



HAL
open science

A supervised patch-based approach for human brain labeling

François Rousseau, Piotr Habas, Colin Studholme

► **To cite this version:**

François Rousseau, Piotr Habas, Colin Studholme. A supervised patch-based approach for human brain labeling. IEEE Transactions on Medical Imaging, 2011, 30 (10), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5771116. 10.1109/TMI.2011.2156806 . hal-00631458

HAL Id: hal-00631458

<https://hal.science/hal-00631458>

Submitted on 12 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A supervised patch-based approach for human brain labeling

François Rousseau
LSIIT, CNRS - Univ. Strasbourg
Strasbourg, France
rousseau@unistra.fr

Piotr A. Habas
BICG, Dept. of Pediatrics
Univ. of Washington, Seattle
p.habas@gmail.com

Colin Studholme
BICG, Dept. of Pediatrics
Univ. of Washington, Seattle
colin.studholme@ieee.org

Abstract

We propose in this work a patch-based image labeling method relying on a label propagation framework. Based on image intensity similarities between the input image and an anatomy textbook, an original strategy which does not require any non-rigid registration is presented. Following recent developments in non-local image denoising, the similarity between images is represented by a weighted graph computed from an intensity-based distance between patches. Experiments on simulated and in-vivo MR images show that the proposed method is very successful in providing automated human brain labeling.

1 Introduction

Automated brain labeling [28, 7, 24, 23] is a crucial step in neuroimaging studies since it provides a segmentation map of contiguous meaningful brain regions. Label propagation (also called label fusion) is a way to segment an image and it is usually performed by registering one or several labeled images to the image to be segmented. The labels are then mapped and fused using classification rules to produce a new label map of the input image [30, 15, 3, 2, 16, 8, 33, 35, 26, 38, 34, 25, 19]. The principle of such registration-based labeling approach is shown in Figure 1. This segmentation approach is highly versatile since the main prerequisite is an anatomy textbook, *i.e.* a set of measured images (such as Magnetic Resonance (MR) imaging or Computerized Tomography (CT)) and the corresponding label maps. The use of label propagation has been extensively investigated for automatic brain MR segmentation, especially for structures such as hippocampus, caudate, putamen, amygdala, etc., and cortex areas. The keypoints of registration-based label propagation approaches concern the accuracy of the non-rigid registration, the fusion rules [11, 30, 36, 15, 3], the selection of the labeled images [37, 2, 35] and the labeling errors in primary manual segmentation.

Because of the use of a registration algorithm, one makes the implicit (and strong) assumption that there exists a one-to-one mapping between the input image and all the anatomical images of the textbook. In the case where no one-to-one mapping is possible, a registration-based labeling framework propagates incorrect labels. Moreover, local incorrect matching due to inherent registration errors can also lead to segmentation errors. Finally, as shown in a recent evaluation study [22], there is a non-negligible discrepancy in term of quality of matching between non-rigid registration techniques. Thus, even if several algorithms are now freely available, the non-rigid registration procedure remains a complex step to setup and it is usually quite CPU time consuming.

In this work, following recent developments in non-local image denoising [6, 17], we propose an alternate strategy for label propagation which does not require any non-rigid registration. The proposed algorithm makes use of local similarities between the image to be labeled and the images contained in the anatomy textbook. The key idea is similar to a fuzzy block matching approach which avoids the constraint of

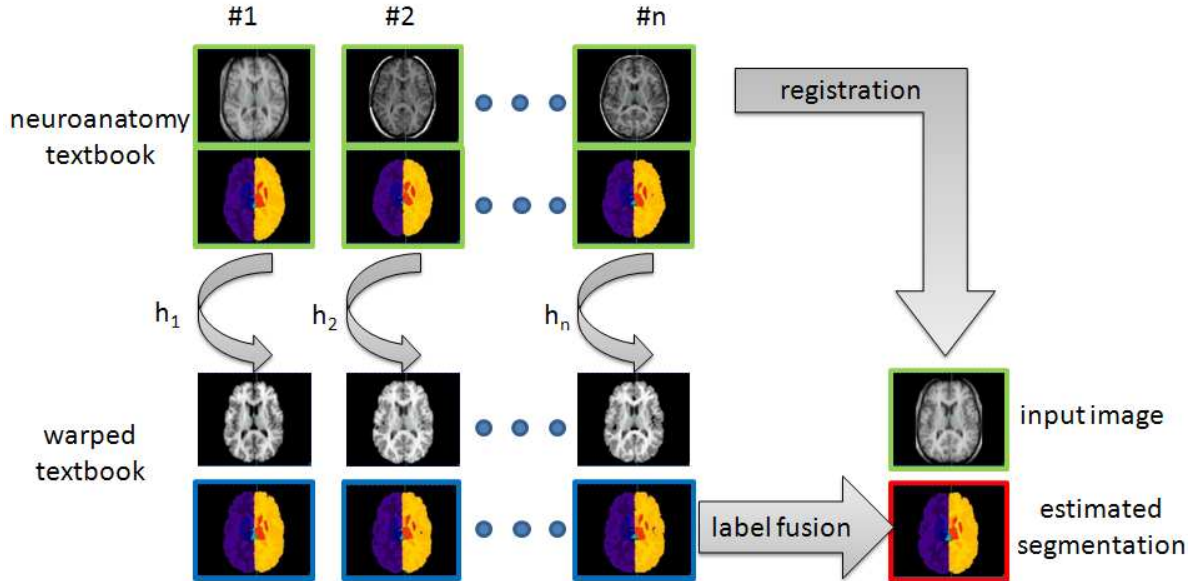


Figure 1: Principle of registration-based label propagation methods. The input data (shown with green borders) are an anatomy textbook (*i.e.* a set of n anatomical images with the corresponding label maps), and one anatomical image I . The set of anatomical images of the textbook is (non-linearly) registered to the input image I , and each label map is deformed with respect to the estimated transformation h_i . The final image segmentation (shown with red borders) is then obtained by fusing all the deformed label maps (shown with blue borders).

a strict one-to-one mapping. The method described in this article has been developed irrespective of a similar patch-based approach recently proposed by Coupé *et al.* [10]. The contributions of our work can be summarized as follows: 1) a patch-based framework for automated image labeling, 2) investigation of several patch aggregation strategies (pair-wise vs group-wise labeling and pointwise vs multipoint estimation), 3) comparison of fusion rules (majority voting rule vs STAPLE [36]), 4) application to human brain labeling using publicly available simulated (Brainweb) and *in-vivo* (IBSR and NA0-NIREP) datasets, 5) comparison with a non-rigid registration-based technique.

2 Patch-based Label Propagation

2.1 Patch-based Principle

Recently, Buades *et al.* [6] have proposed a very efficient denoising algorithm relying on a non-local framework. Since then, this non-local strategy has been studied and applied in several image processing applications such as non-local regularization functionals in the context of inverse problems [20, 27, 14, 29, 31] or medical image synthesis [32].

Let us consider, over the image domain Ω , a weighted graph w that links together the voxels of the input image I with a weight $w(\mathbf{x}, \mathbf{y})$, $(\mathbf{x}, \mathbf{y}) \in \Omega^2$. This weighted graph w is a representation of non-local similarities in the input image I .

In [6], the non-local graph w is used for denoising purpose using a neighborhood averaging strategy (called

non-local means (NLM):

$$\forall \mathbf{x} \in \Omega, I_{nlm}(\mathbf{x}) = \frac{\sum_{\mathbf{y} \in \Omega} w(\mathbf{x}, \mathbf{y}) I(\mathbf{y})}{\sum_{\mathbf{y} \in \Omega} w(\mathbf{x}, \mathbf{y})} \quad (1)$$

where w is the graph of self-similarity computed on the noisy image I , $I(\mathbf{y})$ is the gray level value of the image I at the voxel \mathbf{y} and I_{nlm} is a denoised version of I .

The weighted graph reflects the similarities between voxels of the same image. It can be computed using a intensity-based distance between patches [9]:

$$w(\mathbf{x}, \mathbf{y}) = f \left(\frac{\sum_{\mathbf{x}' \in \mathcal{P}_I(\mathbf{x}), \mathbf{y}' \in \mathcal{P}_I(\mathbf{y})} (I(\mathbf{x}') - I(\mathbf{y}'))^2}{2N\beta\hat{\sigma}^2} \right) \quad (2)$$

where $\mathcal{P}_I(\mathbf{x})$ is a 3D patch of the image I centered at voxel \mathbf{x} ; f is a kernel function ($f(x) = e^{-x}$ in [6]), N is the number of voxels of a 3D patch; σ is the standard deviation of the noise and β is a smoothing parameter. With the assumption of Gaussian noise in images, β can be set to 1 (see [6] for theoretical justifications) and the standard deviation of noise σ is estimated via pseudo-residuals as defined in [13].

2.2 Label Propagation

In this work, we propose to investigate the use of such a non-local patch-based approach for label propagation. Let \mathcal{T} be an anatomy textbook containing a set of T1-weighted MR images \mathcal{I} and the corresponding label maps \mathcal{L} : $\mathcal{T} = \{(\mathcal{I}_i, \mathcal{L}_i), i = 1, \dots, n\}$.

2.2.1 Weighted Graph and Label Propagation

Let us consider, over the image domain Ω , a weighted graph w_i that links together voxels \mathbf{x} of the input image I and voxels \mathbf{y} of the image \mathcal{I}_i with a weight $w_i(\mathbf{x}, \mathbf{y})$, $(\mathbf{x}, \mathbf{y}) \in \Omega^2$. w_i is computed as follows:

$$w_i(\mathbf{x}, \mathbf{y}) = f \left(\frac{\sum_{\mathbf{x}' \in \mathcal{P}_I(\mathbf{x}), \mathbf{y}' \in \mathcal{P}_{\mathcal{I}_i}(\mathbf{y})} (I(\mathbf{x}') - \mathcal{I}_i(\mathbf{y}'))^2}{2N\beta\hat{\sigma}^2} \right). \quad (3)$$

This weighted graph w_i is a representation of non-local similarities between the input image I and the image \mathcal{I}_i of the textbook (see Figure 2).

The assumption on which the proposed method relies is the following one: *if patches of the input image I are locally similar to patches of the anatomy textbook, they should have a similar label*. The label propagation procedure is then performed using the set of graphs $\{w_i\}_{i=1, \dots, n}$ which reflects the local similarities between I and $\{\mathcal{I}_i\}_{i=1, \dots, n}$.

In the context of denoising [6], Buades *et al.* assume that every patch in a natural image has many similar patches in the same image. Thus, the graph w is computed between each voxel of the input image (which has led to the term ‘‘non-local’’ algorithm): $w = \{w(\mathbf{x}, \mathbf{y}), \forall (\mathbf{x}, \mathbf{y}) \in \Omega^2\}$. This is because of the assumption that similarities can be found in the entire image I . However, in the context of human brain labeling, the location of the brain structures is not highly variable and it is then not desirable to go through the entire image domain Ω to find good examples for the segmentation of a specific part of a brain. In this work, the graph is computed using a limited number of neighbors: $w = \{w(\mathbf{x}, \mathbf{y}), \forall \mathbf{x} \in \Omega, \mathbf{y} \in \mathcal{N}(\mathbf{x})\}$, where $\mathcal{N}(\mathbf{x})$ is the neighborhood of the voxel \mathbf{x} . The size of the considered neighborhood \mathcal{N} is directly related to the brain variability. The influence of this parameter is evaluated in the Section 3.2.

2.2.2 Pair-wise Label Propagation

For clarity sake, let start with an input image I and a textbook \mathcal{T} which contains only one anatomical image and the corresponding labeled image: $\mathcal{T} = (\mathcal{I}, \mathcal{L})$ (this is the basis of pair-wise label propagation

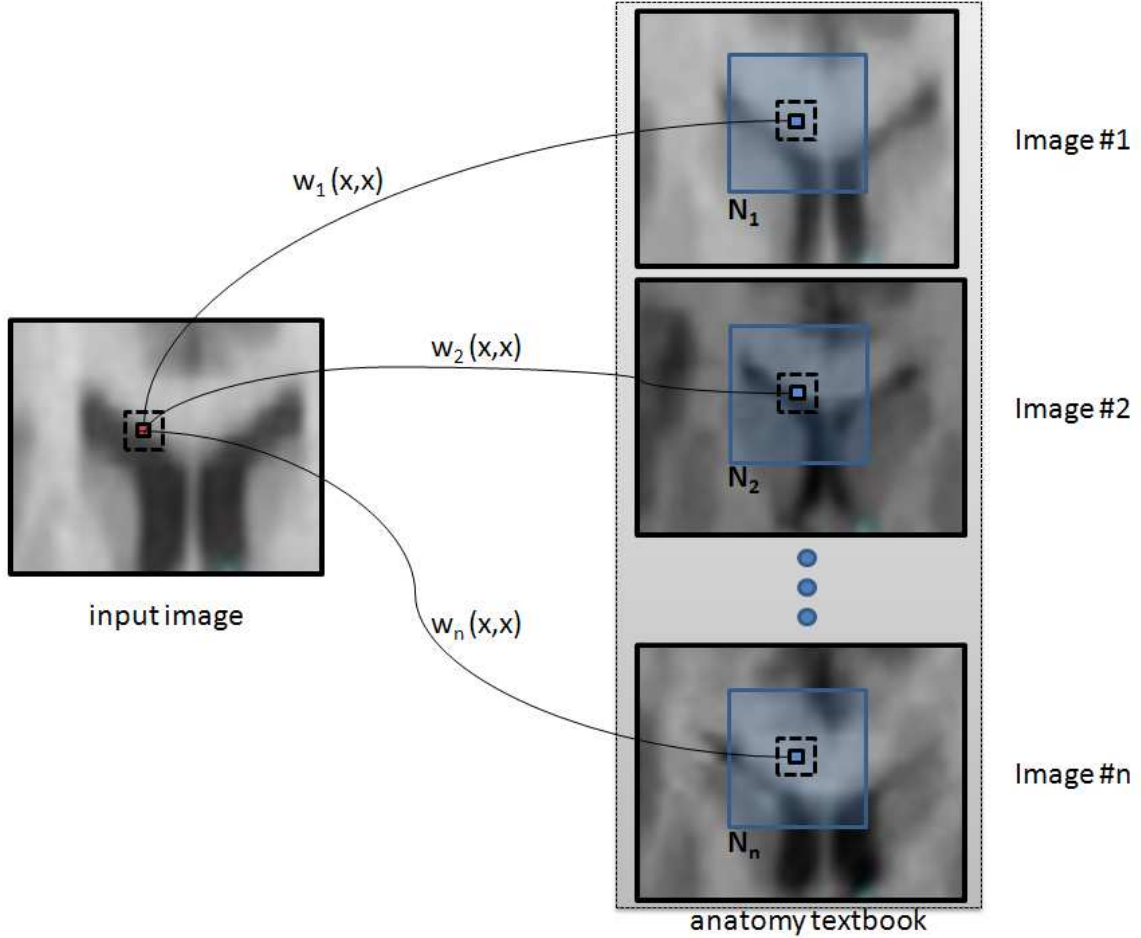


Figure 2: Weighted graph building. The set of graphs $\{w_i\}_{i=1,\dots,n}$ is a representation of non-local interactions between the input image I and the images $\{\mathcal{I}_i\}_{i=1,\dots,n}$ of the textbook. $\mathcal{N}_i(\mathbf{x})$ is the neighborhood of the voxel \mathbf{x} in the image \mathcal{I}_i .

techniques). Using the patch-based approach described previously, the image I can be labeled using the following equation:

$$\forall \mathbf{x} \in \Omega, L(\mathbf{x}) = \frac{\sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w(\mathbf{x}, \mathbf{y}) \mathcal{L}(\mathbf{y})}{\sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w(\mathbf{x}, \mathbf{y})} \quad (4)$$

where $\mathcal{L}(\mathbf{y})$ is a vector of $[0, 1]^M$ (M is the total number of labels) representing the proportions for each label at the voxel \mathbf{y} in the image \mathcal{L} (this notation unifies the case where the anatomy textbook contains hard or fuzzy label maps). Thus, $L(\mathbf{x}) = (l_1(\mathbf{x}), l_2(\mathbf{x}), \dots, l_M(\mathbf{x}))$ are the membership ratios of the voxel \mathbf{x} with respect to the M labels, such that $\sum_k l_k(\mathbf{x}) = 1$ and $l_k(\mathbf{x}) \in [0, 1], \forall k \in [1, M]$. Then, for each voxel \mathbf{x} of I , the equation (4) leads to a fuzzy labeling of the input image I since $L(\mathbf{x}) \in [0, 1]^M$. A hard segmentation H of the image I can be obtained by taking the component of $L(\mathbf{x})$ with the highest value. The hardening of the fuzzy label vector $L(\mathbf{x})$ to get a binary label vector $H(\mathbf{x})$ is done as follows:

$$h_k(\mathbf{x}) = \begin{cases} 1 & \text{if } l_k(\mathbf{x}) = \|L(\mathbf{x})\|_\infty \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Now, let us consider a textbook \mathcal{T} containing n pairs of images: $\mathcal{T} = \{(\mathcal{I}_i, \mathcal{L}_i), i = 1, \dots, n\}$. The most straightforward approach is to perform n times the pair-wise procedure previously described, leading to n

fuzzy segmentations $\{L_i(\mathbf{x})\}_{i=1,\dots,n}$ (or n hard segmentations $\{H_i(\mathbf{x})\}_{i=1,\dots,n}$). Then, these n label maps can be fused using a classifier combination strategy [21]. Thus, for instance, under the assumption of equal priors and by hardening the fuzzy label maps L_i to get a set of hard segmentations $\{H_i\}_{i=1,\dots,n}$, the final labeling $\Lambda(\mathbf{x})$ at the voxel \mathbf{x} can be obtained by applying the majority voting (MV) rule:

$$\lambda_k(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^n h_{ik}(\mathbf{x}) = \max_{j=1}^M \sum_{i=1}^n h_{ij}(\mathbf{x}) \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

where h_{ij} is a binary value corresponding to the label j of the image i at the voxel \mathbf{x} and $\Lambda(\mathbf{x}) = (\lambda_1(\mathbf{x}), \lambda_2(\mathbf{x}), \dots, \lambda_M(\mathbf{x}))$.

The pair-wise labeling approach using the majority voting rule to fuse the labels is described in Algorithm 1. A version of this pair-wise technique using STAPLE as the final label fusion is also evaluated in the Section 3.

Algorithm 1 Pair-wise labeling method using the majority voting rule

inputs: an image I and an anatomy textbook $\mathcal{T} = \{(\mathcal{I}_i, \mathcal{L}_i), i = 1, \dots, n\}$

output: a label image Λ

for all $\mathbf{x} \in \Omega$ **do**

for $i = 1$ to n **do**

 Compute $L_i(\mathbf{x})$ using Equation (4)

 Compute $H_i(\mathbf{x})$ using Equation (5)

end for

 Compute $\Lambda(\mathbf{x})$ by aggregating the set of labels $\{H_i\}_{i=1,\dots,n}$ (Equation (6))

end for

2.2.3 Group-wise Label Propagation

We propose to study a group-wise combination strategy which takes all the images of the textbook to produce a fuzzy labeling. Indeed, the Equation (4) can be applied for an arbitrary number of labeled images:

$$\forall \mathbf{x} \in \Omega, L(\mathbf{x}) = \frac{\sum_{i=1}^n \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w_i(\mathbf{x}, \mathbf{y}) \mathcal{L}_i(\mathbf{y})}{\sum_{i=1}^n \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w_i(\mathbf{x}, \mathbf{y})}. \quad (7)$$

Again, this leads to a fuzzy labeling which can be thresholded to obtain a hard labeling by taking the maximum of each vector $L(\mathbf{x})$:

$$h_k(\mathbf{x}) = \begin{cases} 1 & \text{if } l_k(\mathbf{x}) = \|L(\mathbf{x})\|_\infty \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

In this case, the final segmentation Λ can be set to L (resp. H) if a final fuzzy (resp. hard) labeling is desired. Unlike the pair-wise approach, the weight of each label map \mathcal{L}_i is automatically set by using the local patch-based similarity measure (see Equation (3)) with the input image I . There is no need to apply a classifier combination strategy. This is described in Algorithm 2.

Algorithm 2 Group-wise labeling method

inputs: an image I and an anatomy textbook $\mathcal{T} = \{(\mathcal{I}_i, \mathcal{L}_i), i = 1, \dots, n\}$

output: a label image Λ

for all $\mathbf{x} \in \Omega$ **do**

 Compute $L(\mathbf{x})$ using Equation (7)

 Compute $H(\mathbf{x})$ using Equation (8)

$\Lambda(\mathbf{x}) \leftarrow H(\mathbf{x})$

end for

2.3 Pointwise and Multipoint Estimation

In the previous section, according to the classification proposed by Katkovnik *et al.* in [17], the patch-based labeling techniques (both pair-wise and group-wise) provide pointwise estimates (Equations (4) and (7)). This means that the pair-wise and group-wise methods estimate one label vector L for every voxel \mathbf{x} . However, since the patch-based similarity measure is the core of the proposed labeling methods, one can obtain a label patch estimate at each considered voxel. This is called a multipoint label estimator. In contrast to the pointwise estimator, a multipoint estimator gives the estimate for a set of points (in our case, a patch).

For instance, a multipoint estimate for the group-wise labeling method is given by:

$$\forall \mathbf{x} \in \Omega, P_L(\mathbf{x}) = \sum_{i=1}^n \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w_i(\mathbf{x}, \mathbf{y}) P_{\mathcal{L}_i}(\mathbf{y}). \quad (9)$$

where $P_{\mathcal{L}_i}(\mathbf{y})$ is a 3D patch of the label map \mathcal{L}_i centered at the voxel \mathbf{y} , and $P_L(\mathbf{x})$ is one multipoint label estimate centered at the voxel \mathbf{x} . Thus, by going through the entire image domain Ω , one label patch estimate is obtain for each voxel. Since there is an overlap between all these label patches (each containing N voxels), for every voxel $\mathbf{x} \in \Omega$, we have N label estimates. Indeed, the label patch estimate of each voxel \mathbf{y} which belongs to the patch $P(\mathbf{x})$ contributes to the final label estimate of the voxel \mathbf{x} . These N estimates can then be aggregated using a combination classifier. In this work, we have used the majority voting rule to fuse these N estimates.

As suggested in [9], it is possible to speed-up the algorithm by considering a subset Ω^* of the image domain Ω with the constraint that there is at least one label estimate for each voxel. For instance, by using patches of size $3 \times 3 \times 3$ voxels, one can consider only voxels with even spatial index (in each dimension) which leads to a speed-up of 8. Such an approach (denoted as fast multipoint in this article) is evaluated in the Section 3. Moreover, for pointwise and multipoint estimation, based on the mean and variance of patches [9], a voxel preselection can be used to avoid useless computation.

Finally, several patch-based label propagation algorithms can be derived from the proposed framework depending of the following choices: pair-wise (PW) *vs* group-wise (GW), point-wise *vs* (possibly fast) multipoint estimates, majority voting (MV) rule *vs* STAPLE.

3 Experiments

3.1 Evaluation Framework

In this work, experiments have been carried out on three publicly available image datasets: Brainweb [4], the Internet Brain Segmentation Repository (IBSR) database and the NA0 database developed for the Non-rigid Image Registration Evaluation Project (NIREP). These three complementary image datasets provide different challenges for label propagation techniques: Brainweb dataset is used to mainly evaluate the separation power between 2 principal brain structures (white matter and gray matter), IBSR is a well known dataset for segmentation algorithms evaluation since it contains 32 brain structures (white matter, cortex, internal gray structures such as hippocampus, caudate, thalamus, putamen, etc.), and the NA0 dataset which has been originally created for registration evaluation purpose, provide also a good evaluation framework for cortical parcellation algorithms.

The Brainweb dataset¹ is a set of 20 T1-weighted simulated data with these specific parameters: SFLASH (spoiled FLASH) sequence with TR=22ms, TE=9.2ms, flip angle=30 deg and 1 mm isotropic voxel size. Each anatomical model consists of a set of 3-dimensional tissue membership volumes, one for each tissue class: background, cerebro-spinal fluid (CSF), gray matter (GM), white matter (WM), fat, muscle, muscle/skin, skull, blood vessels, connective (region around fat), dura matter and bone marrow. Example images from the Brainweb dataset are shown in Figure 3 (first column).

¹<http://mouldy.bic.mni.mcgill.ca/brainweb>

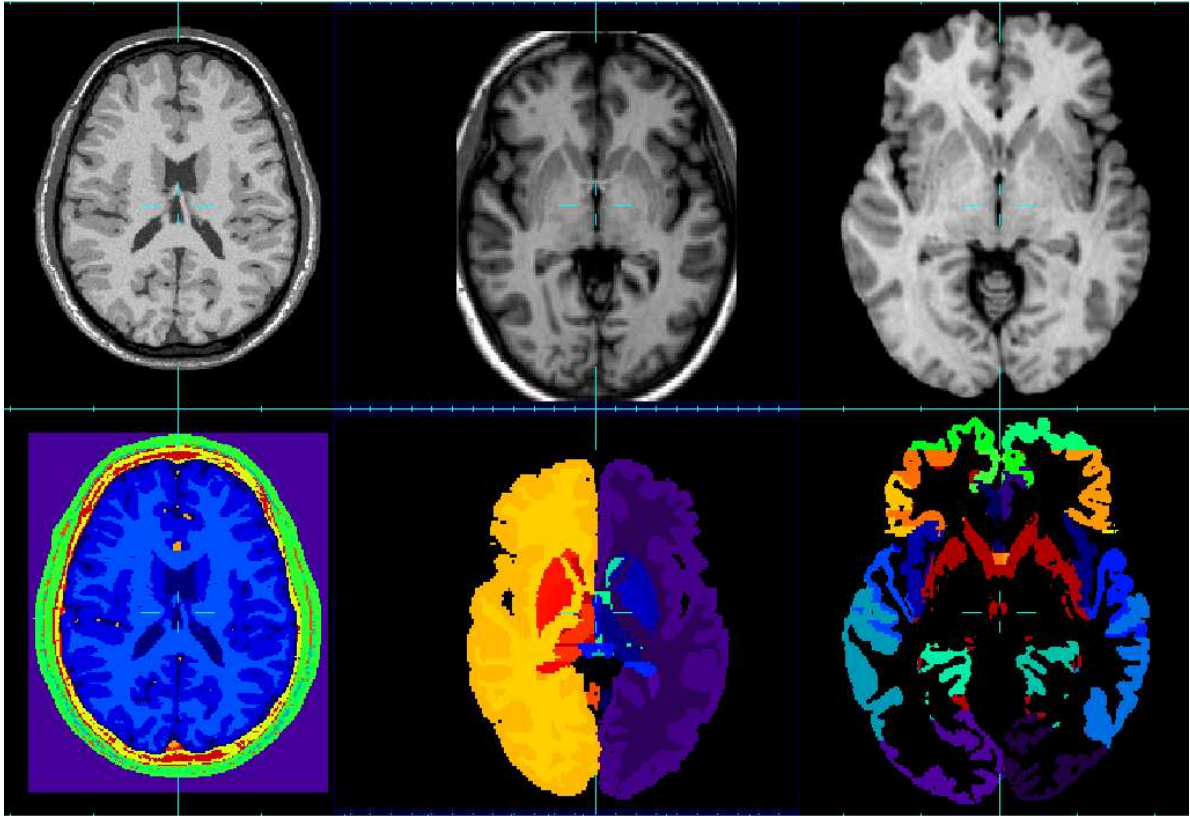


Figure 3: MR image datasets used for the evaluation and the corresponding segmentation. First column: Brainweb database, second column: IBSR database, third column: NA0-NIREP database.

For the IBSR dataset², the MR brain data sets and their manual segmentations are provided by the Center for Morphometric Analysis at Massachusetts General Hospital. It contains 18 images of healthy brains and the corresponding segmentation of the whole brain into 32 structures. Example images from the IBSR dataset are shown in Figure 3 (second column). The following pre-processing has been applied on IBSR images: 1) N3-based bias field correction using MIPAV³, 2) ITK-based histogram matching using 3DSlicer⁴, 3) affine registration using ANTs⁵ [5]. For each image to segment, all the images of the anatomy textbook are registered to this current image. Then, the label maps are also transformed (using a nearest neighbor interpolation). A total of 18*17 affine registrations have been performed.

The evaluation database NA0⁶ consists of a population of 16 annotated 3D MR image volumes corresponding to 8 normal adult males and 8 females. The 16 MR data sets have been segmented into 32 gray matter regions of interest (ROIs) (see Figure 3 (third column)). MR images were obtained in a General Electric Signa scanner operating at 1.5 Tesla, using the following protocol: SPGR/50, TR 24, TE 7, NEX 1 matrix 256×192 , FOV 24 cm. 124 contiguous coronal slices were obtained, with 1.5 or 1.6 mm thick, and with an interpixel distance of 0.94 mm. Three data sets were obtained for each brain during each imaging session. These were coregistered and averaged post hoc using Automated Image Registration (AIR 3.03, UCLA)⁷. The final data volumes had anisotropic voxels with an interpixel spacing of 0.7 mm and interslice

²<http://www.cma.mgh.harvard.edu/ibsr>

³<http://mipav.cit.nih.gov>

⁴<http://www.slicer.org>

⁵<http://www.picsl.upenn.edu/ANTs>

⁶<http://www.nirep.org>

⁷<http://bishopw.loni.ucla.edu/air5>

spacing of 1.5 - 1.6mm. All brains were reconstructed in three dimensions using Brainvox⁸. Before tracing ROIs, brains were realigned along a plane running through the anterior and posterior commissures (i.e., the AC-PC line); this ensured that coronal slices in all subjects were perpendicular to a uniformly and anatomically defined axis of the brain. The following pre-processing has been applied on NA0 images: 1) N3-based bias field correction using MIPAV, 2) ITK-based histogram matching using 3DSlicer.

As it is usually used for label propagation method evaluation, for each aforementioned dataset, a leave-one-out study is performed. Each image is separately selected as the image to be segmented. Using the provided segmentations of the remaining images as the anatomy textbook, label propagation is performed to obtain a segmentation of the considered image. In all cases, the Dice Index (DI) overlap is used as a segmentation quality measure:

$$DI = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (10)$$

where TP is the number of true positives, FP is the number of false positives and FN , the number of false negatives. The Dice index is computed for hard segmentations.

3.2 Neighborhood \mathcal{N} , Patch P and β

Experiments were carried out on the Brainweb dataset to determine the influence of the size of the neighborhood \mathcal{N} (which can be viewed as a local search area for similar patches), the size of the patches P and the smoothing parameter β (Equation 2). A pair-wise label propagation approach with the MV rule to fuse label maps has been used to determine the optimal parameter set. We found that the highest overlap value for both gray matter and white matter is obtained with the smallest size of patches (i.e. $3 \times 3 \times 3$ voxels), the lower value of β (0.5) and large neighborhood \mathcal{N} (the dice index stabilizes at a neighborhood size of $11 \times 11 \times 11$ voxels). These findings are similar to some results obtained for image denoising [9] or image reconstruction [31]. β is a noise level dependent parameter. While the lowest value of β provides the best overlap results on Brainweb images (which are almost noiseless), we have found that a value of 1 is more appropriate for in-vivo brain MR images. In the next sections, the following setting has been used: a patch size of $3 \times 3 \times 3$ voxels, a neighborhood size of $11 \times 11 \times 11$ voxels, and $\beta = 1$.

3.3 Influence of the size of the anatomy textbook

The evolution of the mean Dice index with respect to the size of the anatomy textbook has also been studied. This experiment has been carried out on the IBSR dataset, using up to 30 image permutations for each image to label. The behavior of the fast multipoint patch-based approach is compared with a non-rigid registration-based method (using ANTs [5], with the following command line: `ANTS 3 -m PR[target.nii, source.nii, 1, 2] -i 100x100x10 -o output.nii -t SyN[0.25] -r Gauss[3,0]` and the majority voting rule as label fusion strategy). In both cases, it has been found that increasing the size of the anatomy textbook provides higher overlap values (which stabilize quickly around 8 images). Figure 4 shows the evolution of the mean Dice index with respect to the number of images contained in the anatomy textbook. For all the results shown in the next sections, the maximum number of images has been used.

3.4 Pointwise vs blockwise

Comparison of pair-wise (using the majority voting rule for label fusion) *vs* group-wise approaches is shown in Table 1 for the Brainweb images, in Table 2 for IBSR and in Table 6 for NA0-NIREP. We have also compared three types of estimators: pointwise *vs* multipoint *vs* fast multipoint. As mentioned in [17] (“*in multipoint image estimation a weighted average of a few sparse estimates is better than single sparse estimate alone*”), the results obtained on the three datasets suggest that the use of a multipoint estimator leads to the best segmentation (i.e. the highest overlap measures). Finally, the pair-wise approach and the group-wise approach lead to very similar results.

⁸<http://www.nitrc.org/projects/brainvox>

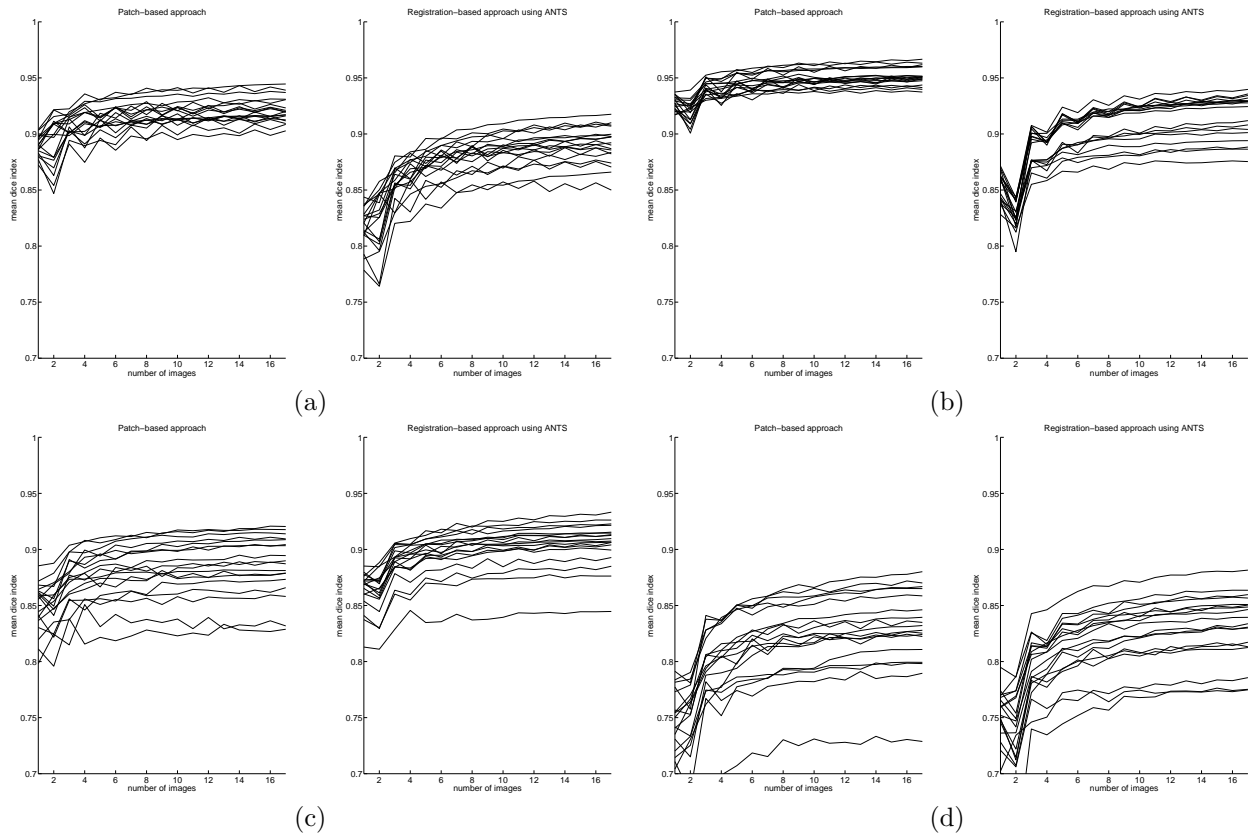


Figure 4: Mean dice index with respect to the number of images used for label propagation Left: patch-based approach, right: registration-based approach using ANTS (using, for each case, majority voting to fuse labels). (a) : left white matter, (b): left cortex, (c): left thalamus, (d): left hippocampus.

3.5 Aggregation strategy

For the label fusion step, the majority voting rule is compared to the simultaneous truth and performance level estimation (STAPLE) [36]. STAPLE estimates the performance of each classifier iteratively and weights it accordingly, relying on an expectation-maximization (EM) optimization approach. We used the implementation provided by T. Rohlfing⁹. As already reported in [3], STAPLE based fusion rule does not necessary lead to higher Dice coefficients compared to the majority voting rule.

3.6 Comparison with previously reported results

The results obtained on the IBSR dataset are reported in Table 5 (highest Dice coefficients are in bold). In this table are also shown results reported in the literature, including non-rigid registration-based approaches and segmentation techniques. This experiment clearly shows that the proposed patch-based framework is very competitive with respect to recently published methods.

3.7 Computational Time

Table 7 shows the comparison of the computational time for a non-rigid registration-based approach (using ANTs [5]) and the proposed patch-based techniques. We choose ANTs as it has been ranked as one of the best among fourteen non-rigid registration algorithms [22]. The use of pointwise or multipoint estimator leads to higher computational time than using non-rigid registration. However, the fast multipoint approach provides a very time efficient algorithm. Moreover, the multi-threading of the proposed label propagation method is particularly adapted due its blockwise nature. As eight processors were used for our experiments, the computational time with multi-threading is about 8 times smaller. In this case, the propagation of one label image with the fast multipoint estimator can be performed in about 1mn.

3.8 Visual Evaluation

Figures 5 and 6 show visual comparison between the proposed patch-based approach with a non-rigid registration-based approach (ANTs + majority voting). The first row of this figure corresponds to the mean image obtained using the T1-weighted MR images of the anatomy textbook. Such mean image is one way of visual evaluation used for non-rigid registration algorithms. The proposed patch-based method can also provided such mean image by simply using intensities (instead of labels) of the images contained in the textbook. The second row corresponds to obtained brain segmentation. This figure illustrates well the fact that the proposed technique allows one to get a better delineation of brain structures (and especially the cortex for instance).

3.9 Combining patch-based strategies and registration-based techniques

The main purpose of the proposed patch-based method is to avoid long time computation due to non-rigid registration. However, one can wonder if the two techniques should cooperate to reach better results. To one side, one of the advantages of the patch-based approach is to possibly consider multiple examples within the same image of the anatomy textbook. The consequence of the local search windows use is to relax the one-to-one constraint usually involved in non-rigid registration-based label propagation approaches. Experiments on the IBSR dataset have clearly shown that the use of the patch-based approach leads to very satisfactory segmentation results for brain structures with sharp contrast. On the other side, the use of one-to-one constraint leads to a topologically regularized segmentation. We have experimentally observed that this aspect is important for cortex parcellation. Indeed, the delineation of cortical areas does not rely on intensity contrast within the cortex. It appears then that the intensity similarity assumption which the patch-based approach relies on is not sufficient to provide the highest quality cortical parcellation.

⁹<http://www.stanford.edu/~rohl fing/software/index.html>

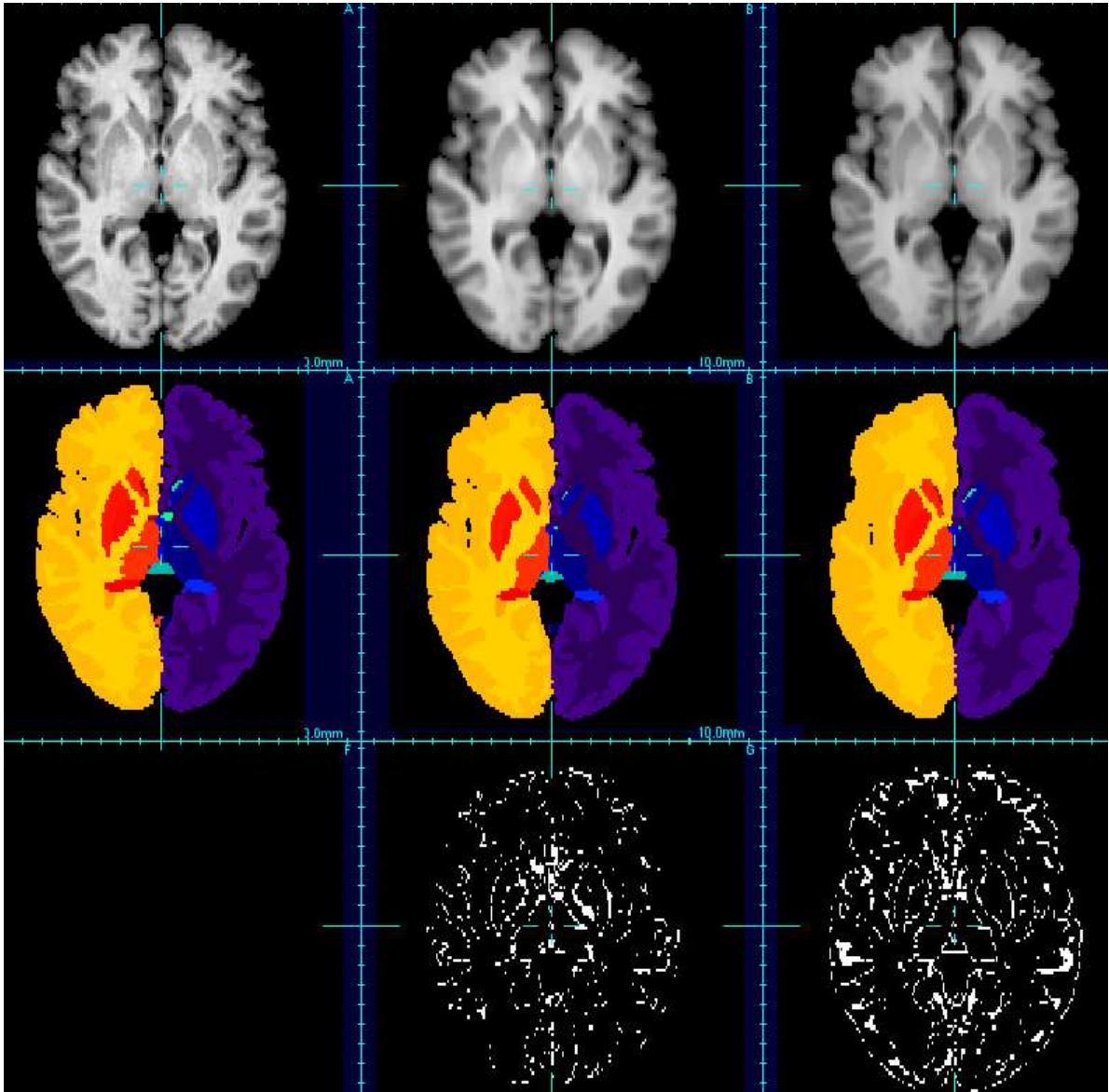


Figure 5: Visual evaluation of brain segmentation results (IBSR, image #7). First row: T1-weighted images, second row: corresponding segmentation, third row: misclassified voxels. Left: ground-truth, middle: patch-based technique, right: non-rigid registration-based approach (using ANTs).

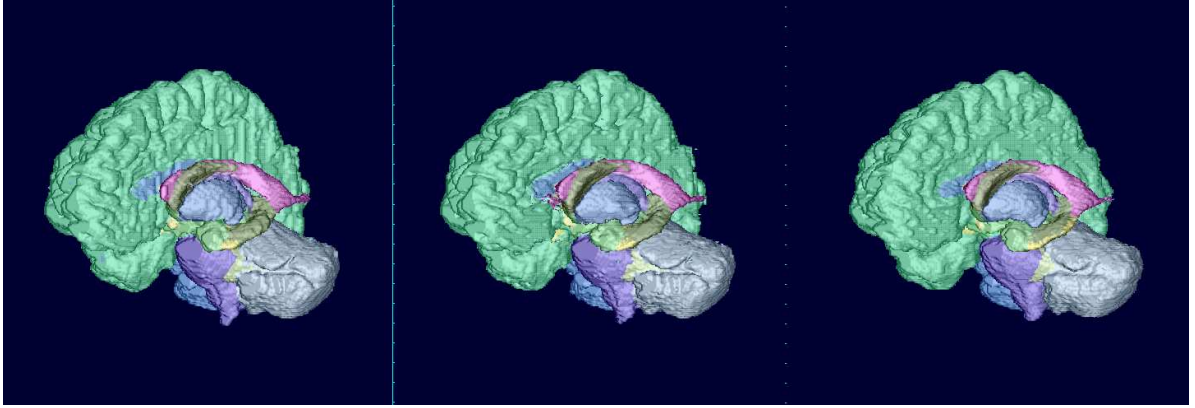


Figure 6: Surface rendering of segmentation results (IBSR, image #7). Left: ground-truth, middle: patch-based technique, right: non-rigid registration-based approach (using ANTs).

Table 8 shows the mean Dice index obtained on the NAO-NIREP dataset using the multipoint pairwise patch-based approach, a non-rigid registration-based technique using ANTs (using majority voting or STAPLE as fusion rule) and the mixed approach. This latter technique simply consists in fusing the labeling obtained with a non-rigid registration algorithm, with the hard cortical mask obtained with a patch-based technique. The mixed approach clearly leads to higher Dice index compared to the two other techniques (which provide similar overlap scores separately). It is also interesting to note that in the combining approach, STAPLE seems to provide a better label fusion framework than the majority voting rule. Figure 7 presents a cortex parcellation result using the combined approach with STAPLE.

4 Discussion

Label propagation is a versatile image segmentation technique which can be applied to a large variety of images. In this work, we focus on the development of a new label propagation framework applied to automated human brain labeling. A patch-based label propagation has been proposed whose purpose is to relax the one-to-one constraint existing in non-rigid registration-based techniques. Indeed, this image similarity-based approach can be seen as a one-to-many block matching technique. It allows the use of several good candidates (*i.e.* the most similar patches) to estimate label patches. Several patch-based algorithms have been derived depending on the patch aggregation strategies and the label fusion rules (if needed). Comparison with previously proposed non-rigid registration-based methods on publicly available in-vivo MR brain images has shown a great potential of this method.

The proposed approach is also related to supervised learning methods which try to learn the link between image intensity patches and label patches. In our work, instead of learning the link between image intensity and labels, the label estimation relies on a weighted graph which represents directly the similarity between intensity of the input image and intensity of the anatomy textbook images. As mentioned in the Introduction, the work presented in this article shares similarities with the one done by Coupé *et al.* [10]. Indeed, they have proposed a patch-based strategy relying on a group-wise technique for label fusion. The main methodological differences concern the smoothing parameter in the definition of patch weights and the voxel pre-selection. About validation, they have investigated one aggregation strategy (group-wise) on two regions of interest: hippocampus and lateral ventricles (on a non-publicly available dataset).

The experiments on three freely available T1-weighted MR image datasets (Brainweb, IBSR, NAO-NIREP) have shown that the proposed framework can lead to high quality segmentation of a large number of brain structures. It has been shown that CPU time consuming non-rigid registration steps can be avoided. The proposed patch-based technique does not require an accurate correspondence between the input image

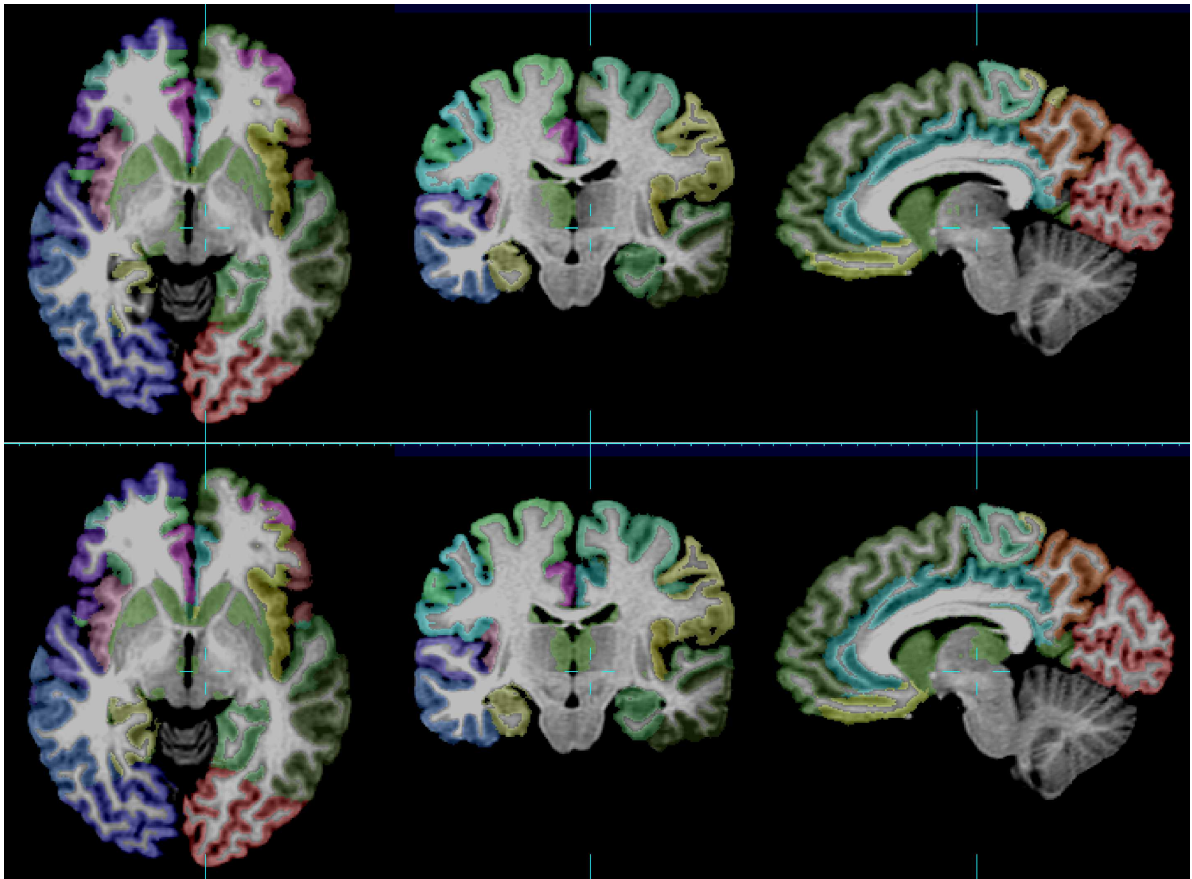


Figure 7: Visual evaluation of cortex parcellation results (NA0-NIREP, image #6). First row: ground truth, second row: combination technique (ANTs + patch + STAPLE).

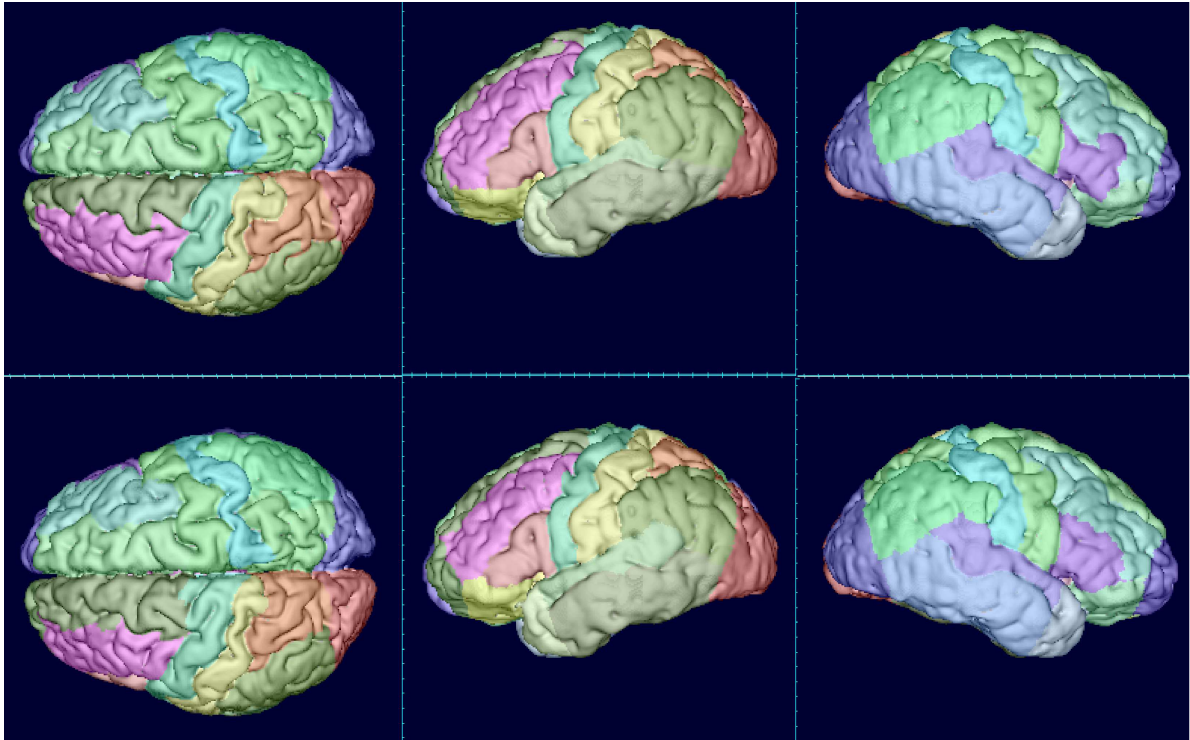


Figure 8: Surface rendering of cortex parcellation results (NA0-NIREP, image #6). First row: ground truth, second row: combination technique (ANTs + patch + STAPLE).

and the anatomy textbook. The usual assumption of one-to-one mapping is relaxed by using local search windows. Further work may investigate the use of more complex similar patch search steps. For instance, using a rotation-invariant patch match is one of the possible extensions of the proposed method (in particular when considering large patches), while keeping in mind that it might increase the CPU time. However, the results obtained with translation-based patch match are already very competitive with respect to existing methods. Some possible methodological extensions are local adaptivity of patch size and search window size, choice of kernel function f and the use of graph operators.

As the algorithm relies on the assumption of similar intensities between the input image and the anatomy textbook, the segmentation accuracy may depend on the contrast of the structure to segment or intensity variations (such as intensity bias). Intensity variations might disturb the search process of similar patches, which would lead to a final labeling obtained with less relevant patch examples. The sensitivity of intensity variations can be captured by the smoothing factor β . However, using a fixed global value of β in conjunction with common correction techniques such as intensity bias correction and histogram matching, the proposed algorithm provides very satisfactory results. Also for the IBSR data, it can be noticed that the segmentation of the pallidum is not as good as the segmentation of other brain structures. The reason for failure is that this structure has no clear contrasted boundaries in MR T1-weighted images. Thus, contrary to registration-based techniques which can propagate spatial relations of the structures, a pure intensity-based method such as the one proposed in this work cannot provide a satisfactory delineation of the pallidum. This assertion is corroborated by the experiments performed on the NA0-NIREP dataset. Because of the key assumption of intensity similarity, a non-rigid registration-based label propagation approach outperforms the proposed patch-based method. In this context of cortex parcellation, we have shown that these two techniques can be complementary. In particular, the joint use of non-rigid registration and patch-based strategy can significantly improve the segmentation result. Thus, further work may explore the incorporation of a regularization term and prior information such as spatial relations or a topological atlas into the patch-based framework.

Contrary to registration-based techniques, especially for the group-wise approach using a patch selection, if there is no correspondence between the input and the anatomy textbook (*i.e.* the image similarity is lower than the threshold used in the patch selection step), no label is propagated. This prevents an incorrect label being introduced in the estimation of the final label image. Thus, further work is to evaluate the robustness of this label propagation method to images which contain lesions or tumors, and to use this patch-based technique to potentially detect these pathological patterns. Another research direction concerns the use of a database of images as IBSR to segment pathological images. Such a "cross-site" study will analyze the robustness of the method in the presence of large anatomical variability, including the initial affine registration, intensity and shape variations, and the research of examples in the anatomy textbook.

Finally, as mentioned in other works [37, 2, 35] and as shown in the experiments performed on the Brainweb dataset, the label image can be estimated using a subset of the anatomy textbook (the observed convergence rate is about 10 example images). This remark leads to the issue of atlas selection, *i.e.* identifying a subset of representative examples in the textbook with respect to the input image. This can be particularly important when using a large anatomy textbook to avoid useless heavy computational burden.

Acknowledgment

The research leading to these results has received funding from the European Research Council under the European Communitys Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 207667). This work is also funded by NIH/NINDS Grant R01 NS 055064.

References

- [1] A. Akselrod-Ballin, M. Galun, J. M. Gomori, A. Brandt, and R. Basri. Prior knowledge driven multiscale segmentation of brain MRI. *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, 10(Pt 2):118–126, 2007.

| brain structure | pair-wise | | | group-wise | | |
|-----------------|-------------|-------------|-----------------|-------------|------------|-----------------|
| | pointwise | multipoint | fast multipoint | pointwise | multipoint | fast multipoint |
| CSF | 0.87 | 0.88 | 0.87 | 0.87 | 0.86 | 0.86 |
| Gray Matter | 0.95 | 0.95 | 0.94 | 0.95 | 0.93 | 0.94 |
| White Matter | 0.95 | 0.95 | 0.94 | 0.95 | 0.94 | 0.94 |
| Fat | 0.83 | 0.87 | 0.85 | 0.83 | 0.85 | 0.84 |
| Muscle | 0.66 | 0.70 | 0.67 | 0.66 | 0.67 | 0.66 |
| Muscle/Skin | 0.93 | 0.94 | 0.93 | 0.93 | 0.93 | 0.93 |
| Skull | 0.92 | 0.93 | 0.91 | 0.92 | 0.92 | 0.92 |
| vessels | 0.61 | 0.59 | 0.56 | 0.61 | 0.56 | 0.57 |
| around fat | 0.37 | 0.36 | 0.30 | 0.37 | 0.34 | 0.36 |
| dura matter | 0.51 | 0.53 | 0.47 | 0.51 | 0.49 | 0.50 |
| bone marrow | 0.81 | 0.83 | 0.79 | 0.81 | 0.81 | 0.80 |

Table 1: Comparison (using the mean Dice index) of several possible combinations of patch-based approaches (using optimal global parameter setting) on the Brainweb dataset. For the pair-wise results, labels have been fused using the majority voting rule. Highest overlap rates are shown in bold.

- [2] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726–738, July 2009.
- [3] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano. Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Transactions on Medical Imaging*, 28(8):1266–1277, August 2009.
- [4] B. Aubert-Broche, A. C. Evans, and L. Collins. A new improved version of the realistic digital brain phantom. *Neuroimage*, 32(1):138145, August 2006.
- [5] B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, February 2008.
- [6] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, January 2005.
- [7] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans. Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3):190–208, 1995.
- [8] D. L. Collins and Jens C. Pruessner. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *NeuroImage*, 52(4):1355–1366, October 2010.
- [9] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, April 2008.
- [10] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954, January 2011.
- [11] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2 edition, October 2000.

| brain structure | pair-wise | | | group-wise | | |
|-------------------|-------------|--------------------|-----------------|--------------------|--------------------|--------------------|
| | pointwise | multipoint | fast multipoint | pointwise | multipoint | fast multipoint |
| Cerebral WM | 0.92 - 0.92 | 0.92 - 0.92 | 0.92 - 0.91 | 0.92 - 0.91 | 0.93 - 0.92 | 0.92 - 0.92 |
| Cerebral Cortex | 0.94 - 0.93 | 0.95 - 0.95 | 0.94 - 0.94 | 0.93 - 0.93 | 0.95 - 0.95 | 0.94 - 0.94 |
| Lateral Ventricle | 0.92 - 0.91 | 0.93 - 0.92 | 0.92 - 0.91 | 0.90 - 0.89 | 0.93 - 0.92 | 0.91 - 0.91 |
| Inferior Lat Vent | 0.58 - 0.53 | 0.61 - 0.56 | 0.58 - 0.55 | 0.50 - 0.49 | 0.60 - 0.56 | 0.56 - 0.54 |
| Cerebellum WM | 0.87 - 0.87 | 0.88 - 0.88 | 0.88 - 0.87 | 0.87 - 0.87 | 0.88 - 0.88 | 0.88 - 0.88 |
| Cerebellum Cortex | 0.93 - 0.94 | 0.95 - 0.95 | 0.94 - 0.94 | 0.93 - 0.93 | 0.95 - 0.95 | 0.94 - 0.94 |
| Thalamus | 0.88 - 0.87 | 0.89 - 0.88 | 0.88 - 0.88 | 0.89 - 0.89 | 0.89 - 0.89 | 0.89 - 0.89 |
| Caudate | 0.87 - 0.87 | 0.87 - 0.87 | 0.86 - 0.86 | 0.87 - 0.87 | 0.88 - 0.89 | 0.88 - 0.88 |
| Putamen | 0.87 - 0.86 | 0.88 - 0.87 | 0.87 - 0.87 | 0.88 - 0.88 | 0.89 - 0.89 | 0.88 - 0.88 |
| Pallidum | 0.69 - 0.66 | 0.71 - 0.69 | 0.66 - 0.63 | 0.77 - 0.77 | 0.79 - 0.79 | 0.78 - 0.79 |
| 3rd Ventricle | 0.79 | 0.81 | 0.79 | 0.74 | 0.80 | 0.77 |
| 4th Ventricle | 0.83 | 0.85 | 0.84 | 0.80 | 0.84 | 0.82 |
| Brain Stem | 0.92 | 0.94 | 0.93 | 0.92 | 0.93 | 0.92 |
| Hippocampus | 0.80 - 0.81 | 0.83 - 0.83 | 0.81 - 0.81 | 0.79 - 0.79 | 0.83 - 0.83 | 0.81 - 0.81 |
| Amygdala | 0.73 - 0.71 | 0.76 - 0.74 | 0.73 - 0.70 | 0.71 - 0.70 | 0.75 - 0.75 | 0.73 - 0.72 |
| CSF | 0.66 | 0.69 | 0.65 | 0.64 | 0.68 | 0.66 |
| Accumbens area | 0.63 - 0.61 | 0.66 - 0.63 | 0.58 - 0.56 | 0.66 - 0.65 | 0.68 - 0.66 | 0.67 - 0.65 |
| Ventral DC | 0.79 - 0.79 | 0.82 - 0.82 | 0.79 - 0.78 | 0.79 - 0.80 | 0.82 - 0.82 | 0.80 - 0.81 |

Table 2: Comparison (using the mean Dice index) of several possible combination of patch-based approaches (using optimal global parameter setting) on the IBSR dataset. When structures are separated into left/right, two overlap scores are reported. For the pair-wise results, labels have been fused using the majority voting rule. Highest overlap rates are shown in bold.

- [12] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, January 2002.
- [13] T. Gasser, L. Sroka, and C. Jennen-Steimetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73(3):625–633, December 1986.
- [14] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, January 2008.
- [15] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, October 2006.
- [16] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken. Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. *IEEE Transactions on Medical Imaging*, 28(7):1000–1010, July 2009.
- [17] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From local kernel to nonlocal Multiple-Model image denoising. *Int. J. Comput. Vision*, 86(1):1–32, 2010.
- [18] A. R. Khan, M. K. Chung, and M. F. Beg. Robust atlas-based brain segmentation using multi-structure confidence-weighted registration. *Medical Image Computing and Computer-Assisted Intervention: MIC-CAI*, 12(Pt 2):549–557, 2009.
- [19] A. R. Khan, N. Cherbuin, W. Wen, K.J.nstey, P. Sachdev, and M. F. Beg. Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): Validation on hippocampus segmentation. *Neuroimage*, 56(1):126–139, 2011.

| brain structure (left/right) | pair-wise patch + MV | pair-wise patch + STAPLE | group-wise patch |
|------------------------------|----------------------|--------------------------|--------------------|
| Cerebral WM | 0.92 - 0.92 | 0.92 - 0.92 | 0.93 - 0.92 |
| Cerebral Cortex | 0.95 - 0.95 | 0.95 - 0.95 | 0.95 - 0.95 |
| Lateral Ventricle | 0.93 - 0.92 | 0.92 - 0.91 | 0.93 - 0.92 |
| Inferior Lat Vent | 0.61 - 0.56 | 0.53 - 0.51 | 0.60 - 0.56 |
| Cerebellum WM | 0.88 - 0.88 | 0.88 - 0.88 | 0.88 - 0.88 |
| Cerebellum Cortex | 0.95 - 0.95 | 0.95 - 0.95 | 0.95 - 0.95 |
| Thalamus | 0.89 - 0.89 | 0.88 - 0.88 | 0.89 - 0.89 |
| Caudate | 0.87 - 0.87 | 0.86 - 0.86 | 0.88 - 0.89 |
| Putamen | 0.88 - 0.87 | 0.86 - 0.85 | 0.89 - 0.89 |
| Pallidum | 0.71 - 0.69 | 0.70 - 0.68 | 0.79 - 0.79 |
| 3rd Ventricle | 0.81 | 0.80 | 0.80 |
| 4th Ventricle | 0.85 | 0.78 | 0.84 |
| Brain Stem | 0.94 | 0.94 | 0.93 |
| Hippocampus | 0.83 - 0.83 | 0.79 - 0.80 | 0.83 - 0.83 |
| Amygdala | 0.76 - 0.74 | 0.72 - 0.70 | 0.75 - 0.75 |
| CSF | 0.69 | 0.66 | 0.68 |
| Accumbens area | 0.66 - 0.63 | 0.65 - 0.63 | 0.68 - 0.66 |
| Ventral DC | 0.82 - 0.82 | 0.83 - 0.83 | 0.82 - 0.82 |

Table 3: Comparison (using the mean Dice index) of aggregation strategies (majority voting *vs* STAPLE) on the IBSR dataset, using a multipoint estimation approach. When structures are separated into left/right, two overlap scores are reported. Highest overlap rates are shown in bold.

- [20] S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling & Simulation*, 4(4):1091–1115, January 2005.
- [21] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.
- [22] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. -C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, July 2009.
- [23] A. Klein and J. Hirsch. Mindboggle: a scatterbrained approach to automate brain labeling. *NeuroImage*, 24(2):261–280, January 2005.
- [24] J.L. Lancaster, L.H. Rainey, J.L. Summerlin, C.S. Freitas, P.T. Fox, A.C. Evans, A.W. Toga, and J.C. Mazziotta. Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping*, 5(4):238–242, 1997.
- [25] T. Langerak, U. van der Heide, A. Kotte, M. Viergever, M. van Vulpen, and J. Pluim. Label fusion in Atlas-Based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Transactions on Medical Imaging*, July 2010.
- [26] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3):2352–2365, February 2010.
- [27] M. Mignotte. A non-local regularization strategy for image deconvolution. *Pattern Recogn. Lett.*, 29(16):2206–2212, 2008.
- [28] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander. Mathematical textbook of deformable neuroanatomies. *Proceedings of the National Academy of Sciences of the United States of America*, 90(24):11944–11948, December 1993.

| brain structure (left/right) | pair-wise patch + MV | pair-wise patch + STAPLE | group-wise patch |
|------------------------------|----------------------|--------------------------|--------------------|
| Occipital Lobe | 0.79 - 0.81 | 0.75 - 0.74 | 0.78 - 0.80 |
| Cingulate Gyrus | 0.77 - 0.80 | 0.77 - 0.78 | 0.76 - 0.79 |
| Insula Gyrus | 0.85 - 0.88 | 0.84 - 0.87 | 0.84 - 0.87 |
| Temporal Pole | 0.83 - 0.87 | 0.81 - 0.86 | 0.82 - 0.86 |
| Superior Temporal Gyrus | 0.75 - 0.74 | 0.75 - 0.74 | 0.74 - 0.73 |
| Infero Temporal Region | 0.84 - 0.84 | 0.82 - 0.81 | 0.83 - 0.83 |
| Parahippocampal Gyrus | 0.82 - 0.84 | 0.80 - 0.83 | 0.81 - 0.83 |
| Frontal Pole | 0.82 - 0.82 | 0.81 - 0.81 | 0.81 - 0.80 |
| Superior Frontal Gyrus | 0.80 - 0.81 | 0.79 - 0.80 | 0.80 - 0.80 |
| Middle Frontal Gyrus | 0.78 - 0.77 | 0.78 - 0.76 | 0.78 - 0.76 |
| Inferior Gyrus | 0.70 - 0.75 | 0.70 - 0.74 | 0.69 - 0.74 |
| Orbital Frontal Gyrus | 0.84 - 0.84 | 0.83 - 0.83 | 0.83 - 0.83 |
| Precentral Gyrus | 0.76 - 0.74 | 0.74 - 0.73 | 0.75 - 0.73 |
| Superior Parietal Lobule | 0.75 - 0.73 | 0.74 - 0.73 | 0.74 - 0.72 |
| Inferior Parietal Lobule | 0.78 - 0.74 | 0.77 - 0.74 | 0.77 - 0.73 |
| Postcentral Gyrus | 0.70 - 0.65 | 0.64 - 0.59 | 0.69 - 0.63 |

Table 4: Comparison (using the mean Dice index) of aggregation strategies (majority voting *vs* STAPLE) on the NA0-NIREP dataset, using a multipoint estimation approach. When structures are separated into left/right, two overlap scores are reported. Highest overlap rates are shown in bold.

- [29] G. Peyré, S. Bogleux, and L. Cohen. Non-local regularization of inverse problems. In *Proceedings of the 10th European Conference on Computer Vision: Part III*, pages 57–68, Marseille, France, 2008. Springer-Verlag.
- [30] T. Rohlfing, D. B Russakoff, and C. R Maurer. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Transactions on Medical Imaging*, 23(8):983–994, August 2004.
- [31] F. Rousseau. A non-local approach for image super-resolution using intermodality priors. *Medical Image Analysis*, 14(4):594–605, August 2010.
- [32] S. Roy, A. Carass, N. Shiee, D.L. Pham, and J. Prince. MR contrast synthesis for lesion segmentation. *Proc IEEE Int Symp Biomed Imaging*, 932-935, June 2010.
- [33] M. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging*, June 2010.
- [34] M. Sdika. Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote. *Medical Image Analysis*, 14(2):219–226, April 2010.
- [35] E. M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. A Viergever, J. P. W. Pluim, and B. van Ginneken. Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus. *Medical Image Analysis*, 14(1):39–49, February 2010.
- [36] S. K Warfield, K. H Zou, and W. M Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, July 2004.
- [37] M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein. Optimum template selection for atlas-based segmentation. *NeuroImage*, 34(4):1612–1618, February 2007.
- [38] P. A. Yushkevich, H. Wang, J. Pluta, S. R. Das, C. Craige, B. B. Avants, M. W. Weiner, and S. Mueller. Nearly automatic segmentation of hippocampal subfields in in vivo focal t2-weighted MRI. *NeuroImage*, 53(4):1208–1224, December 2010.

| brain structure (left/right when available) | Patch-based | LWV-MSD 2009 [3] | Fischl et al. 2002 [12] | Khan et al. 2009 [18] | MA-IDMIN+ 2010 [26] | Sdika 2010 [34] | Akselrod 2007 [1] |
|--|--------------------|---------------------|----------------------------|--------------------------|------------------------|--------------------|----------------------|
| Cerebral WM | 0.93 - 0.92 | 0.78 - 0.78 | - | - | - | 0.91 | 0.87 |
| Cerebral Cortex | 0.95 - 0.95 | 0.81 - 0.81 | - | - | - | 0.94 | 0.86 |
| Lateral Ventricle | 0.93 - 0.92 | 0.83 - 0.82 | 0.78 | 0.85 | - | 0.90 | - |
| Inferior Lat Vent | 0.60 - 0.56 | 0.22 - 0.22 | - | - | - | 0.44 | - |
| Cerebellum WM | 0.88 - 0.88 | 0.79 - 0.79 | - | - | - | 0.86 | - |
| Cerebellum Cortex | 0.95 - 0.95 | 0.86 - 0.86 | - | - | - | 0.95 | - |
| Thalamus | 0.89 - 0.89 | 0.87 - 0.88 | 0.86 | 0.89 | 0.89 | 0.89 | 0.84 |
| Caudate | 0.88 - 0.89 | 0.83 - 0.83 | 0.82 | 0.83 | 0.85 | 0.86 | 0.80 |
| Putamen | 0.89 - 0.89 | 0.86 - 0.86 | 0.81 | 0.87 | 0.90 | 0.88 | 0.79 |
| Pallidum | 0.79 - 0.79 | 0.78 - 0.79 | 0.71 | 0.72 | 0.83 | 0.79 | 0.74 |
| 3rd Ventricle | 0.80 | 0.74 | - | - | - | 0.77 | - |
| 4th Ventricle | 0.84 | 0.77 | - | - | - | 0.81 | - |
| Brain Stem | 0.93 | 0.91 | - | - | - | 0.93 | 0.84 |
| Hippocampus | 0.83 - 0.83 | 0.74 - 0.76 | 0.75 | 0.76 | 0.80 | 0.81 | 0.69 |
| Amygdala | 0.75 - 0.75 | 0.72 - 0.72 | 0.68 | 0.66 | 0.75 | 0.74 | 0.63 |
| CSF | 0.68 | 0.61 | - | - | - | 0.70 | 0.83 |
| Accumbens area | 0.68 - 0.66 | 0.67 - 0.67 | 0.58 | 0.61 | - | 0.72 | - |
| Ventral DC | 0.82 - 0.82 | 0.82 - 0.82 | - | - | - | 0.85 | 0.76 |

Table 5: Comparison (using the mean Dice index) of the group-wise multipoint patch-based approach with other segmentation methods. “Patch-based” denotes multipoint group-wise patch-based approach. Highest Dice coefficients are in bold.

| brain structure (left/right) | pair-wise | | | group-wise | | |
|---------------------------------|-------------|--------------------|-----------------|-------------|--------------------|-----------------|
| | pointwise | multipoint | fast multipoint | pointwise | multipoint | fast multipoint |
| Occipital Lobe | 0.74 - 0.77 | 0.79 - 0.81 | 0.76 - 0.78 | 0.74 - 0.77 | 0.78 - 0.80 | 0.76 - 0.78 |
| Cingulate Gyrus | 0.73 - 0.76 | 0.77 - 0.80 | 0.76 - 0.77 | 0.73 - 0.76 | 0.76 - 0.79 | 0.75 - 0.77 |
| Insula Gyrus | 0.82 - 0.85 | 0.85 - 0.88 | 0.83 - 0.86 | 0.82 - 0.85 | 0.84 - 0.87 | 0.83 - 0.86 |
| Temporal Pole | 0.80 - 0.84 | 0.83 - 0.87 | 0.81 - 0.85 | 0.80 - 0.84 | 0.82 - 0.86 | 0.81 - 0.85 |
| Superior Temporal Gyrus | 0.71 - 0.70 | 0.75 - 0.74 | 0.73 - 0.71 | 0.71 - 0.70 | 0.74 - 0.73 | 0.73 - 0.71 |
| Infero Temporal Region | 0.80 - 0.81 | 0.84 - 0.84 | 0.81 - 0.82 | 0.80 - 0.81 | 0.83 - 0.83 | 0.82 - 0.82 |
| Parahippocampal Gyrus | 0.79 - 0.81 | 0.82 - 0.84 | 0.80 - 0.82 | 0.79 - 0.81 | 0.81 - 0.83 | 0.80 - 0.82 |
| Frontal Pole | 0.78 - 0.77 | 0.82 - 0.82 | 0.80 - 0.79 | 0.78 - 0.77 | 0.81 - 0.80 | 0.79 - 0.78 |
| Superior Frontal Gyrus | 0.77 - 0.77 | 0.80 - 0.81 | 0.78 - 0.78 | 0.77 - 0.77 | 0.80 - 0.80 | 0.79 - 0.79 |
| Middle Frontal Gyrus | 0.75 - 0.73 | 0.78 - 0.77 | 0.76 - 0.74 | 0.75 - 0.73 | 0.78 - 0.76 | 0.76 - 0.75 |
| Inferior Gyrus | 0.65 - 0.71 | 0.70 - 0.75 | 0.67 - 0.72 | 0.65 - 0.71 | 0.69 - 0.74 | 0.67 - 0.72 |
| Orbital Frontal Gyrus | 0.80 - 0.80 | 0.84 - 0.84 | 0.81 - 0.81 | 0.80 - 0.80 | 0.83 - 0.83 | 0.81 - 0.81 |
| Precentral Gyrus | 0.72 - 0.69 | 0.76 - 0.74 | 0.73 - 0.70 | 0.72 - 0.69 | 0.75 - 0.73 | 0.73 - 0.71 |
| Superior Parietal Lobule | 0.71 - 0.69 | 0.75 - 0.73 | 0.73 - 0.70 | 0.71 - 0.69 | 0.74 - 0.72 | 0.72 - 0.70 |
| Inferior Parietal Lobule | 0.74 - 0.70 | 0.78 - 0.74 | 0.75 - 0.71 | 0.74 - 0.70 | 0.77 - 0.73 | 0.75 - 0.72 |
| Postcentral Gyrus | 0.65 - 0.60 | 0.70 - 0.65 | 0.65 - 0.60 | 0.65 - 0.60 | 0.69 - 0.63 | 0.67 - 0.61 |

Table 6: Comparison (using the mean Dice index) of several possible combination of patch-based approaches (using optimal global parameter setting) on the NA0-NIREP dataset. When structures are separated into left/right, two overlap scores are reported. For the pair-wise results, labels have been fused using the majority voting rule. Highest overlap rates are shown in bold.

| data | method | 1 core | | 8 cores | |
|----------|---|----------------|------------|----------------|------------|
| | | time per image | total time | time per image | total time |
| Brainweb | non-rigid registration (using ANTs [5]) | 50 | 950 | - | - |
| | pointwise / multipoint | 67 | 1200 | 9 | 170 |
| | fast multipoint | 10 | 185 | 1.3 | 25 |
| IBSR | non-rigid registration (using ANTs [5]) | 40 | 680 | - | - |
| | pointwise / multipoint | 60 | 1000 | 8 | 130 |
| | fast multipoint | 10 | 165 | 1.2 | 22 |
| NIREP | non-rigid registration (using ANTs [5]) | 78 | 1170 | - | - |
| | pointwise / multipoint | 58 | 840 | 7.5 | 110 |
| | fast multipoint | 10 | 145 | 1.3 | 19 |

Table 7: Approximate mean computational time (in minutes) using an Intel Xeon E5420 2.5GHz. The time per image corresponds to the label propagation using one example. Total time stands for the image segmentation process using the entire textbook (the Brainweb textbook contains 19 images with 11 labels, the IBSR textbook contains 17 images with 32 labels and the NA0-NIREP textbook contains 15 with 33 labels).

| brain structure (left/right when available) | pair-wise patch + MV | ANTs + MV | ANTs + STAPLE | Combination (MV) | Combination (STAPLE) |
|--|-------------------------|--------------|------------------|---------------------|-------------------------|
| Occipital Lobe | 0.79 - 0.81 | 0.80 - 0.81 | 0.70 - 0.72 | 0.81 - 0.83 | 0.83 - 0.86 |
| Cingulate Gyrus | 0.77 - 0.80 | 0.81 - 0.81 | 0.78 - 0.78 | 0.81 - 0.82 | 0.82 - 0.84 |
| Insula Gyrus | 0.85 - 0.88 | 0.85 - 0.87 | 0.82 - 0.84 | 0.86 - 0.88 | 0.86 - 0.89 |
| Temporal Pole | 0.83 - 0.87 | 0.83 - 0.85 | 0.79 - 0.82 | 0.84 - 0.86 | 0.84 - 0.87 |
| Superior Temporal Gyrus | 0.75 - 0.74 | 0.78 - 0.77 | 0.73 - 0.73 | 0.79 - 0.78 | 0.80 - 0.79 |
| Infero Temporal Region | 0.84 - 0.84 | 0.84 - 0.84 | 0.79 - 0.79 | 0.86 - 0.85 | 0.86 - 0.86 |
| Parahippocampal Gyrus | 0.82 - 0.84 | 0.83 - 0.85 | 0.80 - 0.82 | 0.84 - 0.86 | 0.84 - 0.86 |
| Frontal Pole | 0.82 - 0.82 | 0.82 - 0.80 | 0.77 - 0.75 | 0.84 - 0.82 | 0.85 - 0.84 |
| Superior Frontal Gyrus | 0.80 - 0.81 | 0.80 - 0.80 | 0.75 - 0.75 | 0.82 - 0.81 | 0.82 - 0.83 |
| Middle Frontal Gyrus | 0.78 - 0.77 | 0.79 - 0.76 | 0.73 - 0.68 | 0.81 - 0.78 | 0.82 - 0.80 |
| Inferior Gyrus | 0.70 - 0.75 | 0.75 - 0.75 | 0.70 - 0.69 | 0.76 - 0.76 | 0.78 - 0.78 |
| Orbital Frontal Gyrus | 0.84 - 0.84 | 0.84 - 0.83 | 0.81 - 0.80 | 0.85 - 0.84 | 0.86 - 0.85 |
| Precentral Gyrus | 0.76 - 0.74 | 0.77 - 0.75 | 0.71 - 0.69 | 0.78 - 0.76 | 0.80 - 0.79 |
| Superior Parietal Lobule | 0.75 - 0.73 | 0.75 - 0.75 | 0.65 - 0.64 | 0.77 - 0.76 | 0.80 - 0.80 |
| Inferior Parietal Lobule | 0.78 - 0.74 | 0.76 - 0.75 | 0.69 - 0.67 | 0.78 - 0.77 | 0.80 - 0.79 |
| Postcentral Gyrus | 0.70 - 0.65 | 0.72 - 0.70 | 0.67 - 0.62 | 0.73 - 0.70 | 0.76 - 0.74 |

Table 8: Complementarity between patch-based strategies and registration-based techniques on the NA0-NIREP dataset, using a multipoint estimation approach. When structures are separated into left/right, two overlap scores are reported. Highest overlap rates are shown in bold.