Improving alignment in Tract-based spatial statistics: Evaluation and optimization of image registration

Marius de Groot [a,b,*], Meike W. Vernooij [a,c], Stefan Klein [a,b], M. Arfan Ikram [a,c], Frans M. Vos [d,e], Stephen M. Smith [f], Wiro J. Niessen [a,b,d], Jesper L.R. Andersson [f]

[a] Department of Radiology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[b] Department of Medical Informatics, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[c] Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[d] Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands
[e] Department of Radiology, Academic Medical Center, Amsterdam, The Netherlands
[f] FMRIB Centre, University of Oxford, UK

ARTICLE INFO

ABSTRACT

Anatomical alignment in neuroimaging studies is of such importance that considerable effort is put into improving the registration used to establish spatial correspondence. Tract-based spatial statistics (TBSS) is a popular method for comparing diffusion characteristics across subjects. TBSS establishes spatial correspondence using a combination of nonlinear registration and a "skeleton projection" that may break topological consistency of the transformed brain images. We therefore investigated feasibility of replacing the two-stage registration-projection procedure in TBSS with a single, regularized, high-dimensional registration.

To optimize registration parameters and to evaluate registration performance in diffusion MRI, we designed an evaluation framework that uses native space probabilistic tractography for 23 white matter tracts, and quantifies tract similarity across subjects in standard space. We optimized parameters for two registration algorithms on two diffusion datasets of different quality. We investigated reproducibility of the evaluation framework, and of the optimized registration algorithms. Next, we compared registration performance of the regularized registration methods and TBSS. Finally, feasibility and effect of incorporating the improved registration in TBSS were evaluated in an example study.

The evaluation framework was highly reproducible for both algorithms (R^2 0.993; 0.931). The optimal registration parameters depended on the quality of the dataset in a graded and predictable manner. At optimal parameters, both algorithms outperformed the registration of TBSS, showing feasibility of adopting such approaches in TBSS. This was further confirmed in the example experiment.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Diffusion imaging of the brain provides insight into architectural properties, and developmental and degenerative processes of the white matter (Basser et al., 1994; Beaulieu, 2002; Lebel et al., 2010). Quantitative features derived from diffusion imaging, such as fractional anisotropy (FA) and mean diffusivity (MD), allow for comparison of diffusion properties across different subjects (Basser and Jones, 2002). This can be achieved in a number of ways, for example region of interest-based or voxel-based.

Voxel-based analyses offer a fast and automated means of analyzing diffusion data (Büchel et al., 2004; Buchsbaum et al., 1998; van Hecke et al., 2010). They do however require the images to be in a common space in which anatomical correspondence across subjects is assured.

Establishing correspondence by bringing images into a common space is a non-trivial task, for which image registration techniques are commonly employed. However, image registration approaches in general do not achieve perfect anatomical correspondence due to anatomical variability. In an attempt to account for the residual misalignment, increase sensitivity and to satisfy the assumptions of parametric tests (if applied), voxel-based analyses often rely on smoothing. The extent of this smoothing ideally needs to be matched to the expected effect size, which can be spatially varying and not known a-priori (Jones et al., 2005). In 2006, an alternative approach for anatomical alignment of diffusion data was proposed. Tract-based spatial statistics (TBSS) (Smith et al., 2006, 2007) was introduced to mitigate the influence of residual misalignment in registration of diffusion data, and to overcome the need to set smoothing extent in voxel-based analyses. In TBSS, following an initial nonlinear registration step (of "medium" dimensionality), voxels that are local maxima for FA are mapped onto a skeleton composed of sheets of maximum FA voxels, and statistical analysis is performed on skeleton voxels.

* Corresponding author at: Erasmus MC, Biomedical Imaging Group Rotterdam, room Ee2102, PO box 2040, 3000 CA Rotterdam, The Netherlands.
E-mail address: marius.degroot@erasmusmc.nl (M. de Groot).

Constraining the analysis to the white matter skeleton results in a dimensionality reduction, ameliorating the issue of multiple testing. Over the past years, TBSS has been widely adopted, aided by its availability within FSL (Smith et al., 2004; Woolrich et al., 2009) and ease of use. The projection stage in TBSS however, is a spatially local operation, with the voxels containing locally maximal FA projected onto the skeleton independently; therefore it does not enforce spatial consistency of the warped images. This may result in an undesirable loss of anatomical topology of tracts in the projection stage. The main aim of this work is to investigate if it is feasible to replace the two registration + projection stages by a single regularized high-dimensional registration approach inside the TBSS method (while still aiming to carry out cross-subject voxelwise testing on the skeleton, to help minimize correspondence errors).

Since even small errors in correspondence may substantially influence results (Smith et al., 2006), considerable effort has been put in improving the registration of diffusion data (Jones et al., 2002; Park et al., 2003; van Hecke et al., 2007; Yap et al., 2009; Yeo et al., 2009; Zhang et al., 2006). In registration, a spatial transformation is determined by optimizing a similarity metric. For evaluating registration performance across algorithms, such as performed for diffusion imaging by Wang et al. (2011), or to optimize different registration parameters, a similarity metric must be employed as well. This is necessary since we do not know the ground truth anatomical correspondence of two images. To objectively measure registration performance however, we cannot use the same similarity metric that was optimized in the registration process, since this would bias the evaluation.

Similarity metrics in diffusion image registration can be based on scalar images such as FA or structural images. Metrics can, alternatively, be based on higher dimensional image features, e.g., on the full diffusion tensor or a number of its components. A third category of similarity metrics is defined on the results of white matter tractography. These three classes of similarity metrics have all been used in the objective functions of image registration approaches for diffusion images (Guimond et al., 2002; Park et al., 2003; Xu et al., 2003; Yeo et al., 2009; Zhang et al., 2006; Zvitia et al., 2010). Analogously, similarity metrics in all three categories have been employed in order to evaluate registration performance (Park et al., 2003; Wang et al., 2011; Yap et al., 2009; Yeo et al., 2009; Zhang et al., 2007; Zöllei et al., 2010).

An important advantage of a performance measure based on similarity of tractography results is that it is independent of any particular similarity metric, defined on a scalar or higher order image, which is employed in most registration approaches. Also, optimal white matter tract alignment is most closely linked to the eventual registration aim of obtaining anatomical correspondence in white matter (Lawes et al., 2008). We therefore developed a framework to evaluate scalar or higher-order similarity metric based registrations using tractography. Previous work using white matter tractography for this purpose was based either on whole brain tractography (Park et al., 2003) or only on a small selection of tracts (Jia et al., 2011; Xue et al., 2010; Yap et al., 2009; Zhang et al., 2006; Zöllei et al., 2010). Furthermore, all previous work depended on deterministic tractography, which has more difficulty in coping with complex fiber architecture (e.g., crossing fibers) and signal noise than probabilistic tractography (Behrens et al., 2007).

In this work, we extended the use of tractography for image registration evaluation to a broader range of white matter tracts, and we used a probabilistic model for tractography. Parameters for two nonlinear registration algorithms were optimized using similarity of different subjects' warped tracts as the registration performance measure. The optimization was performed on two datasets acquired at different institutions with different spatial resolution. Registration performance for these optimized approaches was then compared to the registration performance of the TBSS method on a white matter skeleton. We show that the optimized registration reproducibly improved the alignment of white matter structures compared to TBSS.

Methods

The evaluation framework consists of an automated approach to perform probabilistic tractography and a tract-based evaluation metric. A schematic overview of the process is provided in Fig. 1.

Tractography

Tractography was performed with PROBTRACKX (Behrens et al., 2003, 2007), a Bayesian approach to probabilistic tractography available in FSL.

Tractography was initialized by defining standard space "seed", "target", "stop" and "exclusion" ROIs (masks). These masks were based on the protocols described by Mori et al. (2002), Stieltjes et al. (2001), and Wakana et al. (2004, 2007), but had to be adapted to cope with the more dispersing nature of probabilistic tractography. Most importantly, exclusion masks were added, e.g., the mid-sagittal slice was added in all but the commissural tracts. All masks were transferred to subject native space using nonlinear registrations obtained with FNIRT (Andersson et al., 2008) with default settings for FA images as available in FSL.

Tracts that could robustly be identified and which would lead to a reasonably uniform sampling across brain regions were selected. These tracts are listed in Table 1. Two tracts, the posterior thalamic radiation and the inferior fronto-occipital fasciculus, were excluded from the final set because of considerable overlap with other tracts. Exclusion of these tracts prevented uneven weighting of different regions in the registration evaluation. The final set therefore consisted of 23 tracts.

Tractography was performed in subject native space while recording tract density at a 1 mm^3 resolution and using between 2000 and 30,000 samples per seed ROI voxel to account for differences in the number of seed voxels and tract geometry. These parameter settings were selected to aim for robust extraction of the tracts, and were based on the observed number of fiber-particles that were included in the tract together with visual inspection of tractography outputs. Commissural tracts and the middle cerebellar peduncle were tracked a second time (adding both runs) with inverted seed-target ROIs to ensure symmetry of the resulting tract. The acoustic radiations and the superior longitudinal fasciculus were also tracked in both directions to increase robustness. After tracking, the tract density image was normalized by dividing with the total number of particles.

An example of an individual subject's tracking result, thresholded for the purpose of visualization, for all tracts is shown in Fig. 2. Tractography was performed for each subject and for each structure. The resulting maps of white matter structures reside in subject native space, and were used for all evaluations.

Tract-based evaluation metric

The registration performance measurement was based on cross-subject similarity of the warped tract maps. Non-thresholded tract density images in subject native space were warped to common space, and then tract similarity was assessed.

To avoid differences in image characteristics between individual and group mean tract maps influencing the results, tract similarity was evaluated on a subject-to-subject basis. Tract similarity was assessed for each structure individually, and then averaged for all structures in each pair of subjects. In order to provide an even weighting over tracts in this averaging, similarity of left–right homologue structures was jointly given an equal weight as that of the commissural tracts and the middle cerebellar peduncle. If a particular tract could not be identified in one of the subjects with the automated tractography approach (i.e. no particles fulfilled the criteria imposed by the protocol masks), the tract was omitted in the aggregation of the subject–subject similarity score.