Management of massive point cloud data: wet and dry

Thursday, November 26, 2009
Oracle, De Meern, The Netherlands

Seminar, 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)

Abstracts of presentations

Jan Schaap (Dienst der Hydrografie)

Needle in a haystack

The Hydrographic Office of the Royal Netherlands Navy (NLHO) is responsible for charting the Dutch sea areas in Europe and the Caribbean, with the primary purpose of safe navigation. The products of NLHO include Electronic Navigational Charts (ENC's), paper charts and other digital and paper Nautical Publications.

The two focuses of this presentation are (1) how to filter out the relevant "needle" from the huge amount of data from the multi-beam echo-sounder (MBES), and also (2) how to discover what else is in the "haystack". With respect to focus (1), filtering the “needles” in an automated way that is relevant for navigation requires expert systems. These should be fed with a huge amount of soundings, and they should have knowledge about features like depth contours and wrecks/obstructions on the seafloor.

Focus (2) deals with systematic artifacts, like heave problems, that are analyzed better when more data are loaded during analysis of MBES data. (Nowadays processing software starts with thinning and binning the data, for the sake of diminishing the size of the datasets.) Combining several surveys in full detail gives better understanding of sea floor dynamics. During the processing of MBES data, the limits of the hardware configuration are felt, when attempts are made to visualize the data and to extract features (semi-) automatically.

In the future, the amount of collected data under water will increase further. The NLHO is attentively following the developments in water column imaging, backscatter-based bottom classification, and artificial aperture processing of AUV (Autonomous Underwater Vehicle) data.

Martin Kodde (Fugro-Inpark)

Pointillism - The art of capturing the earth in points and the challenge of exhibiting them

Fugro collects and interprets data related to the earth's surface and the soils and rocks beneath and provides advice, for purposes related to the oil and gas industry, the mining industry and the construction industry. Fugro operates around the world at sea, on land and from the air, using professional, highly-specialized staff and advanced technologies and systems.

The division Geospatial Services within Fugro operates world wide to produce accurate geospatial datasets by utilizing a wide variety of sensors. Much of Fugro´s work is related to capturing datasets of the bottom of the ocean, especially in deep waters. Onshore, Fugro applies both aerial and terrestrial sensors, ranging from satellite remote sensing to millimetre-precision point measurements with tachyometry.

Many of Fugro's sensors deliver point clouds as an end product. Although the source of these datasets may differ considerably, they share the characteristic that data volumes are huge and ever increasing. In addition, all point clouds pose interesting challenges for efficient processing and viewing of the data. Since a point cloud is seldom a product on itself, the generation of derivative products becomes more and more important.

In this presentation, Fugro will show the audience the art of surveying with some of its most modern point cloud sensors. It will be shown how Fugro currently processes, visualizes and distributes point clouds. Finally, some current challenges and wishes for handling large point clouds are given from the perspective of a company where terabytes of data are produced every day.

Rens Swart (Swart*vast* - Het Waterschapshuis)

How to handle the Up-to-date Height Model of the Netherlands: detailed, precise, but so huge!

The Up-to-date Height Model of The Netherlands (Actueel Hoogtebestand Nederland, AHN) is a digital terrain model of the Netherlands, owned by the 27 water boards and Rijkswaterstaat. It was commenced in 1997 as a joined initiative to obtain a countrywide height model primarily suitable for water management, using the newly available remote sensing technique of laser altimetry. In 2004 it was complete, covering The Netherlands with a density of one height measurement per 1 to 16 square meters.

Partly induced by the increasing requirements of the water boards, partly by offerings of the business, in 2006 the steering committee decided for a new AHN with upgraded specifications. In a cycle of five years, the whole of The Netherlands will be covered by a new height model with a precision in height of 5 cm standard deviation and 5 cm systematic error and a density of 8 to 10 points per sq. m. Also new is the way the requirements are defined in terms of user needs, and the tendering of both the data acquisition and the quality control of the products.

Till the advent of the new AHN-2, many water boards undertook separate corridor laser mapping in order to acquire the high-precision high-density height data necessary for dike management. The upgraded specifications of AHN-2 made it possible to unify the acquisition of height data for both water management and dike management. However, this also causes the amount of data to increase dramatically. In practice often the 0.5 m-grid is used instead of the point cloud. Even then many users must be convinced that this data is, in a way, better than profiles acquired with terrestrial techniques and offer unique opportunities. Apart from that, distribution and use of the distinguished products, including aerial photography, poses a demanding challenge for the users. This challenge will even increase if the AHN is supplied for free and eventually becomes a core spatial dataset.

Dick Simons (TU Delft, Acoustic Remote Sensing)

Exploiting the full potential of the multi-beam echo-sounder for on-line surveying

The multi-beam echo-sounder (MBES) system allows for unprecedented performance in mapping the sea- and river-floors with a 100% coverage. It measures with a single acoustic ping the water depths along a wide swathe perpendicular to the ship track, using the travel times of the echo signals received in the acoustical beams. MBES ping rates depend on the water depth, but typically are several 10’s of Hz. The MBES opening angle is about 150 degrees and contains as much as several 100 narrow beams, thereby providing high-resolution bathymetric maps. However, the amount of data acquired per day of surveying is at least a few gigabytes. Currently the bathymetry is obtained on-line, allowing for establishing a bathymetric map on the fly.

Frequently, however, knowledge about the water column sound speed profile is insufficient for correctly converting the measured travel times to depths. We have developed methods that allow for estimating both the correct bathymetry and the prevailing sound speeds from the MBES measurements by searching for those sound speed profiles that minimize the differences in water depths along overlapping parts of adjacent swathes. This method can be applied as soon as data along overlapping swathes become available. For this semi-online processing, efficient minimization approaches have been implemented.

In the presentation we show that when the available redundancy in the depth measurements is exploited, improved bathymetric maps and water the sound speed profile can be estimated simultaneously. The performance of the method is demonstrated by applying it to real MBES data acquired in the Maasgeul.

In addition to the travel times, the MBES also provides measurements of the backscatter strengths in each beam, which are known to contain information about the sediment types. However, the process of extracting sediment type from the backscatter measurements requires dedicated processing steps. A series of MBES classification approaches have been developed both by commercial companies and universities, which currently can only be applied in a post-processing step and are not yet automated.

We show results of the application of a model-based classification method that employs the MBES backscatter data and discriminates between sediments in the most optimal way. The method has been applied for classification of sediments in a large number of areas. Here, we will show classification results for parts of the river Waal and the North Sea. A future development in this field might be to fully exploit the entire time series per beam. The advantage of obtaining more information, however, is counteracted by again a significant increase in data rate.

Hugo Ledoux (TU Delft, GIS Technology)

Storage and analysis of massive TINs in a DBMS

One of the main problems with LIDAR datasets is that while they provide us with unprecedented precision, computers have great many problems dealing with very large datasets that exceed the capacity of their main memory. With the software tools currently available to practitioners, it is indeed very difficult to simply visualize complete data sets, and the manipulation and processing of the data is nearly impossible. We can summarize the situation by stating that, as far a point cloud datasets are concerned, advances in technologies to collect datasets are by far superior to our ability to process them. This obviously hinders the use of the data by practitioners. Examples of LIDAR processes useful for many applications are: derivation of slope/aspect, conversion to a grid format, control of double points, calculations of area/volumes, viewshed analysis, creation of simplified DTM, extraction of basins, etc.

While simply storing millions of unconnected points in a DBMS is no problem, the processing of LIDAR datasets needs more: we must be able to reconstruct the surface represented by the points, and we must also be able to access this surface to manipulate it (e.g. adding/removing new samples), and also to derive values from it. The fastest way to reconstruct such a surface is arguably with a triangulated irregular network (TIN).

I will discuss briefly the main challenges involved when one wants to create and store TINs in a DBMS, review the possible solution, and present the solution that the GISt group at TU Delft plans on testing.

Mark Masry, Peter Schwarzberg (CARIS)

Marine high density data management and visualization

Several different engineering disciplines make use of massive collections of point cloud data. Both LiDAR and multi-beam sonar systems, among others, generate these types of data. Because the points can be distributed randomly within a volume, it can be difficult to spatially index, store and visualize this type of data. Furthermore, the relative novelty of point cloud data means that workflows for processing it have not yet been firmly established in the industry. CARIS has developed a robust and flexible system for managing massive point clouds. The system can spatially index well over 1 billion points in such a way that they can be queried and processed efficiently. Furthermore, the technology allows the stored points to be visualized interactively, even over a network, and is already integrated with our bathymetric data processing pipeline. This presentation will present some aspects of the point cloud storage system and provide several practical examples of its use in existing applications. The use of different storage backends, including both file systems and relational databases will also be discussed.

Dirk Voets (Imagen)

Visualization and analysis of massive point clouds; tackling the issues, now and in the future

Point clouds obviously represent enormous amounts of data, and therefore analyzing and visualizing these datasets is a true challenge. ERDAS and Leica have many years of experience creating these point clouds, but visualizing them is something that we have taken up fairly recently. The new approach there is not to see point clouds as 'an irregular raster model', or as a triangulated irregular network, but as a true own datatype in itself. This has proven to be more fruitful than squeezing point clouds in a TIN structure.

Another thing that ERDAS has actively picked up is the intense use of open source activities in this field. The Open Source community has actively worked towards solutions in this field. The LIBLAS library for storing point clouds in the LAS format is a very good example of that. Instead of coming up with own solutions, thereby stove piping the field of play, ERDAS has adopted the LAS format, and is working with the open source community to better tackle the implicit issues of handling and visualizing these point clouds.

Gerwin de Haan (TU Delft, Computer Graphics and CAD/CAM)

Scalable visualization of massive point clouds

Graphical renderings of "raw" point clouds can be visually attractive. When combined, the many individual points and their attributes convey spatial structures, especially when viewed in a fluent 3D fly-through and on a stereoscopic display.

For such interactive visualization, a sufficiently responsive graphics update is essential. Where many software packages support simple rendering of smaller (subsets of) point clouds, the size of point clouds currently acquired easily surpasses the rendering capabilities, even on a modern graphics card.

We addressed this issue in recent experiments on exploring Aerial LiDAR datasets in our Virtual Reality systems. We apply out-of-core terrain rendering techniques from graphics engines used in games and flight-simulators.

In this short talk I will highlight some of these techniques and results, and discuss challenges in balancing rendering performance versus visual quality.

Sven Coppens (Tele Atlas / TomTom)

Data management requirements of large terrestrial point clouds, supporting feature extraction, quality information and data maintenance

To map the world Tele Atlas is continuously looking for new source material to improve the quality of the database, to extend the coverage more quickly, to keep it up to date and all this with an acceptable cost. Hereby point cloud data is becoming more and more a source of interest to reach those goals. Mapping reality means modelling hundreds and thousands types of features/attributes and in this session an overview will be given where point clouds can become an added value and we'll also link them to the business areas requesting them. In the last part we'll indicate that cloud data is not a silver bullet and brings along some challenges ones a process needs to industrialized and commercialized, which of course introduces new business opportunities and challenges academically.

Rens van den Bergh (Deltares)

Serious gaming and the need for geodata

A few years ago Deltares started using serious gaming techniques as a way to transfer knowledge. Several delta technology subjects have already profited from this approach after the first game was established.

This first game was made especially for levee patrollers. Levee patrollers survey the levees in a certain area when it is prone to flooding. Often there is a mix of professional and voluntary levee patrollers. As flood risks are quite rare in The Netherlands, it is difficult to be prepared for such a specific situation, especially when there are voluntary levee patrollers concerned. The game helps the patrollers to recognize when a levee shows signs of weakness and what procedures they should follow in order to communicate in an efficient and proper way to the so-called Action Centre. The Action Centre will take the right measures to prevent a breach when the right information is delivered in time.

The game has different types of polders in which levee patrollers are trained. These are all fictional. Sometimes a Water Board asks for a level that is identical to their own region. A Water Board Region is usually much larger than a game level. Therefore, the question is not only whether it is possible to copy a specific area into the game, but whether the game is still playable and has the same learning effects as well.

Stian Broen (Kongsberg)

Handling large amounts of multi-beam data in real time

The new multi-beam echo-sounders from Kongsberg Maritime can output up to 8000 depth samples per second. The Real Time Logging and Quality Assurance Software, SIS, is capable of making a terrain model from this data and presenting the results in a 3D map display also in real time.
The main tool used in this processing is named the GridEngine. This GridEngine will accept all these depths in real time, in addition to seabed image data which is 3-5 times more, and make a bin model out of them. This bin model is then processed in real time, and a Display Model is created.
This Display Model will create and maintain several grids in different resolution and detail. At the highest resolution, this Display Model presents the terrain model at full resolution. The operator can quickly change the map scale and get another view of the whole area in a lower resolution.
The GridEngine is primarily used to create terrain models from multi-beam data, but there are no restrictions so it has also been used to create terrain models from laser data. In this project the height of trees was displayed using the GridEngine and the Display Model.

This talk will explain in further detail how Kongsberg Maritime has developed this system, its use today and its capabilities in tomorrow's applications.

Koen Verbruggen, Archie Donovan (Geological Survey of Ireland),
Thomas Furey (Marine Institute, Ireland)

The INFOMAR project: mapping a seabed area 10 times the size of Ireland

Between 1999 and 2005, the Geological Survey of Ireland and the Marine Institute worked together on the € 32 M Irish National Seabed Survey (INSS) project with the purpose of mapping the Irish marine territory using a suite of remote sensing equipment, from multi-beam to seismic, achieving over 430,000 km.sq. of 100% map coverage. Ireland was the first country in the world to carry out such an extensive mapping project of their extended EEZ. The INSS was succeeded by the multiyear INFOMAR Programme, which is concentrating on mapping twenty-six selected priority bays, three sea areas and the fisheries-protection 'Biologically Sensitive Area'. It will then proceed to complete 100% mapping of the remainder of the EEZ. Designed to incorporate all elements of an integrated mapping programme, the key data acquisition includes hydrography, oceanographic, geological and heritage data. These datasets discharge Ireland's obligations under international treaties to which she is signatory and the uses of these data are vast and multipurpose: from management plans for inshore fishing, aquaculture, coastal protection and engineering works, to environmental impact assessments related and integrated coastal zone management. Airborne LiDAR (Light Detection And Ranging) and inshore- vessel surveys have also been carried out, giving detailed bathymetric, topographic and habitat information for the shallower waters and inshore areas. INFOMAR is a programme that is data-rich. Storage solutions were developed primarily as tape based systems under INSS to cope with c.16 TB of data and this has subsequently been migrated to a disk based SAN with over 100 TB of storage. The INFOMAR website http://www.infomar.ie provides open and free access to the data to anyone with an interest in the marine in a variety of formats and data can be accessed, viewed and downloaded in a variety of ways, for instance through a web mapping viewer, an Interactive Web Data Delivery System (IWDDS) and also via our Web Map Service (WMS). In line with the EU INSPIRE directive, INFOMAR is providing data in an interoperable manner, which strives to meet the relevant OGC and INSPIRE standards. In the last 3 years the project has delivered over 500 GB of data to over 700 clients via the web.

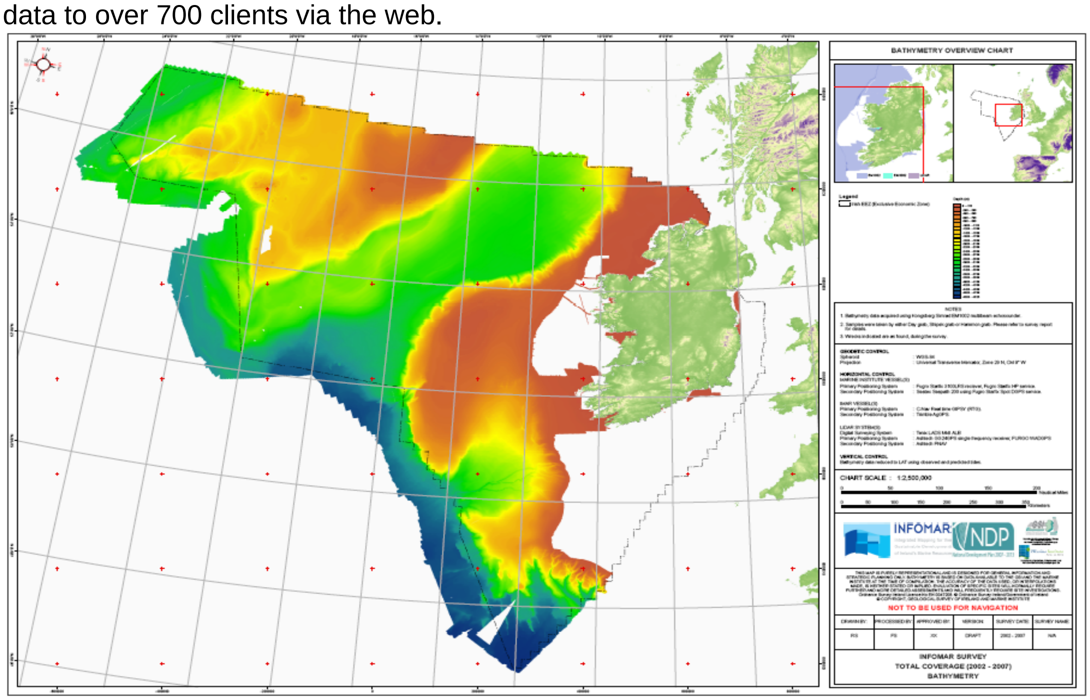

Chart showing Irish marine mapping coverage to end 2008 and the Irish Designated Area.

George Spoelstra (Atlis)

Virtualizing large digital terrain models

Chart producing agencies like Hydrographic Offices face major challenges today in keeping up with the ever-growing amount of data that is produced by modern echo sounders. At the same time users are demanding products faster and yet more reliable.

To support Hydrographic Offices and other chart producers, new innovative technologies need to be developed that will bring a great deal of efficiency into the current survey to chart work processes ensuring better products and a much faster time to market.

Organizations that manage bathymetric data often use the data for various reasons: safe navigation, morphology, off-shore planning to name a few. The challenge with managing large volumes of bathymetric data is to keep everybody happy without the need to build and manage models for each of these communities. A nautical cartographer needs a navigational safe model that is as up-to-date as possible whilst a morphologist might be interested in a series of historical models to analyze sediment transport. The problem even gets more complex as most organizations also have to archive their bathymetric models for liability reasons and also have to make all this data available as part of national and international data infrastructures.

Above challenge can be met by the introduction of virtual digital terrain models. The ATLIS SENS Bathymetry product is utilizing this concept in such a way that data only has to be stored once. All users use the same data and can define their own terrain models without the need of copying the often large volumes of data to an individual workspace. The number of models (both up-to-date and historical) is virtually unlimited thus providing a maximum of flexibility to expert organizations.
Using SENS Bathymetry, large digital terrain models are virtualized by only storing the model's definition. The underlying Oracle Spatial technology that holds the archive of survey data ensures fast retrieval of seamless models that can be used for a wide variety of applications.

Albert Godfrind (Oracle)

Using Oracle's new point cloud data type and the Oracle cluster functionality to manage massive point clouds

Making spatial data an integral part of an IT infrastructure, and applying mainstream IT technologies have been core goals of the Oracle Spatial technology.

After implementing dedicated data types for vector data (SDO_GEOMETRY) and raster data (SDO_GEORASTER), Oracle now also includes types dedicated to the storage of point clouds (SDO_PC) and TINs (SDO_TINs).

This session will examine the design of those data types and how their design, combined with standard database facilities such as clusters offers good scalability to applications.

Martien Ouwens, Aad Koppenhol (Sun Microsystems)

Exploiting parallel hardware (computer clusters / grids) and the link to massive data management software

One of the main problems with massive point cloud datasets is their sheer volume. Whilst they provide us with unprecedented amount of precise information, they challenge IT technologies with their massive volume. As a result a strong focus on high performance computing (HPC) evolved, initially most of the focus of HPC was centered around CPU performance. Present day HPC clusters are demanding increasingly higher rates of aggregate data throughput. Also today's clusters feature larger numbers of nodes with increased compute speeds. The higher clock rates and operations per clock cycle create an increased demand for local data on each node. In addition, InfiniBand and/or other high-speed, low-latency interconnects increase the data throughput available to each node.

During this session I will discuss briefly the CPU, memory, flash, interconnect Grid and I/O technology challenges and how they can contribute to the processing of these datasets.