Improvement Of Mobile Lidar Data Classification Of Urban Road Environment Using Machine Learning Algorithms

By

Mahmoud Abdeltawwab Abdelhamid Mohamed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
In
CIVIL ENGINEERING - PUBLIC WORKS

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

Abstract

3D road mapping is essential for intelligent transportation system in smart cities. Road features can be utilized for road maintenance, autonomous driving vehicles, and providing regulations for drivers. Currently, 3D road environment receives its data from Mobile Laser Scanning (MLS) systems in the format of LiDAR point clouds, which are distinguished with their accuracy and high density. MLS systems provide an effective surveying of long road corridors. MLS huge amount of point clouds require automatic algorithms for point classification without time consuming. Machine learning (ML) algorithms are able to classify road features with high efficiency with acceptable processing time .

The methodology of this research is divided into four main steps. First, subsampling step is proposed and applied to reduce the amount of the dataset without any loss in the information. Second, three neighborhood types, including k-nearest neighbor (KNN), spherical and cylindrical neighborhood methods are presented to determine the neighborhood of each point. Third, geometric point features are derived and used as input for ML algorithms. These features are divided into three main subsets; covariance, moment and height. The raw Z coordinate of points is modified and used as a point feature with its new definition. Thus, the Z coordinate was not included in most of previous research as it had a negative impact on results. Finally, three ML algorithms, namely Random Forest (RF), Gaussian Naïve Bayes, and Quadratic Discernment Analysis are applied to classify the data.

A part of (Paris-Lille-3D) mobile LiDAR point clouds belonging to NPM3D Benchmark suite research project were used in this research. Lille part is a scanning of 1.5 km long road in Lille with more than 98 million points and available as two parts; Lille 1 and Lille 2. The ML algorithms were used to label the LiDAR data into nine classes, namely ground, building, pole, bollard, trashcan, barrier, pedestrian, car, and natural/vegetation.

The results of each step in this research were evaluated. The cylindrical neighborhood method achieved high scores compared with spherical and KNN methods with overall

accuracy of 92.39%, 90.84% and 82.46%, respectively considering all point features and using RF classifier. The classification of the subsampled dataset achieved close results to the full dataset with overall accuracy of 90.26% and 92.39%, respectively considering the cylindrical neighborhood method, all point features and using RF classifier. Generally, the overall classification accuracy is promising and ranging from 78% to 95% according to different neighborhood selection methods, point features set and the ML algorithm. Random forest was the best ML algorithm among the three classifiers with high scores reaching to 95% overall accuracy considering the modified Z coordinate.