Geophysical Research Abstracts
Vol. 17, EGU2015-7689, 2015
EGU General Assembly 2015

A quality control system for digital elevation data

Thomas Knudsen, Simon Kokkendorf, Andrew Flatman, Thorbjørn Nielsen, Brigitte Rosenkranz, and Kristian Keller

Danish Geodata Agency, Copenhagen, Denmark (thokn@gst.dk)

In connection with the introduction of a new version of the Danish national coverage Digital Elevation Model (DK-DEM), the Danish Geodata Agency has developed a comprehensive quality control (QC) and metadata production (MP) system for LiDAR point cloud data.

The architecture of the system reflects its origin in a national mapping organization where raw data deliveries are typically outsourced to external suppliers. It also reflects a design decision of aiming at, whenever conceivable, doing full spatial coverage tests, rather than scattered sample checks.

Hence, the QC procedure is split in two phases: A *reception* phase and an *acceptance* phase.

The primary aim of the reception phase is to do a quick assessment of things that can typically go wrong, and which are relatively simple to check: Data coverage, data density, strip adjustment.

If a data delivery passes the reception phase, the QC continues with the acceptance phase, which checks five different aspects of the point cloud data:

1. Vertical accuracy
2. Vertical precision
3. Horizontal accuracy
4. Horizontal precision
5. Point classification correctness

The vertical descriptors are comparatively simple to measure: The vertical accuracy is checked by direct comparison with previously surveyed patches. The vertical precision is derived from the observed variance on well defined flat surface patches. These patches are automatically derived from the road centerlines registered in FOT, the official Danish map data base.

The horizontal descriptors are less straightforward to measure, since potential reference material for direct comparison is typically expected to be less accurate than the LiDAR data.

The solution selected is to compare photogrammetrically derived roof centerlines from FOT with LiDAR derived roof centerlines. These are constructed by taking the 3D Hough transform of a point cloud patch defined by the photogrammetrical roof polygon. The LiDAR derived roof centerline is then the intersection line of the two primary planes of the transformed data.

Since the photogrammetrical and the LiDAR derived roof centerline sets are independently derived, a low RMS difference indicates that *both* data sets are of very high accuracy.

The horizontal precision is derived by doing a similar comparison between LiDAR derived roof centerlines in the overlap zone of neighbouring flight strips.

Contrary to the vertical and horizontal descriptors, the point classification correctness is neither geometric, nor well defined. In this case we must resolve by introducing a human in the loop and presenting data in a form that is as useful as possible to this human. Hence, the QC system produces maps of suspicious patterns such as

- Vegetation below buildings
- Points classified as buildings where no building is registered in the map data base

- Building polygons from the map data base without any building points
- Buildings on roads

All elements of the QC process is carried out in smaller tiles (typically 1 km $\times$ 1 km) and hence trivially parallelizable. Results from the parallel executing processes are collected in a geospatial data base system (PostGIS) and the progress can be analyzed and visualized in a desktop GIS while the processes run.

Implementation wise, the system is based on open source components, primarily from the OSGeo stack (GDAL, PostGIS, QGIS, NumPy, SciPy, etc.). The system specific code is also being open sourced. This open source distribution philosophy supports the parallel execution paradigm, since all available hardware can be utilized without any licensing problems.

As yet, the system has only been used for QC of the first part of a new Danish elevation model. The experience has, however, been very positive. Especially notable is the utility of doing full spatial coverage tests (rather than scattered sample checks). This means that error detection and error reports are exactly as spatial as the point cloud data they concern. This makes it very easy for both data receiver and data provider, to discuss and reason about the nature and causes of irregularities.