Robust GM/WM Segmentation of the Spinal Cord with Iterative Non-local Statistical Fusion

Andrew J. Asman[1], Seth A. Smith[2], Daniel S. Reich[3], and Bennett A. Landman[1,2]

[1] Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235
[2] Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA 37235
[3] Translational Neuroradiology Unit, National Institutes of Health, Bethesda, MD, USA 37235
{andrew.j.asman,seth.smith,bennett.landman}@vanderbilt.edu,
daniel.reich@nih.gov

Abstract. New magnetic resonance imaging (MRI) sequences are enabling clinical study of the *in vivo* spinal cord's internal structure. Yet, low contrast-to-noise ratio, artifacts, and imaging distortions have limited the applicability of tissue segmentation techniques pioneered elsewhere in the central nervous system. Recently, methods have been presented for cord/non-cord segmentation on MRI and the feasibility of gray matter/white matter tissue segmentation has been evaluated. To date, no automated algorithms have been presented. Herein, we present a non-local multi-atlas framework that robustly identifies the spinal cord and segments its internal structure with submillimetric accuracy. The proposed algorithm couples non-local fusion with a large number of slice-based atlases (as opposed to typical volumetric ones). To improve performance, the fusion process is interwoven with registration so that segmentation information guides registration and vice versa. We demonstrate statistically significant improvement over state-of-the-art benchmarks in a study of 67 patients. The primary contributions of this work are (1) innovation in non-volumetric atlas information, (2) advancement of label fusion theory to include iterative registration/segmentation, and (3) the first fully automated segmentation algorithm for spinal cord internal structure on MRI.

Keywords: Spinal Cord Parcellation, Multi-Atlas Segmentation, Non-local Correspondence Models, Registration Refinement.

1   Introduction

The spinal cord is an essential and vulnerable component of the central nervous system [1, 2]. Differentiating and localizing pathology/degeneration of the gray matter (GM) and white matter (WM) plays a critical role in assessing therapeutic impacts and determining prognoses [3, 4]. Automated methods have localized the cord [5] and semi-automated segmentation has been used for internal segmentation [6]. Yet, automated GM/WM delineation has not been reported. Increased automation is necessary for routine volumetric assessment of the cord structures. Given the small size and artifacts of spinal cord MRI, the feasibility of an approach has only recently come to light using magnetization transfer (MT) MRI of the spinal cord *in vivo* [2].
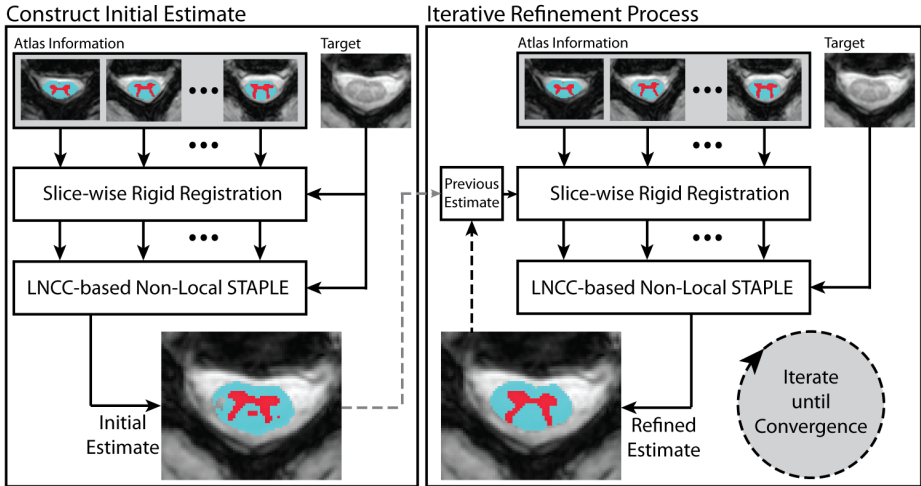
Fig. 1. Flowchart of the proposed iterative Non-Local STAPLE (iNLS) multi-atlas framework. Using an iterative atlas-target registration refinement framework, we expand the range of anatomical variability that can be reliably segmented.

Over the past decade, multi-atlas segmentation has come to prominence for its ability to rapidly and robustly generalize from labeled examples (i.e., atlases) [7, 8]. Unfortunately, as medical imaging researchers move out of the cranial vault towards more highly variable anatomical structures, the traditional registration followed by label fusion multi-atlas model [9-13] becomes increasingly problematic, as we are dependent upon reasonable atlas-target registrations.

Herein, we present the first fully-automated approach for GM/WM segmentation of the spinal cord through extension of a recently proposed non-local statistical fusion algorithm (Non-Local STAPLE – NLS [14]). We demonstrate submillimetric accuracy and show statistical improvement over other state-of-the-art approaches. The primary theoretical contributions of this work are: (1) we apply slice-based — as opposed to volumetric — atlases; (2) we adapt the NLS non-local correspondence model to use the locally normalized correlation coefficient (LNCC) to reduce the need for accurate intensity normalization [9]; and (3) we apply iterative registration refinement to lessens the impact of registration failures (**Figure 1**).

2   Theory

First, we describe the theoretical basis for the iterative non-local STAPLE (iNLS) framework and how it differs from the original NLS. Consider a target gray-level image represented as a vector, $I \in \mathbb{R}^{N \times 1}$. Let $T \in L^{N \times 1}$ be the latent representation of the true target segmentation, where $L = \{0, \dots, L-1\}$ is the set of possible labels. Consider a collection of R registered atlases with associated intensity values, $A \in \mathbb{R}^{N \times R}$, and label decisions, $D \in L^{N \times R}$. Let $\theta \in \mathbb{R}^{R \times L \times L}$ parameterize the raters (registered atlases) performance level. Each element of θ, $\theta_{js's}$, represents the probability that rater j observes label s' given that the true label is s at a given target voxel and the *corresponding* voxel on the associated atlas —

i.e., $\theta_{js's} \equiv p(D_{i^*j} = s', \boldsymbol{A}|T_i = s, \boldsymbol{I})$, where i^* is the voxel on atlas j that corresponds to target voxel i. Throughout, the index variables i, i^* and i' will be used to iterate over the voxels, s and s' over the labels, and j over the registered atlases.

Building upon the seminal Simultaneous Truth And Performance Level Estimation (STAPLE) algorithm [11], NLS reformulates the statistical fusion framework from a non-local means perspective. The goal of any non-local correspondence model is to estimate $f(A_{i'j}|I_i)$ – the probability that voxel i' on atlas j directly corresponds to the target image at voxel i. Originally, NLS used a Gaussian difference model [10, 12, 13] which has been shown to be highly successful for whole-brain segmentation [10, 12, 14]. Herein, we modify this correspondence model to be:

$$f(A_{i'j}|I_i) \equiv \alpha_{ji'i} = \frac{1}{Z_\alpha} \exp\left(\frac{LNCC_{\mathcal{N}_p}(A_{i'j}, I_i)}{\epsilon}\right) \exp\left(-\frac{\mathcal{E}_{ii'}^2}{2\sigma_d^2}\right) \tag{1}$$

where the first distribution is the intensity similarity model governed by locally normalized correlation coefficient between a patch on atlas j centered at voxel i' and the target image centered at voxel i, the second distribution is the spatial compatibility model, and Z_α is the partition function. In the intensity similarity model, we use the notation $LNCC_{\mathcal{N}_p}(\cdot,\cdot)$ to indicate the locally normalized coefficient using a patch window defined by $\mathcal{N}_p$, and ϵ is the weight factor for the exponential similarity. In the spatial compatibility model, $\mathcal{E}_{ii'}$ is the Euclidean distance between voxels i and i' in image space, and σ_d is the corresponding standard deviation. The partition function Z_α enforces the constraint that $\sum_{i' \in \mathcal{N}_s(i)} \alpha_{ji'i} = 1$, where $\mathcal{N}_s(i)$ is the set of voxels in the *search neighborhood* of a given target voxel. Through this constraint, $\alpha_{ji'i}$ can be directly interpreted as the probability that voxel i' on atlas j is the latent corresponding voxel, i^*, to a given target voxel, i.

Using Eq. 1, we can estimate the latent performance level parameters based upon the assumed lack of atlas-target correspondence. By taking the expected value across the search neighborhood, $\mathcal{N}_s(i)$, and assuming conditional independence between the intensity-label relationships, the performance level parameters can be approximated as

$$f(D_{i^*j} = s', \boldsymbol{A}_j|T_i = s, I_i, \theta_{js's}^{(k)}) \approx E[f(\boldsymbol{D}_j|T_i = s, \theta_{js}^{(k)})f(\boldsymbol{A}_j|I_i)]$$
$$= \sum_{i' \in \mathcal{N}_s(i)} \alpha_{ji'i} \theta_{js's}^{(k)} \tag{2}$$

We can then integrate the approximation provided in Eq. 2, directly into the Expectation-Maximization (EM) algorithm governing the statistical fusion framework. First, in the E-step, we estimate $\boldsymbol{W} \in \mathbb{R}^{L \times N}$, where W_{si} represents the probability that the true label associated with voxel i is label s, given the provided information. Using a Bayesian expansion and conditional independence between the atlases, the solution for $\boldsymbol{W}$ on iteration k is

$$W_{si}^{(k)} \equiv f(T_i = s|\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{I}, \boldsymbol{\theta}^{(k)})$$
$$= \frac{f(T_i = s) \prod_j \sum_{i' \in \mathcal{N}_s(i)} \alpha_{ji'i} \theta_{js's}^{(k)}}{\sum_n f(T_i = n) \prod_j \sum_{i' \in \mathcal{N}_s(i)} \alpha_{ji'i} \theta_{js'n}^{(k)}} \tag{3}$$

Finally, the resulting performance level parameters (M-step) are obtained by maximizing the expected value of the conditional log likelihood function.

$$\theta_{js's}^{(k+1)} = \frac{\sum_i \left(\sum_{i' \in \mathcal{N}_s(i):D_{i'j}=s'} \alpha_{ji'i} \right) W_{si}^{(k)}}{\sum_i W_{si}^{(k)}}. \tag{4}$$

2.1   Iterative Global Refinement Using the Previously Estimated Segmentation

In the first iteration of iNLS, registration is based on normalized correlation between the atlas and target image intensity using a 3 degree-of-freedom rigid body transform. For subsequent iterations, we iteratively refine the registration by maximizing overlap between the atlas segmentation and the current segmentation estimation. Specifically,

$$\Phi_j^R = \arg \max_{\Phi_j^R} \sum_i \delta \left(D_{\Phi_j^R(i)j}, \Psi_i \right) \tag{5}$$

where Φ_j^R represent the parameters associated with the rigid transformation (i.e., translation and rotation) between the current estimated segmentation, Ψ_i, and the transformed atlas labels, $D_{\Phi_j^R(i)j}$, and $\delta(\cdot, \cdot)$ is the kronecker delta function.

2.2   Initialization and Convergence

For all experiments, iNLS was initialized using a 2mm isotropic search neighborhood, a 1mm isotropic patch neighborhood, and the weight factor, ϵ, was set to 0.2. These parameter values were obtained by performing leave-one-out cross-validation using the provided atlases. The remaining parameters remain identical to the original NLS approach [14]. For the iterative global refinement procedure, convergence was detected when the rigid transformation parameters ceased to change across the atlases (less than 5 iterations for all presented results).

3   Methods and Results

3.1   Experimental Design

We study a dataset consisting of 67 MR images of the cervical spinal cord. All data were obtained on a 3T Philips Achieva (Philips Medical Systems, Best, The Netherlands) using a single channel body coil for transmission and a 16 channel neurovascular coil for signal reception. The center of the imaging volume was aligned to the space between the 3rd and 4th cervical levels. T2*w data were obtained using a 3D gradient echo (TR/TE/a = 121/12ms/9°) with a 3-shot EPI covering a field of view of 190 x 224 x 90 mm^3 with nominal resolution of 0.6 x 0.6 x 3 mm^3. Fat saturation was implemented by using a 1331 binomial excitation (ProSet), 2 signal averages, and a SENSE factor of 2. This acquisition was a part of an MT experiment where the same parameters would be performed with the addition of an MT prepulse.

Due to the highly variable nature of the spinal cord and the difficulty in performing consistent high degree-of-freedom registration [15], all multi-atlas segmentations were performed on a slice-by-slice basis. A collection of 85 slice atlases were randomly selected from the 2,010 (67 volumes × 30 slices) available slices. Note that atlas slices from the volume of interest were excluded during the leave-one-out segmentation to prevent biasing the results.

As benchmarks, we compare iNLS to a majority vote (MV) [8], an LNCC-weighted locally weighted vote (LWV) [9], STAPLE [11], non-local voting (NLV) [13], and a single iteration of NLS. For STAPLE, NLS and iNLS, "consensus voxels" (i.e., voxels where all registered atlases agree) were ignored. For fairness of comparison, the same non-local correspondence model was used for NLV, NLS, and iNLS. For all benchmarks, the presented results use all available atlases (up to 85) with a pairwise rigid 3 degree-of-freedom alignment using FLIRT [16]. Quantitative accuracy of each of the benchmarks was assessed on a volumetric basis using the Dice Similarity Coefficient (DSC), bi-directional mean surface distance error (MSDE), the bi-directional Hausdorff distance error (HDE).

3.2   Experimental Results

iNLS demonstrated statistically significant improvement over each of the considered benchmarks in terms of DSC, MSDE, and HDE for both gray matter and white matter segmentation (**Figure 2**). Importantly, iNLS is the only algorithm that provides a MSDE of less than 0.5 mm for all 67 subjects. In addition, iNLS results in a substantial decrease in outliers – particularly for the surface distance based metrics.
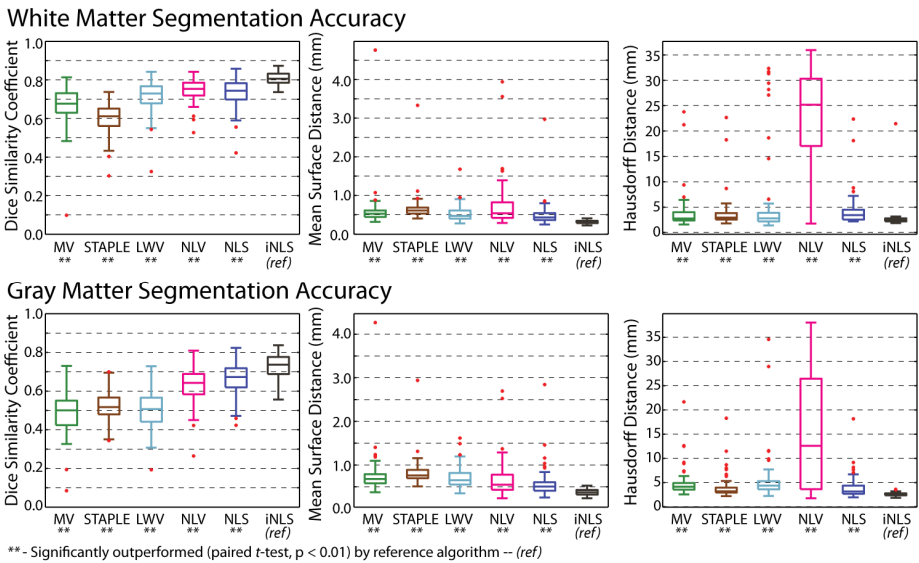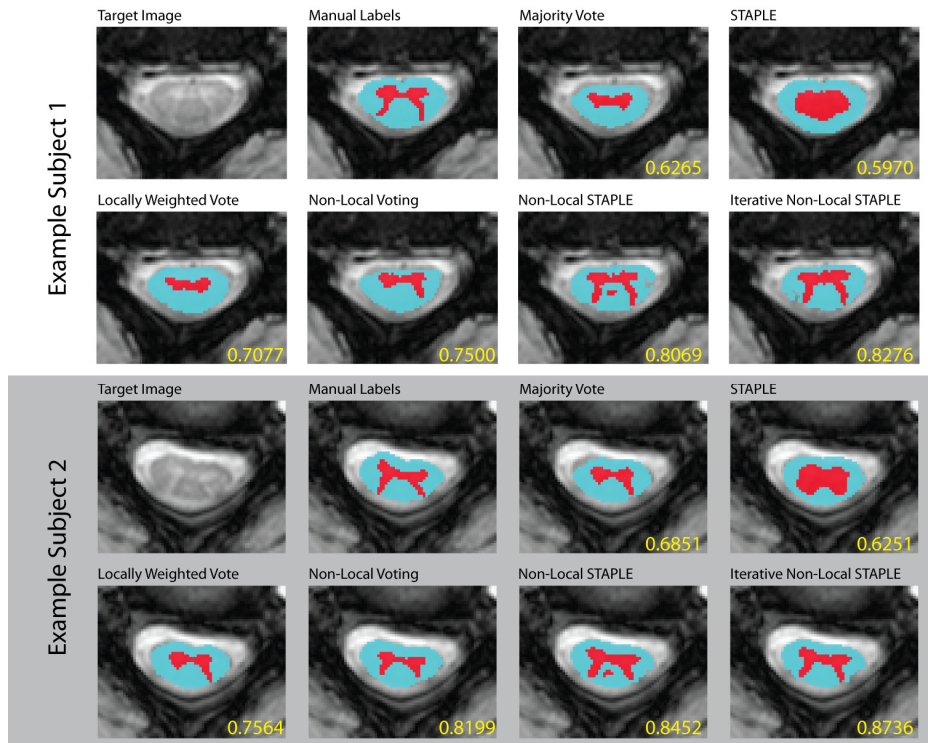

Fig. 2. Quantitative analysis on GM/WM segmentation of the spinal cord. iNLS provides significant improvement over all of all considered benchmarks.

As an aside, despite highly competitive results in terms of DSC and MSDE, NLV results in highly sub-optimal HDE values due to its susceptibility to outliers in the estimation process (i.e., outlier registrations result in "speckle noise" in the background of the estimate).

Qualitative results in terms of slice-wise accuracy (**Figure 3**) and volumetric surface distance error (**Figure 4**) support the quantitative improvement. iNLS provides visual (along with numeric) improvements over the initial NLS estimate. These can be appreciated in the precision with which the convoluted shape of the GM/WM boundary within the spinal cord. In **Figure 4**, it is evident that only iNLS provides estimates that are consistently less than 2mm on a voxelwise basis.

Note: Inlay shows white matter Dice Similarity Coefficient for subject volume.

Fig. 3. Slice-wise qualitative analysis of GM/WM segmentation of the spinal cord. Due to the lack of non-rigid registration, all of the non-local methods provide valuable accuracy improvements. However, only the proposed method, iNLS, is able to consistently maintain the complex shape of the GM/WM structures within the spinal cord.
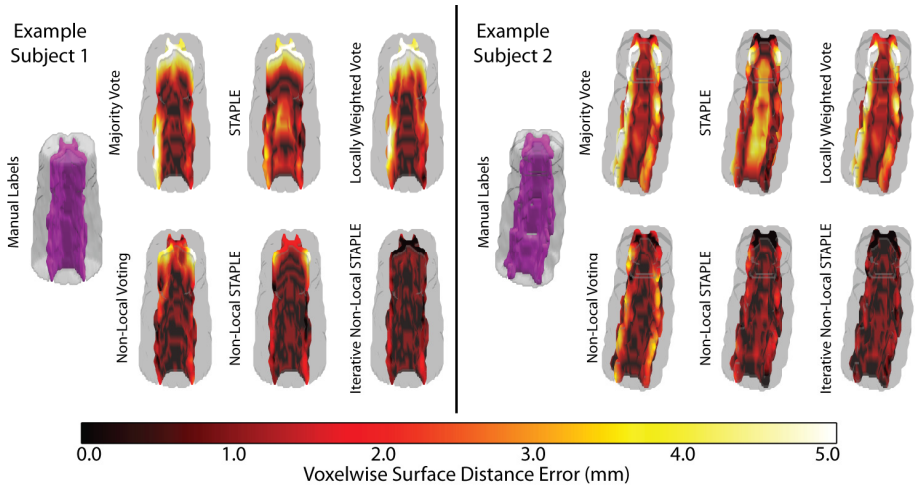
Fig. 4. Voxelwise surface distance error for spinal cord GM segmentation accuracy. The proposed iterative Non-Local STAPLE algorithm provides consistent volumetric improvement over the considered benchmarks in terms of voxelwise surface distance error.

4   Discussion

Accurate GM/WM delineation of the spinal cord plays a critical role in understanding the pathophysiological nature of spinal cord disease and assessing therapeutic interventions. Herein, we demonstrate an effective segmentation framework specifically targeting spinal cord GM/WM. We extend the recently proposed NLS algorithm with two critical advancements that enable robust GM/WM segmentation of the spinal cord. First, we reformulate the non-local correspondence model using the LNCC similarity metric to limit the need for accurate intensity normalization and to minimize the impact of imaging artifacts (e.g., intensity inhomogeneity). Second, we describe a new iterative atlas-target registration refinement process. Together, these advancements dramatically reduce the impact of initial registration failures, and, thus, significantly increase the robustness and accuracy of the resulting segmentation. We assessed the accuracy of the proposed iNLS framework against several of the current state-of-the-art benchmark algorithms and demonstrated statistically significant improvement in terms of DSC, MSDE, and HDE (**Figure 2**). Additionally, we provide both slice-wise (**Figure 3**) and volume-wise (**Figure 4**) qualitative examples that demonstrate the type of improvement exhibited by the proposed framework.

While the proposed framework is not the first algorithm to use the LNCC similarity metric (e.g., [9]) or the first approach to use segmentation-based registration refinement (e.g., [17, 18]), the provided joint-framework is novel. For example, [18] used iterative segmentation/registration to form group-wise consistent atlas representations while [17] used segmentation information in a deformable registration cost function. A fortunate consequence of moving to slice-wise registration is the speed of the individual registrations (i.e., seconds per slice as opposed to

minutes/hours per volume); hence, many more simple registrations were evaluated in the iNLS framework than would have been pragmatic in a volumetric one. In fact, we evaluated the use of high degree-of-freedom registration tools, but these consistently resulted in catastrophic failures (i.e., no label overlap) when applied to the raw MRI of the spinal column (data not shown). In conclusion, our efforts demonstrate that we can achieve submillimetric segmentation accuracy in spite of the severe distortion, inhomogeneity, low-contrast, and small-scales involved in spinal cord MRI.

References

1. Dietz, V., Curt, A.: Neurological aspects of spinal-cord repair: promises and challenges. The Lancet Neurology 5, 688–694 (2006)
2. Yiannakas, M., et al.: Feasibility of Grey Matter and White Matter Segmentation of the Upper Cervical Cord In Vivo: A pilot study with application to Magnetisation Transfer Measurements. Neuroimage 63, 1054–1059 (2012)
3. Gilmore, C.P., et al.: Spinal cord gray matter demyelination in multiple sclerosis—a novel pattern of residual plaque morphology. Brain Pathol. 16, 202–208 (2006)
4. Jarius, S., Wildemann, B.: AQP4 antibodies in neuromyelitis optica: diagnostic and pathogenetic relevance. Nature Reviews Neurology 6, 383–392 (2010)
5. Chen, M., et al.: Topology preserving automatic segmentation of the spinal cord in magnetic resonance images. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1737–1740. IEEE (2011)
6. Horsfield, M.A., et al.: Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis. Neuroimage 50, 446–455 (2010)
7. Rohlfing, T., et al.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Transactions on Medical Imaging 23, 983–994 (2004)
8. Heckemann, R.A., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115–126 (2006)
9. Artaechevarria, X., et al.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. IEEE Trans. Med. Imaging 28, 1266–1277 (2009)
10. Sabuncu, M.R., et al.: A generative model for image segmentation based on label fusion. IEEE Transactions on Medical Imaging 29, 1714–1729 (2010)
11. Warfield, S.K., et al.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23, 903–921 (2004)
12. Wang, H., et al.: Multi-Atlas Segmentation with Joint Label Fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 611–623 (2012)
13. Coupé, P., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage 54, 940–954 (2011)
14. Asman, A.J., Landman, B.A.: Non-Local STAPLE: An Intensity-Driven Multi-Atlas Rater Model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 426–434. Springer, Heidelberg (2012)

15. Commowick, O., Wiest-Daesslé, N., Prima, S.: Automated diffeomorphic registration of anatomical structures with rigid parts: Application to dynamic cervical MRI. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 163–170. Springer, Heidelberg (2012)
16. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Medical Image Analysis 5, 143–156 (2001)
17. Heckemann, R.A., et al.: Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage 51, 221 (2010)
18. Jia, H., et al.: Iterative multi-atlas-based multi-image segmentation with tree-based registration. Neuroimage 59, 422–430 (2012)