Automatic atlas-based segmentation of brain white matter in neonates at risk for neurodevelopmental disorders

Fonseca, Lúcia; van Pul, C.; Lori, N.; van den Boom, R.; Andriessen, P.; Buijs, J.; Vilanova Bartroli, Anna

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Automatic Atlas-Based Segmentation of Brain White Matter in Neonates at Risk for Neurodevelopmental Disorders

L. Fonseca, C. van Pul, N. Lori, R. van den Boom, P. Andriessen, J. Buijs, and A. Vilanova

Abstract Very preterm infants, <32 weeks gestation, are at high risk for brain injury. Cognitive deficits are often diagnosed at a later stage, since there are no available predictive biomarkers in the neonatal period. The maturation of specific white matter (WM) brain structures is considered a promising early-stage biomarker. With Diffusion Tensor Imaging (DTI) tractography, an in vivo and non-invasive evaluation of these anatomical structures is possible.

We developed an automatic tractography segmentation pipeline, which allows for maturation assessment of the different segmented WM structures. Our segmentation pipeline is atlas-based, specifically designed for premature neonates at term equivalent age. In order to better make use of global information from tractography, all processing is done in the fiber domain. Segmented fiber bundles are further automatically quantified with respect to volume and anisotropy. Of the 24 automatically segmented neonatal tractographies, only three contained more than 30% mislabeled fibers. Results show no dependency to WM pathology. By automatically segmenting WM, we reduced the user-dependency and bias characteristic of manual methods. This study assesses the structure of the neonatal brain based on an automatic WM segmentation in the fiber domain method using DTI tractography data.

L. Fonseca • R. van den Boom
Medical Image Analysis, Eindhoven University of Technology, Eindhoven, The Netherlands

C. van Pul
Clinical Physics, Máxima Medical Center, Veldhoven, The Netherlands

School of Medical Physics and Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

N. Lori
LANEN, INECO, INCYT (Favaloro-CONICET), Rosario, Argentina

Centre Algoritmi, University of Minho, Braga, Portugal

P. Andriessen • J. Buijs
Neonatology Maxima Medical Center, Veldhoven, The Netherlands

A. Vilanova (✉)
Computer Graphics and Visualization, Delft University of Technology, Delft, The Netherlands

Medical Image Analysis, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: a.vilanova@tudelft.nl

1   Introduction

Each year, over half a million babies in Europe are born prematurely. Very preterm infants, <32 weeks gestation, are at high risk for cognitive deficits without major motor deficits [33]. Therapy can be effective, especially when started at early postnatal age, which is a stage characterized by high brain plasticity. However, many patients are diagnosed at a later and more rigid neurdevelopmental stage. In order to offer patients the best possible chances of rehabilitation, early postnatal predictive biomarkers are necessary [9, 13, 16, 38].

Deviations in the development of white matter structures are considered promising early-stage biomarkers. Cognitive development is related to the maturation of the brain networks, the WM structures. These structures can be in vivo and non-invasively reconstructed by Diffusion Tensor Imaging (DTI) tractography [2, 22, 23]. Diffusion of water molecules in the brain is restricted by the underlying anatomical structure, becoming anisotropic for voxels containing WM. Even in newborns, despite their lower brain maturation, tractography can be used to reconstruct WM structures. Analysis of the WM structures can be used to show abnormalities in diffusion tensor parameters and in fibers being associated with maturational problems [3, 12, 14, 29, 30, 32].

In order to assess maturation of specific WM anatomical structures, i.e., bundles, these structures need to be first identified among the complete fiber dataset. This is made possible by a segmentation process that classifies the fibers into different bundles. Although segmentation can be done manually [14, 30, 39], it requires extensive knowledge about complex WM fiber anatomy, introduces user bias, and can become too time consuming for practical use. Furthermore, some structures have challenging shapes that are difficult to segment manually.

Automatic segmentation methods have been developed for adult tractography [6, 11, 17, 26, 27, 40], however, they need to be redesigned to account for the under-developed stage of the neonatal brain, since neonatal tractography results in a lower number of fibers, displays smaller and broken fibers that cannot be ignored, misses anatomical structures yet to develop, and has a higher sensitivity to partial volume effects due to their smaller brain size [24, 29, 45].

Improvement of the tractography algorithm can be used to mitigate some of the previous referred problems. Nevertheless, such improvements involves making assumptions that will generate other issues. Some DTI atlas-based methods use voxel level scalar-data [1, 8, 31], which is mainly based on local information. On the other hand, fiber-wise methods employ global information, which can allow for a better overview of the WM anatomical structures, and eventual abnormalities on these structures [6, 15, 18, 20, 25, 26, 42, 46].

In addition, with the eventual goal of modeling and analyzing WM fibers, it is advantageous to process the fibers themselves, instead of DTI scalar-data. Therefore, we developed an automatic tractography segmentation pipeline tailored for neonatal brains, which allows for subsequent maturation assessment. Segmentation is atlas-based, specifically designed for premature neonates at term equivalent age (TEA).

Our automatic tractography segmentation pipeline neonate data is based on 3D distances between fibers like in O'Donnell et al. [26], and it is specifically tailored for dealing with the neonatal tractography characteristics.

2 Material and Methods

2.1 Subjects

Máxima Medical Center (MMC) provided clinical data for this study. In total, 30 patients with an MRI-scan were included. All MRIs were conducted as part of routine clinical practice. According to Dutch Law on Medical Research with Humans (WMO) a waiver for ethical assessment was provided by the local Medical Ethical Committee of MMC, considering that anonymous data collection was performed. The preterm neonates (n = 13, of whom 11 born at a gestational age less than 32 weeks) were all scanned at term equivalent age. Full-term neonates were scanned (n-17) in the second week postnatally. For each patient the relevant clinical information was available.

For preterm infants, the MRI injury was scored according to Woodward et al. [44]. The Woodward scoring system takes into account deviations or abnormalities in WM, subarachnoidal space, basal ganglia, and determines the presence of cysts ventricular dilation, and hemorrhage as described in Kooij [14] and van Pul et al. [30]. The scores were classified in normal (Woodward score for WM of 5–6), mild (Woodward score for WM of 7–9), moderate (Woodward score for WM of 10–12), and severe (Woodward score for WM of 13–15) pathology. For full-term newborns suspected from hypoxic-ischemic encephalopathy the Shankaran method [34] was used, and pathology was defined again to the groups normal (Shankaran score normal), mild (Shankaran score 1A), moderate (Shankaran score 1B) and severe (Shankaran score 2).

2.2 Data Acquisition

The data was acquired on a Philips Achieva 3.0 T MRI-scanner. The DTI-sequence was performed with b-values 0 and 800 s/mm^2 in 32 directions using a single-shot-EPI sequence, with $TE/TR = 48/7745$ ms using a SENSE factor 2, in less than 5 min scan time. The dataset consists of 50 adjacent slices, each slice with 128×128 voxels, each voxel corresponds to a size of $1.44 \times 1.41 \times 2$ mm. Correction of movement artifacts and eddy current distortions in the DWI images was performed with the Philips software from the scanner workstation. The DTI sequence was part of the regular MRI series also including T1 and T2 series.

2.3 Tractography

2.3.1 Masking

Previous to tractography, a skull-stripping method is applied in order to reduce the number of spurious fibers. For the neonatal DTI data, best results are found for a hybrid method which combines a threshold clipping approach with morphological operators.[1] The skull-stripping mask is created by processing of the $b0$ image from DTI. An empirical investigation of the optimal parameters for neonatal datasets is conducted. Optimal parameters do not result in holes in the mask, and contain no non-brain regions, like facial and neck area and regions outside the skull.

First, a histogram of the $b0$ image is computed, for which values above a lower and upper threshold are set to 0. Best results for our dataset are found for upper and lower thresholds of 85% and 99.9% of the intensity value distribution. This mask is then further processed by application of morphological operators. First, satellites are removed. We apply a connected component analysis for 26-connectivity. Components below 100 voxels are considered satellites and deleted. Second, holes are filled by using a smart closing operator. A smart closing operator is defined similar to a normal closing operator, i.e., dilation followed with erosion, having in addition the property of conserving the initial mask outside contour. It does so by comparing the after-closing mask with a complementary image of the before-closing mask [37]. The structural element comprises a 3-D 6-connectivity.

Third, the image is smoothed by a 3-D Gaussian kernel. For the tested dataset, best results were empirically found for a standard deviation, σ, of 10 voxels, which is smaller than typically used in adults because of the smaller brain size of neonatal patients. The smoothed result is again converted into a mask by assigning the value true to voxels higher than 0.5; this can be seen as a second erosion procedure. After smoothing, the mask gets its satellites removed and holes filled once more. The mask is then applied to the multiple DW images.

2.3.2 Tractography

Tractography is performed with a deterministic streamline based full brain seeding algorithm [36], implemented in the software developed at the Biomedical Image Analysis Group from the University of Eindhoven, vIST/e.[2] The tracking parameters used are specific for the patient type of neonates at term age, as suggested by van Pul et al. [29, 30]: minimum fiber length of 20 mm, C_l [43] for minimum anisotropy index of 0.12, maximum fiber angle of 10° and minimum seed distance of 0.5 voxels. The results of the tractography are tractograms, i.e., full brain fiber tracts, which are in the order of thousands of fiber tracts.

---

[1]http://vistalab.stanford.edu/.

[2]http://sourceforge.net/projects/viste/.

2.4 Atlas Creation

The preterm neonatal atlas is built from the tractograms of three preterm neonates without pathology imaged with DTI at TEA from a previous study (Kooij [14], van Pul [30], permission granted). These datasets were acquired using the same type of scanner and protocol as the data of this current study. The atlas represents the common fiber patterns among healthy neonatal DTI tractography results. The tractograms from two of the patients are aligned with the third patient by affine registration of the linear anisotropy, C_l maps, with the software package SPM8.[3]

After registration, all fibers from the three patients are clustered by applying an Affinity Propagation Method [5] to facilitate the manual atlas labeling, and distance metric described in Sect. 2.5.1. The parameters of the clustering are set to maximize cluster correctness. A cluster is correct if it contains fibers that belong just to one bundle, i.e., anatomical structure, however, it does not necessarily need to contain all fibers from that bundle. A bundle can contain a combination of two or more clusters.

All clusters were visually inspected by three experts (two neonatologists and an MR physicist with >10 year experience with DTI) and labeled according to anatomical structure based on the WM atlases by Mori et al. [23] and Wakana et al. [39]. In this way all fibers were divided into the following bundle labels: CC (corpus callosum), CR (corona radiata, left and right), SS (sagittal stratum, left and right), CG (cingulum), FX (fornix), MCP (middle cerebellar peduncle), cheeks, artifacts, and other. An image of the atlas is presented in Fig. 1.

2.5 Data Processing

The processing pipeline of the DTI data to segment has the following steps: Tractography; Clustered-based sampling; Registration; Labeling; and Propagation—as schematically represented in Fig. 2. The tractography set is first sampled using a

Fig. 1 Atlas of preterm neonates imaged at TEA—lateral view. Label of segmented structure is depicted by a *color-key*, at *right side* of the image

Color key:
- Corpus Callosum
- Cingulum
- Fornix
- Corona Radiata (left)
- Corona Radiata (right)
- Inf. Fronto Occipital (left)
- Inf. Fronto Occipital (right)
- Middle Cerebellar Peduncle

---

[3] http://www.fil.ion.ucl.ac.uk/spm/software/spm8/.
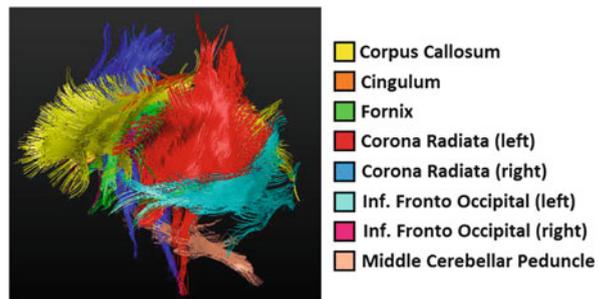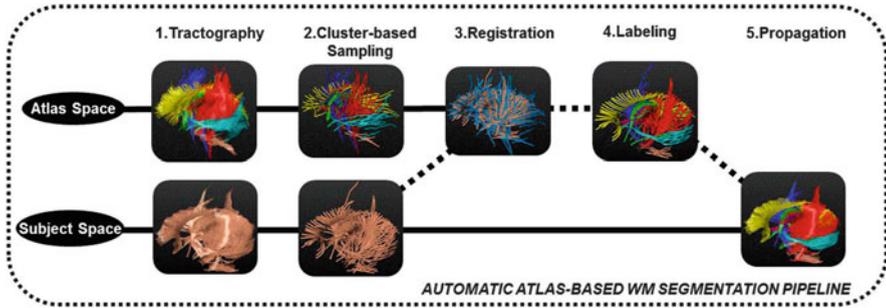
Fig. 2 Schematic representation of the processing pipeline. Columns depict the processing step and the *two lines* depict if the fibers are on subject coordinates space or atlas coordinates space

clustering algorithm, selecting the cluster fiber representatives. This first step is applied to both atlas and subject tractography sets. After, the subject sampled fiber set is registered to the atlas sampled fiber set, passing from the subject space to the atlas space. Still in the atlas space, the subject sampled fiber set is labeled. The final step of propagation is in the subject space, labeling the complete subject fiber set. Each of these steps will now be explained further. The tractography step corresponds to the method described already in Sect. 2.3.2.

2.5.1 Cluster Based Sampling

Sampling of the fibers is applied in order to add robustness and reduce computational costs in subsequent processing steps. Fibers from the same anatomical bundle can have different lengths and the amount of fibers can differ across bundles. Due to this, sampling cannot be performed randomly across structures, since we need all important bundles represented in the downsampled set. Our sampling method starts by computing similarities, i.e., within similarities, for each pair of fibers within the subject. Next, these similarities are used for clustering by Affinity Propagation [5], consequently finding the cluster fiber representatives. The sampled dataset is then downsampled to these cluster fiber representatives. A scheme of this data flow is depicted in Fig. 3.

We developed a distance measure adapted for neonatal data which is based on the Hausdorff distance. Distances in fibers are usually based on point-wise spatial distance between the reconstructed fibers points. We encountered problems with the common distance measures used for adult tractograms, e.g., average mean distance [21], due to the presence of broken reconstructed fibers in neonatal tractography. This is mainly due to the presence of low anisotropy areas. Adaptation of the tracking algorithm can mitigate the issue of broken fibers, but not without introducing assumptions, and, therefore, other issues. Ignoring short fibers, as it is commonly done for adult brains, would discard a large part of the tractogram.
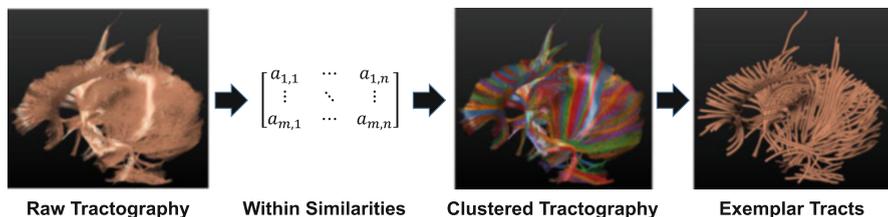
Fig. 3 Sampling data flow. Within similarities are stored in a matrix which the cluster algorithm makes use of
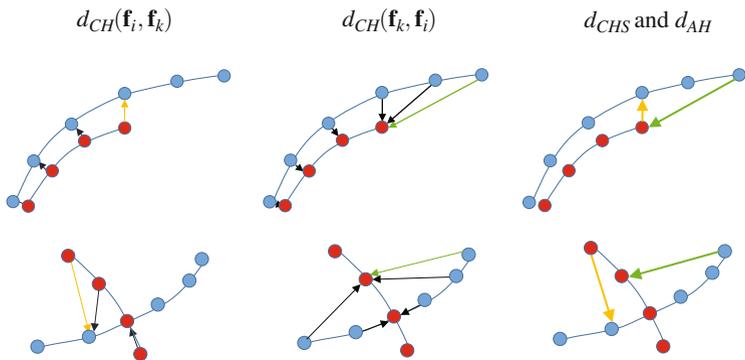
Fig. 4 Illustration of classic Hausdorff distance and our proposed adaptation. *First row* shows parallel fibers of different lengths that belong to the same bundle. *Second row* crossing fibers that belong to different bundles. First two columns shows d_{CH} from *red to blue fiber* and vice versa. Last column shows in *green* the d_{CHS} giving similar distance for parallel and crossing fiber configurations. d_{AH} is shown in *orange* with larger distance for crossing fibers than for parallel

Therefore, it is not desirable to penalize the distance between fibers due to the fiber length. On the other hand, we want to penalize crossing or diverging fibers. Most measures proposed for the adult brain [6, 15, 19, 21], e.g., average mean or Hausdorff do not fulfill our requirements. Some measures [6, 15, 19] avoid finding point to point correspondence which dominates the computational costs by mapping fiber-curve parametrizations. These measures penalize short fibers, diverging or crossing fibers equally and, therefore, these different cases cannot be discerned. Wassermann et al. [42] have a different approach where the fibers are represented implicitly. Wassermann et al. [42] distance measures looks mainly at the overlapping area, so fibers that are diverging and have a small overlap have the same distance than fibers with a full overlap but different lengths.

We developed a simple adaptation of the classic Hausdorff distance (see Fig. 4). The classic Hausdorff distance, $d_{CH}(\mathbf{f}_i, \mathbf{f}_k)$, is the maximum distance of the point-wise minimum distances between the fibers $\mathbf{f}_i$ and $\mathbf{f}_k$. In other words, it is the greatest of all the distances between each point in one fiber to its closest point in the other fiber.

This distance can be defined as:

$$d_{CH}(\mathbf{f}_i, \mathbf{f}_k) = max_{\mathbf{p}_r \in \mathbf{f}_i}(min_{\mathbf{p}_l \in \mathbf{f}_k} ||\mathbf{p}_r - \mathbf{p}_l||) \tag{1}$$

where $\mathbf{p}_r$ and $\mathbf{p}_l$ stand for coordinate point belonging to the fibers $\mathbf{f}_i$ and $\mathbf{f}_k$ respectively notice that $d_{CH}(\mathbf{f}_i, \mathbf{f}_k) \neq d_{CH}(\mathbf{f}_k, \mathbf{f}_i)$. In order to make the distance symmetric the maximum is taken as shown:

$$d_{CHS}(\mathbf{f}_i, \mathbf{f}_k) = max(d_{CH}(\mathbf{f}_i, \mathbf{f}_k), d_{CH}(\mathbf{f}_k, \mathbf{f}_i)) \tag{2}$$

Our adapted distance measure simply takes the minimum instead:

$$d_{AH}(\mathbf{f}_i, \mathbf{f}_k) = min(d_{CH}(\mathbf{f}_i, \mathbf{f}_k), d_{CH}(\mathbf{f}_k, \mathbf{f}_i)) \tag{3}$$

The behaviour of d_{AH} is similar to the Hausdorff distance when fibers have similar length. On the other hand, using this adapted measure, crossing or diverging fibers of different lengths will have a relatively large distance to parallel or overlapping fibers of different lengths (see Fig. 4). This facilitates the inclusion of broken parallel fibers in the corresponding bundle, and at the same time discarding crossing or diverging fibers. Despite not being a metric and also less robust than other measures proposed in literature, we still decide to use d_{AF} since it suits our requirements.

In the next step, for the clustering we use the Affinity Propagation algorithm (AP) [5, 17]. Besides the subject fiber set, also the atlas fiber set is downsampled to improve the registration process. An image of the downsampled atlas is shown in Fig. 5 for the anatomical significant labels.
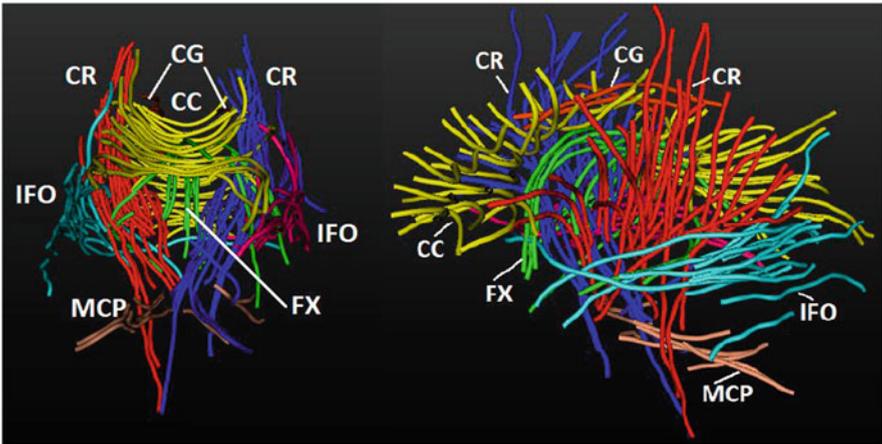
Fig. 5 Downsampled Atlas. Same *color code* as in Fig 1

Fig. 6 Correctness study for cluster preference parameter (x-axis). Y-axis depicts correctness and number of clusters, *left* and *right* respectively
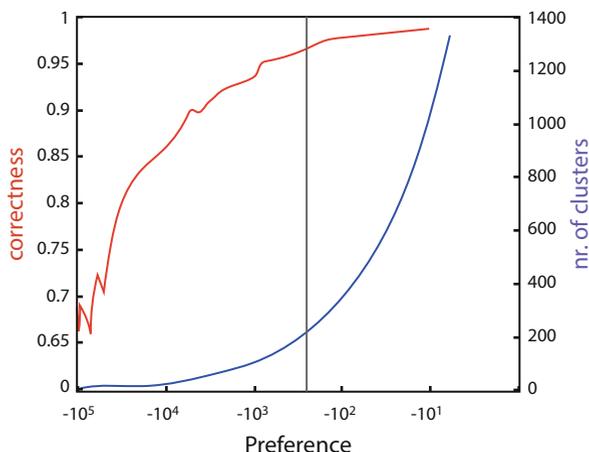
Affinity Propagation was chosen from the studied clustering methods as it allows producing clusters containing reconstructed fibers from not more than one WM anatomic structure [17]. For this clustering technique, the number of clusters is not a predefined value, but depends on previously chosen preference values. Data points with a high preference are more likely to be chosen by the clustering algorithm as a cluster center, than data points with a low preference. Because there is no a priori knowledge about which fibers are more appropriate to become cluster centers and constitute the down-sampled dataset, preference is made equal for all fibers. In order to optimize the preference value, i.e., produce cluster containing fibers from no more than one anatomical structure of interest, a correctness test is conducted (Fig. 6), for the atlas fibers, for which the ground truth of the anatomic labels is known. A balance was found between cluster correctness and the number of clusters, resulting in a choice for the preference value of −200, as it produces slightly more than 200 clusters with a correctness of 95%. After clustering, the cluster fiber representatives form the downsampled tractogram. For comparison of the subject tractogram with the atlas, also a downsampled atlas tractogram was made, shown in Fig. 5 with the anatomical labels.

2.5.2 Registration

Registration involves applying a transformation model to the subject fiber points, in order to correct for head position inside the scanner and variation of head size and shape between subjects and the atlas. Our transformation model is based on an affine transformation [20, 26, 46]. The optimal affine transformation is found in the fiber domain, i.e., fiber-wise, based on 3D fiber distances minimization.

In order to accelerate convergence, a first initialization is achieved by aligning the center of mass of the downsampled subject's tractogram and the center of mass of the atlas. After the initialization, the used registration algorithm is based on the work from O'Donnell et al. [26] with adult tractograms. O'Donnell also developed a Hausdorff inspired distance via a probability density distribution. This probability density function describes how probable it is for that subject fiber to have close-by neighbor atlas fibers. The equation is as follows:

$$\delta(\mathbf{f}_i, \mathbf{f}_k) = \frac{e^{-\left(\frac{d(\mathbf{f}_i, \mathbf{f}_k)}{\sigma}\right)^2}}{Z} \tag{4}$$

where σ is defined as a radius of interest where distances outside this radius will have similarities close to 0. Z is a normalization constant, and therefore will not influence the optimization procedure. Similarities of the subject fiber to all the atlas fibers are then combined in one similarity probability density function [26]. In order to maximize this similarity probability its entropy is minimized [35], using a cost function. Last iteration cost-function value can be further used for quantitative evaluation of segmentation results. The optimization algorithm applied is a direction set method named COBYLA, which stands for Constrained Optimization by Linear Approximation [28]. For our neonatal data, the best results are empirically found for the value of σ equal to 3 mm. This is smaller than in adults due to the smaller size of the neonatal brain.

In our application of the O'Donnell algorithm, best registration results are obtained while optimizing iteratively between four distinct deformation-types: translation, rotation, scaling and shearing. For each of these deformation types, a transformation is probed in a domain of three degrees of freedom (DoF), i.e., one for each spatial coordinate. Due to the difficulty of registering the cluster center fibers corresponding to the CG (cingulum) anatomy, translation and rotation are performed once more.

2.5.3 Labeling

After registration, computation of similarity between every registered cluster fiber representative and every cluster atlas fiber representative is calculated. The adapted Hausdorff distance presented in Sect. 2.5.1 is used again for the similarity metric. Each cluster center registered fiber receives the label of the atlas cluster center fiber towards which it has the smallest distance.

2.5.4 Label Propagation

The label of the cluster fiber representative is propagated to all the fibers that belong to the cluster it is representing. It is therefore possible to return to the

full tractography dataset, without any deformation performed to the fibers. Volume and averaged values of anisotropies can be calculated per segmented structure automatically using vIST/e.[4] Averaged anisotropy values correspond to a weighted average of the anisotropy of all voxels in that segmented structure: the anisotropy value of a voxel is included as many times as the number of fibers passing through it.

3 Evaluation

Of 30 available neonatal datasets, six contained large MRI artifacts which highly disrupted the DTI anatomical structure information and failed in the first step of the pipeline. For the remaining 24 datasets, the complete pipeline could be executed. The processed data was from 13 term borns and 11 preterms at term equivalent age, of whom eight patients were classified as having no abnormalities, seven had mild abnormalities, eight moderate abnormalities and one with severe pathology. On average, the processing took 2–8 h to produce an automatically labeled data. This time varies with the number of fibers in the subject tractograms which is in the order of thousands. The main computational costs are related to the calculation of the within similarities measurement for the cluster-based sampling which has not been optimized.

A qualitative evaluation of the segmented tractography was performed by two MRI experts. For each segmented structure, the number of incorrectly labeled fibers is evaluated, based on visual inspection. The ratio between incorrectly labeled fibers and the total number of fibers per segmented structure is further referred to as the percentage error. For minor structures, FX, CG, MCP, IFO, segmentation errors are considered less severe than segmentation errors at major structures in the neonatal brain: the CRs and CC. The segmentation performance is divided in four classes:

4 *Good*: All segmented structures have less than 10% of error;
3 *Sufficient*: Major structures until 10% of error, but minor structures until 50% of error;
2 *Moderate*: Major structures with error between 10% and 30%, minor structures until 50% of error;
1 *Bad*: All the structures, major and minor, with more than 30% of error.

Of the 24 analyzed datasets, the segmentation results were scored in these four performance classes: three were classified as *Bad* segmentation, four as *Moderate* segmentation, seven as *Sufficient* segmentation, and ten as *Good* segmentation. An example of a segmentation for each segmentation class, is shown in Fig. 7. The main structures in neonatal datasets (corpus callosum and corona radiate) are easily recognized in the patient with classification *Good*. For the *Moderate* and *Bad* examples, the automatic fiber clustering and segmentation contains visible errors.

---

[4] http://sourceforge.net/projects/viste/.
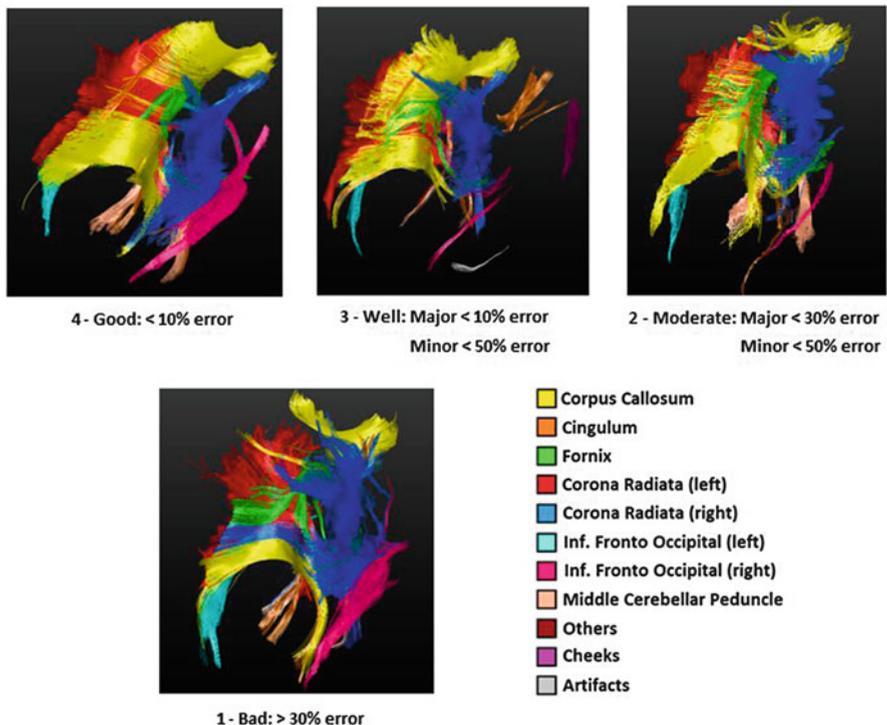
Fig. 7 Tractography segmentation examples for each quality class, posterior view. From *left to right* and *top to bottom*: *Good* segmentation performance example; *Sufficient* segmentation performance example, *Moderate* segmentation performance example; *Bad* segmentation performance example; the color of the structure and its anatomical label are displayed in the legend

Table 1 Distribution of segmentation performance classes and pathology scores

| | Segmentation | | | | |
Pathology	1-Bad	2-Moderate	3-Sufficient	4-Good	
No	1	2	3	2	8
Mild	1	2	1	3	7
Moderate	1	0	3	4	8
Severe	0	0	0	1	1
Total	3	4	7	10	24

As shown in Table 1, there is no correlation between segmentation performance and presence of pathology. We can observe that patients in whom segmentation perform badly (score 1), the pathology was not severe. Patients with no abnormalities and mild abnormalities are distributed among all four classes of segmentation performance. Patients with no abnormalities are in greater number in the class *Sufficient*, and patients with mild abnormalities in the class *Good*. Patients with
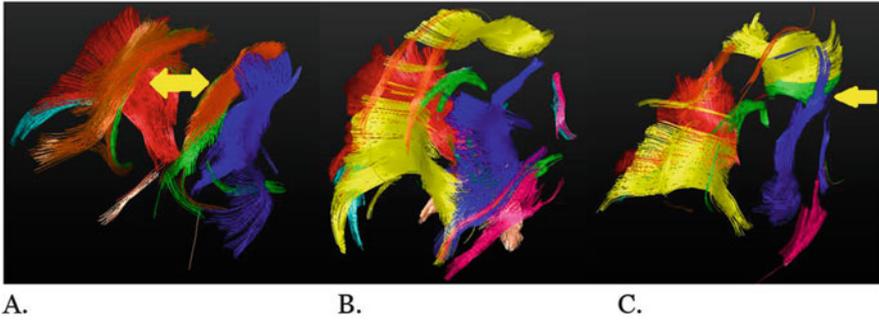
Fig. 8 Segmentation results for patients with severe WM abnormalities, posterior views. From *left to right*: (**a**) patient with CC agenesis with no further WM pathology, classified with *Good* segmentation performance; (**b**) patient with large WM abnormalities (signal intensity) and widened ventricles classified with *Good* segmentation performance; (**c**) patient with severe WM signal abnormalities classified with *Bad* segmentation performance, anterior part of the CC is segmented as FX

moderate abnormalities are also in greater number in the class *Good*. The patient with severe abnormality also had segmentation performance classified as *Good*.

Figure 8 presents the segmentation results for three patients with severe WM pathology. The first patient had no corpus callosum (CC), this is called CC-agenesis. The segmentation pipeline still processed well the tractogram and no CC was segmented for this patient but all the other structures could be observed. An arrow was inserted in the typical CC region in the image Fig. 8, for ease of interpretation. This patient had large areas of diffuse white matter signal intensity abnormalities but could still be segmented with our pipeline and was classified as having a *Good* segmentation performance. The third patient, also with white matter signal abnormalities, shows poor segmentation performance and the segmentation performance was scored as *Bad*, since the CC (a main structure) is clearly showing mislabeled fibers. For this last patient the frontal part of the CC, in yellow, is also erroneously labeled as FX, in green (also pointed out by an arrow).

Looking in more detail to the registration process, it shows that the cost-function value (CF value) obtained at the end of the registration process has a lower average value in the group with *Good* performance. A plot of the distribution of last cost-function value per segmentation performance class is presented in Fig. 9. The last cost value gives an indication of the quality of the registration, the lower the value the better the registration was matched.

Patients with a last CF-value bigger than 35 were individually assessed; this corresponded to three patients. These three patients had no severe white matter pathology: one had even no abnormalities, the other two had mild abnormalities due to residuals from a hemorrhage. Therefore, it can be observed that the deviation from a normal brain was not the reason for the lower quality results.
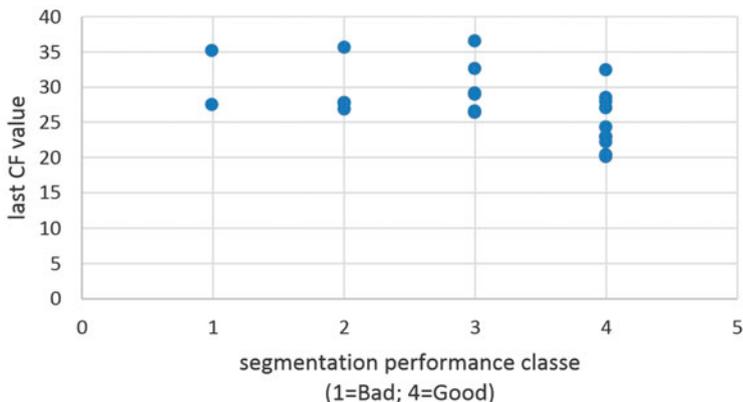
Fig. 9 Distribution of last CF value per segmentation performance class

For the complete analyzed dataset, only 30% of the patient tractography results contained spurious fibers, i.e., fibers that did not belong to any anatomical structure and only arise due to scanning or processing errors. Without skull-stripping, all tractography results contained spurious fibers. Existence of spurious fibers is almost evenly present for all the four performance classes, so segmentation performance is not dependent on this step.

4   Discussion

We have implemented an automatic pipeline that segments Diffusion Tensor Images of neonates automatically into images displaying the main WM structures in the neonatal brain. The method is automatic, without user interaction, and tailored for the evaluation of the neonatal brain. Main pipeline embedded methods and their relation to the segmentation results are here discussed: (1) the tensor fitting and tractography algorithm, (2) the tractography neonatal TEA atlas, (3) the skull-stripping method, (4) the cluster-based sampling method, and (5) the fiber-wise registration and correspondent fiber similarity metrics.

1. The tractography algorithm used is a simple FACT based method [36]. Subjects with poor segmentation performance often show partial volume effects in the clustering results (e.g., crossing over of fibers from one structure to another). Partial volume effects have not been addressed by our pipeline. Crossing over of fibers from different structures occurred usually between the CC and FX, or between CC and CG. For some patients a partial volume effect is also observed between CC and the CR. Improvement of tensor fitting accuracy is constrained by practical reasons in neonatal clinical practice. The DTI protocol on many clinical scanners is still limited to a maximum of 32 diffusion directions. If additional

packages are purchased scanning in more directions become possible, however, that requires longer scan time, making it not easy to apply in the neonatal population in a routine clinical setting. Improvement of the tractography algorithm itself can be used to mitigate but not solve some of the previous referred problems. Nevertheless, such improvements also involves making assumptions that might not suit our other constraints of our problem. An interesting direction would be to explore approaches like the one presented by Durrleman et al. [4] or Wang et al. [41] where segments of fibers are considered rather than the full length fibers. This will, however, increase the complexity of the pipeline given the considerable different nature of the proposed approaches.

2. Regarding the used atlas, it is built based on three datasets of neonates. Though, we consider these datasets as quite representative of neonatal TEA DTI data, using more data sets for the atlas construction might improve the accuracy of the results. In addition, it is important to note that although the atlas contains classes of spurious fibers, i.e., fibers that did not belong to any anatomical structure, these can only arise due to scanning or processing errors. Future atlas versions should consider removal of spurious fibers categorization and address them in the segmentation pipeline only with a good skull-stripping method or other outlier detection.

3. The applied skull-stripping reduces spurious fibers to 70%, therefore, reducing calculation of similarities for fibers without anatomical interest. However, there is empirical parameter tuning needed that should be addressed in the future.

4. The cluster-based sampling is still quite a computationally costly, as the cluster method in itself computes similarity metrics. Still, performing registration and label attribution using as input a sampled set was found to improve the accuracy of these pipeline steps, in contrast with using the complete tractogram for representing the atlas and the patient fibers. Computational costs were not the focus of this chapter, and we believe that computation costs can still be further reduced by using sparse distance matrices calculations, e.g., space subdivisions strategies like octrees.

5. For all factors probed to investigate what is affecting segmentation performance, only the registration accuracy seemed to matter. In particular a peculiar curved shape of the CC often lead to an excessive shrinking of the fibers in the registration step. Excessive shrinking might be caused due to an ineffective tuning of the registration constraint parameters, and/or due to the fact this kind of CC shape is not represented in the atlas, and/or even due to an inadequate similarity measure for the registration between atlas and subject fibers. Possible ways to tackle this can include applying a different deformation model which allows for local transformation, like elastic and fluid deformations—with the need of then defining optimal constraints for these models. Another direction would be to tune the similarity metric being used for registration specifically for dealing with neonatal data, like the one being used for the cluster-based sampling. It would be interesting to investigate other registration approaches in this context, for example, the registration approach proposed by Garyfallidis et al. [7].

As an end note, it is also important to refer that FX segmentation results still do not allow for an accurate investigation of this structure anisotropy. There are still too many erroneous fibers present. Guevara et al. [10] and O'Donnell et al. [25] also found for their methods that the association type fibers, FX and CG for the used atlas labels, were the ones more difficult to segment.

5 Conclusions

Our study shows that is feasible to automatically segment WM structures in the neonatal brain, by using an atlas-based and fiber-wise processing of DTI data. We observe that our fiber-wise method, i.e., making use of global information, allows for the radiologist and neonatologist to have a better overview of the patient's WM anatomical structure and eventual abnormalities. An automatic segmentation also means less user-dependency and a less time-consuming analysis, thus allowing to study WM maturation in an easier and more objective way. Segmentation performance showed not to be influenced by presence of WM pathology in subjects, even when anatomical structures were missing due to severe WM deviations. Indicating that is a good segmentation approach to be used when pathology is present.

In order to improve further registration quality, we believe that future work should aim to increase atlas variability but probably even more important to probe similarity metrics between fibers that are less computational expensive but still adequate for neonatal tractography. In addition, it might be of interest to extend the atlas for representing all gestational ages, for allowing study of full-term neonates at risk of neurodevelopmental disorders.

References

1. Anjari, M., Srinivasan, L., Allsop, J., Hajnal, J., Rutherford, M., Edwards, A., Counsell, S.: Diffusion tensor imaging with tract-based spatial statistics reveals local white matter abnormalities in preterm infants. NeuroImage **35**(3), 1021–1027 (2007)
2. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**(1), 254–267 (1994)
3. Dubois, J., Dehaene-Lambertz, G., Perrin, M., Mangin, J., Cointepas, Y., Duchesnay, E., Le Bihan, D., Hertz-Pannier, L.: Asynchrony of the early maturation of white matter bundles in healthy infants: quantitative landmarks revealed noninvasively by diffusion tensor imaging. Hum. Brain Mapp. **29**, 14–27 (2008)

4. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models of sets of curves and surfaces based on currents. Med. Image Anal. **13**(5), 793–808 (2009)
5. Frey, B., Dueck, D.: Clustering by passing messages between data points. Science **315**, 972 (2007)
6. Garyfallidis, E., Brett, M., Correia, M.M., Williams, G.B., Nimmo-Smith, I.: Quickbundles, a method for tractography simplification. Front. Neurosci. **6**, 175 (2012)
7. Garyfallidis, E., Ocegueda, O., Wassermann, D., Descoteaux, M.: Robust and efficient linear registration of white-matter fascicles in the space of streamlines. NeuroImage **117**, 124–140 (2015)
8. Geng, X., Gouttard, S., Sharma, A., Gu, H., Styner, M., Lin, W., Gerig, G., Gilmore, J.H.: Quantitative tract-based white matter development from birth to age 2 years. NeuroImage **61**(3), 524–557 (2012)
9. Glass, H., Bonifacio, S., Shimotake, T., Ferriero, D.: Neurocritical care for neonates. Pediatr. Neurol. - Curr. Treat. Opt. Neurol. **13**, 574–589 (2011)
10. Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Fillard, P., Le Bihan, D., Mangin, J.: Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. NeuroImage **61**, 1083–1099 (2012)
11. Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D., Calabresi, P., Pekar, J., van Zijl, P., Mori, S.: Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. NeuroImage **39**, 336–347 (2008)
12. Huppi, P., Dubois, J.: Diffusion tensor imaging of brain development. In: Seminars in Fetal & Neonatal Medicine, vol. 11, pp. 489–497 (2006)
13. Jong, M., Verhoeven, M., Baar, J.: School outcome, cognitive functioning, and behavior problems in moderate and late preterm children and adults: a review. In: *Seminars in Fetal & Neonatal Medicine*, vol. 17, pp. 163–169 (2012)
14. Kooij, B.: MRI analysis and neurodevelopmental outcome in preterm infants. PhD thesis, Utrecht University (2011)
15. Labra, N., Guevara, P., Duclap, D., Houenou, J., Poupon, C., Mangin, J.F., Figueroa, M.: Fast automatic segmentation of white matter streamlines based on a multi-subject bundle atlas. Neuroinformatics **15**(1), 71–86 (2017). doi: 10.1007/s12021-016-9316-7
16. Latal, B.: Prediction of neurodevelopmental outcome after preterm birth. Ped. Neuro **40**, 413–419 (2009)
17. Leemans, A., Jones, D.: A new approach to fully automated fiber tract clustering using affinity propagation. In: *ISMRM*, p. 17 (2009)
18. Leemans, A., Sijbers, J., De Backer, S., Vandervliet, E., Parizel, P.: Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. Magn. Reson. Med. **55**, 1414–1423 (2006)
19. Maddah, M., Grimson, W.E.L., Warfield, S.K., Wells, W.M.: A unified framework for clustering and quantitative analysis of white matter fiber tracts. Med. Image Anal. **12**(2), 191–202 (2008)
20. Mayer, A., Zimmerman-Moreno, G., Shadmi, R., Batikoff, A., Greenspan, H.: A supervised framework for the registration and segmentation of white matter fiber tracts. IEEE Trans. Med. Imag. **30**, 1 (2011)
21. Moberts, B., Vilanova, A., van Wijk, J.J.: Evaluation of fiber clustering methods for diffusion tensor imaging. In: IEEE Visualization, pp. 65–72 (2005)
22. Mori, S., Zijl, P.: Fiber tracking: principles and strategies - a technical review. NMR Biomed. **15**, 468–480 (2002)
23. Mori, S., Wakana, S., Nagae-Poetscher, L., Zijl, P.: MRI Atlas of Human White Matter. Elsevier, Amsterdam (2005)
24. Neil, J., Miller, J., Mukherjee, P., Hüppi, P.S.: Diffusion tensor imaging of normal and injured developing human brain - a technical review. NMR Biomed. **15**, 543–552 (2002)
25. O'Donnell, L., Westin, C.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Trans. Med. Imag. **26**(11), 1562–1575 (2007)

26. O'Donnell, L., Wells, W., Golby, A., Westin, C.: Unbiased groupwise registration of white matter tractography. In: MICCAI, vol. III, pp. 123–130 (2012)
27. Olivetti, E., Nguyen, T.B., Garyfallidis, E.: The approximation of the dissimilarity projection. In: 2012 International Workshop on Pattern Recognition in NeuroImaging (PRNI), pp. 85–88. IEEE, New York (2012)
28. Powell, M.J.D.: Direct search algorithms for optimization calculations. Acta Numer. **7**, 287–336 (1998)
29. Pul, C., Buijs, J., Vilanova, A., Roos, G., Wijn, P.: Infants with perinatal hypoxic ischemia: feasibility of fiber tracking at birth and 3 months. Radiology **240**, 203–214 (2006)
30. Pul, C., Kooij, B., Vries, L., Benders, M., Vilanova, A., Groenendaal, F.: Quantitative fiber tracking in the corpus callosum and internal capsule reveals microstructural abnormalities in preterm infants at term-equivalent age. AJNR Am. J. Neuroradiol. **33**, 678–684 (2012)
31. Ratnarajah, N., Qiu, A.: Multi-label segmentation of white matter structures: application to neonatal brains. NeuroImage **102**, 913–922 (2014)
32. Rose, J., Vassar, R., Cahill-Rowley, K., Guzman, X., Stevenson, D., Barnea-Goraly, N.: Brain microstructural development at near-term age in very-low-birth-weight preterm infants: an atlas-based diffusion imaging study. NeuroImage **86**, 244–256 (2014)
33. Saigal, S., Doyle, L.: An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet **371**(9608), 261–269 (2008)
34. Shankaran, S., Barnes, P., Hintz, S., Laptook, A., Zaterka-Baxter, K., McDonald, S., Ehrenkranz, R., Walsh, M., Tyson, J., Donovan, E., Goldberg, R., Bara, R., Das, A., Finer, N., Sanchez, P., Poindexter, B., Van Meurs, K., Carlo, W., Stoll, B., Duara, S., Guillet, R., Higgins, R.: Eunice kennedy shriver national institute of child health and human development neonatal research network.brain injury following trial of hypothermia for neonatal hypoxic-ischaemic encephalopathy. Arch. Dis. Child Fetal Neonatal Ed. **97**(6), F398–404 (2012)
35. Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
36. Vilanova, A., Berenschot, G., van Pul, C.: DTI visualization with streamsurfaces and evenly-spaced volume seeding. In: Joint Eurographics - IEEE TCVG Symposium on Visualization, pp. 173–182 (2004)
37. Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans. Image Process. **2**(2), 176–201 (1993)
38. Volpe, J.: Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. Lancet Neurol. **8**, 110–124 (2009)
39. Wakana, S., Jiang, H., Nagae-Poetscher, L., van Zijl, P., Mori, S.: Fiber tract-based atlas of human white matter anatomy. Radiology **230**, 77–87 (2004)
40. Wang, R., Benner, T., Wedeen, V.: Diffusion toolkit: a software package for diffusion imaging data processing and tractograph. In: ISMRM, vol. 15, p. 3720 (2007)
41. Wang, X., Grimson, W.E.L., Westin, C.-F.: Tractography segmentation using a hierarchical dirichlet processes mixture model. NeuroImage **54**(1), 290–302 (2011)
42. Wassermann, D., Bloyb, L., Kanterakisb, E., Vermab, R., Derichea, R.: Unsupervised white matter fiber clustering and tract probability map generation: applications of a gaussian process framework for white matter fibers. NeuroImage **15**(1), 228–241 (2010)
43. Westin, C.-F., Maier, S., Khidhir, B., Everett, P., Jolesz, F., Kikinis, R.: Image processing for diffusion tensor magnetic resonance imaging. In: MICCAI. Lecture Notes in Computer Science, vol. 1679, pp. 441–452. Springer, New York (1999)
44. Woodward, L., Anderson, P., Austin, N., Howard, K., Inder, T.: Neonatal mri to predict neurodevelopmental outcomes in preterm infants. New Engl. J. Med. **355**(7), 685–694 (2006)
45. Zhai, G., Lin, W., Wilber, K., Gerig, G., Gilmore, J.: Comparisons of regional white matter diffusion in healthy neonates and adults performed with a 3.0-t head-only mr imaging unit. Radiology **229**, 673–681 (2003)
46. Zvitia, O., Mayer, A., Shadmi, R., Miron, S., Greenspan, H.: Co-registration of white matter tractographies by adaptive-mean-shift and Gaussian mixture modeling. IEEE Trans. Med. Imag. **29**(1), 132–145 (2010)