Massive Point Clouds

The progress of computer technology still obeys the half-a-century old Moore's law. Likewise, the capacity of sensors, the number and size of pixels in cameras and even the number of nozzles on print heads doubles nearly every two years. Yet the amount of data today's sensors produce is growing faster than the processing and storage capacity of Data Base Management Systems (DBMS). As a result, the full potential of point cloud data such as the x,y-related heights from airborne Lidar or the x,y,z coordinates from terrestrial laser scanners remains unexploited.

Lidar Digital Terrain Models (DTM) may be so detailed that their volume may exceed hundreds of billions of points. For example, the DTM of The Netherlands (AHN 2) has a density of one point per $50cm_2$. To deal with the limits of their software, users may order reduced and aggregated datasets. They might be happy with high-definition point clouds but face in-house limitations which prevent the data details being fully exploited. Furthermore, transport from depository to the user may end up as a lengthy and cumbersome voyage since dissemination facilities are out of whack with massive data volumes.

To balance massiveness and limits due to processing and dissemination, clever algorithms are urgently needed. Several approaches can be distinguished. One is to lower the number of digits needed to represent the x,y,h triplet. For example, the points can be arranged in geographical blocks in which the heights vary only slightly. When the height values vary between 12.07 and 12.92, the x values between 174,200 and 174,850 and the y values between 361,700 and 362,600, the first digits may be omitted and stored just once for each block. When using the data, these digits can be added by the software. When the density is 4 points/m_2, this enables huge savings in terms of storage space and transport time.

Another approach is to convert the point cloud into a raster and to use JPEG compression just as is done for imagery, i.e. the raster is divided into tiles of eight by eight cells and then transformed to the frequency domain. This can be done either as lossy – when heights cannot be restored exactly – or lossless compression.

A third method is to perform a segmentation of the point cloud, i.e. areas with similar or smoothly varying heights are detected using region growing, a well-known method in pattern recognition. Next the vertices of boundaries are stored, together with the parameters of the polynomial describing the height variations. Meadows, streets and football fields may all be represented by a horizontal or a tilted plane. Such a parametric description allows areas containing thousands or even millions of points to be efficiently represented by a few values. Suppliers who provide such intelligently crafted point clouds as well as the associated software may be doing their clients a huge favour.