



The fiber-density-coreset for redundancy reduction in huge fiber-sets



Guy Alexandroni^a, Gali Zimmerman Moreno^b, Nir Sochen^c, Hayit Greenspan^{b,*}

^a Department of Electrical Engineering, Tel Aviv University, 69978, Israel

^b Department of Biomedical Engineering, Tel Aviv University, 69978, Israel

^c Department of Applied Mathematics, Tel Aviv University, 69978, Israel

ARTICLE INFO

Keywords:

DWMRI
HARDI
Coreset
Redundancy reduction
Tractogram simplification
Fiber clustering

ABSTRACT

State of the art Diffusion Weighted Magnetic Resonance Imaging (DW-MRI) protocols of white matter followed by advanced tractography techniques produce impressive reconstructions of White Matter (WM) pathways. These pathways often contain millions of trajectories (fibers). While for several applications the high number of fibers is essential, other applications (visualization, registration, some types of across-subject comparison) can achieve satisfying results using much smaller sets and may be overburdened by the computational load of the large fiber sets. In this paper we propose a novel, highly efficient algorithm for extracting a meaningful subset of fibers, which we term the Fiber-Density-Coreset (FDC). The reduced set is optimized to represent the main structures of the brain. FDC is based on an efficient geometric approximation paradigm named coresets, an optimization scheme showing much success in tasks requiring large computation time and/or memory. FDC was compared to two commonly used methods for selecting a reduced set of fibers: fiber-clustering and downsampling. The reduced sets were evaluated by several methods, including a novel structural comparison to the full sets called 3D indicator structure comparison (3D-ISC). The comparison was applied to High Angular Resolution Diffusion Imaging (HARDI) scans of 15 healthy individuals obtained from the Human Connectome Project. FDC produced the most satisfying subsets, consistently in all 15 subjects. It also displayed low memory usage and significantly lower running time than conventional fiber reduction schemes.

1. Introduction

Diffusion Weighted Magnetic Resonance Imaging (DW-MRI) is a powerful tool for noninvasive characterization of brain tissues. It maps the architecture of neural connections by measuring the local orientation of water molecule diffusion (Prčková et al., 2012). Magnitude and orientation of the diffusion of water molecules in each voxel can be inferred by computing a diffusion model that best fits the DW-MRI data. A review of common diffusion model reconstruction techniques can be found in the literature (e.g., Assemlal et al., 2011). The reconstructed diffusion function is used by tractography algorithms in order to create a set of streamlines representing the main pathways of neuron bundles in the brain (Yamada et al., 2009). Such sets of streamlines are usually called tractograms or fiber-sets. Modern reconstruction techniques, in conjunction with sophisticated tractography algorithms allow the creation of highly dense fiber-sets comprised of up to several millions of fibers. In recent years the use of such

huge fiber-sets has become more common due to several reasons: First and most important, as the spatial resolution of MRI scanners increases, so does the number of voxels in the scan result. Hence, millions of seeds are needed during the tractography process to achieve good spatial coverage. The result is a fiber-set size of several million fibers. The coverage issue is relevant also to brain connectivity analysis, a relatively new field of study in brain science, as was extensively reviewed by Rubinov and Sporns (Rubinov and Sporns, 2010). In order to achieve an accurate connectivity matrix on high resolution DW-MRI, one needs to produce very dense tractograms containing millions of fibers (Hagmann et al., 2008). Another technique which has recently gained popularity is the application of Tract-Density Imaging (TDI) (Calamante et al., 2010). TDI produces super-resolved qualitative maps of the streamline count (density) at each voxel. In order to achieve TDI maps with low variability, sets of 5–10 millions of fibers are usually required (Calamante et al., 2012).

While relatively efficient applications, such as TDI and connectivity

Abbreviations: DW-MRI, Diffusion Weighted Magnetic Resonance Imaging; QSDR, Q-space Diffeomorphic Reconstruction; TDI, track density imaging; HC, Hierarchical Clustering; QB, QuickBundles; FDC, Fiber-Density-Coreset; HCP, Human Connectome Project; PMF, probability mass function; CSD, constrained spherical deconvolution; UNC, Uncinate tracts; CS, Corticospinal tracts; IFO, Inferior-Fronto-Occipital tracts; CC, Corpus Callosum tract; MCP, Middle Cerebellar Peduncle tract; 3D-ISC, 3D indicator structure comparison; DS, down-sampling; RS, random-sampling

* Corresponding author. Tel.: +972 3 6407398.

E-mail addresses: guy.alexandroni@gmail.com (G. Alexandroni), hayit@eng.tau.ac.il (H. Greenspan).

<http://dx.doi.org/10.1016/j.neuroimage.2016.11.027>

Received 12 July 2016; Accepted 12 November 2016

Available online 14 November 2016

1053-8119/© 2016 Elsevier Inc. All rights reserved.

analysis, are able to process huge sets in reasonable time, other common operations on fibers cannot handle this load of work. This includes applications which are based on dissimilarity measurements between sets of fibers, as well as other complex computations on individual fibers. Another acute problem accompanying the use of huge fiber-sets is the overwhelming memory space needed to store these sets. This also makes rendering of a full fiber-set virtually impossible on most PCs. Hence, huge fiber-sets, which are becoming more widespread, cannot be used for many brain research applications, without some sort of initial simplification.

There have been some efforts in recent years to address the issue of handling huge fiber-sets. One possible approach is to find more efficient representations for individual fibers. Among others, efficient representations have been studied by Zvitia et al. (2010) and Presseau et al. (2015) which performed optimization on the number of samples along a given fiber and their locations. This optimization significantly reduces the computation load; however, when dealing with multiple sets of millions of fibers, this may not be enough for many complex operations. Presseau et al. also suggested a novel compression format that in addition to the efficient representation, also includes quantization and encoding steps, which greatly reduce the amount of memory needed to store full fiber-sets. This may be suitable to some of the fiber-based applications, but since calculations cannot be done on the compressed form, complex operations still pose a challenge.

Dense fiber-sets contain multiple redundancies in the form of nearly identical fibers, as shown in Fig. 1. It is therefore possible to remove redundant fibers and to keep only unique representatives, as a means to construct a reduced fiber-set.

The reduced fiber-set can be utilized for various purposes:

- It can assist complex algorithms with handling huge number of fibers, for example in the construction of atlases (Guevara et al., 2012; Yoo et al., 2015), or in the registration of fiber-sets.
- Visualization purposes: The tremendous amount of memory needed for handling huge sets inhibits their visualization due to Random Access Memory (RAM) limitations on most computers. Currently this is resolved by using aggressively downsampled sets, but this may lead to a loss of various parts of WM structures, as demonstrated in Fig. 2.
- Training purposes: Supervised learning frameworks, which involve complex operations on individual fibers and thus have a limit on the training set size, may benefit by training on reduced sets that were intelligently inferred from multiple brain scans.
- Reduced sets can facilitate multi-subject comparisons, in which the fiber load is multiplied by the number of subjects (e.g. in a multi-subject comparison framework, (Zimmerman-Moreno et al., 2016)).

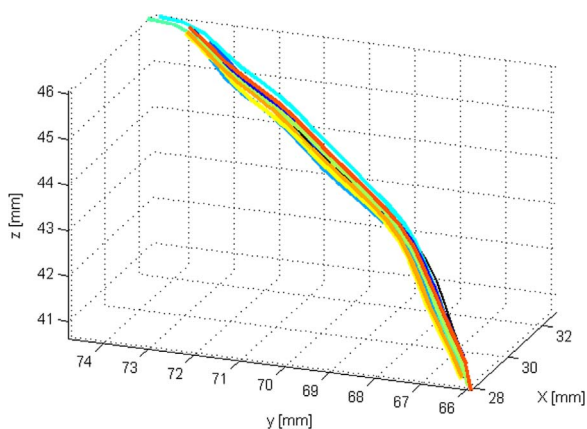


Fig. 1. A fiber and 10 of its closest, almost identical, neighbors. The 10 neighboring fibers were derived by computing a Cosine distance between the given fiber and all the other fibers in a fiber-set and collecting the 10 closest ones. The presence of multiple identical fibers is typical in dense fiber-sets.

Usually, smaller sets are achieved by inferring a subset of fibers from the full set. The simplest and hence, the most common approach is to downsample the fiber-set. Downsampling is equivalent to randomly choosing a subset. It may result in losing important information about the connectivity of the brain. In extreme conditions, parts of tracts can vanish altogether. This phenomenon is illustrated in Fig. 2.

A zoom-in view on a portion of the original fiber-set is shown in Fig. 2 (left). Fig. 2 (right) is the result of downsampling the full set by a factor of 10. Note that the horizontal bundle of fibers (circled) was lost in the downsampling process.

Recently, several more sophisticated approaches for selecting a subset of representative fibers were proposed. Gori et al. (Gori et al., 2016) use an approximation scheme for fiber bundles which results in a parsimonious representation of weighted prototypes. Prototypes are chosen to represent groups of similar fibers, and both the fibers and the prototypes are modeled as weighted currents. Two streamlines are considered similar if their endpoints are close to each other and if their pathways follow similar trajectories. The selection of the prototypes is based on minimization of an approximation error. The resulting representation was shown to preserve bundle shapes and can be used to accurately reconstruct the original structural connectivity. While providing a representation that stores the connectivity information, this process is still quite complex. The reported computation time is almost 3 h for the reconstruction of an 80,000 prototype set.

Other sophisticated frameworks are often based on clustering. Guevara et al. (2011) presented a preprocessing step to be used before the analysis of huge fiber-sets. First, the set is split according to location and fiber length. The voxels are clustered into parcels using k-means and a parcel connectivity matrix is computed according to the number of tracts passing through each pair of parcels. The parcels are clustered using Hierarchical Clustering (HC) and fiber bundles are inferred. Those bundles are further split; centroids are computed and merged using HC and Hausdorff distance. HC was also used by Wassermann et al. (2010). They propose to represent the fibers and fiber bundles as Gaussian Processes, with a similarity measure based on inner product between Gaussian Processes couples. This similarity measure does not rely on point correspondences and hence is suitable for fibers of different lengths. Garyfallidis et al. (2012) developed the QuickBundles (QB), an efficient clustering method which claims to overcome the complexity of large fiber-sets and provides informative clusters. Each QB cluster can be represented by a single centroid streamline and thus simplify the fiber-set representation. The number of clusters in QB depends on a predefined maximal intra cluster pairwise fiber distance; the method has a linear complexity with respect to the number of fibers. Reichenbach et al. (2015) came up with the V-Bundles clustering which is also linear in the number of line segments in the fiber data and can cluster large datasets without the use of random sampling or complex multi-pass procedures. It copes with interrupted streamlines and allows multi-subject comparisons.

Several recent works propose methods for selection of subsets of fibers, not for fiber-set reduction, but rather in order to make the tractograms more consistent with the underlying neuroanatomy (e.g., SIFT (Smith et al., 2013) and COMMIT (Daducci et al., 2015)). These fiber-filtering methods assign weights to candidate fibers in a large fiber super-set. The weights are used to filter out unwanted fibers such that the final fiber density will reflect the true density of underlying neuronal connections.

In the current work we propose a quick and efficient method for the reduction of the number of fibers in large, whole-brain tractograms. The method is based on an innovative paradigm in data science, called Coresets. The term coreset is used to indicate a set of size c , which is a representative subset of the full data set. Finding a small representative subset of objects can be viewed as a geometric approximation problem. Agarwal et al. (2005) have reviewed several works which employ the idea of coresets to develop efficient approximation algorithms for various geometric problems. Among others, coresets were used for

Download English Version:

<https://daneshyari.com/en/article/5631335>

Download Persian Version:

<https://daneshyari.com/article/5631335>

[Daneshyari.com](https://daneshyari.com)