Marine high-density data management and visualization

M.A. Masry, P. Schwartzberg
CARIS Ltd., Canada TU

Abstract

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 3D 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, as well as a workflow for procesing them. The system can efficiently store and index well over 1 billion points in such a way that they can be queried and processed efficiently. 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 paper will present aspects of CARIS's point cloud storage system and explain its utility in a workflow for processing marine data.

1. Introduction

Modern sensor systems produce huge volumes of data. Sonar and LiDAR systems and other acquisition systems generate large quantities of unorganized 3D points, each of which can have multiple attributes and can be treated as an independent location in 3D space. This paper is intended to provide an overview of the way in which this data type is used in the bathymetric data processing workflow and, in particular, the new point cloud storage system used by CARIS software applications to store and process this data.

Bathymetry is typically measured using multibeam systems – large, powerful sonar systems that cover large areas (swaths), and are typically deployed using survey patterns that cause their measurements to overlap. Modern multibeam systems are capable of generating vast quantities of data: a Kongsberg Simrad EM3002 single head system, for example, can record 40 samples per second, with 254 beams per sample. The sonar signals are typically processed by an online bottom-detection algorithm that returns only those signals that represent the intersection between the sonar pulse and the seafloor.

Sonar returns are typically stored in terms of range and angle. These pairs are then processed using software such as CARIS HIPS that combines them with time-stamped position, heading and attitude data recorded from GPS and motion sensors. The output is a series of multi-attributed, georeferenced 3D points. The horizontal positions are typically recorded in terms of latitude and longitude, and the vertical position in terms of depth. The attributes stored with each point often include measurements of the uncertainty of the point's horizontal and vertical locations. A typical, week-long shallow water survey can record well over 256 million 3D points, and surveys that produce well over 1 billion points are not uncommon. Storing and visualizing this collection of data is challenging.

In this paper we describe CARIS's system for handling high-volume point clouds, provide some visualization examples, and discuss our future goals for this work.

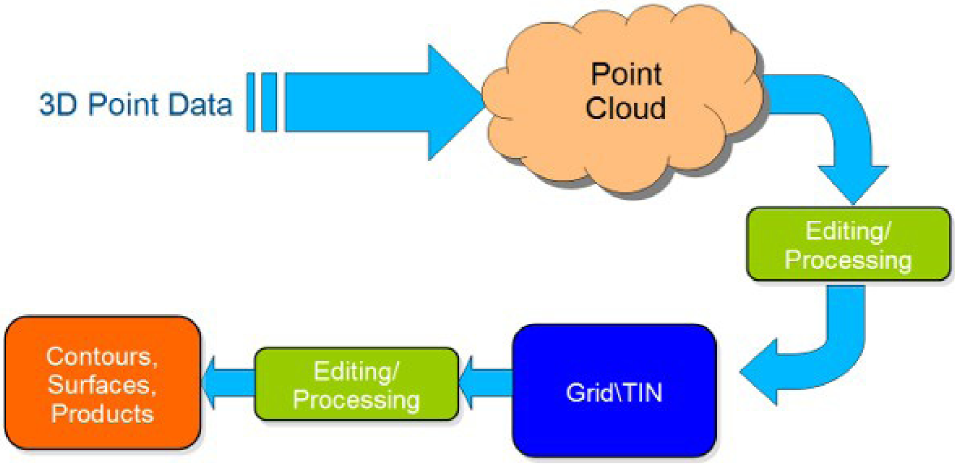

Figure 1. A typical bathymetric point post-processing workflow.

2. Downstream processing

The generation of georeferenced 3D point data from raw sonar data and time-stamped sensor information is largely automated, though user intervention is often required to correct poor input parametrizations or sensor noise. The next-generation 3D Point Cloud storage system described below, allows for sophisticated querying and visualization that can aid in this process. Three-dimensional points can be imported into either format directly from open file formats such as GSF, LAS, or any number of formatted text files, as well as through the above mentioned processing workflow for sonar data.

Bathymetric data processing typically focuses on the generation of 2.5 dimensional surfaces from processed 3D points. These surfaces can either be regular grids or Triangulated Irregular Networks (TINs). While they may be delivered as a final product, these often form the inputs for a range of subsequent processes. These processes include the contouring and sound selection processes that form an important part of the production process for nautical charts, as well as volume computations, interpolation tools and surface generalization algorithms. A typical post-processing workflow is shown in Figure 1.

3. The CARIS 3D point cloud

CARIS has implemented a sophisticated 3D point cloud data structure that can store billions of multi-attributed 3D points. The point cloud is the current data storage mechanism for 3D point data in CARIS Bathy DataBASE and will be incorporated into CARIS HIPS in a future release.

3.1 Data structure

Each point in the point cloud is stored using 64 bit double precision values, and can carry with it multiple attributes and flags; the attributes are grouped into bands and stored independently from the positional data to minimize I/O accesses, while the flags are stored along with the positional information. While the point cloud supports independent points, several points can also be grouped together into a single "multiple return" data point. The point cloud is bounded by a 3D spatial volume that is partitioned into multiple overlapping sub-volumes during the creation process. The points within each sub-volume are also ordered in a way that facilitates efficient visualization.

To create the point cloud, an insertion program analyzes an incoming stream of points to determine their aggregate spatial volume, sub-divides the volume appropriately, then determines which points lie within each sub-volume. The inserter can split each sub-volume recursively if the number of points in the sub-volume exceeds a given tolerance. The inserter also structures the points in such a way that new points can be added to the point cloud post-creation.

The points in each sub-volume are written to a storage device using the CARIS Spatial Archive (CSAR) framework. A diagram of a typical technology stack developed on top of the CSAR framework is shown in Figure 2. The framework provides developers with a set of tools that can be used to handle large volumes of multi-dimensional data by partitioning it into pieces called "chunks", each of which is given a unique key that can be used to retrieve it from storage devices such as standalone files or relational databases. The CSAR framework provides CARIS with a device-independent way to read, write and cache large data volumes in memory; it is used as part of the infrastructure of the CARIS Bathy DataBASE, allowing users to visualize and process data stored in the database directly, without having to first import the data.

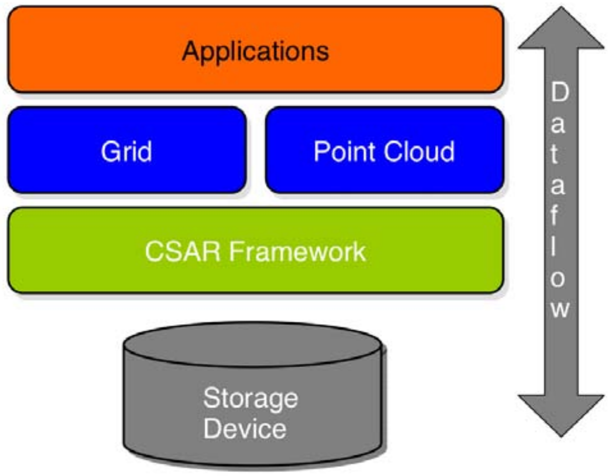

Figure 2. The application technology stack.

Chunks can be of different types, though in practice the chunks in the point cloud are homogeneous in type. Chunks are first created in a memory cache, then written to disk as the cache begins to fill. Each chunk passes through an I/O layer that serializes it into a contiguous memory block using a predefined serialization scheme, and optionally compresses it to minimize the size on disk. Because it is important that point locations be returned exactly, the system only implements lossless compression. When the chunk is read back from disk as part of a read call, its endian type is checked, then adjusted to the machine endian type if necessary. The chunk is then deserialized and loaded into the chunk cache, which may, depending on the amount of data currently loaded into the cache, cause another chunk to be written to disk to free up some available memory.

The point cloud creation process can specify the size of each chunk and the mapping of the points in each sub-volume into one or more chunks. Because I/O system performance generally varies with the number of I/O operations required to move data to and from storage, the number of points in each sub-volume generally has a significant effect on system performance, and must therefore be chosen carefully. The importer divides the stream spatially into sub-volumes using an octree-based volume subdivision scheme that begins with a tree of a single level. When the number of points in a sub-volume reaches a limit, it is split into 8 sub-volumes at the next lower level. This process continues until the number of points in a sub-volume is such that it can be stored in a single chunk.

The point cloud importer has been tested on point sets of larger than 1 billion multi-attributed 3D points, all of which were successfully imported into a single CARIS point cloud; preliminary results suggest that much higher numbers of points can be stored without further modifications to the structure.

3.2 Visualization and interaction

Structured indexing and spatial query mechanisms were implemented to allow subsets of the clouds (both points and attribution) to be retrieved from storage with a minimal number of I/O operations. The query mechanisms can return the subset of points within a constrained 3D volume, 2D projection, attribute range, flag setting, or resolution constraint. Once the 3D points have been imported into the point cloud structure, it is often necessary to select a subset for removal or editing. The CARIS Subset point editing utility in both CARIS HIPS and Bathy DataBASE uses these query mechanisms to select a subset of the point data from the point cloud storage formats and allow users to modify it interactively. The utility can be used to identify outliers, modify point attributes and flag points for exclusion from later downstream post-processing.

The sub-volumes in the point cloud are organized hierarchically to facilitate rapid 2D and 3D visualization and interaction with well over a billion points: overviews can be drawn in less than a second. Each point can be coloured using the value of any attribute, allowing for rapid and in-depth investigation of trends in the data. Figure 3 shows an example of use of different attributes to colour a dataset from the 2008 Shallow Survey sample set. Because only points within a specified visual tolerance are loaded onto screen, the visualization engine can sustain frame rates of 20-30 frames per second on a Pentium Core 2 Duo notebook computer with 256MB of discrete graphics memory.

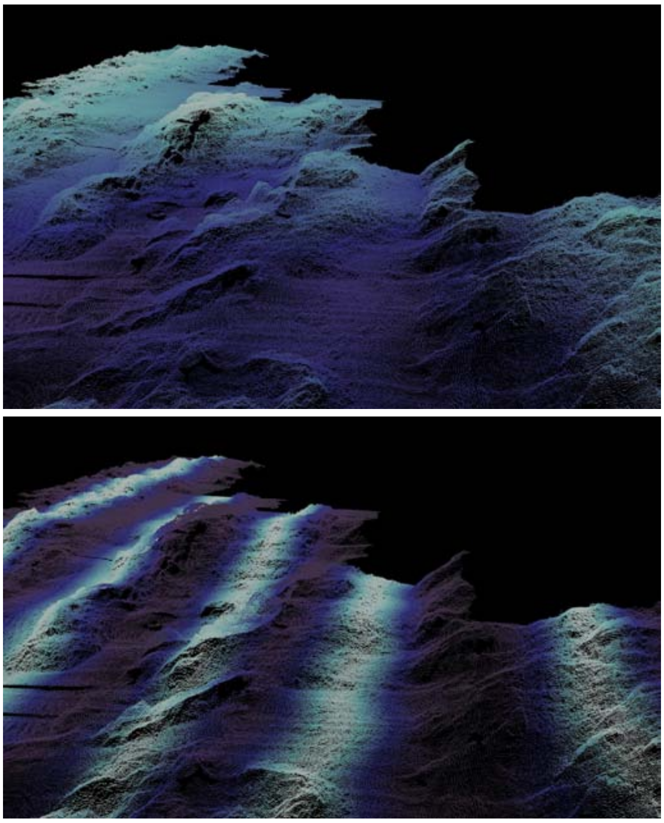

Figure 3. Two views of part of a 3D point cloud containing several million points. The top view shows the data colored by depth, while the bottom view shows the data colored by horizontal uncertainty. This data was taken from the Shallow Survey 2008 sample data set (http://shallow survey2008.org/).

Points from nearby sub-volumes are loaded into the 3D scene using a background thread, so the 3D visualization system maintains interactive frame rates while manoeuvring around the point volume, even when the data set size exceeds the size of main memory. Data can be loaded into the 3D scene at a rate of 5 – 10 sub-volumes per second, though this rate depends greatly on the performance of the underlying I/O subsystems; fewer sub-volumes per second can be loaded into memory over a network connection than from a local disk, for example.

4. Conclusions and future work

This paper presented a typical workflow for marine bathymetric point data acquired from multi-beam sonar systems, and CARIS's 3D point cloud system, which is capable of storing very high volume sets of unorganized 3D point data. Several important issues remain to be addressed, however. The amount of computation required to create a point cloud remains high and, ideally, it would be possible to create and process a point cloud of several billion points in real-time while preserving those aspects the structure that facilitate rapid query and visualization. Furthermore, because of the sheer volume of data, editing the points in the cloud is a challenge, so the development of structures that allow for rapid modification of collections of points within the cloud, or the addition of new points to the cloud will be important. Finally, the increasing volume of points generated by today's sensor systems will pose scalability challenges: compressing the point data, and improving the I/O efficiency and cache performance of point cloud data structures will help address this issue.

References

[1] Wimmer, M and Scheiblauer, C, Instant Points: Fast Rendering of Unprocessed Point Clouds, Proceedings of Europgrahics Symposium on Point-Based Rendering (2006), pp 129-136.

[2] Wand, M., Berner, A., Bokeloh, M., Jenke, P., Fleck, A., Hoffmann, M., Maier, B., Staneker, D., Schilling, A., Seidel, H.P., Processing and Interactive Editing of huge point clouds from 3D Scanners, Computers and Graphics 32, 2 (2008), pp 204-220.

[3] Rusinkieweicz S, Levoy, M., Qsplat: A Multiresolution point rendering system for large meshes, Proceedings of ACM SIGGRAPH (2000), pp 343-352.