Content

Management of massive point cloud data: wet and dry

Editorial

P.J.M. van Oosterom, TU Delft, the Netherlands
M.G. Vosselman, University of Twente/ITC, the Netherlands
Th.A.G.P. van Dijk, Deltares/University of Twente, the Netherlands
M. Uitentuis, ATLIS, Bunnik, the Netherlands

Seminar on 26 November 2009

This publication contains a selection of papers that result of the seminar ‘Management of massive point cloud data: wet and dry’ on Thursday 26 November 2009 at Oracle, De Meern, the Netherlands. This seminar was jointly organized by the subcommissions 'Marine Geodesy' and 'Core Spatial Data' of the Netherlands Geodetic Commission (part of the Royal Netherlands Academy of Arts and Sciences) and the SIM (Spatial Information Management) commission of the OGh (Oracle gebruikersclub Holland).

The theme of the seminar was about the challenges caused by the ever increasing amount of data that is generated by modern sensor systems, both in the wet and dry sectors. To reach a broader audience, both marine (multi-beam echo) and land (LiDAR) data was included. The audience did originate from the Netherlands government/authorities (Netherlands Hydrographic Office, Rijkswaterstaat, Kadaster), research organizations (institutes, universities), and industry (wet and dry data acquisition, Geo-ICT).

It was decided to combine both the wet and dry perspective, with the focus on data management of large point clouds, in order to discover the common challenges to be included in the research agenda. Current solutions may not be sufficient for future needs and therefore new software (data structures, algorithms) and hardware (parallel computing, clusters, grids) need to be investigated.

Prior to the seminar, authors submitted an extended abstract. All 15 presentations as presented are on-line at http://www.gdmc.nl/events/pointclouds. After the seminar the authors were asked to submit full papers and include and reflect the discussions during the seminar. The full papers were then reviewed by the editors and finally the authors submitted revised papers. This publication contains the final selection papers, 10 in total.

Overview of selected papers

The paper by authors of the Hydrographic Office of the Royal Netherlands Navy (NLHO), *Righolt, Schaap, Dorst, and Vos*, focuses on the need for improved and more automated processes, algorithms and visualisation techniques for the validation of massive point cloud data in the Hydrographic field. Large multi beam point clouds may contain erroneous points. The process to find these points is very time consuming. First the authors give an overview of the historic and current use of bathymetric data at the NLHO and the general aspects of data validation. This is followed by an explanation how artefacts in the data are handled. This leads to recommendations for research were both the academic as private sector should cooperate to improve current processing and bathymetric data management solutions.

The combination of both wet and dry point clouds is addressed in the paper by *Kodde* of Fugro-Inpark. An overview of various land-based methods for the acquisition of point cloud data is presented, with some additional attention for Fugro’s own DRIVE-MAP and FLI-MAP systems.

That the point cloud data model is a bit more complicated than just x, y, z values becomes clear when the LAS-format is presented, which allows in addition to store properties such as: intensity, return number, scan direction, classification, user data, r, g, b values, etc. For better dissemination of point cloud data this paper proposes a standardized Web Point Clouds Service (WPCS) in analogy with existing OGC web services. This service should support data streaming and the concept of Level of Details.

The paper by *Swart* describes the properties of the Dutch digital terrain model AHN-2 and its potential use for water management purposes. With some 8 – 10 points per square meter, the AHN-2 is likely to be the most detailed nationwide digital terrain model (DTM). Clearly, this also implies that the data amounts are enormous. The paper explains the need for such high point densities for various tasks of the water boards and the progress that has been made so far in the realisation of this DTM. Profiles drawn across dikes are used for stability analysis as well as for mapping toe lines. Several applications are reported to still work with height images instead of the original point clouds due to a lack of suitable software for handling the massive point clouds.

Simons, Amiri-Simkooei, Siemes and Snellen (TU Delft) describe the recent developments in the processing of multi-beam echo sounder data by focussing on two applications, supported by case study results. Firstly, they emphasise the importance of correcting the MBES measurements for errors caused by the unknown sound speed in the water column (i.e. beam steering and conversions of beam angle and travel-time combinations), which may cause ‘droopies’ or ‘smileys’ in the across track direction. Their method exploits the redundancy of echo soundings in the overlap region between two adjacent swathes, so that no additional sound speed profile measurements are required. Simons et al. present a model that predicts the two-way travel times. By minimising the difference between the measurements and model predictions, optimised depths (or sounds speeds to recalculate the bathymetry) are obtained per ping. A case study of the Meuse near Rotterdam Harbour demonstrates that the artifacts due to unknown velocity profiles are successfully removed from the data (in real time) with an accuracy by which shallow beam-trawl tracks remain visible. Secondly, the variation in backscatter intensity, depending on incidence angle and thereby representing variations in seabed morphology and sediment characteristics, is used in the classification of sea bed sediments. Simons et al. present a new approach that uses backscatter measurements per angle and accounts for the ping-to-ping variability in backscatter intensity. Linear curve-fitting is used to obtain the best model to fit to the backscatter measurements of different sediment types and the difference between measurements and model results are minimised. Acoustic classes are identified statistically (multiple-hypothesis testing). Case studies of the Cleaver Bank, North Sea, Netherlands and a small section of the river Waal show that sediment classification results are in good agreement with the sea bed samples and represent the grain size distributions mapped by geologists.

Ledoux (TU Delft) proposes a new “Triangle” data type to store TINs in spatial databases. In this star-based data structure only the star of each point is stored. The introduction of the paper is focussed on the aspect that computers have great difficulties dealing with very large datasets that exceed the capacity of their main memory. In addition the need to reconstruct surfaces with a triangulated irregular network (TIN) respecting the Delaunay criterion is mentioned. Then an overview is given how TINs are handled in software that is on the market and used in the academic world and their current limitations. Hugo Ledoux introduces the Triangle data type and compares this with the currently used way to manage TINs. The data structure is significantly different from what is usually available in a DBMS. It appears that the Triangle data structure is quite promising in terms of data handling and data space efficiency.

Masry and Schwartzberg of CARIS Ltd describe their system for the management and visualisation of massive point clouds resulting from bathymetric surveys. Point clouds are imported in several open file formats and stored in an octree like structure. The point clouds in the octree leaves can be loaded, processed and put back sequentially. A hierarchical data structure allows visualising arbitrary subsets of the point cloud with 20 – 30 frames per second. Data compression is implemented lossless to maintain the high point location accuracy.

One of the great advantages of massive point clouds is the easy insight in spatial (three-dimensional) structures within the data, especially when visualised in interactive 3D fly-through displays. Because of the extremely large point clouds acquired with state-of-the-art techniques, even modern software and hardware cannot handle these amounts of data well. *De Haan* (TU Delft) highlights several rendering techniques that originate from graphic rendering engines used in games and flight simulators, underpinned by experimental visualisation results with LiDAR (Light Detection And Ranging) data. Real-time visualisation of these amounts of data is a constant play between smart structuring of the data, optimising levels of detail to viewing distances and using efficient ways of paging data in and out of memory. Examples of the AHN2 data set are used to show how point clouds can be used to best extract the terrains, thereby avoiding artefacts such as ‘draping’ and holes. The balance between rendering performance and visual quality is discussed. Future challenges are to develop automated approaches for more flexible data sampling and selection of levels of detail.

For a data producer such as Tele Atlas (part of TomTom) point cloud data offer a number of opportunities to create and maintain their data sets. *Coppens* argues that these opportunities are there for collecting and maintaining the road geometry (better accuracy), road attribution (pavement, lanes, tunnels), road furniture (signs, curbs, trees), and especially for creating 3D models. For various applications, such as personal navigation, car navigation, GIS, Internet LBS, the importance of the above mentioned data components vary. However, in order to make the opportunities realities, there is a number of challenges which will have to be addressed and solved: to keep representations in line with reality (huge amount of data to be processed in timely manner), to have different representations of reality (consistent multiple level of details), to have high positional accuracy (even sufficient for advanced driving assistance systems, ADAS) and to have usable data for applications (not too big, not too detailed).

The paper of *Broen* (Kongsberg Maritime) also addresses the requirement more sophisticated and automated methods of processing, i.e. validating, correcting and analysing the measurements, of the large point cloud data acquired by modern multibeam echo sounders. Broen presents their Seafloor Information System (SIS) of which the Grid-Engine is a core technology that processes and displays the data in real-time. Digital terrain models (DTMs) are built, whereby different resolutions are coupled to certain scales of the terrain maps. The paper includes a full overview, from multibeam data acquisition, including additional parameters, and real-time ping processing via depth processing to displaying and interpreting terrain models. Grid-Engine is successful in the application of multibeam data but was also applied to DTMs of laser data. The near-future release of the Dynamic Grid-Engine will automatically optimise settings in the gridding process and aims at even smarter memory allocations based on predicted vessel heading.

In the final paper, the concept of the Virtual Continuous Model (VCM) is introduced by *Spoelstra* of ATLIS. The VCM technology is intended to support Hydrographic Offices in better managing the ever-growing amount of data produced by multi-beam echo sounders, while at the same time users expect a faster delivery of products (more up-to-date). Using VCM the data is only stored once in the archive of survey data. The users can define their own digital terrain models (DTM), which are all based on the same and shared data, managed in an Oracle spatial database. These models can be defined for various products (including historic data) and different levels of detail, while it always remains possible to trace back the original sources; e.g. for liability.

Closing words

The organizers/editors would like to thank all persons involved in the seminar: first of all the authors of the papers in this publication for their efforts, next the presenters of contributions not included in this publication (but with presentations on-line available), the audience for the active perception and discussion during the seminar, the KNAW/NCG for supporting the seminar and making this publication possible, and last but not least: Oracle, the Netherlands for hosting the seminar. From our side it was a pleasure to be involved in both the seminar and the publication afterwards.

Enjoy reading the contributions in this publication!