SPATIAL ANALYSIS FOR OUTLIER REMOVAL FROM LIDAR DATA

A. A. Matkan [a], M. Hajeb [a*], B. Mirbagheri [a], S. Sadeghian [b], M. Ahmadi [a]

[a] Department of Remote Sensing & GIS, Shahid Beheshti University, Evin, Tehran, Iran.
(a-matkan, M_Hajeb,B_Mirbagheri, mos.ahmady)@sbu.ac.ir
[b] Geomatics College of National Cartographic Center, Meraj street, Azadi square, Tehran, Iran. Sadeghian@ncc.org.ir

KEY WORDS: Spatial analysis, Cross-validation, LiDAR, Outlier detection.

ABSTRACT:

Outlier detection in LiDAR point clouds is a necessary process before the subsequent modelling. So far, many studies have been done in order to remove the outliers from LiDAR data. Some of the existing algorithms require ancillary data such as topographic map, multiple laser returns or intensity data which may not be available, and some deal only with the single isolated outliers. This is an attempt to present an algorithm to remove both the single and cluster types of outliers, by exclusively use of the last return data. The outliers will be removed by spatial analyzing of LiDAR point clouds in a hierarchical scheme that is uses a cross-validation technique. The algorithm is tested on a dataset including many single and cluster outliers. Our algorithm can deal with both the irregular LiDAR point clouds and the regular grid data. Experimental results show that the presented algorithm almost completely detects both the single and cluster outliers, but some inlier points are wrongly removed as outlier. An accuracy assessment indicated 0.018% Error α and, 0.352% Error β that are very satisfactory.

1. INTRODUCTION

LiDAR is a relatively new technology which is an alternative to field surveying and photogrammetric techniques to collect elevation data. This technology is able to provide high accuracy three-dimensional data with reasonable cost and time. 3-D data acquired by this technology are applicable in 3-D urban modeling, DTM generation, mapping and etc. Although LiDAR data present high height accuracy, there are some defects in them leading to some disadvantages in output of next processes. Hence, these errors should be removed before performing any process. One of the most important errors in LiDAR data is the outlier points. Measured elevation for these points is unreasonably more or less from their neighboring points. The outliers are mainly measurements that do not obey the local surface geometry and do not belong to the topography of the interested area. In some references in the literature, the points with too high elevation values are named "positive outliers" and the points with too low elevation values are named "negative outliers", so we used these terms, too. The outliers can be caused from different sources. Positive outliers are resulted from hitting laser beams to birds or other suspended objects at high altitude. In connection with the negative outliers, it is believed that the laser beams be reflected several times among the glasses of buildings before they are detected, just like the multi-path effect of GPS. These specular reflections result in a longer travel time of the laser beam, and thus a lower elevation is calculated during post-flight processing. The negative outliers are often located at a few spots beside which there are tall buildings (Hu, 2003). If a dataset contains many outliers or unreliable points reflected from water features, the subsequent processes will not produce reliable results at neighboring regions (Hu, 2003). It is better to detect outlier points prior to the filtering procedure because they may have bad effects on the output of filtering algorithms (Baligh et al., 2011).

Positive outliers are easily detected and removed during the filtering process due to their high elevation. So, it is not necessary to remove them in a separate process. Despite this, the negative outlier detection and removal should be done before point cloud filtering because there is an assumption in this field that treats the lowest point in a local neighborhood as a bare-earth point (Sithole, 2005). However, this assumption will not work properly for low outlier points. Using it results in losing bare-earth points near to those outlier points and consequently causes a conical hole in obtained DTM.

Many outlier points can be discarded promptly if a prior knowledge about the terrain relief of the region of interest is available. These information can be found in topographic maps. Mabudi (2005) used minimum and maximum elevation of the interested area to remove both positive and negative outliers. This strategy does not work well in high relief area. In addition, these information may not always be available. Hence, some strategies have been considered in which no prior knowledge is required. Maximum acceptable difference between first and last laser returns is one of them. Logically, for a LiDAR point, the first return value should be more than the last one. Therefore, the result of subtracting the last return from first return should not be greater than a threshold which is determined considering the height accuracy of the LiDAR system. Matkan et al. (2014) used this strategy to remove the outliers. They set the threshold to twice the height accuracy of the LiDAR data. A major weakness of this approach is that it is not able to remove the negative outliers from the last return data as well as the positive outliers from the first return. Surface fitting is another approach in which a surface is fitted on a certain number of nearest neighbors of a point. The considered point is treated as an outlier if its distance from the surface is greater than a predefined value (Jiang and Ming 2006). Presented method in (Leslar et al., 2011) fits a quadratic curved surface to sections of LiDAR data. Generally, the surface fitting method does not work well in high relief area (Jiang and Ming 2006). In (Pang, 2011) a minimum covariance determinant (MCD) based multiple attributes model is proposed which used both range and intensity LiDAR data to extend traditional data processing methods from single attribute to multiple attribute. The connectivity based outlier factor is then defined which indicates the spatial neighborhood relationship of a point to its neighbors as an attribute. This attribute and the height attribute are extracted to organize a 2-D space. In the