COMPRESSION ALGORITHMS FOR MULTIBEAM SONAR RECORDS

ANDRZEJ CHYBICKI[1], MAREK MOSZYŃSKI[1], PAWEŁ POĆWIARDOWSKI[2]

[1]Gdańsk University of Technology
Narutowicza 11/12 Street, Gdańsk-Wrzeszcz , Poland
Andrzej.Chybicki@eti.pg.gda.pl
[2]Reson Inc.,
100 Lopez Road Goleta, California 9311, USA
pap@reson.com

Operational requirements of multibeam sonar systems result in very large volumes of datasets stored on local hard drives of operator's station. In this context, the process of archiving acquired data becomes a crucial problem. The paper investigates various lossy and lossless compression methods that can be applied to multibeam sonar data to reduce the size of acquired files without loosing relevant information. The specific character of MBS data allows applying various signal, image and video compression methods to achieve better results than when using standard compression tools. Various techniques of reordering the data were analyzed to achieve best possible compression ratio.

INTRODUCTION

Seafloor mapping and monitoring is one of the most current issues concerning chart production and dredging industry. Modern multibeam echosounder systems are capable of recording backscatter data for the whole water column, not just for the seabed, as is the case with standard hydrographic systems [1]. This new functionality is of particular interest to the diverse groups such as fisheries acoustics community, navies or hydrographers, for a variety of reasons. Firstly, it is expected that much more detailed information about fish distributions can be derived from multibeam echosounder data, because multibeam systems offer 3-dimensional data compared to the conventional 2-dimensional data sets collected using single beam echosounders [2]. Furthermore, the fact that the same instrument and same data sets can be shared between researchers and hydrographers offers an interesting new perspective, leading to savings in instrumentation and survey costs. In this context, archiving data collected in surveys becomes a crucial problem. The article presents various lossless compression and data reorganizing methods for multibeam sonar records that may lead to reducing quantity of acquired data without loosing relevant information.

1. THEORY

In recent years, multibeam echosounders capable of logging not only bathymetry focused data, but also the full water-column information have become available. Unlike bathymetric multibeam sonars, which only capture the seafloor, full water-column multibeam systems produce very large data sets [3] and allow for visualization and analysis of objects other than the seabed such as single fish, fish schools or pollution. Water column data is a 2-D array where the value of point defined by x and y is the 16-bit integer value of beam numbered x and sample numbered y. Typically the maximal number of beams is between 128 to 3520 per sonar head. Number of samples depends on the depth and probing frequency and can reach up to 32 thousands samples [4]. Since each value is stored as 16-bit value, the size of water column data can exceed up to 95% of all data collected by multibeam system what is shown in Table 1. Therefore, methods of appropriate processing and compressing water-column data seem to be very important in the context of reducing the size of MBS records. Moreover, since water column data is a series of 2-D arrays it can be treated as a series of n-bit gray-scale images, therefore methods of image or video compression methods can be applied.

Tab.1 Content of water-column data in tested datasets

20060719_204657__7125 (400kHz).s7k	**85.72%**
20070315_184108.s7k	**97.96%**
20070720_170721.s7k	**87.51%**
20070831_185543.s7k	**87.87%**

Image compression is achieved by eliminating various types of redundancy that exist in the pixel values. Individual gray-scale images contain interpixel and coding redundancy. However, similar images have similar pixel values in corresponding areas, comparable edge distributions and similar histograms. Sets of similar images contain an additional type of redundancy [8], which is the redundancy resulting from the common information found in more than one image in the set and is called set redundancy. Set redundancy can be considered in the context of compression of similar images. It has been shown [5] that in sets of comparable images, the image entropy H relates to similarity S as follows:

$$H = -S\log(S) - (1-S)\log((1-S)/n)$$

where $n = 65535$ for 16-bit gray-scale images. According to this formula, the value of entropy H decreases as the similarity S increases. The general image compression scheme that consists of three stages: pixel mapping for reducing interpixel redundancy, quantization and symbol encoding for coding redundancy, does not capitalize on set redundancy that exists in sets of similar images.

The research presented in the paper was based on data acquired from RESON 7k series sonars which are capable of logging amplitude and phase water column data what is shown in Fig.1.

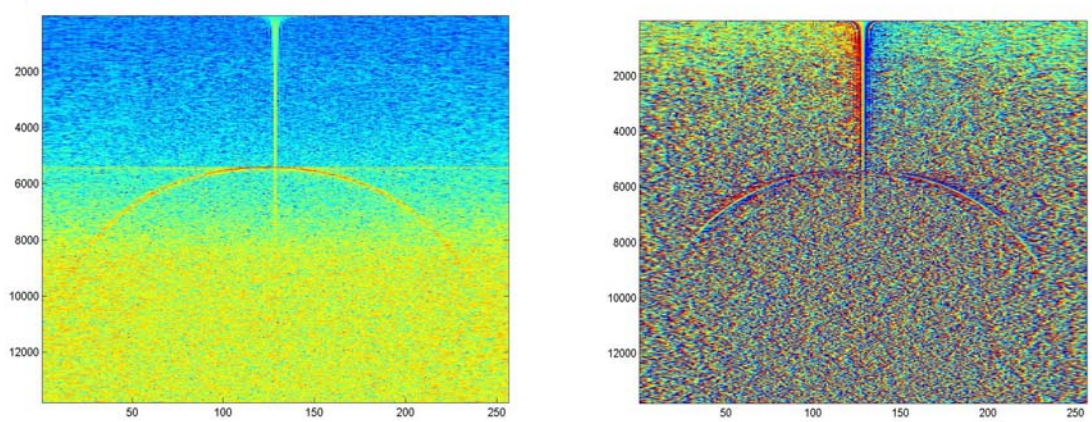

Fig.1 Sample visualization of water column containing amplitude and phase data

2. REORGANIZING THE DATA AND COMPRESSION ALGORITHMS

The compression of water column data has been developed in the research. The overall diagram of compression process is shown in Fig.1.

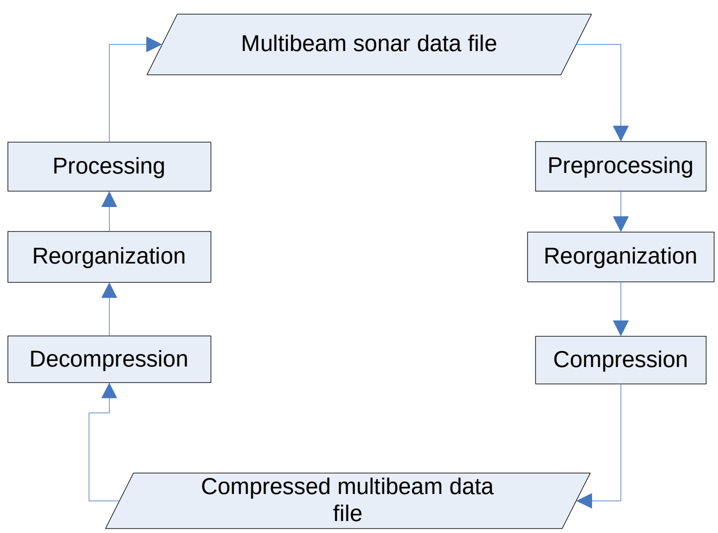

Fig.2 Overall diagram of multibeam data compression process

The compression process of MBS data is divided into three stages: preprocessing, reorganization and lossless compression. The preprocessing block is responsible for reading and processing the MBS data which includes RESON s7K file parsing and data transformation described below. The transformed data is passed to Reorganization Block where one of the reorganization algorithms (Make Matrix, Beam to Beam or Sample to Sample) is applied. The reorganizing and preprocessing blocks are designed to make data more compressible by considering the specific character of applied compression methods.

A closer investigation of beam samples amplitude and phase records revealed that its values are stored in linear scale. Since, the accuracy of transducers used in tested multibeam systems (RESON s7k series) is more than 0.5dB [6], therefore, the appropriate conversion from linear scale to logarithmic scale with 0.1 . 0.5 or 1 dB resolution was also applied in Preprocessing Block.

To make data more compressible, algorithms of water column data reorganization were developed. Many solutions have been tested although three basic mechanisms gave the most satisfying results. To remove redundancy resulting from similar pixel values in the same areas in subsequent water column datasets Beam to Beam (B2B) algorithm was developed. The water-column data is available as a three-dimensional dataset (WaterColumnData) where the first dimension is the number of water column-datasets in compressed. dataset, and second and third is beam and sample number. B2B is a reorganization algorithm which transforms 3D water column dataset to one 2-D matrix with the following procedure:

```
void B2B(WaterColumnData)
for (i=1;i<nWc;i++)
    for (j=1;j<noBeams)
        for (k=1;k<noSamples;k++)
            {
             B2B(k,nWc*(j-1)+i) =WaterColumnData(i,k,j);
             }
```

Where *nWc* – number of water-column datasets in compressed dataset, *noBeams* – number of beams and *noSamples* is the number of samples recorded by multibeam system.

Sample-to-Sample (S2S) is a modification of B2B method where water column data is reorganized with the procedure presented below:

```
void S2S(WaterColumnData)
for (i=1;i<nWc;i++)
    for (j=1;j<noSamples;j++)
        for (k=1:k<noBeams;k++)
            {
             S2S(nWc*(j-1)+i,k) = WaterColumnData(i,j,k);
            }
```

The best results were acquired using Make-Matrix (MM) method, which is a development of previous algorithms. MM is a data reorder method where water-column data is reordered with the procedure presented below:

```
void MM(WaterColumnData)
 for (i=1;i<nWc;i++)
      for (j=1;j<noBeams;j++)
           for (k=1;k<noSamples;k++)
                {
                MM(k,(i-1)*j) = WaterColumnData(i,k,j);
                }
```

The basic idea behind all these methods is to capture and exploit the common information found in similar images, in order to improve image compressibility. All the reorganizing methods presented above are designed to remove set redundancy between subsequent water column datasets.

To achieve highest possible compression ratio lossless compression methods were applied. Several algorithms were tested at the beginning of the research such as (lzw, rle, wavelet transform, lz 77/78). Best results of lossless compression were obtained when using deflate/inflate algorithm that is a combination of lz77/78 and Huffman coding compression. The basic concept of this method is to find repeating series of bytes within the input byte stream. An encoded match to an earlier string consists of a length (3-258 bytes) and a distance (1-32768 bytes). Relative back-references can be made across any number of blocks, as long as the distance appears within the last 32kB of uncompressed data decoded. The second compression stage of deflate consists of replacing commonly used symbols with shorter representations and less commonly used symbols with longer representations. The method used is Huffman coding which creates an unprefixed tree of non-overlapping bit-sequences, where the length of each sequence is inversely proportional to the likelihood of that symbol needing to be encoded [8]. The more likely symbol is to be encoded, the shorter its bit-sequence will be. The deflate/inflate algorithm is a part of PKZIP archiving tool and it is specified in RFC 1951. It is also used in Portable Network Graphics Format (PNG) files to compress the image data [8].

To evaluate the effect of reconstruction process, selecting an appropriate criterion for optimality to meet a specific goal is crucial. It was shown that the Peak Signal to Noise Ratio (PSNR) defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

Where, *MSE* is a mean square error defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}||I(i,j) - K(i,j)||^2$$

and MAX_i is the maximum possible pixel value of the image, *I(i,j)* corresponds to original pixel *(i,j)* of the original image and *K(I,j)* corresponds to pixel *(i,j)* of reconstructed image, is a good criterion to measure digital image reconstruction quality [9-12].

3. RESULTS

The tables below present obtained results for set of testing data acquired from RESON Seabat 7k series sonars. As it is shown, the results obtained by developed reorganizing and compression methods are far better then using standard zip compression of data collected in water column datagram. The first column describes the name of tested dataset; second column is a compression ratio of standard deflate/inflate method used in ZIP defined as:

$$CompressionRatio = \frac{UncompressedSize}{CompressedSize}$$

Following columns contain the compression ratio obtained when using MM, B2B and S2S methods. The results are divided into three datasets – each one for desired accuracy: 0.1dB, 0.5dB, and 1dB. Reorganization and compression algorithms were developed using MATLAB 6.5.

Tab.2 shows the results of applied algorithms using 0.1dB accuracy for phase and amplitude data. Obtained compression ratios using MM, B2B and S2S methods are much better then when using standard ZIP compressor. The best results were obtained when using Make Matrix reorganization algorithm, although Sample-to-Sample revealed to be fairly good.

Tab.2 Results of compression using MM, B2B and S2S algorithms with 0.1 dB accuracy

Data Set name	Zip Ratio	MM Ratio	B2B Ratio	S2S Ratio	PSNR
20060719_204657__7125 (400kHz).s7k	2,07	2,86	2,25	2,84	51,8
20070315_184108.s7k	1,26	2,25	2,19	2,24	49,1
20070720_170721.s7k	2,46	2,83	2,5	2,76	54,0
20070831_185543.s7k	1,88	2,84	2,59	2,6	46,9

Results of compression with 0.5dB accuracy for amplitude and beam data are shown in Tab.3.

Tab.3 Results of compression using MM, B2B and S2S algorithms with 0.5 dB accuracy

Data Set name	Zip Ratio	MM Ratio	B2B Ratio	S2S Ratio	PSNR
20060719_204657__7125 (400kHz).s7k	2,07	4,26	3,3	4,24	46,2
20070315_184108.s7k	1,26	3,33	3,23	3,28	43,0
20070720_170721.s7k	2,46	3,76	3,46	3,78	48,3
20070831_185543.s7k	1,88	4,13	3,89	3,78	40,3

Obtained compression ratios are much better than when using 0.1dB resolution. In one dataset (20070720_170721.s7k) the results from S2S were better then MM.

Tab.4 Results of compression using MM, B2B and S2S algorithms with 1 dB accuracy

Data Set name	**Zip Ratio**	**MM Ratio**	**B2B Ratio**	**S2S Ratio**	**PSNR**
20060719_204657__7125 (400kHz).s7k	2,07	5,25	4,11	5,22	43,5
20070315_184108.s7k	1,26	4,11	3,99	4,04	40,2
20070720_170721.s7k	2,46	4,55	4,21	4,5	45,7
20070831_185543.s7k	1,88	4,98	4,68	4,57	37,4

Tab.4 presents results when using 1dB accuracy. In most of the cases MM and S2S algorithms gave the same or almost the same results. The compression methods were applied separately for the amplitude phase beam data.

4. CONCLUSIONS

It is clear, that multibeam water-column imaging provides a viable means of examining mid-water and near-bottom echoes. Investigation results show that most of the information acquired from water-column data may be reduced with proposed methods, yielding much better results than using standard compression tools. Moreover, it was shown that the compression ratio also depends on applied reordering methods and, obviously, on applied accuracy.

Presented results show that using dedicated methods of re-ordering and compressing multibeam records can lead to significant reduction of data acquired by multibeam systems and can improve performance of multibeam sonars. The future work will focus on online compression algorithms of multibeam sonar records and improvement of presented methods.

REFERENCES

1. E. Hammerstad, “Advanced multibeam echosounder technology,” Sea Technology, vol. 36, pp. 67-69,1995.
2. B. Buelens, R. Williams, A. Saley, T. Pauly, “Model Inversion for Midwater Multibeam Backscatter Data Analysis”, Oceans - Europe 2005
3. B. Buelens, R. Williams, A. Sale, T. Pauly “Midwater acoustic modeling for multibeam sonar simulation”
4. Data Format Definition, Seabat 7k Data Format , Volume I. Version 0.54
5. K. Karadimitriou, “Set Redundancy, the Enhanced Compression Model, and Methods for Compressing Sets of Similar Images”, Ph.D. Dissertation, Louisiana State University, Baton Rouge, LA, August 1996.
6. http://reson.com
7. K. Karadimitriou, M. Fenstermacher “Image Compression In Medical Image Databases Using Set Redundancy”, Division of Diagnostic Imaging, M.D. Anderson Cancer Center, Houston, TX
8. Antaneus Feldspar, “An Explanation of the Deflate Algorithm”, http://zlib.net/
9. Atsuro Ichigaya, Masaaki Kurozumi, Naohiro Hara, Yukihiro Nishida, and Eisuke Nakasu , “A Method of Estimating Coding PSNR Using Quantized DCT Coefficients”, IEEE Transactions on circuits and systems for video technology, vol. 16, no. 2, Februrary 2006
10. Tommy C. L. Chan, Tai-Chiu Hsung, Member, IEEE, and Daniel Pak-Kong Lun, Member, IEEE, “Improved MPEG-4 Still Texture Image Coding Under Noisy Environment” IEEE transactions on image processing, vol. 12, no. 5, May 2003
11. Qian Du, Member, IEEE, and Chein-I Chang, Senior Member, IEEE, “Linear Mixture Analysis-Based Compression for Hyperspectal Image Analysis”, IEEE Transactions on Geoscience and remote sensing, vol 42, no. 4, April 2004
12. R. A. Schowengerdt, “Remote Sensing: Models and Methods for Image Processing”, 2nd ed. Orlando, FL: Academic, 1997, pp. 470–471.