POLISH MARITIME RESEARCH 1(77) 2013 Vol 20; pp. 77-86
10.2478/pomr-2013-0009

A novel method for archiving multibeam sonar data with emphasis on efficient record size reduction and storage

Marek Moszynski, Assoc. Prof.
Andrzej Chybicki, Ph.D.
Marcin Kulawiak, Ph.D.
Zbigniew Lubniewski, Ph.D.
Gdansk University of Technology, Poland

ABSTRACT

Over the past few years considerable advances in sonar technology, spatial positioning capabilities and computer processing power have lead to significant improvements in mapping, imaging and technologies of seafloor exploration. Recently, modern multibeam echosounder systems (MBES) capable of recording backscatter data for the whole water column, not just for the seabed, have become available thus providing data allowing for visualization and analysis of objects other than the seabed such as single fish, fish schools or pollution. Unlike bathymetric sonars, which only capture the seafloor, multibeam systems produce very large amounts of data during surveys. Because of this, storing the data collected during hydrographic or scientific cruises becomes a crucial problem. In this context, the paper proposes a new approach for efficient reduction and storage of MBES records. The results of a sample implementation of the algorithm being tested on several different sets of MBES data are also discussed.

Key words: sonar; compression; storage; reduction; efficiency

INTRODUCTION

Modern multibeam echosounder systems (MBES) are capable of recording backscatter data for the whole water column, which allows for analysis and visualization of not only the seabed, but objects such as fish schools and single fish as well. They produce very large amounts of data during surveys, as it has been shown eg. by Buelens *et al*. (2005). Current trends in sonar development involve the use of innovative transducer materials as well as application of sophisticated processing techniques including focusing algorithms that dynamically compensate for the curvature of the wavefront in the nearfield and thus allow narrower beam widths (higher lateral resolution) at close ranges. Future developments will probably focus on "hybrid", phase-comparison/beam-forming sonars, the development of broad-band "chirp" multibeam sonars, and perhaps synthetic aperture multibeam sonars (Demkowicz & Bikonis, 2006). Not long ago, Buelens *et al*. (2006) considered storage and compression as one of the most important computational challenges in processing of MBES data. However, since Ferguson's and Chayes's (1995) proposal of a binary file format for multibeam sonar data storage, there has been little development in the field of efficient MBES data processing which would allow for fast, semi-real-time sharing of the results between diverse groups of interest such as fishermen, hydrographers or researchers (Chybicki 2008). Apart from works of Wu and Zielinski (1997), not much attention has been given to research of algorithms for storing and archiving MBES data that would allow for efficient browsing, analysis and visualization of collected information.

Even the prospect of having a common file format in which the same data sets can be shared between researchers and hydrographers offers an interesting new perspective, leading to savings in instrumentation and survey costs. Efficient MBES data storage algorithms can also make Remotely Operated Vehicles (ROV's) and particularly Autonomous Underwater Vehicles (AUV's) to become more widely adopted as platforms for seafloor mapping systems.

This paper presents the concept of a MBES data reduction and storage algorithm, developed by the authors for RESON Inc., along with the results of its sample implementation.

MATERIALS AND METHODS

MBES data, just like every other type of information, can be encoded using well-known methods. However, given the nature of MBES, a specialized approach could yield substantially better results. During the presented research, the authors designed and utilized a modification of Huffman coding (1952) for reduction of MBES record sizes according to the rules of information theory (1948), as described in the following section.

Entropy coding

Entropy coding is a lossless data compression method based on information probability schemes. This area of knowledge was first mathematically described by Claude E. Shannon. In 1948 he defined the basics of information theory in which he proposed concepts of measuring the amount of information. He assumed that a set of possible events $S = \{s_1, s_2, \ldots, s_n\}$ is called the alphabet if every element x_i is a symbol used to construct the communique. He also assumed that the probability of every event is known as $p(x_i) = p_i$. The set of probabilities is described as $P = \{p_1, p_2, \ldots, p_n\}$ assuming that for every $I(p_i) > 0$ and $\sum_i p_i = 1$. In this case, the function: $I(x_i) = -\log_k(x_i)$ for $k > 0$ and $k \neq 1$ is called the self-information associated with event x_i.

According to Shannon (1948), self-information is the function for measuring the information brought by the particular event x_i, which is reversely proportional to the probability of the event.

The self-information thesis was extended by the definition of entropy which says that the function defined as:

$$H(p_1, p_2, \ldots, p_n)$$

is equal to:

$$H(S) = -\sum p_{i*} \log_k(p_i) \tag{1}$$

with the set of n independent events $S = \{s_1, s_2, \ldots, s_n\}$ and the set of corresponding probabilities $P = \{p_1, p_2, \ldots, p_n\}$. Entropy is the mean auto-information of the set of particular events and, when assuming a binary source, can also be considered as the smallest average number of bits required for a symbol to be encoded, as shown by Balakrishnan et al. (2007) and Titchener (2000).

Huffman coding

The first optimized method of compression coding based on the information theory was given by David Huffman (1952). The term refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol.

Huffman coding uses a specific method for assigning the n-bit representations for each entry symbol, known as a prefix code. The method consists of replacing commonly used symbols with shorter representations and less commonly used symbols with longer representations. To achieve this goal Huffman coders create a prefix-free binary tree of non-overlapping bit-sequences, where the length of each sequence is inversely proportional to the likelihood of the symbol to be encoded.

Huffman binary code is *biunique*, meaning that the code is *uniquely decodable* and the sum of the probabilities associated with all symbols is always less than or equal to one. If the sum of probabilities is strictly equal to one, the code is termed a *complete* code. In general, a Huffman code need not be unique, but it is always one of the codes minimizing the complete code length.

Huffman coding is a particularly good compression method for datasets with non-linear distribution, and can be applied to many different types of data such as voice, sound, text and others. The Huffman algorithm is currently one of the most popular and efficient compression methods, and it can work in linear time if input probabilities (also known as *weights*) are sorted. It has been shown (eg. by Larmore, 1986) that when the actual symbol frequencies agree with those used to create the code, no other mapping of individual source symbols to unique strings of bits will produce a smaller average output size.

Although Huffman coding is optimal for a symbol-by-symbol coding with a known input probability distribution, its efficiency has since been surpassed by several compression algorithms developed later. Huffman coding is also used as one of the stages for more complex compression techniques, eg. Cosine Transform based compression methods such as JPEG (Hashemian, 2003), Wavelet Transform Compression methods such as JPEG 2000 (Chang *et al.*, 2006) or Deflate/Inflate algorithms (Khademi & Krishnan, 2006). These coding techniques often have better compression capability although they are characterized by greater computing complexity and larger amount of memory that needs to be allocated during data compression and decompression.

MBES and it's data structure

Multibeam sonars are a type of sonar systems widely used for surveillance and sensing in seas, oceans and other water containers (Mayer *et al.*, 2007). They provide a wide swath survey of the ocean floor at each ping by pulsing the bottom with a series of soundings normal to the track of the vessel and recording the reflected echoes in an orientation parallel to the vessel.

A typical multibeam sonar system consists of several elements, including transducer, Digital Signal Processor (DSP) and workstation, as shown in Fig. 1. The transducer, which uses linear or cylindrical arrays of receivers and projectors mounted at appropriate angles to each other (typically in so called the "Mills cross" arrangement, where the receive array lies in the athwart ship direction and the transmit array lies in the fore-aft direction) is responsible for ping generation and registration of backscattered acoustic wave signal. The registered data is transmitted to the DSP where the beamforming (based on sum-delay or Fourier transform algorithms), bottom detection (amplitude or phase detection algorithms), georeferencing, calibration and stabilization processes are performed. For each receive beam, the bottom echo from the intersection of the transmit and the receive beam footprint is detected. The entire system is managed by

Fig. 1. *The overview of multibeam sonar system structure*

MOST WIEDZY Downloaded from mostwiedzy.pl

the workstation, which is a PC unit operating under Windows or Linux. The workstation also enables storing and archiving of processed data.

The minimum and maximum depth values for data acquisition depend on sonar model and vary between a few meters (min) to over 1000 m (max). Nominal beamwidth ranges from 0.5° to 1.5° x 0.5° to 1.5° and the frequency of performing the transmit-receive cycle (ping-rate) is 40 Hz. The commonly applied operating frequency of mutlibeam systems is between 200 kHz and 400 kHz. Most mutlibeam systems offer integrated modules capable of acoustical imaging, mapping and data storage. The data recorded by these systems is stored in binary files defined by vendors. Although there are several formats of MBES data storage, their syntactic contents are similar due to the common architecture of multibeam systems.

The echo arrival time and the angle of the receive beam provide information for bathymetry, and the backscattering strength is used for seabed imaging (Shah & Talukdar, 1999). The process of forming beams is called beamforming and has been widely described in literature e.g. by Řrnhovde (1999) or Thorner (1990). In this context, sea-bottom charts and images are formed by appropriate processing and combining data from many consecutive swaths.

The result of surveys made with the use of MBES systems is usually stored in two basic configurations:

a) The data is processed and logged by dedicated operating software like QUINSy, HYPACK, PDS 2000 or other. In this case the results of surveys are kept in dedicated databases or database files.
b) The raw records retrieved from MBES system are logged as a binary file in order to perform further analysis.

In both of these configurations the MBES data is received as a set of binary records containing various information delivered by the multibeam system, such as bathymetry, water-column data, beam geometry data and sonar configuration data. Additionally, external sensors data such as CTD probe, GPS, Compass, Gyro and others may be also included. The exact structure of the dataset depends on the particular multibeam sonar model, survey purpose, configuration and connected external sensors. Nevertheless, the following characteristics are common for every multibeam sonar system:

- The multibeam system must be connected to a workstation in order to collect and process the acquired data and present the results;
- Data is delivered as a set of binary datagrams (records) via a connection between the sonar processor and the workstation;
- The MBES produces several types of datagrams, each containing different types of information such as GPS data, bathymetry data, external sensors data (i.e. heave, peach, roll) and water column data;
- No compression of MBES records is performed either by the sonar hardware or the specialized software.

Although collecting water-column data is not mandatory during multibeam hydrographic cruises, it delivers a lot of additional information about the monitored marine environment. It is also used by diverse groups of researchers as an important component of survey results, delivering acoustic information about various underwater objects, as shown by Gerlotto *et al.* (1999) and Mayer *et al.* (2002) and for seafloor characterization and classification (Lubniewski, 2010). The water column data represents a set of signals acquired from multibeam sonar sensors in one sonar ping. The set of beamforming functions depends on the sonar specification. The datasets presented in the paper were acquired by RESON Seabat 7125 mutlibeam system which utilizes standard sum-delay beamforming. Two different visualizations of the water-column data acquired from this device are shown in Fig. 2. In this sonar the water-column is available as a two-dimensional binary array where the value of point defined by indices *m* and *n* is the 16-bit integer value of beam numbered *m* and sample numbered *n* (Marcus, 2007). The presented research was conveyed with the use of eight different datasets acquired by means of the RESON Seabat 7125 MBES. The overall characteristics of this data are shown in Tables 1 and 2.

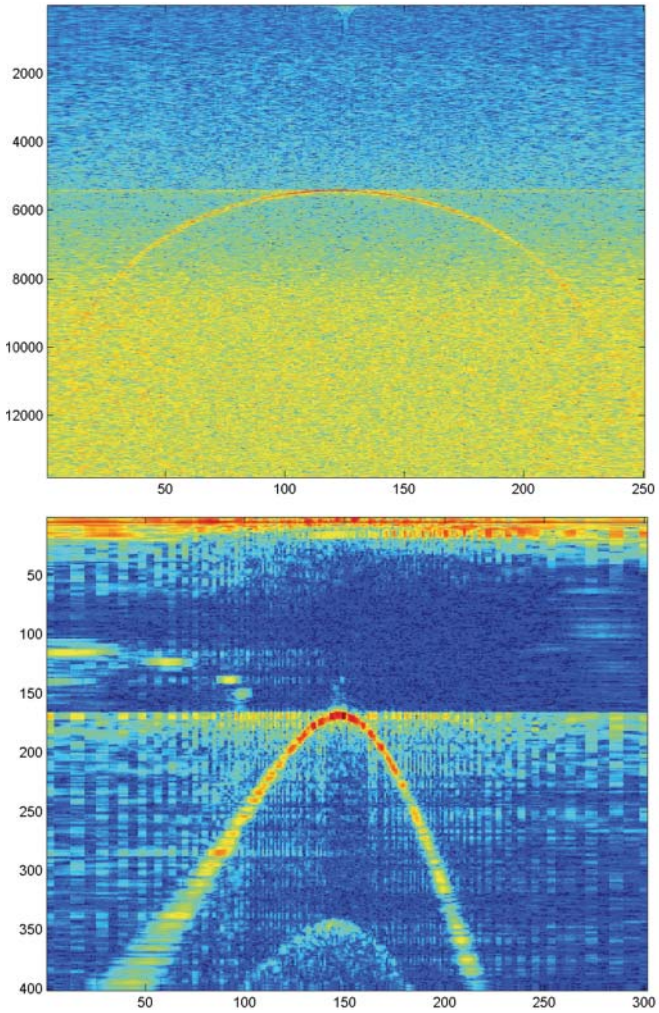


Fig. 2. *Sample water-column data acquired by RESON 7125 Seabat Multibeam System.*

In order to explain the importance of efficient water-column data processing, a more detailed description needs to be given. Although the tested system generates 256 beams, in some multibeam systems the maximum number of beams can reach 3520 per one sonar head. The number of registered samples for individual beams depends on the depth and probing frequency and can reach up to 32 thousands, as shown eg. by Van Buren and Blue (1990), Xiaofeng and Wenjun (1998) or Foote *et al*. (2003). Since each value is stored as a 16-bit float or integer, the size of water column data can easily exceed 95% of all data collected by a multibeam system, as shown in Table 1. This is particularly apparent when the data is stored in linear scale, as it is the case in the datasets utilized in the presented work.

In literature the value of water-column datagram samples is calculated according to the sonar equation:

$$EL = SL + TS - 2TL$$

MOST WIEDZY Downloaded from mostwiedzy.pl

In other words: Received echo level (EL) is equal to transmitted source level (SL) plus target strength (TS) minus two-way transmission loss (2·TL) expressed in dB. The two-way transmission loss is:

$$2TL = 2\,\alpha\,R + 40 \log R \text{ (re } R_0\text{=1m)}$$

where α is absorption coefficient (dB/m) and R is slant range (m).

In a water column datagram, the received echo level is adjusted for system dependent parameters in addition to being amplified with a range dependent TVG function:

$$A_{WC} = EL - SL - 10 \log \Omega_{TX} \cdot \Omega_{RX} + \\ + 2\,\alpha\,R + 40^* \log R + C \tag{2}$$

where Ω_{TX} and Ω_{RX} are along-track (transmit) and across-track (receive) beam widths respectively (expressed in radians). These values are dependent on actual sonar model and operating parameters. For instance, in the RESON 7125 MBES which was used as a source of data during the presented work, the parameter Ω_{RX} was equal to 0.5° and Ω_{TX} equaled 1°. The parameter C represents gain selection of MBES transceiver used to adjust the signal level to system dependent parameters.

In the presented formulas the echo levels calculations are expressed in dB. Also, the transducers accuracy used in the RESON 7125 MBES system is lower than 0.5 dB, as shown eg. by Van Buren and Blue (1990), Xiaofeng and Wenjun (1998), or Foote *et al*. (2003). This is the reason why some MBES manufacturers (eg. Kongsberg) apply conversion from linear scale to dB scale with 0.5dB re 1μP precision. That has been also applied during the presented research.

In this context, the objectives of this work are to achieve improved data reduction ratio and offer reduced time of processing in comparison to standard archiving tools.

The proposed algorithm

The developed algorithm is a modification of static Huffman codes dedicated for multibeam sonar data. It aims to:

- enable efficient coding/decoding of large binary files containing MBES data;
- lay foundations for a standard file format for compressed MBES data;
- deliver an efficient archiving tool for online processing of particular MBES records, fulfilling the following requirement:
  - compression time of every record produced by MBES system must be shorter than the time of creating it by the sonar system;
  - program memory use must be unrelated to the size of compressed files, and be possibly small.

The algorithm aims to improve MBES data processing speed in comparison to standard Huffman coders by assuming that the histograms of records and probabilities for particular symbols (byte values) to be encoded are similar within individual datagram types. Using this knowledge, the algorithm composes several Huffman trees based only on the first datagram of particular type delivered by the MBES system.

According to (1), the entropy of a dataset (datagram) is the minimum average number of bits that can represent an encoded symbol using binary prefix code, which can be expressed as:

$$E_{pattern} = -\sum_{i=0}^{256} p_i \log_2(p_i)$$

where p_i is the probability corresponding to symbol i that can also be defined as n_i/N where n_i is the number of bytes containing value i in first pattern datagram and N is the total number of bytes. The latter is usually constant for every datagram type, as shown by Marcus (2007). According to Huffman theory, $\log_2(p_i)$ represents the number of bits required to encode symbol i. In this case, $E_{pattern}$ is the entropy of the pattern datagram, which defines the Huffman tree. Once a tree is created for the analyzed type of datagram, the calculated Huffman structure is applied for encoding the rest of the records of particular type. In this case the entropy, calculated for a consecutive record is defined as:

$$E = -\sum_{i=0}^{256} (p_i + \Delta p_i) \log_2(p_i) \tag{3}$$

where Δp_i is the difference of probability between the probability symbol i in the pattern datagram and the current datagram, and $\sum_{i=0}^{255} \Delta p_i = 0$.

To make data uniquely decodable, additional information about the compressed records, such as the type of compressed datagram and its original and compressed sizes, must be saved to the result file.

Fig. 3 depicts the entropy of the water-column data calculated by (3) on five sample datasets. As shown in the diagram, the entropy difference of at least 30 consecutive datagrams is approximately constant. Thus, assuming that the entropy is the lower bound for Huffman coding algorithm, compression ratio shouldn't vary much between consecutive swaths. It is also worth mentioning that in some cases an increase of entropy was observed after several pings, which is a result of on-device gain adjustment when forming beam geometry data. This will have a considerable impact on the presented results.

It may be noticed that the number of datagram types produced by the MBES system is constant and, according to technical specification, does not exceed 20-25 in most cases. This issue is described in detail by Marcus (2007) and Van Buren & Blue (1990). The entropy characteristics shown in Fig. 3 allow the algorithm to assume that the consecutive records of the same type generated by the MBES system are similar in content. Thus, all consecutive datagrams of the same type are encoded using the same Huffman tree which was generated for the first datagram. Because the Huffman tree is generated only once for each datagram type, the algorithm is granted a considerable increase in performance without negative influence on compression ratio. This mechanism is depicted in Fig. 4.

As it has been said in the preceding section, in some MBES systems (e.g. RESON 7k series) the water-column data is stored in linear scale, which causes unnecessary waste of storage space. In order to further optimize the size of the resulting data, its precision has been made a user-controlled parameter which can be set in the range of 0.1-1 dB. For best data reduction efficiency without losing information relevant for visualization or analysis of the water-column data, the parameter must be set in such a way as to match the accuracy of the MBES system. During the tests, this parameter was set to 0.5 dB, which is more than the native precision of the RESON 7125 MBES, as shown eg. by Van Buren and Blue (1990), Xiaofeng and Wenjun (1998), or Foote *et al*. (2003).

An additional advantage of this approach is the fact that the proposed compressed file format enables access and decompression of particular records without exigency of decoding the whole dataset, while maintaining the efficiency of decompression. In this case, the only elements that must be read from the compressed file are the appropriate Huffman tree and the position of the compressed data. Reading the structure

MOST WIEDZY Downloaded from mostwiedzy.pl

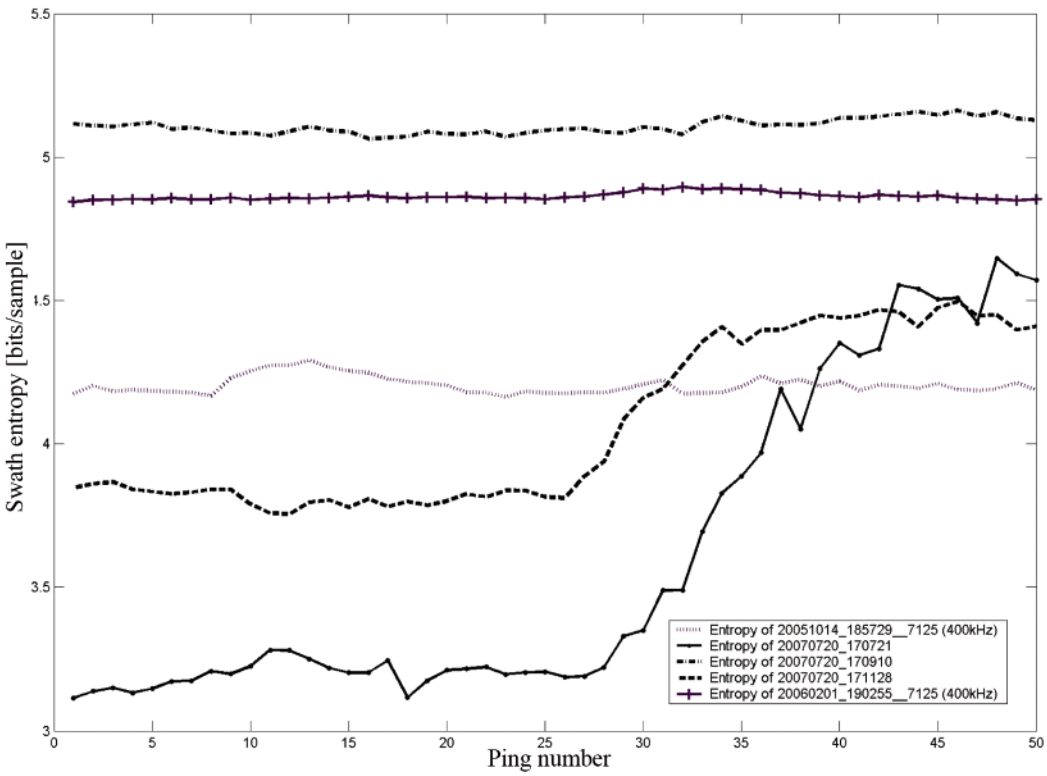


Fig. 3. *Entropy of consecutive water-column records according to (3)*

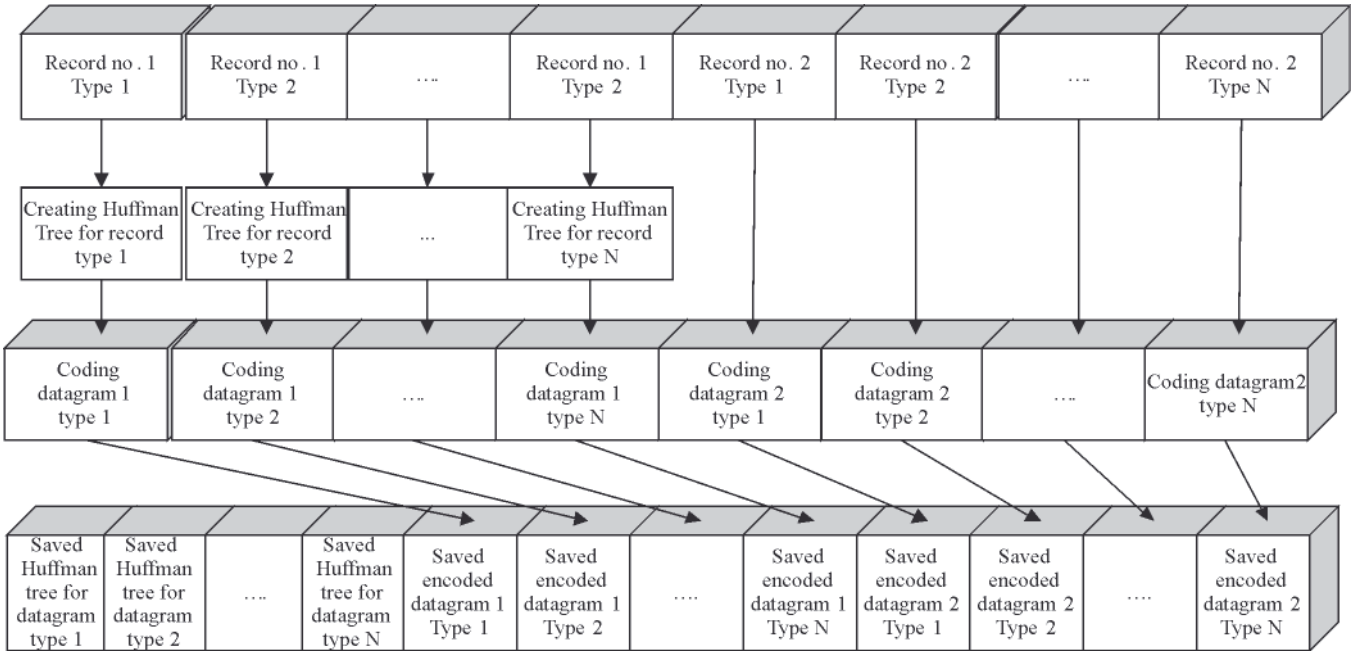


Fig. 4. *Block diagram of the proposed online compression algorithm*

of the compressed file and retrieving the information such as the total number of datagrams, original size of particular datagrams and their types is also possible without the need of decoding the whole compressed dataset. That is particularly useful when processing files bigger than 500 MB. In this case the decompression process would last considerably longer when using standard archiving tools.

RESULTS

A sample application of the proposed algorithm for multibeam data processing was implemented using C++ programming language. The software processes binary files containing 7k series multibeam sonar records. This section contains the results of testing the software as well as the information on the testing methodology and input datasets.

Testing methodology and input data characteristics

The basic acceptance criteria for the proposed method are: time of compression of particular datagram types, time of processing of whole dataset, and compression ratio.

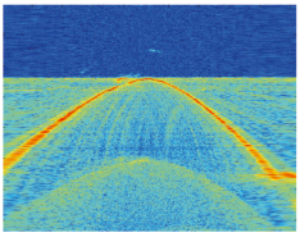
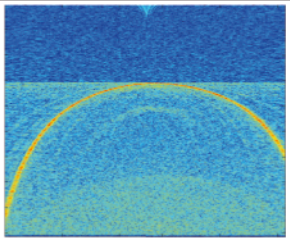
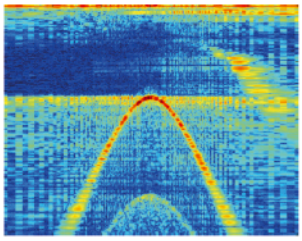
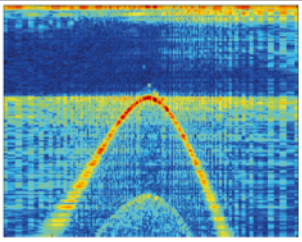
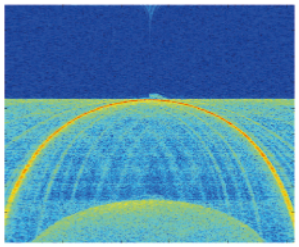
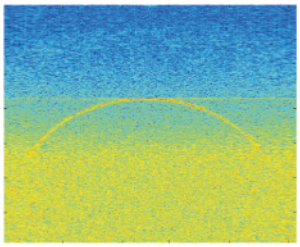
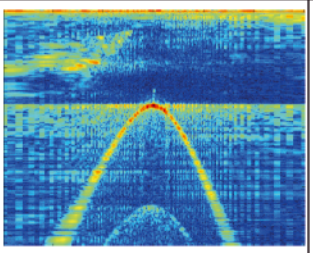
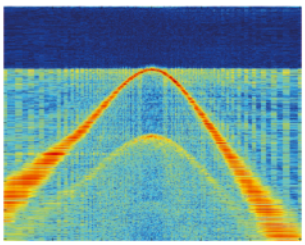
Tables 1 and 2 show the information on the datasets used for testing the compression algorithms. The files were acquired during scientific or research surveys in an area of Santa Barbara, California, USA, which is characterized by relatively flat and homogenous bathymetry. The data was

MOST WIEDZY Downloaded from mostwiedzy.pl

Tab. 1. Description of sample datasets acquired during research in the area of Santa Barbara, California, USA

Dataset name	Percentage of water-column data in entire dataset	Operating frequency [kHz]	Sample rate [kHz]	Result file size[MB]
20060719_204657__7125 (400kHz).s7k	85.72%	396	34	557
20051014_185729__7125 (400kHz).s7k	83.09%	396	34	589
20070720_170910.s7k	82.77%	100	6	1 100
20070720_171128.s7k	82.82%	100	6	477
20060201_190255__7125 (400kHz).s7k	83.07%	400	34	601
20070315_184108.s7k	97.96%	396	34	127
20070720_170721.s7k	82.84%	100	6	937
20070831_185543.s7k	87.87%	100	6	1046

Tab. 2. Acoustic parameters of water-column data processed during the tests

20060719_204657__7125 (400kHz).s7k	**20051014_185729__7125 (400kHz).s7k**	**20070720_170910.s7k**	**20070720_171128.s7k**
			
Number of pings: 146 Number of samples = 1595 Number of beams = 512 Range selection [m] = 35 Power selection [dB] = 218 Gain selection [dB] = 30	Number of pings: 274 Number of samples = 1829 Number of beams = 256 Range selection [m] = 40 Power selection [dB] = 207 Gain selection [dB] = 43	Number of pings: 1013 Number of samples = 401 Number of beams = 301 Range selection [m] = 50 Power selection [dB] = 185 Gain selection [dB] = 83	Number of pings: 429 Number of samples = 401 Number of beams = 301 Range selection [m] = 50 Power selection [dB] = 200 Gain selection [dB] = 25
20060201_190255__7125 (400kHz).s7k	**20070315_184108.s7k**	**20070720_170721.s7k**	**20070831_185543.s7k**
			
Number of pings: 278 Number of samples = 1839 Number of beams = 256 Range selection [m] = 40 Power selection [dB] = 220 Gain selection [dB] = 30	Number of pings: 9 Number of samples = 1595 Number of beams = 512 Range selection [m] = 300 Power selection [dB] = 220 Gain selection [dB] = 19	Number of pings: 843 Number of samples = 401 Number of beams 301 Range selection [m] = 50 Power selection [dB] = 175 Gain selection [dB] = 40	Number of pings: 137 Number of samples = 806 Number of beams = 301 Range selection [m] = 100 Power selection [dB] = 220 Gain selection [dB] = 10

MOST WIEDZY Downloaded from mostwiedzy.pl

acquired using maximum pinging frequency of 40 Hz. It is important to note that the surveys were made under various conditions, with varying place of measurement, depth, bottom type, sonar frequency, probing frequency, etc. This is to clearly show that the research was based on a broad range of datasets, which cover a lot of the possible MBES survey configurations.

Table 1 clearly shows that appropriate methods of processing water-column data are the key to reducing the size of MBES records. This means that an efficient and fast algorithm for processing water-column data will guarantee optimal results of processing the entire file. Table 2 presents the information on acoustic parameters and sample visualizations for every water-column dataset processed during the tests. The images display raw frames of data without any form of post-processing such as fanshape correction.

Measured efficiency of the algorithm

The presented results were obtained on a standard desktop PC equipped with an Athlon X2 3600 + CPU paired with 1 GB of RAM, running Windows XP SP3. For each dataset, Table 3 lists the time of compression for water-column records compared with the time in which the records were generated by the sonar system. Additionally the compression ratio acquired with the use of the proposed method and that acquired by the standard ZIP compression tool are also presented. The time of compression of the remaining elements of a record (bathymetry, position etc.) is usually about 1-2 ms and is always shorter than the time of record generation.

The proposed method can be also applied as an archiving tool for MBES data. Currently, the most commonly used tools for archive volume reduction are the well known file compression tools such as ZIP, RAR or the newly developed 7-ZIP. These tools do not utilize the information about the character of MBES data and consider the input simply as a series of bytes. Table 4 presents a comparison of the proposed archiving tool with other standard data compression applications based on the total time of archiving the whole dataset.

DISCUSSION

As shown in Table 4, the proposed method offers superior archiving speed in all test cases, being on average 68% faster than the next best ZIP algorithm, while offering up to 111% smaller output file size. In most cases the proposed tool also offers greatest reduction of the input file size, with two exceptions (test cases no. 6 and 7) where 7-ZIP performs 15% better, but at a cost of up to 829% longer processing time. There is also one special case where 7-ZIP offers 91% better results. This is due to statistical change of datagram contents during the survey where the pattern datagram doesn't describe the content of records registered at later stages of the survey. In this case additional analysis of the data is required to update Huffman trees describing the statistical information about the content of the records (as seen previously in Fig. 3). Even though the proposed algorithm fails to produce the smallest file size in this test case, it is 384% faster than 7-ZIP.

To better understand the origins of the obtained results, a more in-depth analysis of the data used in test cases no. 6 and 7 is required. Figures 5 and 6 show the differences between the contents of consecutive water-column datagrams in the problematic surveys, in the form of images representing the actual data and corresponding histograms.

Tab. 3. *Compression ratio yielded with proposed methods of particular water-column datagrams compared to time of generation of the records by RESON 7125 MBES system*

Dataset name	Average water-column record compression time [ms]	Time of generation of water-column record [ms]	Compression ratio	ZIP compression ratio
20060719_204657__7125 (400kHz).s7k	80	150-200	0.28-0.32	0.48
20070315_184108.s7k	350-400	450-500	0.31-0.33	0.79
20070720_170721.s7k	15-20	100-120	0.19-0.21	0.41
20070831_185543.s7k	20-25	300-350	0.32-0.35	0.53
20051014_185729__7125 (400kHz).s7k	40	70-80	0.26-0.27	0.47
20070720_170910.s7k	12	120-130	0.32-0.33	0.66
20070720_171128.s7k	11-12	120-130	0.23-0.28	0.45
20060201_190255__7125 (400kHz).s7k	42-43	70-80	0.30-0.31	0.52

Tab. 4. *Compression ratios and the time of archiving of selected compression tools in comparison to the proposed method*

Test case	Dataset name	Result file size[MB]	ZIP ratio/ time [s]	7-ZIP ratio/ time[s]	RAR ratio/ time [s]	Proposed methods ratio/time [s]
1	20060719_204657__7125 (400kHz).s7k	557	0.65 / 155	0.50 / 295	0.54 / 220	0.31 / 65
2	20070315_184108.s7k	589	0.79 / 35	0.63 / 73	0.69 / 42	0.31 / 12
3	20070720_170721.s7k	1 100	0.51 / 255	0.23 / 11 min.	0.44 / 380	0.44 / 172
4	20070831_185543.s7k	477	0.48 / 80	0.40/195	0.41 / 110	0.19 / 29
5	20051014_185729__7125 (400kHz).s7k	601	0.63/178	0.48 / 380	0.57 / 182	0.28 / 90
6	20070720_170910.s7k	127	0.53 / 305	0.40 / 13 min.	0.46 / 430	0.49 / 234
7	20070720_171128.s7k	937	0.52 / 125	0.40 / 13 min.	0.46 / 170	0.46 / 94
8	20060201_190255__7125 (400kHz).s7k	1046	0.66 / 170	0.54 / 340	0.61 / 225	0.31 / 82

MOST WIEDZY Downloaded from mostwiedzy.pl

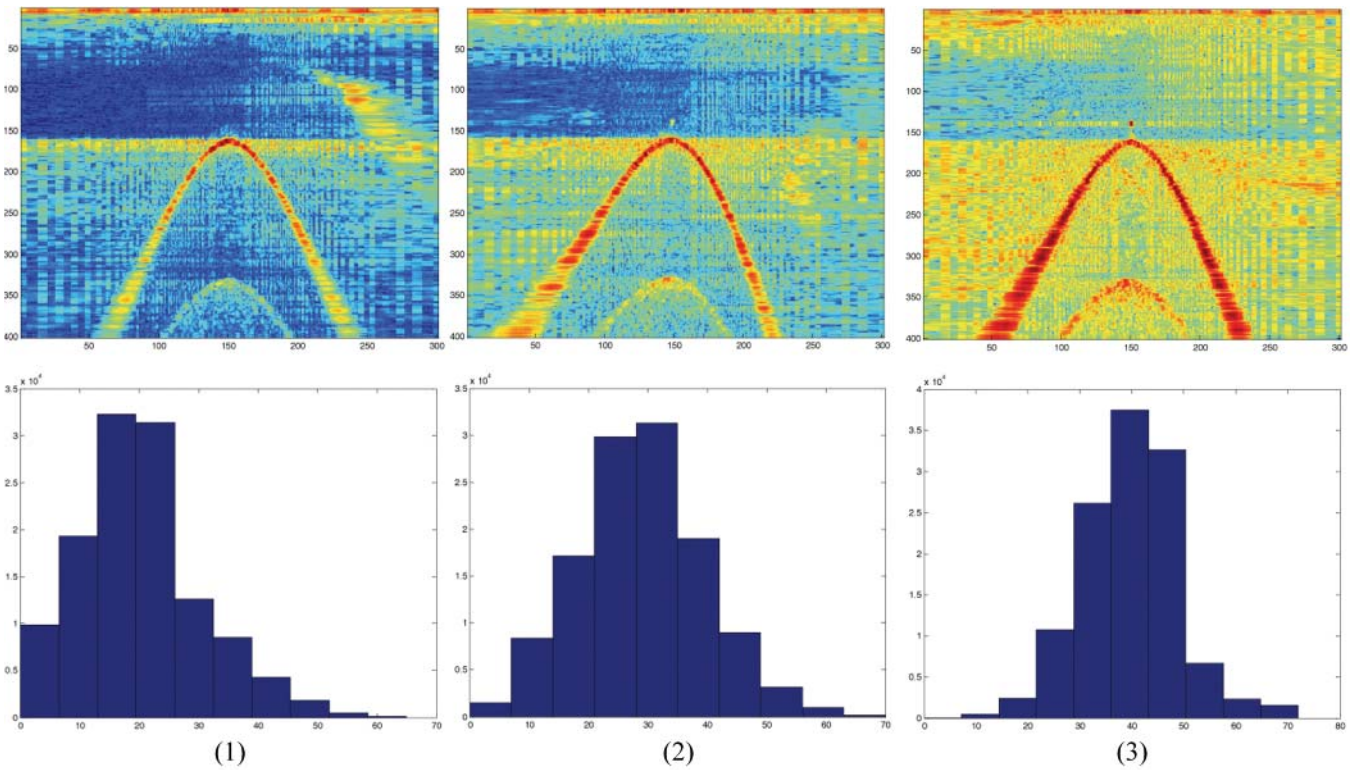


Fig. 5. Water column data and corresponding histogram for consecutive records in 20070720_170910. Left – ping no. 1, middle – ping no. 150, right – ping no. 300

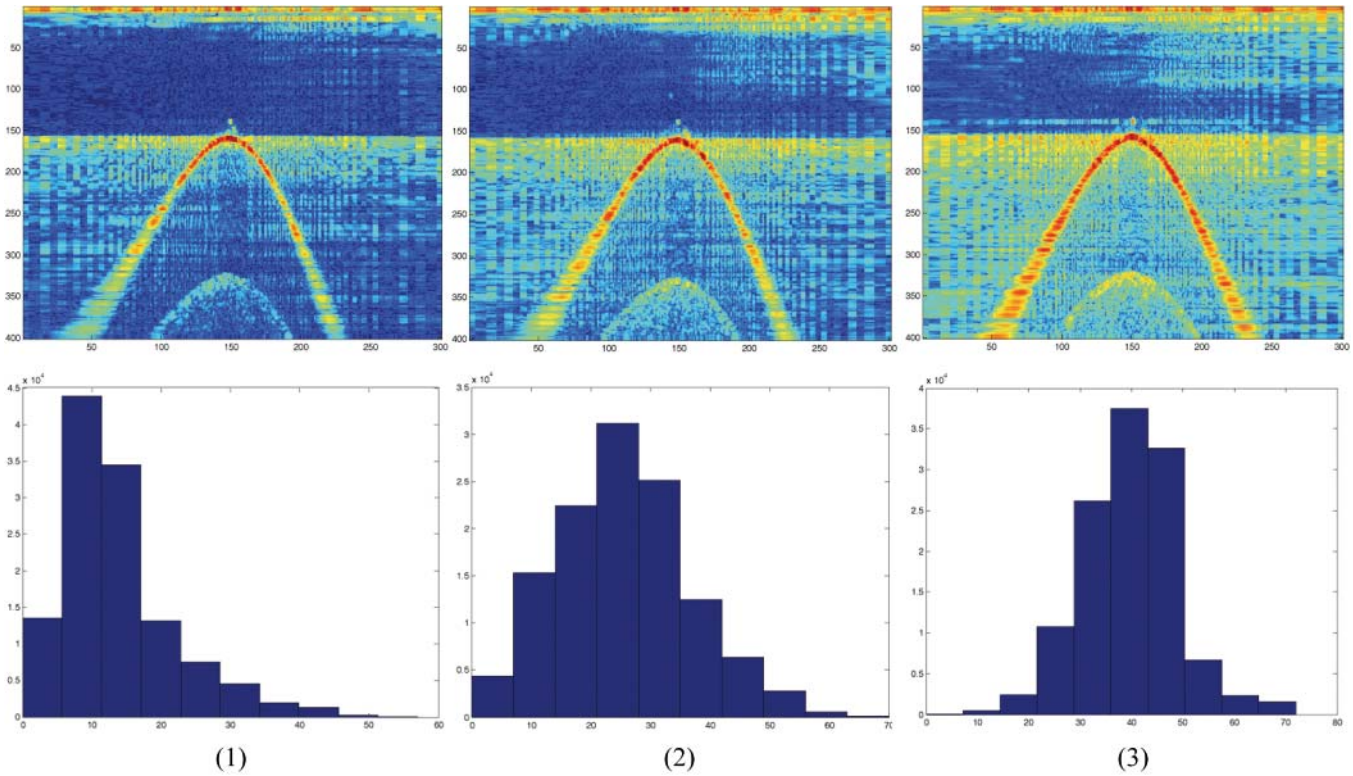


Fig. 6. *Water column data and corresponding histograms of records in 20070720_171128. Left – ping no. 1, middle – ping no. 100, right - ping no. 220*

According to (2), such parameters as TVG, gain, and power of the sonar, as well as others that can be automatically or manually adjusted during the survey, influence the statistical parameters of the data contents and thus can make compression less effective. The change of gain as a function of ping number in the two problematic datasets is shown in Fig. 7. a modification of the algorithm that would automatically update the Huffman trees when the compression of consecutive records becomes less effective is an important issue to consider, especially for dataset structures similar to those of 20070720_170910 and 20070720_171128.

What is also important is the fact that the time of processing particular datagrams is shorter than the time interval between delivering consecutive records by the MBES system. Therefore, the proposed method can be used as on-the-fly record size reduction tool in routine research or commercial surveys, or efficient archiving tool for bigger multibeam sonar data warehouses.

MOST WIEDZY Downloaded from mostwiedzy.pl

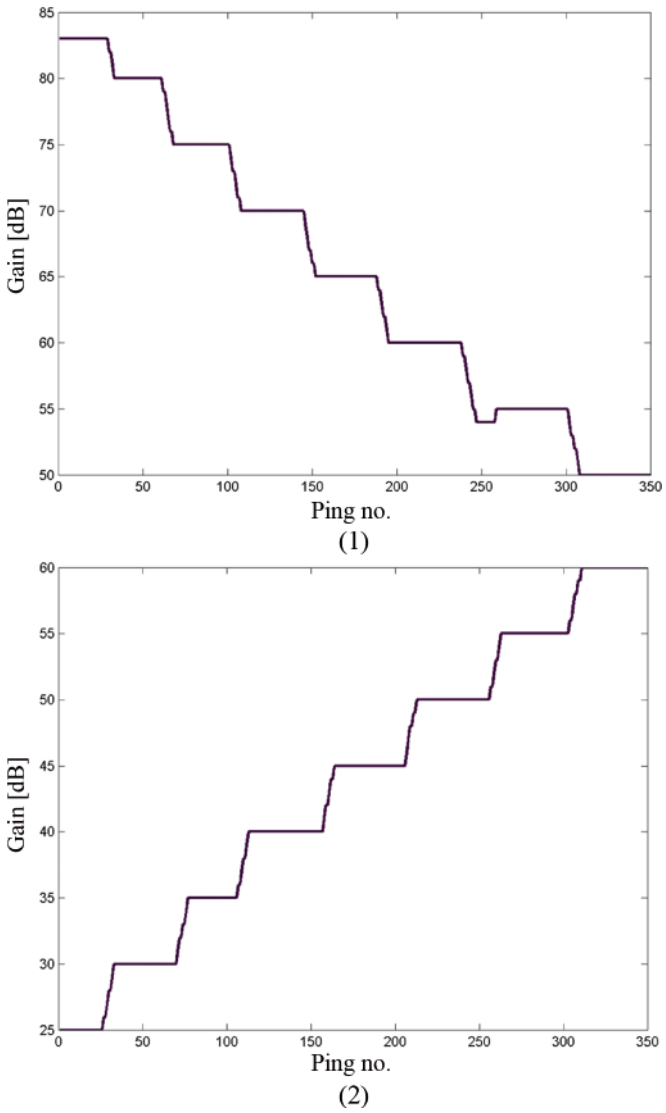


Fig. 7. *Gain selection presented as a function of ping no. in 20070720_170910 (left) and in 20070720_171128 (right)*

An additional advantage of the presented algorithm is that it is dedicated to processing MBES data, and thus it allows unique identification of the type of processed information. As a consequence, the proposed compressed file format enables selective access to data such as individual datagrams and their types without the need of decoding the whole compressed dataset. This ascertains considerable time savings in comparison to standard archiving tools, in particular for larger datasets.

The versatility of the algorithm allows its application to on-the-fly MBES data processing. The presented performance results were obtained on a standard PC-class computer. Implementing and running the algorithm on a dedicated processor such as a CUDA-enabled GPU provides a substantially different level of performance, opening a new range of possible applications of the algorithm. Preliminary research shows that utilizing a GPU allows the algorithm to achieve semi-real-time compression of water column data. This in turn allows the compressed frames to be dynamically visualized in three dimensions via a Web-based Geographic Information System. a sample visualization of animated MBES swath in 3D is shown in Fig. 8.

Because the contents of every MBES frame have been already processed, they may be easily analyzed according to the paradigms of Geovisual Analytics. Application of online analysis algorithms enables extraction of water-column data such as pelagic fish schools on-the-fly as well as their visualization in 3D context alongside digital terrain data

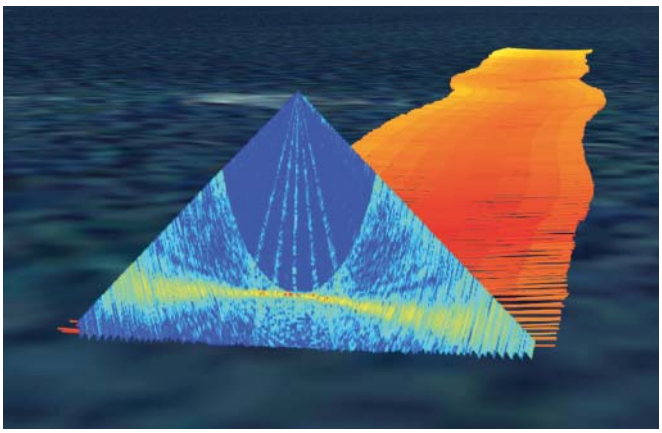


Fig. 8. *A three-dimensional visualization of processed sonar data*

such as bathymetry (Kulawiak 2010). Fig. 9 shows a three-dimensional visual analysis of processed water column data in the context of a standard two-dimensional map via WebGL. For this purpose, GPS location data from the vessel is used to localize the collected MBES information in geographical context in the form of a marker on a two-dimensional map constructed with the use of HTML5. The contents of the data are then visualized inside the marker popup cloud on a three-dimensional canvas. The water-column data is streamed from the server in compressed form, which enables its dynamic visualization on the client. The data presentation has been set up to exclusively depict the ocean bottom as well as objects floating below the water surface. Changes in the shape of the bottom have been reflected by different colors of the three-dimensional model. The backscatter objects in water column data are colored in white.

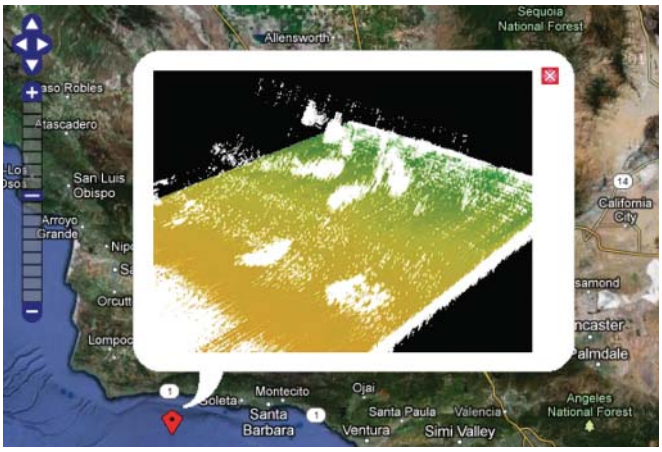


Fig. 9. *Three-dimensional geovisual analysis of water-column backscatter data*

The proposed modification of Huffman coder, optimized for storing and compressing MBES records, enables its application in devices which process MBES data in real-time, while its efficiency rivals that of well-established compression tools like ZIP and RAR. However, the presented study of multibeam sonar data reduction and storage methods has been conveyed in a field that has not yet been properly explored. This will likely change as the amounts of data collected during research, as well as commercial surveys, will continuously increase. Properly designed MBES data processing algorithms may result in opening up new worlds of interactive exploration to a multitude of users. Despite their preliminary nature, the presented results show great promise of fast and efficient MBES data reduction, storage and retrieval. Some important factors influencing the compression ratio such as TVG, gain, and bottom type are yet to be considered. The presented MBES data reduction and storage

MOST WIEDZY Downloaded from mostwiedzy.pl

techniques can become a powerful tool for users of MBES systems, used as an efficient archiving tool for researchers, hydrographers, ecologists, fishers and commercial groups of interest, such as dredging industry or pipeline inspection and chart production companies.

Acknowledgements

Authors of the paper wish to thank Pawel Pocwiardowski, Ph.D. from RESON Inc. for the effort that he put into the presented research.

BIBLIOGRAPHY

1. Balakrishnan, K. J. & Touba, N. A.: *Relationship Between Entropy and Test Data Compression*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 26 (2) 2007: 386 - 395.
2. Buelens, B., Williams, R., Sale, A., Pauly, T.: *Model Inversion for Midwater Multibeam Backscatter Data Analysis*, Oceans 2005 - Europe, 1: 431 – 435.
3. Buelens, B., Williams, R., Sale, A., Pauly, T.: *Computational Challenges in Processing and Analysis of Full-Watercolumn Multibeam Sonar Data*. In: Proceedings of the Eighth European Conference on Underwater Acoustics, 12-15 June 2006. Carvoeiro, Portugal.
4. Chang, Y.-W., Truong, T.-K., Chang, Y.: Direct mapping architecture for JPEG Huffman decoder, Communications, IEE Proceedings, 153 (3) 2006: 333 - 340.
5. Chybicki, A., Moszyński, M., Pocwiardwoski, P.: *Compression algorithms for multibeam sonar records, Hydroacoustics*, vol. 11, 2008, pp. 55-63, 25th Symposium on Hydroacoustic (SHA'25), Jastrzębia Góra, Poland, 25-28 May, 2008.
6. Demkowicz, J. & Bikonis, K.: *Combined Spline Wavelet Decomposition for 3D Seafloor Imaging from Multibeam Sonar Echoes*, Acta Acustica United with Acustica, 92, 2006: 1-187.
7. Ferguson, S. & Chayes, D. A.: *Use of Generic Sensor Format to Store Multibeam Data*, Marine Geodesy, 18 (4) 1995: 299-315.
8. Foote, K. G., Chu, D., Hammar, T. R., Baldwin, K. C., Mayer, L. A., Hufnagle, L. C., Jech, J. M.: *Protocols for calibrating multibeam sonar*, Journal of the Acoustical Society of America, 114 (4) 2003: 2307-2308.
9. Gerlotto, F., Soria, M., Freon, P.: *From two dimensions to three: the use of multibeam sonar for a new approach in fisheries acoustics*, Canadian Joumal of Fisheries and Aquatic Science, 56 1999: 6-12.
10. Hammerstad, E.: *Advanced multibeam echosounder technology*, Sea Technology, 36 1995: 67-69.
11. Hashemian, R.: *Direct Huffman coding and decoding using the table of code-lengths*, Information Technology: Coding and Computing [Computers and Communications] Proceedings: 237-241, 2003.
12. Huffman, D. A.: *a Method for the Construction of Minimum-Redundancy Codes*, Proceedings of the I.R.E.: 1098-1102, 1952.
13. Khademi, A. & Krishnan, S.: *Comparison of JPEG 2000 and Other Lossless Compression Schemes for Digital Mammograms*. pp. 3771-3774. In: Engineering in Medicine and Biology Society, 2005 IEEE-EMBS 2005 27th Annual International Conference. 17-18 Jan. 2006, Shanghai, China.
14. Kulawiak, M., Chybicki, A., Moszynski, M.: *Web-based GIS as a tool for supporting marine research*, Marine Geodesy 33 (2 & 3) 2010: 135-153.
15. Larmore, L. L.: *Height restricted optimal binary trees*, SIAM Journal of Computing, 16 1986: 1115-1123.
16. Lubniewski, Z., Stepnowski, A. Chybicki, A.: *Seafloor characterisation combined approach using multibeam sonar echo signal processing and image analysis*. p. 131-137. In: Proceedings of 10th European Conference on Underwater Acoustics ECUA 2010, Istanbul, Turkey.
17. Marcus, J.: *SeaBat 7K Data Format Definition for Ifremer*, Revision: B, Reson Inc, 2007.
18. Mayer, L., Li, Y. & Melvin, G.: *3D visualization for pelagic fisheries research and assessment*, ICES Joumal of Marine Science, 59, 2002: 216-225.
19. Mayer, L. A., Raymond, R., Glang, G., Richardson, M. D., Traykovski, P., Trembanis, A. C.: *High-Resolution Mapping of Mines and Ripples at the Martha's Vineyard Coastal Observatory*, IEEE Journal of Oceanic Engineering, 32 (1) 2007: 133-149.
20. Řrnhovde, A.: *High Resolution Beamforming of SIMRAD EM3000 Bathymetric Multibeam Sonar Data*, MSc Thesis, University of Oslo, Norway, 1999, 121 pp.
21. Shah, G. S. & Talukdar, K. K.: *a study of enhanced signal processing on multibeam bathymetric data*, OCEANS'99 MTS/IEEE Riding the Crest into the 21st Century, 2 (2) 1999: 904-909.
22. Shannon, C. E.: *a mathematical theory of communications*, Bell Systems Technical Journal, 27, 1948: 379-423.
23. Thorner, J. E.: *Approaches to sonar beamforming*. p. 69 - 78. In: Proceedings of the 1990 IEEE Southern Tier Technical Conference. April 1990, Binghamton, NY, USA.
24. Titchener, M. R.: *a measure of information*, p 353 - 362. In: Data Compression Conference 2000 Proceedings DCC 2000, 28-30 March 2000. Snowbird, USA.
25. Van Buren, A. L. & Blue, J. E.: *Calibration of underwater transducers at NRL*, Journal of the Acoustical Society of America, 86 (S1) 1990: S60.
26. Wu, L. & Zielinski, A.: *Lossless Compression of Hydroacoustic Image Data*, IEEE Journal of Oceanic Engineering, 22 (1) 1997: 93-101.
27. Xiaofeng, J. & Wenjun, Y.: *Transient calibration of underwater acoustic transducers Measurement*, Science and Technology, 9 (11) 1998.

CONTACT WITH THE AUTHORS

Marek Moszynski, Assoc. Prof.
Andrzej Chybicki, Ph.D.
Marcin Kulawiak, Ph.D.
Zbigniew Lubniewski, Ph.D.
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Narutowicza 11/12
80-233 Gdansk, POLAND
e-mail: marmo@pg.gda.pl
e-mail: andrzej.chybicki@eti.pg.gda.pl
e-mail: marcin.kulawiak@eti.pg.gda.pl

MOST WIEDZY Downloaded from mostwiedzy.pl