A new compression format for tractography datasets reconstructed from dMRI

Caroline Presseau[1], Pierre-Marc Jodoin[1], and Maxime Descoteaux[1]
[1]*University of Sherbrooke, Sherbrooke, Quebec, Canada*

Target Audience : This work is addressed to all researchers who are interested in dMRI tractography studies.

Purpose: For a single brain, a fiber tracking dataset may contain hundreds of thousands, and often millions of streamlines and can take up to several gigabytes of memory. This amount of data is not only heavy to compute, but also difficult to visualize and hard to store on disk (especially when dealing with a collection of brains). These problems call for a fiber-specific compression format that simplifies its manipulation. As of today, no fiber compression format has yet been adopted and the need for it is now becoming an issue for future connectomics research. In this work, we propose a new compression format, *.zfib*, for tractography datasets reconstructed from diffusion magnetic resonance imaging (dMRI). Our pipeline consists of five steps that are tested, evaluated and validated under a wide range of tractography configurations (step size, deterministic, probabilistic, tensorline, streamline number) and compression options. Similar to JPEG, the user has one parameter to select: a worst-case maximum tolerance error in millimeter (mm).

Methods: *Dataset:* dMRI images were acquired on a whole-body 3T Magnetom Trio scanner (Siemens, Erlangen) equipped with an 8-channel head array coil. The spin-echo echo-planar-imaging sequence (TE = 100 ms, TR = 12 s, 128x128 image matrix, FOV = 220x220 mm^2) consists of 60 diffusion encoding gradients with a b-value of 1000 s/mm^2 [1]. Tractography was done using state-of-the-art MRtrix [2], with fiber orientation distributions of order 8, step size of 0.1, 0.2, 0.5, 1, number of streamlines 60K, 120K, 240K and deterministic, probabilistic and tensorline tracking with default parameters (see [2] for more details).

Compression pipeline: The proposed method is a five-step processing pipeline containing a linearization, a transformation, a nonlinear coefficient approximation, a rounding and an encoding step. The process of linearization is to remove consecutive streamline points that are within a maximum error threshold set by the user. We then implement two transformations: i) a Discrete Cosine Transform (DCT) and ii) a Fast Wavelet Transform (Daubechies 4 and 6, and Cohen-Daubechies-Fauveau (CDF) 5-3 and 9-7 [3]) along each dimension (x, y, z) of each streamline respectively. This transformation concentrates energy on a small number of coefficients approximated using a nonlinear approximation that keeps the K largest coefficients while others are set to zero [3]. The smaller is the K, the more aggressive is the compression, and the larger is the error in millimeter. Hence, we set the smallest K that gives a maximum error below the maximum error threshold set by the user. We then use uniform and non-uniform rounding that respectively consists to round each non-zero coefficient at its 10^n position, and up to its first *n* non-zero digits [5]. Four quantitative measures were used for analysis - the maximum (max) and the mean errors between the original and the compressed streamlines, the processing time and the compression ratio.

Tab 1. Compression comparison of transformations on streamlines computed with step size of 0.2 mm on a subsampled dataset with 2 500 streamlines

Transformation	Max error (mm)	Mean error (mm)	Comp. ratio (%)
DCT	0.502261	0.233636	97.8278
Daubechies 4	0.515364	0.197055	83.5009
Daubechies 6	0.504040	0.197144	83.2497
CDF 5-3	0.511255	0.198732	83.6523
CDF 9-7	0.510941	0.199029	84.2380

Tab 2. Compression comparison of tractography algorithms with 0.5 step size and 120 000 streamlines

Algorithms	Max error (mm)	Mean error (mm)	Comp. ratio (%)
Determistic	0.502261	0.233636	97.8278
Probabilistic	0.505909	0.222020	97.9625
Tensorline	0.502957	0.262439	98.0292

Tab 3. Compression comparison of step size in deterministic tractography with 120 000 streamlines

Step size (mm)	Max error (mm)	Mean error (mm)	Comp. ratio (%)
0.1	0.504231	0.2488230	99.5641
0.2	0.503730	0.0244969	99.2685
0.5	0.502261	0.233636	97.8278
1.0	0.501944	0.217874	96.7363

Tab 4. Compression comparison of the number of streamlines in deterministic tractography

Number of streamlines	Max error (mm)	Mean error (mm)	Comp. ratio (%)
60 000	0.501515	0.244791	97.7371
120 000	0.503166	0.244739	98.0020
240 000	0.519700	0.244825	99.5539

Results: As seen in Tab.1, the DCT showed a stronger energy compaction, which concentrates almost all the energy of the streamline in a few coefficients. This means that a small number of coefficients are sufficient to represent each streamline and that the DCT is a good choice for compression purposes. As expected, the uniform rounding presented more stable results according to max and mean errors and compression ratio (CR), and arithmetic encoding always gave a bigger CR than Huffman encoding (not shown in tables). This is inline with the theory that states that arithmetic coding is always better or at least identical to Huffman coding [4]. Next, Tab. 2 shows that the tensorline algorithm is more compressible, as expected, because streamlines are smoother when computed from the tensor model. CR is also similar for both deterministic and probabilistic algorithms. Tab. 3 shows that the smaller is the step size, the bigger is the CR. This suggests that a small step size produces a large number of co-linear points that are removed by the linearization step. Then, Tab. 4 shows that the more streamlines there are, the more chance that streamlines closely overlap, which means that they have very similar representations resulting in less symbols to encode and thus, a bigger CR. For example, the 240 000 streamlines file of Tab. 4 was originally 668Mb and was compressed to 3Mb! It is worth noting that the voxel resolution of the dMRI data is often on the order of 2 mm isotropic, or now 1.25 mm with the Human Connectome Project, and that 0.5 mm max error is almost unperceivable to the eye when uncompressing the compressed streamlines from our technique.

Discussion & Conclusion: In this work, we have presented a new compression format (.zfib) and a new pipeline in five steps that shows impressive compression ratios. Overall, with datasets of at least 60 000 streamlines generated from more or less any streamline tractography algorithm with a step size below 1 mm and a resulting max tolerance error of 0.5 mm, a compression ratio of 97% or better can be obtained. This is a huge gain and of great potential for future connectomics and group studies using tractography results with large number of streamlines.

References: [1] Anwander et al., Cerebral Cortex 2007. [2] Tournier et al., International Journal of Imaging Systems and Technology, 2012. [3] Mallat, Stephane, A Wavelet Tour of Signal Processing, 2008. [4] Sayood, Khalid, Introduction to Data Compression, 2006. [5] Cover Thomas M, Elements of Information Theory, 2006.