A HIGHLY SCALABLE DATA MANAGEMENT SYSTEM FOR POINT CLOUD AND FULL WAVEFORM LiDAR DATA

A. V. Vo [1,*] , D. F. Laefer [2], M. Trifkovic [2], C. N. L. Hewage [1], M. Bertolotto [1],
N. A. Le-Khac [1], U. Ofterdinger [3]

[1] School of Computer Science, University College Dublin, Dublin 4, Ireland – anhvu.vo@ucd.ie,
chamin.lokugamhewage@ucdconnect.ie, (michela.bertolotto, an.lekhac)@ucd.ie
[2] Center for Urban Science and Progress, New York University, Brooklyn, New York, USA – (debra.laefer, mt3786)@nyu.edu
[3] School of Natural and Built Environment, Queen's University Belfast, Belfast, Northern Ireland, UK – u.ofterdinger@qub.ac.uk

KEY WORDS: LiDAR, Point Cloud, Full Waveform, Spatial Database, Big Data, Web Service, Web-based Visualisation

ABSTRACT:

The massive amounts of spatio-temporal information often present in LiDAR data sets make their storage, processing, and visualisation computationally demanding. There is an increasing need for systems and tools that support all the spatial and temporal components and the three-dimensional nature of these datasets for effortless retrieval and visualisation. In response to these needs, this paper presents a scalable, distributed database system that is designed explicitly for retrieving and viewing large LiDAR datasets on the web. The ultimate goal of the system is to provide rapid and convenient access to a large repository of LiDAR data hosted in a distributed computing platform. The system is composed of multiple, share-nothing nodes operating in parallel. Namely, each node is autonomous and has a dedicated set of processors and memory. The nodes communicate with each other via an interconnected network. The data management system presented in this paper is implemented based on Apache HBase, a distributed key-value datastore within the Hadoop eco-system. HBase is extended with new data encoding and indexing mechanisms to accommodate both the point cloud and the full waveform components of LiDAR data. The data can be consumed by any desktop or web application that communicates with the data repository using the HTTP protocol. The communication is enabled by a web servlet. In addition to the command line tool used for administration tasks, two web applications are presented to illustrate the types of user-facing applications that can be coupled with the data system.

1. BACKGROUND

Laser scanning or Light Detection And Ranging (LiDAR) is one of the latest technologies for airborne topographic mapping. The most common type of airborne LiDAR uses a pulsed laser mounted on an airplane to measure the range to target objects on the ground at a high frequency (i.e. up to exceeding 1 MHz). The range measurement is based on the time of flight of the laser pulse reaching the target objects and returning. The ranging sensor works in conjunction with a scanning mechanism that sweeps the laser beam across a space to cover the target objects. In addition, a position and orientation system (i.e. POS) is required on an airborne LiDAR platform to capture the platform's motion. Post-flight, these data components from the ranging sensor, the scanning mechanism, and the POS are integrated to derive a three-dimensional (3D) representation of the scanned scene. Data derived from laser scanning is typically in the format of a point cloud, which is a spatially coherent group of discrete sampling points taken from the surfaces being mapped. Each sampling point consists of a tuple of the point's coordinates (e.g. x, y, z), a timestamp recording the acquisition time, and potentially several other scalar and vector attributes (e.g. signal intensity, colours, normal vector). Some LiDAR sensors allow recording and exporting full waveform (FWF) data, which contain more complete signal backscatter emitted from and received by the sensors (i.e. waveforms) (Mallet and Bretar, 2009). The waveforms are often digitised and recorded as individual timeseries of signal amplitude. The waveforms, together with auxiliary datasets (e.g. the sensor's position and orientation) offer valuable insight into the data's origin. Retention of full waveform data is becoming a more common practice in LiDAR data acquisition.

Due to their high potential value and wide range of applications (US Geological Survey, 2020), airborne LiDAR data are being acquired at massive scales. Many countries have fully or partially completed their national LiDAR maps. They include Canada, England, Denmark, Estonia, Finland, Japan, the Netherlands, Slovenia, and Switzerland. In the United States, the US Geological Survey is leading the 3D Elevation Program (3DEP), a decade-long national project that aims to complete the acquisition of nationwide LiDAR mapping by 2023. Those large-scale LiDAR acquisition projects generate huge amounts of LiDAR data. For example, as of 2019, over 12 trillion LiDAR data points were made available to the public by 3DEP. In the Netherlands, the nationwide LiDAR acquisition is repeated every 6 years with increasing point densities. The first national LiDAR scan of the Netherlands (Actueel Hoogtebestand Nederland 1 - AHN1) completed in 2003 with most of the country mapped at a density under 1 points/m^2. The second scan, AHN2, completed in 2012 and resulted in 640 billion data points at a density of 6-10 points/m^2. AHN3 was launched in 2014 and due to complete by 2019 (Riveiro and Lindenbergh, 2020). AHN4 is set to start in 2020 and to complete by 2023 (https://www.ahn.nl/ahn-4). Those few selected examples of large-scale LiDAR projects illustrate the massive amounts of LiDAR data being collected worldwide.

Full waveform LiDAR data are more demanding in terms of storage and processing. FWF LiDAR has mostly been collected at regional scales. In the US, FWF data are collected by the National Ecological Observatory Network (NEON) and the

* Corresponding author