Assessment of the Possibility of Using the SAND Library for Processing Point Clouds in the Big Data Environment on the Example of UAV-LiDAR Data for a Forest

Dorota Marczykowska, Jaroslaw Czajka, Wojciech Ostrowski and Magdalena Pilarska-Mazurek (Poland)

Key words: Engineering survey; Geoinformation/GI; Laser scanning; Remote sensing; Data processing tools; Point clouds; Forestry; UAV-LiDAR

SUMMARY

UAV-LiDAR surveys deliver very dense point clouds, with over 300 pts/m2. This density is close to that of terrestrial laser scanning (TLS) and allows to use UAV data in almost the same way as TLS in some applications. One of the application areas is forest management, where high density of point cloud enables automation of many forest inventory and planning processes, but TLS surveys are nearly impossible to do due to dense vegetation and very low efficiency of measurements. Together with the Dragonfly Vision company, we are carrying out a project in which a solution using UAV-LiDAR data for exactly that purpose is created. Such a dense point cloud for large areas means billions of points to process. It requires a lot of computing power. A tool combining GPU and the benefits of distributed environment has very high potential to improve analysis of this kind of data. SAND library is a part of CENAGIS project, developed by the Faculty of Geodesy and Cartography at the Warsaw University of Technology. This python library allows to analyze and process point clouds using GPU and is designed to work with pySpark. Appropriate experiments were carried out to determine the possibility of using SAND library and the calculation time needed for large data sets. The influence of the set size on the total computation time was investigated. The possibilities of creating canopy height model (CHM) and determining forest structural attributes were also checked. In this study multiple ways to create height model using SAND were tested, such as: maximal raster value, planes fitting, and pit-free-like approach with processing data in layers. Capability of determining single tree features based on point cloud, like height or 2D area of canopy's extent was also investigated. The library implements few clustering methods and statistics for clusters, which can be used to analyze individual trees. All these functions show not only that manipulating a point cloud of forest by SAND and pySpark is possible, but also that these instruments together enable creation of comprehensive tools for foresters. These tools are able to process data obtained for large areas in a reasonable time. Thanks to the variety of methods, algorithms and the availability of a wide parameterization, SAND library also provides an excellent

opportunity for research on the algorithms for creating individual products related to forest management.