Compression Techniques for BioFilm Optical Coherence Tomography (OCT) Data

MAX H. FAYKUS, Clemson University, USA
DR. MELISSA C. SMITH (ADVISOR), Clemson University, USA
DR. JON C. CALHOUN (ADVISOR), Clemson University, USA

Optical Coherence Tomography (OCT) is as a fast and non-destructive technology for bacterial biofilm imaging. However, OCT generates approximately 100 GB per flow cell, which complicates storage and data sharing. Data reduction reduces data complications by reducing the overhead and the amount of data transferred. This work leverages the similarities between layers of OCT images to minimize the data in order to improve compression. This paper evaluates the 5 lossless and 2 lossy state-of-the-art compressors to reduce the OCT data. The reduction techniques are evaluated to determine which compressor has the most significant compression ratio while maintaining a strong bandwidth and minimal image distortion. Results show that SZ with frame before pre-processing is able to achieve the highest CR of 204.6× on its higher error bounds. The maximum compression bandwidth SZ on higher error bounds is $\sim 41MB/s$, for decompression bandwidth, it is able to outperform ZFP achieving $\sim 67MB/s$.

CCS Concepts: • **Information systems** → **Data compression**; *Data layout*.

Additional Key Words and Phrases: compression, lossy compression, high performance computing, optical coherence tomography

ACM Reference Format:
Max H. Faykus, Dr. Melissa C. Smith (advisor), and Dr. Jon C. Calhoun (advisor). 2023. Compression Techniques for BioFilm Optical Coherence Tomography (OCT) Data . 1, 1 (September 2023), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Biofilms are a growth that occur on ships by microorganisms, plants, and algae[7]. They can create drag penalties, causing lower energy efficiency on crafts[7]. Optical coherence tomography (OCT) is a fast and non-destructive method to image biological systems[8][12][10] including biofilms. OCT datasets are generally large[2], which causes difficulties in storage and data transmission. Data reduction is needed for long-term storage and for data transfer to clusters for analysis. This poster evaluates different methods for data reduction utilizing both various methods of pre-processing and compressors (lossless and lossy) on OCT biofilm data to determine the most significant compression ratio (CR) while maintaining strong bandwidth and accuracy.

Authors' addresses: Max H. Faykus, mfaykus@g.clemson.edu, Clemson University, Clemson, South Carolina, USA; Dr. Melissa C. Smith (advisor), Clemson University, Clemson, South Carolina, USA, mcsmith@clemson.edu; Dr. Jon C. Calhoun (advisor), Clemson University, Clemson, South Carolina, USA, jonccal@clemson.edu.

Manuscript submitted to ACM

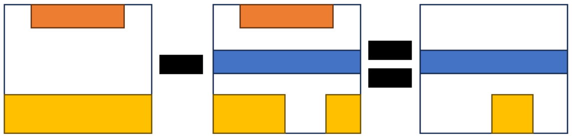


Fig. 1. Lossy Differences Calculation

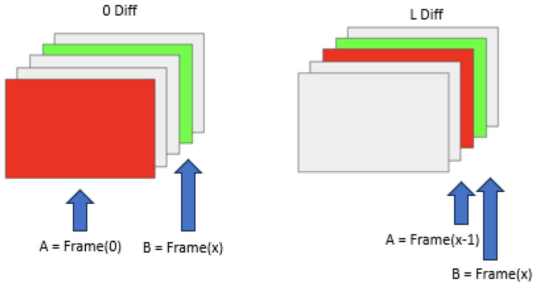


Fig. 2. Differences Iteration

2 METHODOLOGY

In this poster we explore five lossless and two lossy methods for data reduction including ZSTD (lossless)[4], BLOSCLZ (lossless)[1], LZ4/LZ4HC (lossless)[3], ZLIB (lossless)[5], SZ (Lossy) [6], and ZFP (Lossy)[9]. Each of these compressors were tested with unaltered data and two pre-processing methods.

2.1 Evaluation Metrics

In data reduction, the most important metrics are size reduction, speed, and accuracy, which is denoted by CR, bandwidth, and image distortion. The CR is used to show the effectiveness of different compressors and demonstrates how far the data can be reduced. Compression/decompression bandwidth is the elapsed time for data to be compressed and decompressed, respectively. SSIM is a metric for lossy compression that evaluates the structural degradation of image quality during compression [11].

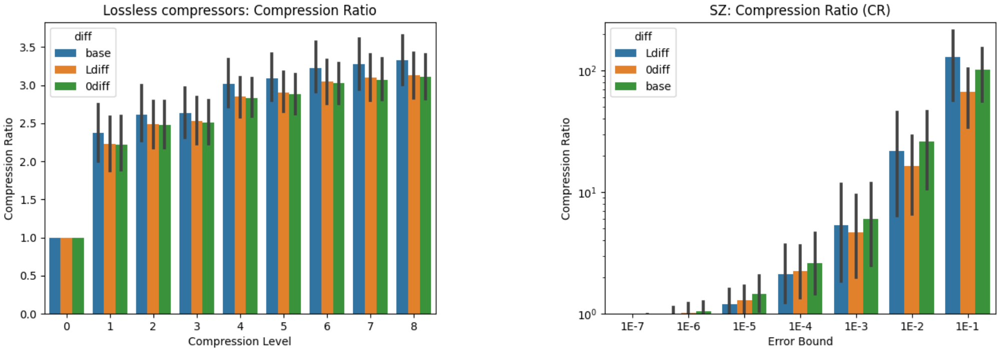
2.2 Data Pre-processing

To improve the CR, the 3D nature of OCT images is leveraged and tested with multiple pre-processing steps to reduce the variation of data points. These methods include:

1. Frame 0 Diff: The difference is calculated between the first frame of the OCT image and every subsequent image frame. The resulting diff values are passed to the compressors (Figures 1 and 2).
2. Frame before Diff: The difference of each frame is taken from the preceding frame. The compressors are given the resulting differences of each frame (Figures 1 and 2).

CPU	Intel® Xeon® Gold 6258R CPU @ 2.70GHz
Architecture	X86_64
GCC	12.1.0
SZ Version	2.1.12
ZFP Version	1.0.0
ZSTD Version	1.5.5
LZ4 Version	1.9.4
ZLIB Version	1.2.13
BLOSC Version	1.21.2
LIBPRESSIO Version	0.94.0

Fig. 3. Software Versions



(a) Lossless

(b) SZ Lossy

Fig. 4. Difference Calculation

3 RESULTS

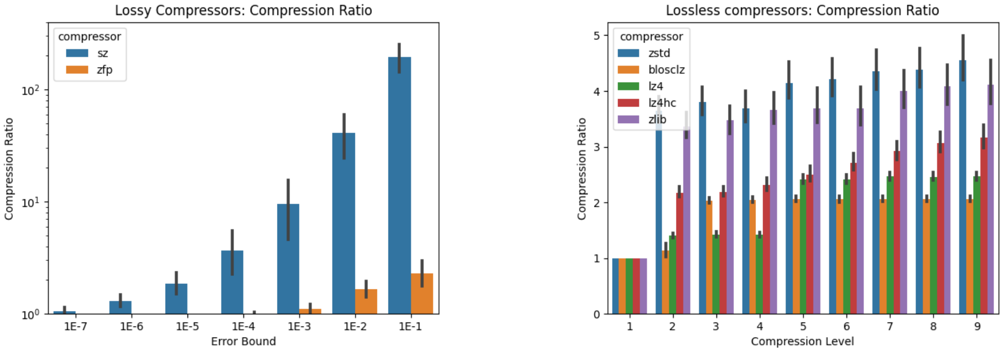
3.1 Pre-Processing

Utilizing pre-processing yielded negative compression on lower error bounds and a higher CR than unaltered data on higher error bounds. Ldiff achieves the highest CR on higher error bounds with SZ but degraded in SSIM (Figure 4b). ZFP and lossless was unable to benefit from pre-processing (Figure 4a).

3.2 Compressibility

Each of the compressors are run over the entire dataset on the Palmetto Cluster using software versions shown in Figure 3, and the average CR are presented in Figures 5. Lossy compressors are evaluated over an error bound which is used to set the accuracy of the data to analyze its effects on CR, compression bandwidth and decompression bandwidth. The absolute error bound is varied between 1e-7 to 1e-1. As the error bound for the compressors increases, so does the CR. The maximum CR achieved by SZ is 927.85x and by ZFP is 10.82x.

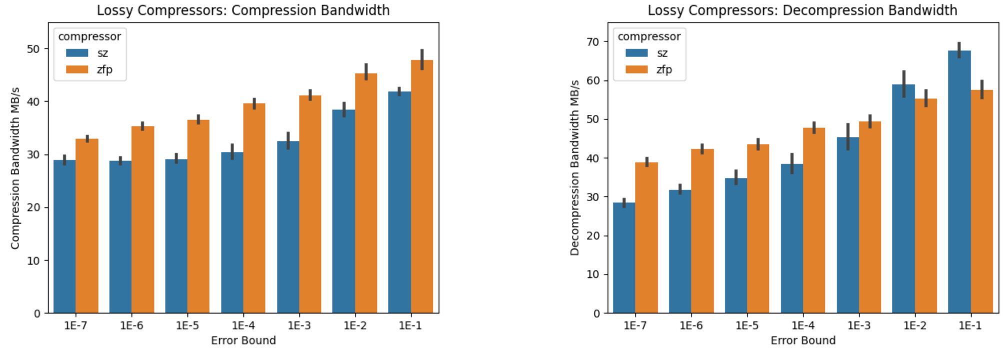
For lossless compressors, the best CR is ZSTD while performing average for compression bandwidth and decompression bandwidth (Figure 6). The frame before pre-processing is able to achieve up to an 8.707% CR increase on lossy SZ



(a) Lossy CR

(b) Lossless CR

Fig. 5. Compression Ratios



(a) Compression Bandwidth

(b) Decompression Bandwidth

Fig. 6. Bandwidths MB/s

and a 4.56% increase on lossless compressors. SZ can achieve a higher CR when compared to ZFP, but its SSIM degrades when the error bounds increase. ZFP has greater compression and decompression bandwidths compared to SZ.

3.3 Image Distortion

Lossless compressors perfectly preserve the data, while lossy compressors will lose information to improve CR. Lossy compression achieves a higher CR at the cost of image distortion. Figure 7 shows the structural similarity index measure (SSIM) over various error bounds. SZ begins degrading at 1e-2 while ZFP remains consistent.

4 CONCLUSION

To effectively analyze biofilm OCT images, data reduction is essential. These results show that although lossless methods are perfectly preserved and reduced the data, they are slower and have less CR compared to lossy methods. We also explored leveraging the 3D nature of the OCT by taking the difference between frames. Improvements on higher error

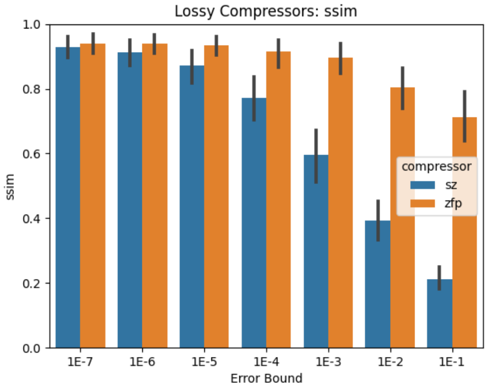


Fig. 7. Lossy Absolute Error Bound on SSIM

bounds shows how the volume data can be leveraged. Ldiff was able to achieve the highest CR on higher error bounds with SZ. ZSTD yields the best CR for lossless compressors, while SZ is able to provide the best compression for lossy compressors. ZFP was seen to have the higher SSIM, compression bandwidth, and decompression bandwidth when compared to SZ. Lossless compressors trade compression time for smaller file size and lossy compressors trade data distortion for smaller file sizes. In conclusion, this study shows SZ with Ldiff pre-processing is the best compressor for the biofilm OCT dataset with higher CRs and bandwidths and also outperforms lossless methods.

ACKNOWLEDGMENTS

Clemson University is acknowledged for their generous allotment of compute time on the Palmetto Cluster. This material is based on work supported by the National Science Foundation under Grant Nos. MRI# 2024205, MRI# 1725573, MRI# 1725573, and CRI# 2010270.

REFERENCES

[1] Francesc Alted. 2017. Blosc, an extremely fast, multi-threaded, meta-compressor library.
[2] Josef F Bille. 2019. High resolution imaging in microscopy and ophthalmology: new frontiers in biomedical optics. (2019).
[3] Yann Collet. 2011. LZ4 lossless compression algorithm. *https://lz4.org/* (2011).
[4] Yann Collet and Murray Kucherawy. 2018. *Zstandard Compression and the application/zstd Media Type.* Technical Report.
[5] Peter Deutsch and Jean-Loup Gailly. 1996. *Zlib compressed data format specification version 3.3.* Technical Report.
[6] Sheng Di and Franck Cappello. 2016. Fast error-bounded lossy HPC data compression with SZ. In *2016 ieee international parallel and distributed processing symposium (ipdps).* IEEE, 730–739.
[7] Andrea Farkas, Soonseok Song, Nastia Degiuli, Ivana Martić, and Yigit Kemal Demirel. 2020. Impact of biofilm on the ship propulsion characteristics and the speed reduction. *Ocean Engineering* 199 (2020), 107033.
[8] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. 1991. Optical coherence tomography. *science* 254, 5035 (1991), 1178–1181.
[9] Peter Lindstrom. 2014. Fixed-rate compressed floating-point arrays. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2674–2683.
[10] Michael Wagner and Harald Horn. 2017. Optical coherence tomography in biofilm research: a comprehensive review. *Biotechnology and Bioengineering* 114, 7 (2017), 1386–1402.
[11] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.
[12] Zahid Yaqoob, Jigang Wu, and Changhuei Yang. 2005. Spectral domain optical coherence tomography: a better OCT imaging strategy. *Biotechniques* 39, 6 (2005), S6–S13.