Proceedings of the Eighth European Conference on Underwater Acoustics, 8th ECUA
Edited by S. M. Jesus and O. C. Rodríguez
Carvoeiro, Portugal
12-15 June, 2006

COMPUTATIONAL CHALLENGES IN PROCESSING AND ANALYSIS OF FULL-WATERCOLUMN MULTIBEAM SONAR DATA

Bart Buelens[1]-[2], Ray Williams[1], Arthur Sale[1] and Tim Pauly[2]
[1] School of Computing, University of Tasmania, Hobart, Tasmania, Australia
[2] SonarData Pty Ltd, Hobart, Tasmania, Australia

Abstract: *Several multibeam sonar systems are now capable of collecting and recording data samples covering the full-watercolumn, not just the seabed. Such systems, while still facing hardware challenges such as limited dynamic range and bandwidth, collect vast quantities of data, generally an order of magnitude more than conventional hydrographic multibeam or scientific single beam sonar systems. In this paper, the challenges faced by data processing systems for analysis of full-watercolumn multibeam sonar data are explored. Full-watercolumn multibeam data sets are valuable to scientists from traditionally diverse fields, providing simultaneous information about bathymetry, seabed type and habitats, and biomass in the watercolumn. Aspects of the data processing pipeline that are considered in this paper include raw data storage, data pre-processing, visualization and exploratory data analysis, statistical data analysis and post-processing, and presentation and interpretation of results. A general framework is outlined, and specific aspects applicable to the kind of data and problems at hand are emphasized. Proposed solutions to some of the challenges are reviewed and placed within an overall framework of multibeam sonar watercolumn data analysis. It will become clear that successful contributions to the field have been made, but that a general analysis method has yet to emerge.*

1. INTRODUCTION

Recently, many hydrographic multibeam systems have been modified to collect data samples from the whole watercolumn, not just from the seabed [1]. In addition, new multibeam systems have been and are being designed and developed specifically for this purpose [2]. Both kinds of multibeam systems collect vast quantities of data, generally an order of magnitude more than conventional high-end sonar systems. Data processing and analysis systems are facing the challenge of handling these large data volumes. In this paper, some issues and possible approaches are considered.

Initially the hardware on the sonar processing boards was the limiting factor in sonar data collection. Some of the engineering challenges have been overcome thanks to new developments in electronics, although there are still limitations, such as the reduced sampling rate of data output due to restricted bandwidth capacity. In this paper, emphasis is put on the data post-processing aspects. It is therefore assumed that full-watercolumn multibeam data can be collected by the sonar system, within the limits dictated by the hardware.

2. RAW MULTIBEAM DATA CHARACTERISTICS

2.1 General characteristics

While the precise detail of data collected by multibeam systems is dependent on the system in question, a data file or data stream generally contains - in addition to the acoustic backscatter data - spatial and temporal information, ideally some meta-information about the sonar system used, and the settings at the time of data logging. Spatial, temporal and meta information is important, but is small in terms of data quantities. The acoustic backscatter data accounts for typically 95% to 99% of the amount of data. In most cases the raw signals have been beamformed already on the sonar's DSP board, and the beamformed data is output and recorded to disk. Some instruments allow for the recording of data prior to beamforming. In such cases the data are the samples as obtained by the A/D conversion of the transducer's electric output signal.
It's the beamformed data that scientists usually work with. Beamformed acoustic backscatter samples are typically stored as magnitude values representing the transducer's output signal magnitude after beamforming. Some manufacturers express these values on a logarithmic scale and store them as such in the file. Depending on the number of quantization levels the instrument A/D converter handles, typically between 7 and 12 bits are used to store a single backscatter sample value. An unbeamformed sample is complex, consisting of an in-phase and a quadrature component, hence it requires twice the amount of space on disk.

2.2 Storage and compresssion

Manufacturers of echosounder and sonar systems in general have their own custom file formats, with varying levels of completeness, reliability and efficiency. From the point of view of a data post-processing system, efficiency is a crucial factor. It must be possible to access data quickly. Since disk access is mostly the limiting factor, it is important to store a given amount of data in as small a file as possible. It is also helpful to reduce file size by simply not storing data that is of no interest, such as sample values below a certain threshold. However, maintaining as much information as possible - and thereby making use of a data compression scheme is the preferred approach. Compression schemes other than Run Length Encoding (RLE) – arguably the simplest of schemes - are unseen in manufacturers' file formats. However, it has been demonstrated that utilizing more sophisticated lossless compression schemes can reduce the file size considerably [3]. Lossy compression can achieve even better result in terms of file size, but is a compromise between file size on disk and information content [4].

3. DATA PRE-PROCESSING

Prior to analysing the multibeam data, the raw data must be represented in a meaningful form. For some applications the raw beamformed data may be usable as it is, for example in some fish behaviour studies. However, in most applications - in fisheries in particular - scientists will want quantitative data, which is calibrated data, in meaningful numerical units. Calibrating multibeam systems is proving to be a challenging task. Significant progress has been made in recent years [5, 6]. When calibration information is available it has to be take into account, and the raw data samples must be represented as calibrated volume backscatter values, commonly denoted by SV. Some further corrections may be required, for example spatial corrections due to specific transducer mounts and vessel attitude, or temporal corrections due to timing issues. The information that is required in order to correct the data for all these effects is obtained by calibrating the various sensors in use: the echosounder, the motion sensor, the compass etc. The data is considered to be pre-processed and ready for analysis when it is represented in calibrated form, correctly placed in time and space.

4. DATA PROCESSING AND ANALYSIS

4.1 Visualizing and scrutinizing data

In the first instance, it is important to simply look through the data off-line, to identify anomalies due to noise, system failure, interference etc. In single beam hydro acoustics, data collected during a period of several hours can be visualized simply as a single image, commonly referred to as an echogram. For multibeam data, this is impossible to do because of the extra spatial dimension. A multibeam data set can be represented in two dimensions (2D) only as a sequence of images, each image representing one ping (Fig 1.a). Alternatively, it's possible to take a cross section, and consider a particular beam of all pings as a virtual single beam system, giving rise to a classic echogram (Fig 1.b). It is important to note that such an image is merely a cross section of all the available data; while informative, it's incomplete.

The true nature of multibeam data can only be visualized in a three-dimensional (3D) space, where time can be added as a 4th dimension, leading to time-varying three dimensional imagery (Fig 2) [7].

4.2 Filtering and processing

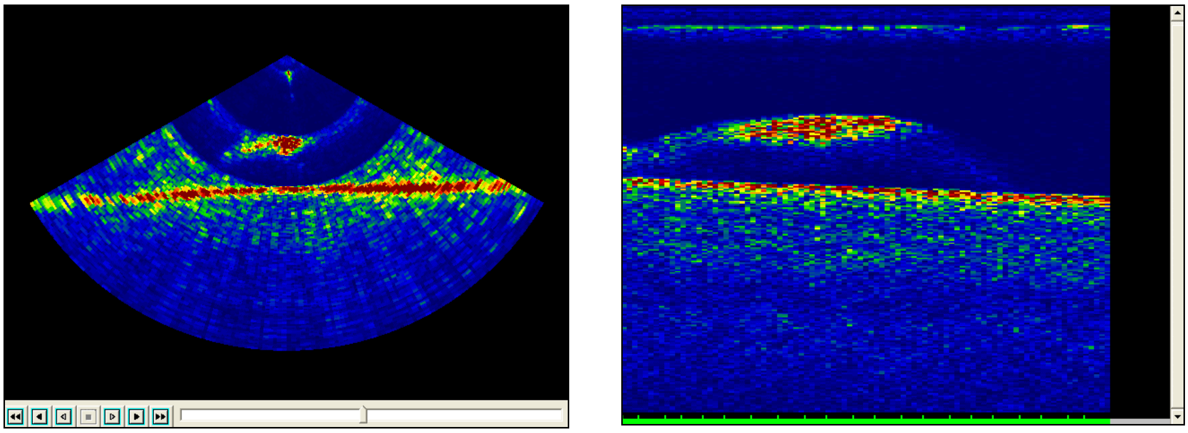
It may be desirable to filter or otherwise process the data, for example to remove specific kinds of noise such as speckle, background noise, or particular artefacts. With high ping rates, it may be sufficient to work with an averaged subset of the data, which can be achieved by applying a smoothing filter and subsequent resampling of the data.

4.3 Object detection

One of the more challenging analysis tasks is the determination of objects in the data, in particular the seabed surface and fish schools. Many of the algorithms that are applied today are either very simplistic, such as threshold-based techniques [8], or are extended versions of algorithms that have proven successful in single beam water column echosounding or multibeam hydrography [7]. Reasonable results are obtained (Fig 2), but there is certainly scope for improving the detection methods, taking full advantage of research outcomes of 3D computational geometry.

4.4 Object analysis

Once objects have been detected, various properties and features can be calculated. In general, surfaces and schools have morphological properties (area, volume, size, location etc.) and energetic properties (backscatter statistics, echo pulse properties etc.). In the context of fisheries, in particular mean backscatter values and volumes of schools are parameters of interest, as they are used for biomass calculations.



(a) (b)

Fig.1: (a) a single ping of multibeam watercolumn data, (b) an echogram composed by taking a cross-section through all pings, along a particular beam.

5. PRESENTATION AND INTERPRETATION OF ANALYSIS RESULTS

5.1 Presentation

The dissemination of an analysis of multibeam watercolumn data will include:

- visual presentation: a visualization of the data under consideration, in conjunction with any objects or features that have been determined. Ideally such a visualisation is delivered as an animated or interactive 3D presentation.
- numerical results: while the main numerical features of objects can be visually represented by color-coding the corresponding objects in a visual presentation, all results of the object analyses are typically presented in tabular form, either for direct use in reports, or for import in database applications for further analysis.

5.2 Interpretation

The interpretation of the results obtained is done by the scientists conducting the research and experiments. In interpreting the presented results, it is important to take into account the sensor hardware specifications, accuracy and limits, and to have an understanding of the data processing scheme that has lead to the final results. This will allow for judging the relevance of the outcomes, and the confidence that one can put in them. In order for the multibeam watercolumn data processing scheme to become a standard procedure, more research is needed, in particular comparisons with existing and accepted technology will prove invaluable, e.g. [9].

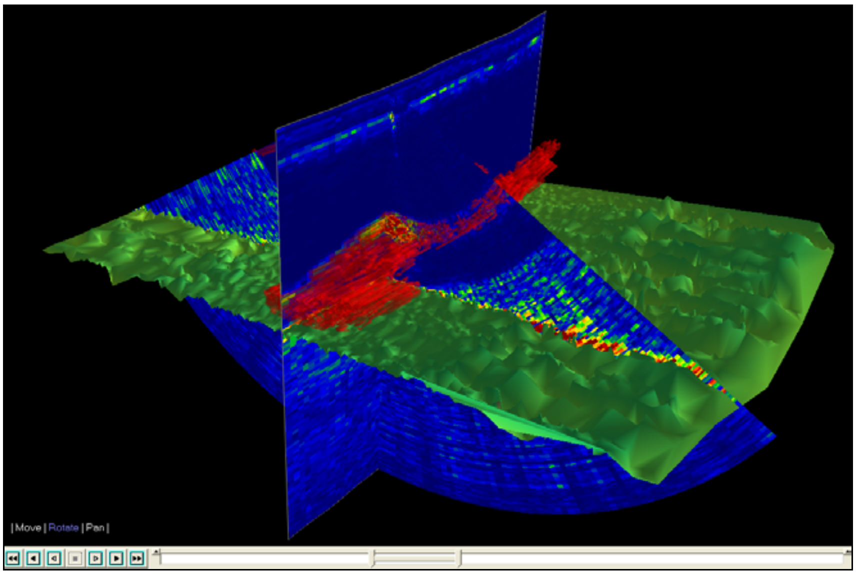


Fig.2: Snap-shot of an animated 3D visualisation of various aspects of the watercolumn multibeam data, and objects detected therein.

6. ECHOVIEW - TOWARDS A GENERAL DATA HANDLING FRAMEWORK

6.1 State-of-the-art software

It has become clear in the previous sections that significant work has been conducted in important areas such as calibration, object detection and multi-dimensional interactive visualisation of multibeam data. The software package Echoview [10], developed by SonarData Pty Ltd, incorporates many research outcomes and state-of-the-art experimental tools in all of the areas discussed. It is the most complete multibeam watercolumn data processing package available today.

6.2 Challenges

While significant progress has been made, and is continuously being made, some major challenges remain, not only for Echoview but for any multibeam water column data analysis software package. These challenges include

- Vast data volumes: it remains impossible to simultaneously analyse or process data covering several days of recordings; even several hours is often not feasible. Data compression and reduction schemes are needed.
- Scrutinizing data: since it is almost impossible to manually consider every data sample, there is a need for automation. Automated algorithms should scrutinize the data to an extent that manual checking is reduced. This is common practice nowadays in bathymetric echosounding [11], but unseen for midwater analysis.
- Object detection algorithms: these are often too simple, custom, ad-hoc methods, breaking down in more complicated situations; furthermore they are often tuned to specific data geometries and need tweaking, tuning or even rewriting for unseen geometries.
- Classification: much has been published on seabed classification, and also on classification of seabed and fish species from single beam acoustic data. No publications about classifying fish schools based on watercolumn multibeam data have emerged yet.
- Visualisation: this is the area in which most progress has been made. Data that is currently available can be visualized on standard monitors using standard input devices such

as keyboard and mouse. Possible improvements to current systems include virtual reality visualisations and stereo displays, to enhance the spatial 3D feeling.

Tackling these challenges in future will be important in order to make full-water column multibeam echosounding a viable technique for fisheries research. The current authors, and SonarData, are more than aware of these challenges and are working towards an enabling framework in which the problems as stated above should become more feasible. This includes a filtering stage based on acoustic inversion, which leads to a data representation that should facilitate further object detection and classification. Early results of this work have been presented before [12, 13].

REFERENCES

[1] Hammerstad E, Advanced Multibeam Echosounder Technology, *Sea Technology* 36, pp. 67-69, 1995.

[2] Andersen LN, Berg S, Gammelsæter OB, Lunde EB, New scientific multibeam systems for fishery research applications, In *Eighth European Conference on Underwater Acoustics*, 8th ECUA, Carvoeiro, Portugal, 2006.

[3] Pitman T, *Development of a multibeam sonar data logging system*, BEng (Hons) thesis, University of Tasmania, 2002.

[4] Wu LX, Zielinski A, Bird JS, Lossless compression of hydroacoustic image data, *IEEE Journal of Oceanic Engineering* 22, pp. 93-101, 1997.

[5] Cochrane NA, Li Y, Melvin GD, Quantification of a multibeam sonar for fisheries assessment applications, *Journal of the Acoustical Society of America*, pp. 745-758, 2003.

[6] Foote KG, Chu DZ, Hammar TR, Baldwin KC, Mayer LA, Hufnagle LC, Jech JM, Protocols for calibrating multibeam sonar, *Journal of the Acoustical Society of America* 117, pp. 2013-2027, 2005.

[7] Wilson MP, Higginbottom IR, Buelens B, Four-dimensional visualization and analysis of water column data from multibeam echosounders and scanning sonars using Sonardata Echoview for fisheries applications, In *1st International Conference on Underwater Acoustic Measurements (UAM)*, Heraklion, Crete, Greece, 2005.

[8] Gerlotto F, Soria M, Freon P, From two dimensions to three: the use of multibeam sonar for a new approach in fisheries acoustics, *Canadian Journal of Fisheries and Aquatic Sciences* 56, pp. 6-12, 1999.

[9] Melvin GD, Cochrane NA, Li YC, Extraction and comparison of acoustic backscatter from a calibrated multi- and single-beam sonar, *ICES Journal of Marine Science* 60, pp. 669-677, 2003.

[10] SonarData Pty Ltd, *Echoview on-line documentation*, http://www.sonardata.com/WebHelp/Echoview.htm, 2006.

[11] Calder BR, Automatic Statistical Processing of Multibeam Echosounder Data, *International Hydrographic Review* 4, pp. 1-16, 2003.

[12] Buelens B, Williams R, Sale A, Pauly T, Model inversion for midwater multibeam backscatter data analysis, In *IEEE Oceans '05 Europe*, Brest, France, pp. 431-435, 2005.

[13] Buelens B, Williams R, Sale A, Pauly T, A scientific data mining approach to midwater multibeam echosounding for fisheries applications, In *1st International Conference on Underwater Acoustic Measurements (UAM)*, Heraklion, Crete, Greece, 2005.