NeuroImage xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Automatic clustering and population analysis of white matter tracts using maximum density paths

Gautam Prasad [a,b], Shantanu H. Joshi [c], Neda Jahanshad [a,b], Julio Villalon-Reina [a,b], Iman Aganj [g], Christophe Lenglet [h], Guillermo Sapiro [i,j], Katie L. McMahon [k], Greig I. de Zubicaray [l], Nicholas G. Martin [m], Margaret J. Wright [l,m], Arthur W. Toga [a,b,d,e], Paul M. Thompson [a,b,c,d,e,f]

[a] Imaging Genetics Center, Institute for Neuroimaging & Informatics, University of Southern California, Los Angeles, CA, USA
[b] Laboratory of Neuro Imaging, Institute for Neuroimaging & Informatics, University of Southern California, Los Angeles, CA, USA
[c] Department of Neurology, University of California Los Angeles, CA, USA
[d] Dept. of Neurology, Psychiatry, Engineering, Radiology, University of Southern California, Los Angeles, CA, USA
[e] Dept. of Ophthalmology, University of Southern California, Los Angeles, CA, USA
[f] Department of Pediatrics, University of Southern California, Los Angeles, CA, USA
[g] Martinos Center for Biomedical Imaging, Radiology Department, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[h] Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN, USA
[i] Dept. of Electrical and Computer Engineering, Computer Science, Duke University, NC, USA
[j] Dept. of Biomedical Engineering, Duke University, NC, USA
[k] Center for Advanced Imaging, University of Queensland, Brisbane, Australia
[l] School of Psychology, University of Queensland, Brisbane, Australia
[m] QIMR Berghofer Medical Research Institute, Herston, Australia

ARTICLE INFO

Article history:
Accepted 8 April 2014
Available online xxxx

Keywords:
HARDI
Tractography
MRI
Brain
Clustering
Atlas
Dijkstra
Shortest path
Geodesic distance
Hough
Connectivity
Maximum density path
Curve registration
Longest path

ABSTRACT

We introduce a framework for population analysis of white matter tracts based on diffusion-weighted images of the brain. The framework enables extraction of fibers from high angular resolution diffusion images (HARDI); clustering of the fibers based partly on prior knowledge from an atlas; representation of the fiber bundles compactly using a path following points of highest density (maximum density path; MDP); and registration of these paths together using geodesic curve matching to find local correspondences across a population. We demonstrate our method on 4-Tesla HARDI scans from 565 young adults to compute localized statistics across 50 white matter tracts based on fractional anisotropy (FA). Experimental results show increased sensitivity in the determination of genetic influences on principal fiber tracts compared to the tract-based spatial statistics (TBSS) method. Our results show that the MDP representation reveals important parts of the white matter structure and considerably reduces the dimensionality over comparable fiber matching approaches.

Introduction

Diffusion weighted imaging (DWI) measures the directional diffusion of water through the brain in vivo. By following the dominant directions of diffusion across the brain, whole-brain tractography algorithms can reconstruct the brain's major white matter pathways, extracting a vast number of fibers that are amenable to statistical analysis. We can then study these white matter regions in individuals and populations to better understand disease effects (Daianu et al., 2013; Jahanshad et al., 2012b; Takahashi et al., 2002), changes in brain microstructure and connectivity with age (Abe et al., 2002; Dennis et al., 2012), hemispheric differences (Jahanshad et al., 2010), sex differences (Peled et al., 1998), and genetic influences (Jahanshad et al., 2013a; Kochunov et al., 2010).

High angular resolution diffusion imaging (HARDI) enables a more accurate representation of fiber directions compared to the more standard single-tensor model (Basser and Pierpaoli, 1996). The single-tensor model does not account for fiber crossing or mixing, but the orientation distribution function (ODF) (Tuch, 2004) can be derived from HARDI images to discriminate multiple fibers with different orientations passing through a voxel (Leow et al., 2009; Zhan et al., 2010).

http://dx.doi.org/10.1016/j.neuroimage.2014.04.033
1053-8119/

The large number of fibers generated by the tractography algorithms first needs to be clustered according to known anatomical pathways before comparing them across subjects. A wealth of clustering methods has been applied to tractography results including fuzzy clustering (Shimony et al., 2002), normalized cuts (Brun et al., 2004), *k*-means (O'Donnell and Westin, 2005), spectral clustering (O'Donnell et al., 2006), Dirichlet distributions (Maddah et al., 2008), hierarchical clustering (Visser et al., 2011), a Gaussian process framework (Wassermann et al., 2010b), and median filtering (Prasad et al., 2011a). Some of these methods readily benefit from prior anatomical information provided by an atlas of likely locations of the tracts in the brain (Yendiki et al., 2011), suggesting when to split or combine clusters to conform to known anatomy. In one approach (Jin et al., 2011a,b, 2013), several labeled atlases are deformed onto a fiber set extracted from a new subject, and a fiber matching and voting process are used to help decide the anatomical bundles to which the fibers belong.

Following clustering, several methods can be used for fiber bundle matching. (Colby et al., 2011) use a parametric curve-based method to resample fibers in a bundle based on shared seed points and then compute correspondences from the resampling to create a representative path for an individual or group. A similar re-sampling approach is used in a method (Yeatman et al., 2012) that filters fiber bundles to match a probabilistic atlas. (Corouge et al., 2006) analyze fiber bundles by resampling and then aligning them across subjects using Procrustes analysis (Goodall, 1991) to generate a mean shape. (Roberts et al., 2005) apply a density measure derived from tractography results. Their measure (fiber density index; FDi) quantifies the average number of detected fiber paths passing through voxels in a ROI. (Wassermann et al., 2010a) use Gaussian processes to create voxel-wise probability maps of white matter structure. The fiber locations in high density regions of the image space are used by O'Donnell et al. (2009) as a template to align other fibers and compute correspondences. Yushkevich et al. (2008) analyze white matter tracts using deformable geometric medial models that allow for integration of nearby tensor-based features to reduce the dimensionality and improve registration. (Patel et al., 2010) use a fast-marching algorithm to encapsulate white matter tracts in voxel based boundaries, which are then matched using variational techniques.

In contrast to the parameterized methods mentioned above, white matter analysis can also be performed using a voxel-based approach. A popular method known as tract based spatial statistics (TBSS) (Smith et al., 2006), uses a skeletonized representation of white matter and uses nonlinear registration for matching the skeletons. Although it is a very popular approach, TBSS does not explicitly represent tracts that would be recognized by anatomists, and therefore is not guaranteed to produce a consistent labeling of tracts from one brain to another (Schwarz et al., 2013). Although voxel-based methods can also be used to analyze DWI, they are often sensitive to the image registration (Tustison et al., 2012). Most existing white matter analysis techniques focus on nonlinear registration of fractional anisotropy (FA) images as in TBSS (Smith et al., 2006) and voxel-based morphometry (VBM), which can be applied to DWI-derived maps such as FA (Jones et al., 2005). Other approaches that focus on diffusion tensor correspondences are usually based on a global image registration (Wang et al., 2011; Yeo et al., 2009), but a high-dimensional registration of tensor fields may also be used, as can tensor-based statistics (Chiang et al., 2008; Lee et al., 2009; Lepore et al., 2008). Given the richness of information provided by tractography, it seems advantageous to directly study the fiber tract bundles rather than simply analyzing voxel-based representation.

Approach

Our work adopts a parameterized approach by refining the representation of white matter structure into compact and localized paths, represented as 3D curves. These paths represent the most influential regions in tractography and are used as compact dimensional representations of the fiber bundle. Our method uses an additional local registration of specific white matter regions to fix biases (Tustison et al., 2012) in voxel-based analysis and many of the problems of registration algorithms (Klein et al., 2009) that work on the entire image. Additionally, our approach may offer increased statistical power as it finds shape homologies across different white matter tracts.

Termed the maximum density path (MDP) approach, it incorporates information from tractography-derived fibers by selecting a subset of fiber bundles from a white matter atlas in the same space. We generate a density image from the fiber bundles and use it to create a graph with voxel locations as nodes and fiber density measures as edges. We implement a widely used graph search algorithm to find the MDP between two pre-specified regions of interest (ROI) in the atlas. The MDPs represent fiber bundles that characterize a tract using points of highest density. These compact descriptions of a tract's scale, location, and high-level geometric information are computed for all subjects in a population. We find correspondences across the paths by bringing them into the same space using geodesic curve registration. Finally, the average MDP for a given population is computed using a nonlinear iterative method. As an example, we use our method to determine genetic influences on white matter tracts based on a large cohort of over 565 twin subjects scanned using HARDI at 4-Tesla. We compare the results to those obtained by the more standard TBSS method.

MDPs have been used as one tool for pilot studies of sex differences and a variety of diseases (Nir et al., 2012; Prasad et al., 2011b). In the current study we explicate the technical details of the method, validate its repeatability, compare it to the widely used TBSS, and use MDPs to study heritability along with genetic associations. The main contributions of this work are as follows:

- Fiber tract bundles are represented by compact reduced dimensional representations known as maximum density paths (MDPs).
- MDPs are represented by vector valued functions and are analyzed in an intrinsic and invariant manner.
- Shape matching between MDPs is achieved using geodesic curve registration that not only yields smooth deformations between MDPs, but also provides shape distances between them.
- Group analysis of MDPs is conveniently performed using an intrinsic statistical framework that enables the computation of shape averages and their first order variations.
- Fiber bundle analysis via MDPs is used to identify highly heritable regions in the white matter tissues in twin subjects and is also used to show genetic associations.

Materials and methods

This section describes important steps starting with the extraction of fibers using HARDI tractography, clustering of fibers using a white matter ROI atlas, representation and matching of fiber bundles using MDPs, and finally, statistical analysis of MDPs in a population. The schematic pipeline outlining the extraction and representation of MDPs is shown in Fig. 1, whereas the workflow for statistical group analysis is shown in Fig. 2.

HARDI tractography using the Hough transform

We use a global tractography algorithm (Aganj et al., 2011) to extract fibers from HARDI images.

The algorithm uses extensive information provided by HARDI at each voxel, parametrized by the orientation distribution function (ODF).

Our tractography method selects fibers in the diffusion image space by generating scores for all possible curves at a seed point. These curves are parameterized using 2nd order polynomials. An additional parameter controls the maximum expected curve length and is set to a value representing the largest dimension of the volume. In practice, the

Please cite this article as: Prasad, G., et al., Automatic clustering and population analysis of white matter tracts using maximum density paths, NeuroImage (2014), http://dx.doi.org/10.1016/j.neuroimage.2014.04.033