modules, the LiDAR las files were easily read and exported into txt, xls for other GIS/CAD, datamining programs like WEKA, SPSS etc. and Tab file especially for Orange, a data mining python program.
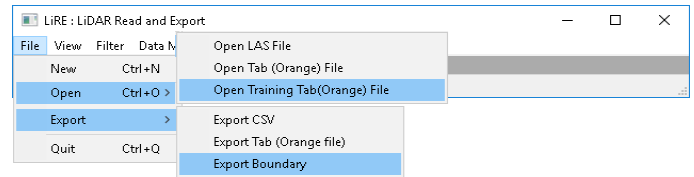
Figure 1: GUI of LiRE program with File menu and submenus.

View menu

Visualization of data is very important in any geospatial data. In LiRE, the View menu provides the option to make the boundary of the scan area as well as RGB scatter plots. The user can plot the data in 2D X-Y plane and 3D X-Y-Z plane. Also, TIN and contour can also be viewed based on elevation. The 3D over view is based on the elevation and intensity of the input data.

**Figure 2:** Submenus of the View menu.

Boundary line extraction in civil engineering is very important aspect as the features such as roads, buildings, river etc. can be easily extracted after classification or filtering. In Feature Boundary Line module, Convex Hull method is applied for boundary extraction. It forms polygon with convex angles such that all the provided points are included, as shown in the Figure 3.
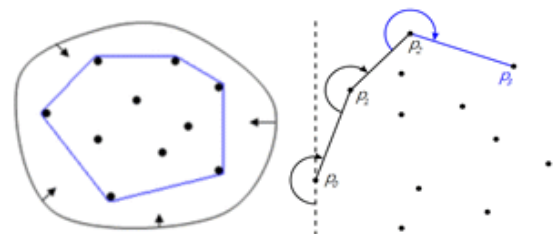

Figure 3: Convex hull.

Filter menu

As the raw 3D laser scan data are highly dense and all points are often not necessary always. Hence, filtering is very essential for the efficient and effective application of the data. Figure 4

shows the submenus of the filtering options. The filter based on the area range is for the cutting of rectangle subset, whereas elevation and intensity can be provided to filter based on the input. Low and High feature points are separated from the mean value which is very useful for filtering offsets or single taller objects like street poles. The density option filters the data randomly based on the provided 75, 50 and 25% of the input into new output file.

Data mining menu

The data mining menu uses both unsupervised and supervised classification of 3D laser scan data as shown in Figure 5. Unsupervised classification uses K-means to form clusters and export that into csv or tab format, whereas supervised methods Classification Trees, K-Nearest Neighbours, Naive Bayesian and Support Vector Machine techniques requires input training data in the File menu and based on which new classes are defined for each points.
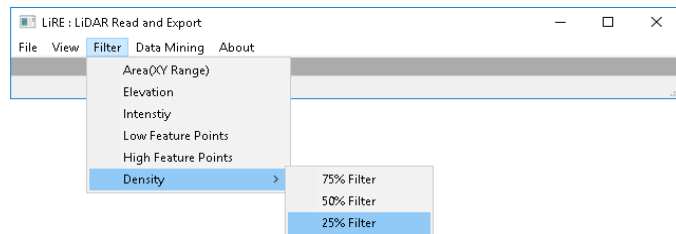
**Figure 4 –** Filter submenus in the program.
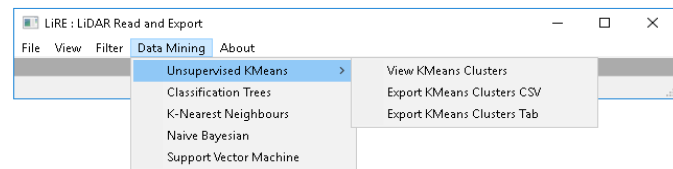
Figure 5 – Filter submenus in the program.

About menu

The menu provides help information as well as information regarding the program.

CLASSIFICATION ON THE REAL TERRESTRIAL 3D LASER SCAN DATA

In order to see the performance of the program, a real field 3D Laser Scan data was prepared in a highway area in Gangneung, Korea with roads, vegetation, concrete walls, street lights. The data was available in Las format and first was opened in the LiRE program to see the plots as shown in Figure 6. The left image shows only plots with bounding box in red whereas the right image is 3D surface where the road and the curb can be seen for the footpath.
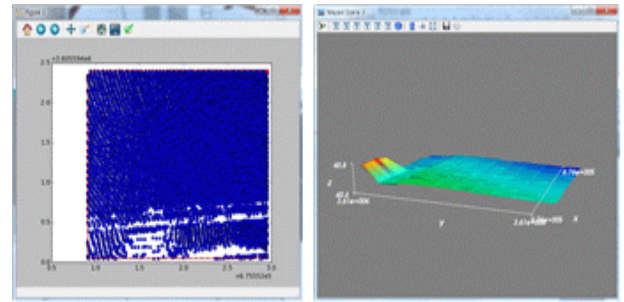

Figure 6: 2D and 3D plot of the highway section in data.

As the data is very dense, the Filter functions were also tested and all seems to have work fine and exported data were less in size and data points. Figure 7 shows the filtering based on 75% density and Figure 8 shows the filtering of high and low data points from the average of the whole scene.
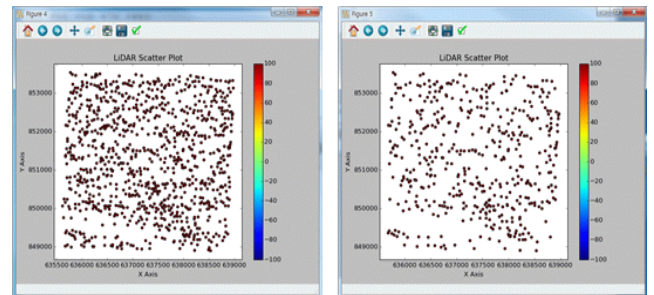
**Figure 7:** Result of 75% density filter.
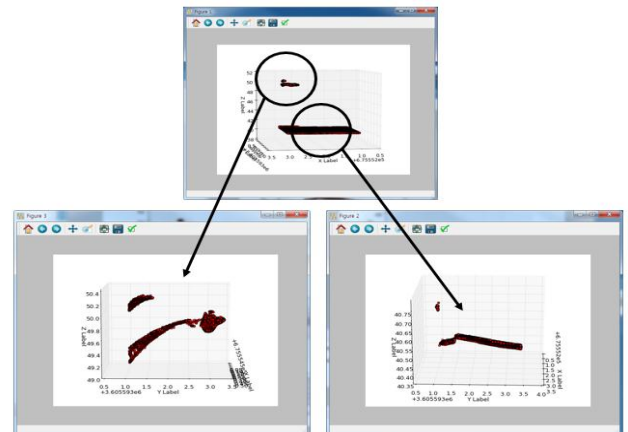
Figure 8: Filtering high and low feature points.

The exported features can be exported in point values or in tiles of selected areas and evenly divide into tiles as per the minimum class in the scene. The scenes can be attributed into various values as shown in the Table 2.

Table 2. Details of the attributes.

S. No.	Attributes	Details
1	Z_avg	Average of elevation
2	Z_var	Variance of elevation
3	Z_skewness	Skewness of elevation
4	Z_kurtosis	Kurtosis of elevation
5	Z_max	Maximum of elevation
6	Z_min	Minimum of elevation
7	Intensity_avg	Average of intensity
8	Intensity_var	Variance of intensity
9	Intensity_skewness	Skewness of intensity
10	Intensity_kurtosis	Kurtosis of intensity
11	Intensity_max	Maximum of intensity
12	Intensity_min	Minimum of intensity
13	Diff_Intensity_maxmin	Different between max. & min. of intensity
14	Diff_Intensity_maxavg	Different between max. & avg. of intensity
15	Diff_Intensity_avgmin	Different between avg. & min. of intensity
16	Separated_Inten_avg1	Average of 1st return intensity
17	Separated_Inten_avg2	Average of 2nd return intensity
18	Separated_Inten_avg3	Average of 3rd return intensity
19	Separated_Inten_avg4	Average of 4th return intensity
20	NorthGradient	North gradient
21	Westgradient	West gradient
22	Eastgradient	East gradient
23	Southgradient	South gradient
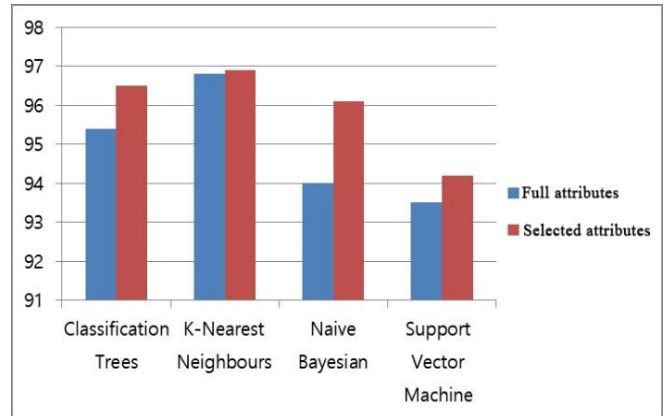
Figure 9: Classification accuracy of supervised classification based on attributes sets.

The complete set of data was tiled into different size 5m × 5m, 10m × 10m, and 20m × 20m in order to see the accuracy of the classification. After the tiles were generated the attributes in Table2. Based on the field inspection training data the supervised classifications were done. During classification two different attribute datasets were used: one full dataset and another selected attributes 1, 2, 6, 7, 8, and 13 from Table 2 based on the classification tree importance. The accuracy of the classification with full attributes and selected attributes are shown in the Figure 9 and 10. With the selected attributes, the accuracy of the classification seems increased. However, decrease in tile size has improved the accuracy as 5m tiles are less accurately classified than 10m ones.
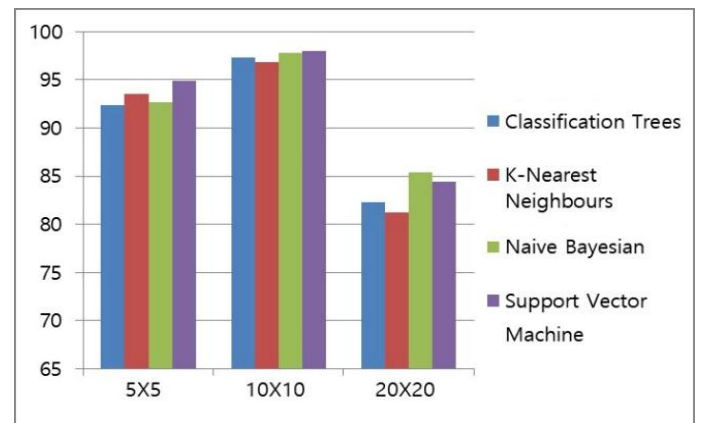
Figure 10: Classification accuracy of supervised classification based on tile size based.

CONCLUSION

In this study, 3D Laser Scan Data management program was develop in python environment using various modules. The program developed is capable of reading, exporting, viewing, filtering and classifying the las files. It was also tested and

applied for a set of data for filtering and supervised classification of data. Filtering data based on area, elevation, intensity and density can reduce the computation and storage cost of the data and can still lead to better classification results. The exported results can be easily used in other civil related programs as well. In future, the program will support more recent LiDAR formats, read more efficiently and classify with more accuracy.

ACKNOWLEDGEMENT

REFERENCES

[1]   Lee In Su; Kang Sang Gu. The Study on Reconnaissance Surveying using Terrestrial Laser Scanner. *Journal of the Korean Society for Geo-spatial Information Science* **2006**, *14*, 79-86.

[2]   Kim, E.; Wie, G.; Cho, H.; Yang, I. A Study for Forest Research using Airborne Laser Scanning. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* **2010**, *28*, 299-304.

[3]   Yang, I.T.; Acharya, T.D.; Lee, D.H.; Shin, M.S. Reverse Engineering of a Bobsleigh Structure using LiDAR. *International Journal of Applied Engineering Research* **2017**, *12*, 976-980.

[4]   Byun, Y.; Lee, J.; Son, J.; Yu, K. A Study on Building Extraction from LiDAR Data using LISA. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* **2006**, *24*, 335-341.

[5]   Kim, N.; Roh, Y.; Jung, K.; Kim, K. Structure Extraction in 3D Cloud Points using Color Information and Hough Transform. *Journal of the Institute of Electronics Engineers of Korea SP* **2009**, *46*, 143-151.

[6]   Lee, J.H.; Yeom, J.H.; Kim, Y.I. Filtering Airborne Laser Scanning Data by Utilizing Adjacency Based on Scan Line. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* **2011**, *29*, 359-365.

[7]   Choi, S.P.; Cho, J.H.; Kim, Y.; Kim, J.S. Classification of Terrestrial LiDAR Data using Factor and Cluster Analysis. *Journal of the Korean Society for Geo-spatial Information Science* **2011**, *19*, 139-144.

[8]   Kim, N.; Kim, Y. A Building Outline Extraction Scheme using Tile-Based Topographical Classification from Aerial LiDAR Data and Building Tile's Airborne Image.

Journal of KIISE : Computing Practices and Letters **2012**, *18*, 916-920.

[9]   Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA Data Mining Software: An Update. *SIGKDD Explor.Newsl.* **2009**, *11*, 10-18.

[10]   Lee, S.; Lee, H.; Sung, C.; Park, C.; Cho, W.; Kim, Y. A Topographical Classifier Development Support System Cooperating with Data Mining Tool WEKA from Airborne LiDAR Data. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* **2010**, *28*, 133-142.