Published in final edited form as:

Connect Neuroimaging (2018). 2018 September ; 11083: 20–28. doi:10.1007/978-3-030-00755-3_3.

GIFE: Efficient and Robust Group-wise Isometric Fiber Embedding

Junyan Wang, Yonggang Shi

Laboratory of Neuro Imaging (LONI), USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of University of Southern California, Los Angeles, CA 90033, USA

Abstract

Tractography is a prevalent technique for in vivo imaging of the white matter fibers (a.k.a. the tractograms), but it is also known to be error-prone. We previously propose the Group-wise Tractogram Analysis (GiTA) framework for identifying anatomically valid fibers across subjects according to cross-subject consistency. However, the original framework is based on computationally expensive brute-force KNN search. In this work, we propose a more general and efficient extension of GiTA. Our main idea is to find the finite dimensional vector-space representation of the fiber tracts of varied lengths across different subjects, and we call it the group-wise isometric fiber embedding (GIFE). This novel GIFE framework enables the application of the powerful and efficient vector space data analysis methods, such as the k-d tree KNN search, to GiTA. However, the conventional isometric embedding frameworks are not suitable for GIFE due to the massive fiber tracts and the registration errors in the original GiTA framework. To address these issues, we propose a novel method called multidimensional extrapolating (MDE) to achieve GIFE. In our experiment, simulation results show quantitatively that our method outperforms the other methods in terms of computational efficiency/tractability and robustness to errors in distance measurements for real fiber embedding. In addition, real experiment for group-wise optic radiation bundle reconstruction also shows clear improvement in anatomical validity of the results from our MDE method for 47 different subjects from the Human Connectome Project, compared to the results of other fiber embedding methods.

1 Introduction

Tractography computes white matter fiber tracts from diffusion MRI and it is a prevalent technique for in-vivo measurement of anatomical connectivity of the brain. However, this technique is also known to be error-prone [1].

Fiber filtering methods have been proposed previously to remove redundant and anatomically invalid fibers based on data fidelity [2], geometrical soundness [3] and anatomical knowledge [4]. Yet, these methods did not explicitly address the inter-subject consistency and the results are not guaranteed to be consistent across subjects, or reproducible. *We hypothesize that the errors in the fiber tracts computed with tractography*

²<http://tosca.cs.technion.ac.il>

³<http://lvdmaaten.github.io/>

are random and they are not anatomically consistent across different subjects. Accordingly, we recently proposed a data-driven framework called Group-wise Tractogram Analysis (GiTA) which identifies the anatomically valid tracts as the common tracts among different subjects [5]. The idea is to measure the *commonness* of each fiber tract with respect to all different subjects and identify those that are *common* among most of the subjects. A major limitation of GiTA is that this framework is computationally very expensive, as it requires comparing all tracts across all subjects. In addition, the efficient and powerful data analysis methods based on vector-space data representation, such as the k-d tree KNN search, are not applicable due to the unequal lengths of the fiber tracts. It is also invalid to resample the fiber curves to the same number of points for computing the Euclidean distance since this generally doesn't reflect the intrinsic geometrical distance between the fiber curves. Besides, the GiTA framework also suffers from inter-subject misalignment due to registration error.

To address the aforementioned issues, we propose a novel group-wise isometric fiber embedding (GIFE) framework. The GIFE framework tries to find the finite dimensional embedding of the fiber tracts of any target subject, given the precomputed embedding of the fiber tracts for the reference subject. This GIFE framework is naturally parallelizable and it does not require computing the full pairwise distance matrices across all pairs of subjects but only the distances between the fibers of the target subjects and the reference subject. Furthermore, we also handle the inter-subject misalignment in the embedding and derive a novel method called multidimensional extrapolating (MDE), as a tribute to the original multidimensional scaling (MDS) framework, to achieve robust and efficient GIFE. MDE can be viewed as a novel variant of the multidimensional unfolding (MDU) framework [6, 7]. Unlike the conventional MDU, MDE allows the embedding of the reference set to be fixed. Besides, MDE deals specifically with the errors in distance measurements such as registration error.

Previously, fiber embedding based on tract affinity has been applied to fiber bundle segmentation [8] without preserving the distance in the embedding space. The embedding based on MDS has been applied to fiber visualization for individual subjects [9]. However, it is not scalable to GiTA.

2 GIFE: group-wise isometric fiber embedding

We adopt the principle of MDS for GIFE because it preserves pairwise distances. In addition, we address the scalability by using only a small amount of tract distances. Moreover, since the inter-subject distances are often inaccurate in our problem, rather than solving the embedding from the inter-subject distances directly, we transform the embedding found using the intra-subject tract distances to fit to the geometry defined by the inter-subject distances and the resultant embedding is robust to the errors in inter-subject distances.

2.1 The classical MDS

Our idea is based on the classical MDS. The basic formulation of the classical MDS can be written as follows [6]:

$$\mathbf{B}^{n \times n} \approx \mathbf{Z}_p \mathbf{Z}_p^T \quad 1$$

where $\mathbf{B}^{n \times n} = -\frac{1}{2} \mathbf{P}^{n \times n} \mathbf{D}^{(2)n \times n} \mathbf{P}^{n \times n}$, $\mathbf{Z}_p$ is the p -dimensional embedding of the fiber tracts, and it denotes the first p columns of the matrix $\mathbf{Z}$, $\mathbf{P}$ is known as the centering matrix defined as $P_{ij} = 1 - \frac{1}{n}$, $\forall i = j$ and $P_{ij} = -\frac{1}{n}$, $\forall i \neq j$, $\mathbf{D}^{(2)}$ is the input squared distance matrix.

The solution of Eq. (1) can be obtained by using the following steps:

$$(a) \mathbf{B} = \mathbf{P} \left[-\frac{1}{2} \mathbf{D}^{(2)} \right] \mathbf{P}, \quad (b) \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T = \text{svd}(\mathbf{B}), \quad (c) \mathbf{Z}_p = \mathbf{E}_p \mathbf{\Lambda}_p^{\frac{1}{2}} \quad 2$$

where $\mathbf{Z}_p$ and $\mathbf{E}_p$ are the first p columns of $\mathbf{Z}$ and $\mathbf{E}$, $\mathbf{\Lambda}_p$ is the top-left $p \times p$ block matrix of $\mathbf{\Lambda}$.

To simplify the derivations and implementation, we make the following assumption.

Assumption 1 Let $-\frac{1}{2} \mathbf{D}^{(2)} = \mathbf{V} \mathbf{A} \mathbf{V}^T$, where $\mathbf{V}$ and $\mathbf{A}$ are the eigenvector and eigenvalue matrices of $-\frac{1}{2} \mathbf{D}^{(2)}$. We assume $\mathbf{V}_p = \mathbf{P} \mathbf{V}$. In fact, we can always impose $\mathbf{V} = \mathbf{P} \mathbf{V}$ as a constraint in the classical MDS factorization framework. We omit this step in this work since we observe that this complication unnecessary as the assumption is often valid in our problem.

Based on this assumption, we have:

$$\mathbf{B} = \mathbf{P} \left[-\frac{1}{2} \mathbf{D}^{(2)} \right] \mathbf{P} = -\frac{1}{2} \mathbf{D}^{(2)} = \mathbf{B}^\# \quad 3$$

which is the simplified Gram matrix that we adopt in the rest of this work.

2.2 The classical multidimensional extrapolating (cMDE)

In our problem, computing, storing and factorizing the massive pairwise distances between all fiber tracts for all pairs of subjects are very costly in general. Alternatively, we propose to consider one of the fiber bundles as a reference. Then, for all other bundles, we propose to *estimate* their embedding based on the precomputed embedding of the reference bundle. We call this problem the *GIFE problem*, and we develop a novel method called multidimensional extrapolating (MDE) to solve it. Since we follow the idea of classical MDS so we call our method the classical MDE, or cMDE. There are three variants of the MDE method: inter-set MDE, intra-set MDE and cMDE. The inter-set MDE is computed using only the distances of fibers from different subjects, the intra-set MDE is computed

using only the distances from the same subjects, and the cMDE is a combination of these two.

In cMDE, we want to find the embedding $\mathbf{Y}_p$ of the target fibers such that:

$$\begin{bmatrix} \mathbf{X}_p \\ \mathbf{Y}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_p^T & \mathbf{Y}_p^T \end{bmatrix} \approx \begin{bmatrix} \mathbf{B}_{XX}^\# & \mathbf{B}_{XY}^\# \\ \mathbf{B}_{YX}^\# & \mathbf{B}_{YY}^\# \end{bmatrix} = \mathbf{B}^\# = -\mathbf{D}^{(2)} = \begin{bmatrix} -\mathbf{D}_{XX}^{(2)} & -\mathbf{D}_{XY}^{(2)} \\ -\mathbf{D}_{YX}^{(2)} & -\mathbf{D}_{YY}^{(2)} \end{bmatrix} \quad 4$$

where the reference embedding $\mathbf{X}_p$ of the reference fibers and the simplified Gram matrix $\mathbf{B}^\#$ of the reference and target fibers are given.

Inter-set MDE: embedding with inter-set distance.—In the ideal case where the registration process in the GiTA is perfect, and the inter-subject distance measure $\mathbf{D}_{XY}$ or $\mathbf{D}_{YX}$ is ideal, we can simply use the following relation to estimate $\mathbf{Y}_p$.

$$\mathbf{B}_{YX}^\# \approx \mathbf{Y}_p (\mathbf{X}_p)^T \Rightarrow \hat{\mathbf{Y}}_p = \mathbf{B}_{YX}^\# \mathbf{X}_p \Lambda_{XX}^{-1} \quad 5$$

where $\mathbf{X}$ and Λ_{XX} are the eigenvector and eigenvalue matrices of $\mathbf{B}_{XX}^\#$, and this formulation is closely related to the *Nyström* approximation [10]. This method might overfit Y to the inter-set distances despite the errors therein.

Intra-set MDE: embedding with intra-set distance.—Since $\mathbf{D}_{YY}$ is also known, we can obtain their eigen-decomposition.

$$\mathbf{B}_{YY}^\# \approx \mathbf{E}_p^Y \Lambda_{YY} (\mathbf{E}_p^Y)^T = \bar{\mathbf{Y}}_p \bar{\mathbf{Y}}_p^T \quad 6$$

The solution found by the above decomposition maybe more reliable than the one found by Eq. (5), since we only use the intra-subject distance measure and no mis-alignment error is present. However, this also does not give us $\mathbf{Y}_p$ directly since

$$\bar{\mathbf{Y}}_p \bar{\mathbf{Y}}_p^T = \bar{\mathbf{Y}}_p \mathbf{R} \mathbf{R}^T \bar{\mathbf{Y}}_p^T \quad 7$$

where $\mathbf{R} \mathbf{R}^T = \mathbf{I}^{p \times p}$, meaning that the solution remains valid up to any arbitrary orthogonal transformation.

cMDE: Finding the optimal orthogonal transformation.—We propose to estimate the optimal $\mathbf{R}$ which transforms the solution of the intra-set MDE $\bar{\mathbf{Y}}$ toward the solution of the inter-set MDE $\bar{\mathbf{Y}}$, such that the geometry of the embedding defined by $\mathbf{B}_{YY}^\#$ is preserved when we try to align the embedding of Y with the inter-set distances. And we propose the following model to solve it:

$$\tilde{\mathbf{Y}} = \bar{\mathbf{Y}}_p R^*, \quad R^* = \operatorname{argmin}_R \frac{1}{2} \|\mathbf{R} - \bar{\mathbf{Y}}_p^T \hat{\mathbf{Y}}_p\|^2, \quad \text{s.t. } \mathbf{R}\mathbf{R}^T = \mathbf{I} \quad 8$$

The rationale of this model lies in that $\bar{\mathbf{Y}}_p^T \hat{\mathbf{Y}}_p$ is actually the least-squares solution of

$$\min_R \|\bar{\mathbf{Y}}_p \mathbf{R} - \hat{\mathbf{Y}}_p\|^2 \text{ without the orthogonality constraint } \mathbf{R}\mathbf{R}^T = \mathbf{I}.$$

The latter formulation is more intuitive. An advantage of Eq. (8) over the latter intuitive formulation is that it admits a closed-form solution [11]

$$\mathbf{R}^* = \mathbf{U}\mathbf{V}^T \quad 9$$

where $\mathbf{U}\mathbf{D}\mathbf{V}^T = \operatorname{svd}(\bar{\mathbf{Y}}_p^T \hat{\mathbf{Y}}_p)$.

3 Experimental results

3.1 Randomly sampled tractograms

In this experiment, we simulate the situation of GIFE problem by subsampling a relatively small real optic radiation bundle reconstructed using the Human Connectome Project (HCP) data. This simulation allows us to quantitatively assess the distance preservability of different methods.

Experiment configuration.—First, we reconstruct the optic radiation fiber bundle for one subject from the HCP data [12, 13] using the method described in [14]. This bundle contains a total of 8170 fibers. Note that we used the unfiltered fiber tracts in this experiment, and some spurious tracts are present. We also compute the Hausdorff distance between all pairs of tracts in the bundle. Then, we randomly sample the bundle 10 times without replacement and each subsample contained 10 percent of the original bundle, and the sub-bundles are denoted as $\{T_0, T_2, \dots, T_9\}$. Finally, we consider T_0 as the reference bundle and extract its intra-set distance matrix from the full distance matrix, and we extract the intra-set and inter-set distance matrices with reference to T_0 for all other sub-bundles. We mainly compare our method with four different comparable methods for MDU: the weighted least-squares MDS (LS-MDS)¹ [6], the Scaling by MAjorizing a COMplicated Function (SMACOF) algorithm² [15] and the Maximum Variance Unfolding (MVU)³ [7]. Only the inter-subject distances with reference to T_0 and the intra-subject distances are used in this comparison. We also compare our method with the inter-set MDE and the intra-set MDE. Lastly, we compute the cMDS computed with the full distance matrix for reference.

We originally computed cMDS with the full distance matrix using 11 dimensions. However, only 7 of them correspond to positive eigenvalues. According to the MDS theory [6], we should only use positive eigenvalues for cMDS so we fix the dimensionality to be 7 in all

---

¹gradient descent implementation

methods. In this experiment, we perturbed the inter-set distance

D_{T_0, T_i} by $D'_{T_0, T_i} = D_{T_0, T_i} \times (1 + n)$ and $n \sim N(0, 0.5)$. Computing the full distance matrix for

this bundle took about 11 hrs on a single CPU core, and computing the subset of distances required by MDE would take only about 4 hrs.

Results.

Some visual results for this simulation experiment are shown in Fig. 1. We observe that our method (cMDE) well approximates the point distribution of cMDS, while the other methods generally fail to preserve the geometrical relationship between the fiber tracts given the partial and imprecise information. From Fig. 2, we can see that the inter-set MDE might be affected by the inter-set misalignment and the intra-set MDE may be oriented arbitrarily while our method optimally restores the point distribution. We are also able to evaluate our method quantitatively by comparing the pairwise distance in the embedding space with the true Hausdorff tract distance. We adopt the signed Pearson correlation coefficient as our distance similarity measure. Note that this measure is invariant to linear scaling which is acceptable in our problem. The results are summarized in Tab. 1. We observe that our method gives very satisfactory distance preservation in the embedding space. In addition, our method compares significantly favorably to other methods in terms of computational efficiency. The computational cost for calculating the distances are not included in this table. This experiment is conducted in MATLAB on Linux with Intel(R) Core(TM) i7-6820HQ CPU @ 2.70GHz and 32GB memory. All iterative methods terminate at convergence or a maximum of 100 iterations. The computational times shown are all the total times. Since the MDE framework is parallelizable, we put (/10) behind the times to indicate the possibility of further breaking down the computational time by parallelization.

3.2 Common optic radiation fiber bundle extraction

Following the GiTA framework, we apply our GIFE framework to extracting common optic radiation fiber bundles [14] using the HCP data for 47 subjects.

We use the raw noisy fiber tracts as the input in this experiment. For this task, we pick the the fibers common in most of the bundles as the common fiber based on a commonness measure defined based on the Euclidean distance in embedding space or Hausdorff distance in the fiber space. We adopt the k-d tree based KNN search to calculate the Euclidean distances. We use 25 dimensions for the embedding. For results, we expect the common bundle to capture the main anatomical characteristics of the optic radiation bundle and we also expect it to be highly organized to follow retinotopy [4]. The total computation time for calculating all pairwise tract distances was around 60056 hrs-core. We also subsample the fibers with a fixed sampling ratio 1/10, which reduces the computation to about 6000 hrs-core. Note that the tract distance calculation involves k-nearest neighbor search which is approximately linear time complexity with k-d tree. Since we implement the original GiTA on a large-scale computing array with thousands of CPUs, the computation time is reduced to a couple of days. By employing the GIFE framework, we reduce the computation time by over 90% to about 500 hrs-core, which is tractable for a small-size cluster with dozens of CPU cores.

We compare our cMDE method with the inter-set MDE and intra-set MDE as well as the original GiTA framework. The results are shown in Fig. 3. The results show that the original GiTA extracts the largest common bundles. However, the tracts appear to be a bit disorganized and this might be due to the inter-subject misalignment. We also observe that the GiTA + intra-set MDE framework failed to extract meaningful common bundle. Both of GiTA + inter-set MDE and GiTA + cMDE extract consistently highly organized bundles with increased organization by raising the commonness, while the bundles extracted by GiTA + cMDE are more anatomically complex and agreeable to the results of the original GiTA. This is an anticipated outcome of the better overlapped reference-target embedding of cMDE over the inter-set MDE. The computational time for the intra-set MDE for 1 subject is about 200 sec-core, and the inter-set MDE took about 0.05 sec-core, solving the optimal R^* and computing the final cMDE mapping took about 0.008 sec-core. We also compare with the MDU solved by LS-MDS and SMACOF in which we fix the reference embedding while iteratively updating only the target embedding for each target subject. However, LS-MDS and SMACOF generally require recalculating all the pairwise distances for the reference and target embeddings at each iteration, and each iteration of them took about 100 secs-core and both methods ran about 50 iterations before convergence. The results of LS-MDS are disorganized and sparse, and SMACOF gives a large amount of disorganized, hence invalid, common optic radiation fiber tracts.

4 Conclusion

In this work, we present a novel GIFE framework to achieve scalable GiTA. We also propose a novel method called MDE to achieve efficient and robust GIFE. The resultant method is highly scalable, parallelizable and robust to inter-subject misalignment. Real experiment shows clearly improved anatomical validity of the results of our proposed MDE method over other methods. This GIFE framework will be generally useful non-exclusively for common bundle reconstruction among all possible GiTA problems.

Acknowledgments

This work was in part supported by the National Institute of Health (NIH) under Grant R01EB022744, R01AG056573, U01EY025864, P41EB015922

References

1. Maier-Hein, et al.: The challenge of mapping the human connectome based on diffusion tractography. *Nature communications* 8(1) (2017) 1349
2. Smith RE, Tournier JD, Calamante F, Connelly A: Sift: Spherical-deconvolution informed filtering of tractograms. *Neuroimage* 67 (2013) 298–312 [PubMed: 23238430]
3. Aydogan DB, Shi Y: Track Filtering via Iterative Correction of TDI Topology. In: MICCAI. (2015)
4. Wang J, Aydogan DB, Varma R, Toga AW, Shi Y: Topographic regularity for tract filtering in brain connectivity. In: International conference on Information Processing in Medical Imaging (IPMI). (2017)
5. Wang J, Shi Y: Gita: Group-wise tractogram analysis. In: OHBM. (2018)
6. Borg I, Groenen PJ: Modern multidimensional scaling: Theory and applications. Springer Science & Business Media (2005)
7. Weinberger KQ, Saul LK: Unsupervised learning of image manifolds by semidefinite programming. *International journal of computer vision* 70(1) (2006) 77–90

8. O'Donnell LJ, Westin CF: Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE transactions on medical imaging* 26(11) (2007) 1562–1575 [PubMed: 18041271]
9. Jianu R, Demiralp C, Laidlaw D: Exploring 3d dti fiber tracts with linked 2d representations. *IEEE transactions on visualization and computer graphics* 15(6) (2009) 1449–1456 [PubMed: 19834220]
10. Drineas P, Mahoney MW: On the nystrom method for approximating a gram matrix for improved kernel-based learning. *Journal of Machine Learning Research* 6(12) (2005) 2153–2175
11. Lai R, Osher S: A splitting method for orthogonality constrained problems. *Journal of Scientific Computing* 58(2) (2014) 431–449
12. Toga AW, Clark KA, Thompson PM, Shattuck DW, Van Horn JD: Mapping the human connectome. *Neurosurgery* 71(1) (2012) 1–5 [PubMed: 22705717]
13. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium WMH, et al.: The WU-Minn human connectome project: an overview. *Neuroimage* 80 (2013) 62–79 [PubMed: 23684880]
14. Kammen A, Law M, Tjan BS, Toga AW, Shi Y: Automated retinofugal visual pathway reconstruction with multi-shell HARDI and FOD-based analysis. *Neuroimage* 125 (1 2016) 767–779 [PubMed: 26551261]
15. De Leeuw J: Applications of convex analysis to multidimensional scaling. *Recent developments in statistics* (1977)

Fig. 1.

3D visualization of the results of isometric embedding for a randomly sampled visual pathway fiber bundle using different methods. (a) shows the original bundle.

Fig. 2.
MDE with intermediate results. (a) a target bundle (b) inter-set MDE (c) intra-set MDE (d) cMDE (e) cMDS with full distance matrix.

Fig. 3.

Common optic radiation bundles. The common bundles from the original GiTA, GiTA +inter-set MDE and GiTA+cMDE generally do not contain spurious tracts.

Table 1.

Quantitative results for distance preservation in the embedding.

	cMDS	LS-MDS	SMACOF	MVU	inter-set MDE	intra-set MDE	cMDE
$\rho(D, D_{GT})$	0.97	0.48	0.36	0.05	0.72	0.38	0.96
time(s)	6.38	1562	1445	2723	0.16(/10)	0.6(/10)	0.73(/10)