Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation.

Pierrick Coupé, José V. Manjón, Vladimir Fonov, Jens C. Pruessner, Montserrat Robles, D Louis Collins

Patch-based Segmentation using Expert Priors: Application to Hippocampus and Ventricle Segmentation

Pierrick Coupé[1], José V. Manjón[2], Vladimir Fonov[1], Jens Pruessner[1,3], Montserrat Robles[2], D. Louis Collins[1]

[1] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada University, 3801 University Street, Montreal, Canada H3A 2B4
[2] Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain
[3] Douglas Hospital Research Center, Department of Psychology, McGill University, Montreal, Canada

Abstract

Quantitative magnetic resonance analysis often requires accurate, robust, and reliable automatic extraction of anatomical structures. Recently, template-warping methods incorporating a label fusion strategy have demonstrated high accuracy in segmenting cerebral structures. In this study, we propose a novel patch-based method using expert manual segmentations as priors to achieve this task. Inspired by recent work in image denoising, the proposed nonlocal patch-based label fusion produces accurate and robust segmentation. Validation with two different datasets is presented. In our experiments, the hippocampi of 80 healthy subjects and the lateral ventricles of 80 patients with Alzheimer's disease were segmented. The influence on segmentation accuracy of different parameters such as patch size and number of training subjects was also studied. A comparison with an appearance-based method and a template-based method was also carried out. The highest median kappa index values obtained with the proposed method were 0.884 for hippocampus segmentation and 0.959 for lateral ventricle segmentation.

Keywords: MRI, brain, hippocampus, lateral ventricles, Alzheimer's disease, image processing, structure segmentation.

1. Introduction

Magnetic resonance (MR) imaging plays a crucial role in the detection of pathology, the study of brain organization, and clinical research. Every day, a vast amount of data is produced in clinical settings, preventing the use of manual approaches for data analysis. Consequently, the development of accurate, robust, and reliable segmentation techniques for the automatic extraction of anatomical structures is becoming an important challenge in quantitative MR analysis. In contrast to brain tissue classification where the intensity of the MR signal can be used to segment different tissue types, anatomical segmentation usually requires information derived from the manual segmentations done by experts (i.e., expert priors), since anatomical structures can be composed of several tissue types and distinct anatomical structures can have the same MR signal properties. To overcome this difficulty, several automatic methods of segmentation have been proposed, such as deformable models or region growing (Chupin et al., 2007; Ghanei et al., 1998; Shen et al., 2002), appearance-based models (Duchesne et al., 2002; Hu and Collins, 2007), and atlas/template-warping techniques (Aljabar et al., 2009; Barnes et al., 2008; Collins et al., 1995; Fischl et al., 2002; Gousias et al., 2008; Hammers et al., 2007; Heckemann et al., 2006; Rohlfing et al., 2004; Zhou and Rajapakse, 2005).

Recently, template-warping techniques that use a library of templates (i.e., MR images with manual expert-based segmentation) have been the subject of intensive investigation for their high accuracy in segmenting anatomical structures. Barnes et al. (2008) proposed to register the most similar template from a library of prelabeled subjects to segment the hippocampus (HC). However, the use of only one template may result in a biased segmentation. To avoid this problem, it is possible to use several similar templates (Aljabar et al., 2009; Collins and Pruessner, 2010; Gousias et al., 2008; Heckemann et al., 2006; Lotjonen et al., 2010), which requires a label fusion strategy (Collins and Pruessner, 2010; Gousias et al., 2008; Hammers et al., 2007; Heckemann et al., 2006; Lotjonen et al., 2010; Rohlfing et al., 2004) to efficiently merge the information derived from the selected templates. In addition, the extensive computational burden required by the nonlinear registration step needs to be reduced, for instance, by preselecting templates (Aljabar et al., 2009).

In template-warping techniques, two main assumptions are made. First, constraints on the shapes of structures are used implicitly because of the one-to-one correspondence between the voxels of the image to be segmented and those of the warped templates. This restriction presents the advantage of forcing the resulting segmentation to have a similar shape to those of expert-labeled structures in the template library. However, according to the regularization used during registration, some details can be lost and local high variability cannot be captured. Second, label fusion techniques usually assign the same weight to all samples during a voting procedure and consider only the absolute number a criterion. This approach is sensitive to registration error, since it does not take into account the relevance of each sample (Lotjonen et al., 2010). In the present work, we propose a patch-based scheme with a weighted label fusion, where the weight of each sample is only driven by the similarity of intensity between patches (i.e., small subvolumes of the image defined as three-dimensional [3D] cubes). In the proposed method, voxels with similar surrounding neighborhoods are considered to belong to the same structure and thus are used to estimate the final label.

As exemplars, patch-based methods are currently the focus of attention of the computer vision community in various domains such as texture synthesis (Efros and Freeman, 2001), in-painting (Criminisi et al., 2004), restoration (Buades et al., 2005), and single-frame super resolution (Protter et al., 2009). In each of these domains, patch-based methods have been the subject of intensive investigation because they exhibit very high performance despite their simplicity. Inspired by the nonlocal means denoising filter (Buades et al., 2005), we propose a nonlocal patch-based approach using expert manual segmentations as priors in the context of anatomical segmentation. The nonlocal means filter has two interesting properties that can be exploited to improve segmentation. First, the natural redundancy of information contained in the image can be used to drastically increase the numbers of samples considered during estimation. Second, the local intensity context (i.e., patch) can be used to produce a robust comparison of samples.

In this study, we describe a fully automated patch-based method using expert priors (i.e., information derived from manual segmentations) and the different steps required for its utilization. Our method is applied to the HC segmentation of healthy subjects and the lateral ventricle segmentation of patients with Alzheimer's disease (AD). During experiments, the influences of different parameters were studied, and a comparison with two other methods was performed. Finally, we discuss further improvements and questions revealed by this new approach.

2. *Materials and Methods*

2.1 Datasets

Two different datasets were used during the experiments to demonstrate the ability of the proposed method to (1) segment complex anatomical structures, (2) address the high variability of pathological structures, and (3) use multi-site training databases.

First, we used our method to segment the hippocampi of healthy subjects. The HC plays an important role in human memory and orientation. Moreover, HC dysfunction is involved in a variety of diseases, including AD (Jack et al., 2000), posttraumatic stress disorder (Bremner et al., 1995), major depression (Bremner et al., 2000), schizophrenia (Buss et al., 2007; Tanskanen et al., 2005), and epilepsy (Bernasconi et al., 2003). This structure is especially difficult to segment because of its small size, high variability, low contrast, and discontinuous boundaries in MR images (Chupin et al., 2007; Siadat et al., 2007). Finally, the HC is composed of several tissue types, which prevents the use of simple intensity-based techniques.

Second, we applied our method to the lateral ventricle segmentation of patients with AD. In such patients, structural variability is increased as a result of the pathology, and this variability represents a challenge for segmentation techniques such as atlas warping. Ventricular volume has been shown to provide a useful marker of neuronal degeneration and thus could be used as an indicator of AD (Nestor et al., 2008). However, despite the high contrast between tissue and cerebrospinal fluid (CSF), various factors render ventricle segmentation difficult. First, partial volume effects can impact segmentation, especially on MR images with limited resolution (Wang and Doddrell, 2001). Moreover, the temporal horns and occipital poles of the ventricles can be disconnected from the main body, which affects appearance-based methods and region-growing techniques. Finally, the choroid plexus appears with similar intensities to gray matter, which prevents the use of simple threshold-based techniques.

- Hippocampus dataset

The HC dataset consists of T1-weighted (T1w) MR images (fast field echo, TR = 17 ms, TE = 10 ms, flip angle = 30 °, 256×256 matrix, 1 mm in plane resolution, 1 mm thick slices) of 80 subjects randomly extracted from a group of 152 young, healthy individuals acquired on a 1.5T Philips GyroScan imaging system (Philips Medical Systems, Best, The Netherlands) in the context of the International Consortium for Brain Mapping (ICBM) project (Mazziotta et al., 1995). The local ethics committee approved the study and informed consent was obtained from all participants. The 80 subjects selected comprised 39 males and 41 females of similar ages (mean age: 25.09 ± 4.9 years). The MR images were manually segmented by an expert directly into stereotaxic space. For each subject, the HC label was manually defined using the protocol described by Pruessner et al. (2000). The resulting segmentations obtained an intraclass reliability coefficient (ICC) of 0.900 for inter-rater reliability (4 raters) and 0.925 for intra-rater reliability (5 repeats).

- Ventricle dataset

The ventricle dataset consists of T1w MR images (gradient-recalled echo, TR = 22 ms, TE = 10 ms, flip angle = 30 °, 250 mm field of view, 256×256 matrix, 110 sagittal partitions 1.5 mm thick, resulting in a voxel size of $0.98 \times 0.98 \times 1.5\ mm^3$) of 80 subjects randomly extracted from a dataset of 271 elderly patients with mild to moderate AD, aged between 50 and 85 years. The images were acquired at 62 different study sites. The manual segmentations were performed on the images in native space. Inter- and intra-rater variability were studied on 10 patients

scanned on the same SIEMENS Sonata 1.5T imaging system (Siemens, Erlangen, Germany). The inter-rater variability (3 raters) was estimated to be 0.987, and the intra-rater (10 repeats) variability to be 0.990. This dataset is not publicly available.

2.2 Method overview

As in template-warping methods, the proposed patch-based method uses expert manual segmentations as priors in order to achieve the segmentation of anatomical structures. However, our method has two main differences compared with template-warping methods: the scale of the considered objects and the label fusion scheme.

First, while template-warping methods work at the level of anatomical structure, our method handles a finer scale by using patches. Therefore, instead of performing the fusion of nonlinearly deformed template structures, the proposed method achieves the labeling of each voxel individually by comparing its surrounding patch with patches in training subjects in which the labels of the central voxels are known. When the patch under study resembles a patch in the training subjects, their central voxels are considered to belong to the same structure, and this training patch is used to estimate the final label. By this method, several samples from each training subject can be used during the label fusion, enabling a drastic increase in the number of sample patches involved in the label estimation.

Second, template-warping methods usually use a majority voting scheme to fuse the labels (Aljabar et al., 2009; Collins and Pruessner, 2010; Heckemann et al., 2006; Rohlfing et al., 2004) that considers the relevance (or weight) of all the samples labeled as similar. In the proposed method, the intensity-based distances between the patch under study and the patches in the training subjects are used to perform a weighted label fusion based on the nonlocal means estimator (Buades et al., 2005). The term *nonlocal* indicates that the spatial distance between the patches' centers is not taken into account; thus, the weight of each sample is only driven by the similarity of intensities between patches. In such an approach, the intensity-based distance between patches decreases as the relevance of the considered sample increases.

In other words, by taking advantage of the redundancy of information present in the image, the patch-based nonlocal means scheme enables the robust use of a large number of samples during estimation. This number will be significantly more important than the number of training subjects, in contrast to in template-based methods (i.e., where the number of warped subjects dominates). Moreover, contrary to classical majority voting schemes that give the same weight to all the samples, the nonlocal means scheme enables the robust distinction of the most similar samples according to their local context (i.e., their surrounding patches). Finally, in the proposed method, a *patch-based* weighting is used to perform a *pixel-based* aggregation of the labels ensuring the independency of the votes.

2.3 Image preprocessing for library construction

The first step of the proposed method involves organizing the library of training subjects to be used for patch comparison. During this step, variability caused by image formation is minimized by performing denoising, an inhomogeneity correction, and an intersubject intensity normalization (see Fig. 1). Moreover, since the anatomical intersubject regularity will be used to drive the search within the library, the training subjects of the database are linearly transformed into stereotaxic space to ensure a coarse correspondence between the anatomical locations of the images (see Fig. 1).

- Denoising

All images in the database were first denoised with the 3D block-wise nonlocal means filter recently proposed for MR images by Coupe et al. (2008). To remove the intensity bias introduced by the Rician nature of noise, a Rician adaptation of nonlocal means (Wiest-Daessle et al., 2008) was also used. The Rician noise level, used as a filtering parameter, was estimated with the object-based method proposed (Coupe et al., 2010).

- Inhomogeneity correction

To ensure that each tissue type has the same intensity within a single image, the well-known N3 intensity nonuniformity correction of Sled et al. (1998) was used.

- Linear registration to stereotaxic space

All the subjects were linearly registered to the MNI-ICBM152 template by using affine registration. For the ventricle dataset, the estimated transformation was applied to the expert-based segmentation using nearest-neighbor interpolation. For the HC dataset, the label interpolation was not performed because the labels are defined in stereotaxic space.

- Intensity normalization

Finally, the intensities of the images were set in [0-100] and were normalized together by following the method proposed by Nyul and Udupa (2000). With this method, we ensure that the contrast and luminance of each tissue type are consistent across the training subjects in the database.

At the end of this procedure, the images were cropped around the structure of interest to reduce the size of the library (see cropped images in Fig. 1). These different preprocessing steps ensure that the tissue intensities are consistent within the images (inhomogeneity correction) and across the subjects of the database (intensity normalization). Finally, the proposed library construction is similar to that used by template-warping techniques. However, while these techniques consider the library at the level of anatomical structure, our approach considers the library at the patch level.

2.4 Search strategy within the library

The search within the library is designed to find the most similar patches, but is also constrained in order to avoid useless computations. Therefore, the search process uses different strategies. First, we constrain the segmentation with an initialization mask. Second, we consider the probability that similar patches can be found in similar subjects. Then, we consider that the anatomical intersubject variability in stereotaxic space is limited; thus, we can define a limited search volume around the location under study. Finally, we consider that two similar patches should have similar luminance and contrast.

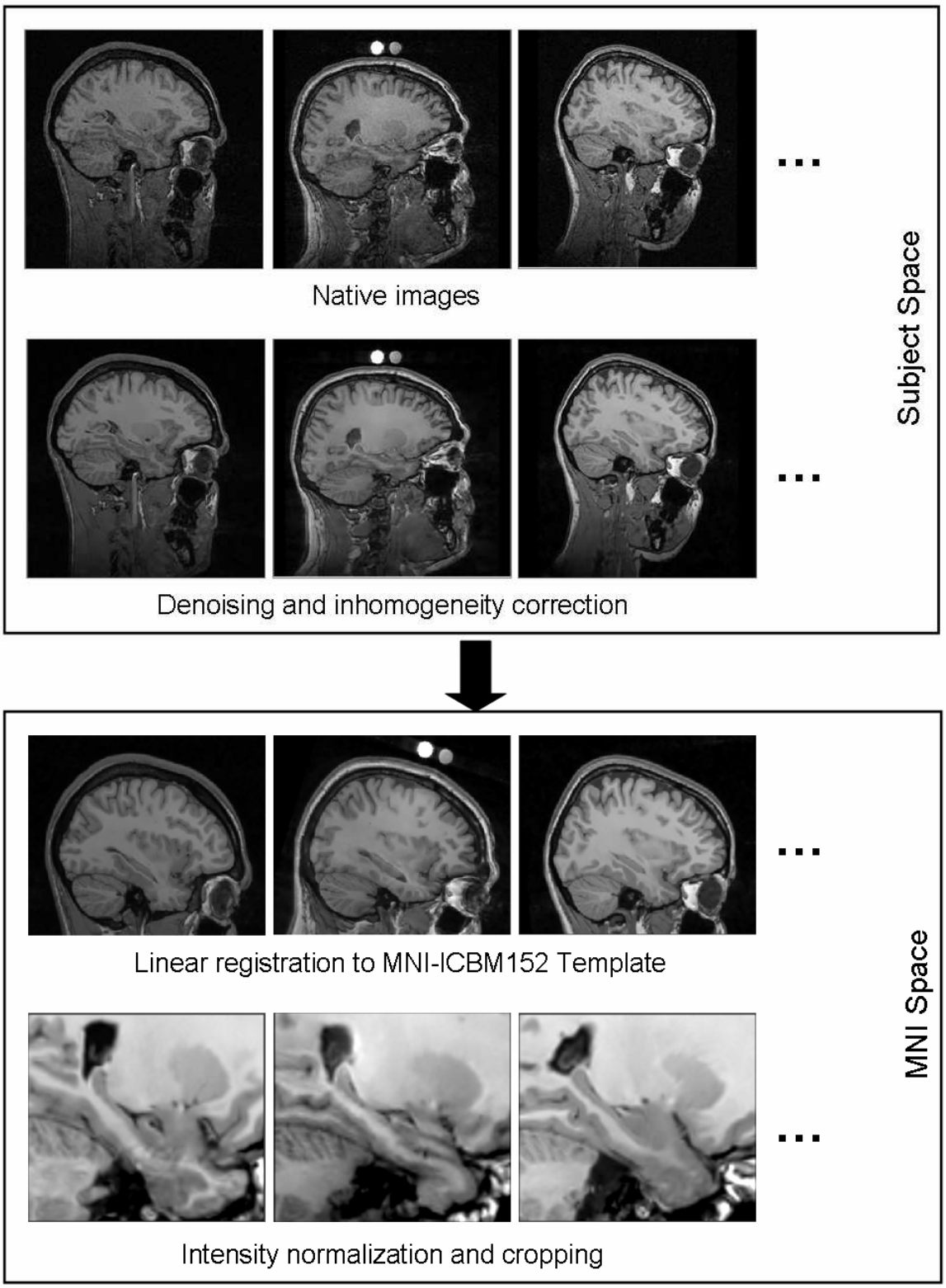

Fig. 1. Preprocessing. Preprocessing workflow used for library construction. First, denoising and inhomogeneity correction steps are performed in the subject space. The subjects are then linearly registered to the MNI-ICBM152 Template in MNI space. Finally, an intensity normalization of the different subjects is applied before cropping the images around the structure of interest

- Initialization mask

Instead of segmenting the entire image under study, we define an initialization mask around the structure of interest. A number of strategies can be used to propose an accurate initialization, such as matching the best subject (Barnes et al., 2008) followed by a morphological dilation of the mask. In this case, we chose a very fast and simple approach that uses the union of all the expert segmentations in the training database as the initial mask. In this way, we ensure that the structure is completely included in the mask and demonstrate the robustness of our method to coarse initialization (see Fig. 2).

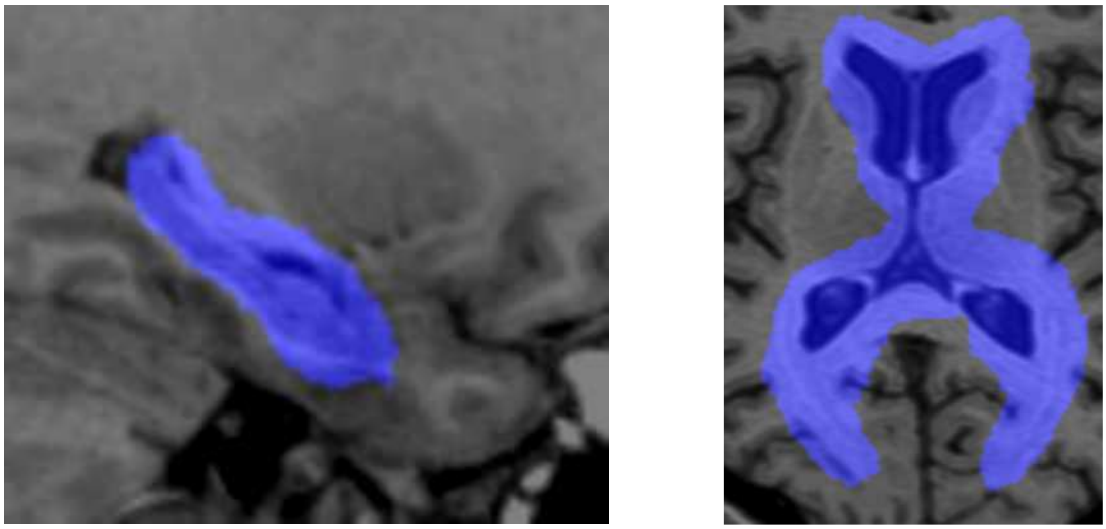

Fig. 2. Initialization masks. Initialization masks used for the hippocampus and ventricle datasets overlaid in blue on one subject.

- Subject selection

A selection is also performed at the subject level that resembles the selection of best subjects in the label fusion method (Aljabar et al., 2009). In our method, we use the sum of the squared difference (SSD) across the initialization mask instead of normalized mutual information over the image, as suggested by Aljabar et al. (2009). This strategy was chosen because SSD is sensitive to variation in contrast and luminance; thus, we expect to find a greater number of similar patches (in the sense of the L2 norm) in subjects with smaller SSDs. The same N closest subjects are retained during the entire segmentation process (see Fig. 3 where the three closest subjects are displayed).

- Search volume definition

Initially, the nonlocal means denoising filter was proposed as a weighted average of all the pixels in the image (Buades et al., 2005). For computational reasons, the entire image cannot be used and the number of pixels involved has to be reduced. As done for denoising (Buades et al., 2005; Coupe et al., 2008), we use a limited search volume V_i, defined as a cube centered on the voxel x_i under study. Thus, within each of the N selected subjects, we search for similar patches in a cubic region around the location under study (see Fig. 3). This search volume can be viewed as the intersubject variability of the structure of interest in stereotaxic space. This variability can increase for a subject with pathology or according to the structure under consideration.

- Patch preselection

Finally, as proposed for denoising purposes (Coupe et al., 2008), we perform a preselection of the patches to be compared in order to reduce the computational time. By using simple statistics such as mean or variance, it is possible to discard a priori the most dissimilar patches. In the proposed approach, we use luminance and contrast criteria to achieve the patch preselection. Based on the first and second terms of the well-known structural similarity measure (SSIM) (Wang et al., 2004), the preselection procedure can be written as follows:

$$ss = \frac{2\mu_i\mu_{s,j}}{\mu_i^2 + \mu_{s,j}^2} \times \frac{2\sigma_i\sigma_{s,j}}{\sigma_i^2 + \sigma_{s,j}^2}, \quad (1)$$

where μ represents the means and σ represents the standard deviations of the patches centered on voxel x_i (voxel under consideration) and voxel $x_{s,j}$ at location j in subject s. If the value of ss is greater than a given threshold th, the intensity distance between patches i and j is computed. The threshold th was set to 0.95 for all the experiments. This value was chosen empirically because it provides a good balance between segmentation accuracy and computational time reduction for both structures under study. Patch mean and variance are precomputed as maps of local means and local variances that avoid multiple computations.

Finally, the proposed search enables only candidates within the most similar training subjects to be considered (SSD-based subject selection), namely, those whose locations are not too far apart in stereotaxic space (search volume) and whose local neighborhoods are similar to the neighborhood of the voxel under study (patch preselection). Hence, the introduction of outliers is limited during the nonlocal patch-based label fusion and the computational burden is drastically reduced.

2.5 Nonlocal means label fusion

The proposed label fusion strategy is based on the nonlocal means estimator (Buades et al., 2005). In such an approach, the intensity-based distance between patches is used to perform a robust weighted average of samples. In our case, the nonlocal means estimator is used to perform the weighted average of the labels.

- Nonlocal means estimator

For all voxels x_i of the image to be segmented, the estimation of the final label is based on a weighted label fusion $v(x_i)$ of all labeled samples inside the search volume V_i for the N selected subjects:

$$v(x_i) = \frac{\sum_{s=1}^{N}\sum_{j\in V_i} w(x_i, x_{s,j}) y_{s,j}}{\sum_{s=1}^{N}\sum_{j\in V_i} w(x_i, x_{s,j})}, \quad (2)$$

where $y_{s,j}$ is the label given by the expert to voxel $x_{s,j}$ at location j in subject s and $w(x_i, x_{s,j})$ is the weight assigned to $y_{s,j}$ by patch comparison. Depending on the similarity between the patch surrounding x_i and that surrounding $x_{s,j}$, the weight $w(x_i, x_{s,j})$ is computed as:

$$w(x_i, x_{s,j}) = \begin{cases} \exp \dfrac{-\left\|P(x_i) - P(x_{s,j})\right\|_2^2}{h} & if \quad ss > th \\ 0 & else \end{cases}, \tag{3}$$

where $P(x_i)$ represents the cubic patch centered at x_i and $\|.\|_2$ is the normalized L2 norm (i.e., normalized by the number of elements) computed between each intensity of the elements of the patches $P(x_i)$ and $P(x_{s,j})$. As explained in the section on the search strategy (section 2.4), if the structural similarity *ss* between the patches is less than the threshold *th*, the weight is not computed and is set directly to zero.

Finally, by considering the labels *y* defined in {0,1}, the final label $L(x_i)$ is computed as:

$$L(x_i) = \begin{cases} 1 & v(x_i) > 0.5 \\ 0 & v(x_i) < 0.5 \end{cases}. \tag{4}$$

In the event that *ss* is less than *th* for all patches in the library, –1 is returned to indicate that the selected library does not allow a decision to be made. Note that our method can be also applied to probabilistic labels *y* defined in [0,1] without any modifications.

Figure 3 presents an overview of the different steps involved in achieving the segmentation of one voxel x_i included in the initialization mask. After the selection of the *N* most similar subjects in the training library (*N = 3* in this example), the patch $P(x_i)$ (in green) is compared with all the patches $P(x_{s,j})$ contained in the search volume V_i within the *N* selected subjects. The most similar patches $P(x_{s,j})$ (in blue) to the patch $P(x_i)$ obtain the highest weights, as shown in the weight maps. For the 2D slice in this illustration, 12 labeled samples have significant weights in subject s_1, the two most similar patches are in subject s_2, and no similar patches are found in subject s_3.

- Local adaptation of *h*

As usual in estimation problems using a robust function, the tuning of the decay parameter *h* plays a crucial role. When *h* is very low, only a few samples are taken into account. When *h* is very high, all samples tend to have the same weight and the estimation is similar to a classical average.

The value of *h* should depend on the distance between the patch under consideration and the library content. In fact, when the library contains patches very similar to the patch under study, *h* needs to be decreased to drastically reduce the influence of the other patches. However, when no similar patches exist in the library, *h* has to be increased to relax the selection. To achieve this local adaptation of *h* automatically, we propose an estimation of $h(x_i)$ based on the minimal distance between $P(x_i)$ and the considered patches $P(x_{s,j})$:

$$h(x_i) = \min_{x_{s,j}} \left\|P(x_i) - P(x_{s,j})\right\|_2 + \varepsilon, \tag{5}$$

where ε is a small constant to ensure numerical stability in case the patch under consideration is contained in the library. This kind of local adaptation has been similarly used for adaptive MRI denoising by Manjón et al. (2010).

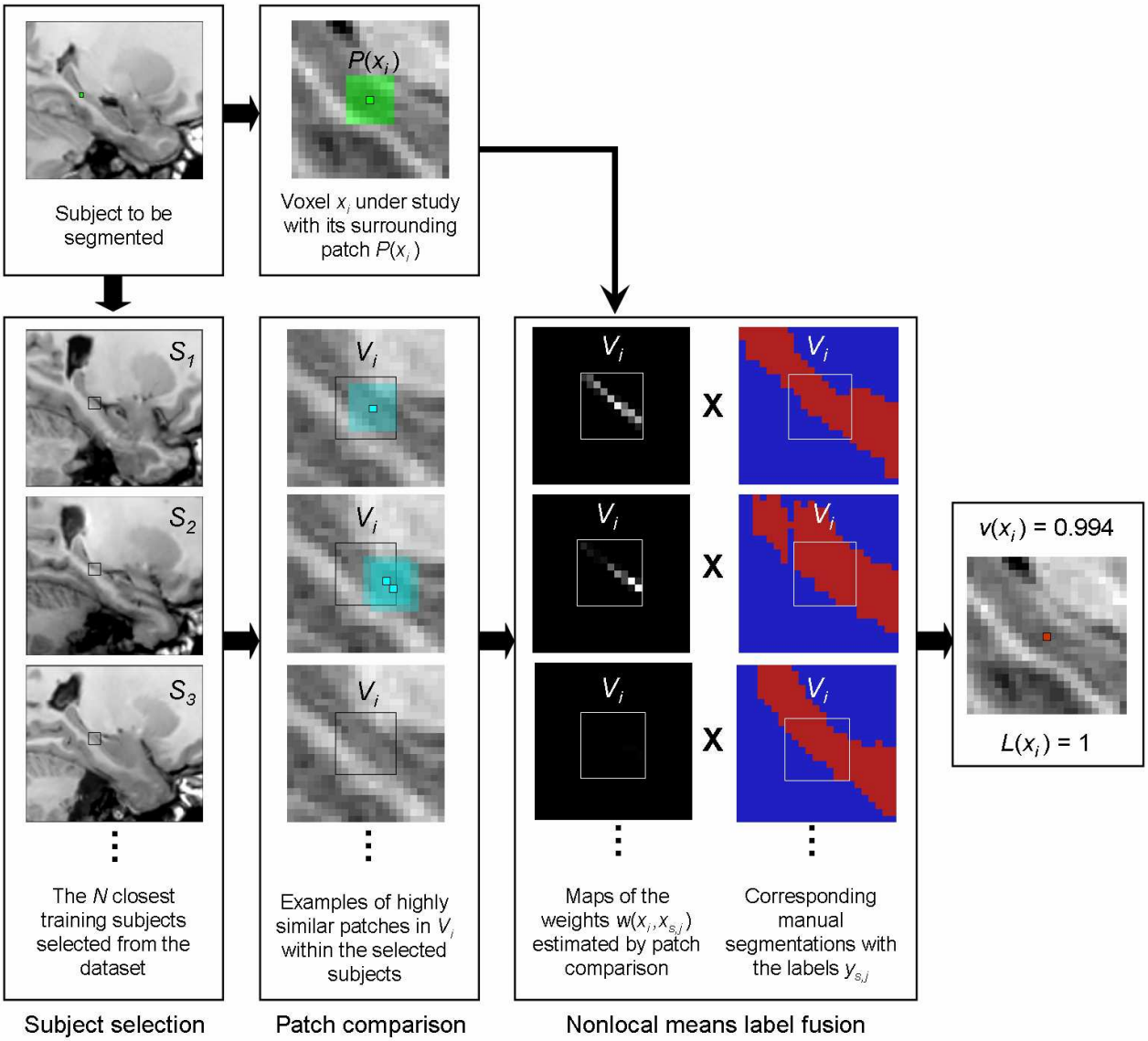

Fig. 3. Global overview. Overview of the different steps involved in achieving the segmentation of one voxel x_i included in the initialization mask. The patch $P(x_i)$ (in green) is compared with all the patches $P(x_{s,j})$ contained in the search volume V_i within the N selected subjects ($N = 3$ in this example). The weight maps show that the highest weights are obtained by the most similar patches $P(x_{s,j})$ (in blue) to the patch $P(x_i)$. After the nonlocal means fusion of the expert-based labels $y_{s,j}$, the resulting estimation is $v(x_i) = 0.994$. Thus, the final label is $L(x_i) = 1$.

Beyond its high denoising performance, the success of the nonlocal means filter is attributable to its algorithmic simplicity. In the proposed segmentation method, we tried to preserve this interesting aspect by keeping the algorithm as simple as possible. However, many improvements on the original nonlocal means denoising filter have been proposed, some of which could be applied to segmentation. The interested reader can find a review of these improvements in (Buades et al., 2010). For instance, a locally adaptive size of the search volume according to the estimator variance, as suggested by Kervrann and Boulanger (2008), could avoid useless computation in large constant areas (e.g., CSF in ventricle segmentation).

2.6 The method through an example

In order to provide an intuitive understanding of our method, we examine the spatial distribution of the variables involved in the segmentation process. Here, we present a detailed example of HC segmentation for $N = 2$ and $N = 20$, illustrated in Fig. 4:

- First, the **normalized number of samples** (see bottom of Fig. 4) shows that the nonlocal means estimation considers around 500 sample patches on average for each training subject (e.g., 10,000 samples on average for $N = 20$). Not all the considered samples have significant weights, but this number is significantly higher than N, the number of training subjects, as shown in Fig. 3. The spatial variation of this number depends on the patch preselection. In fact, since preselection (see Eq. 1) rejects all patches with dissimilar luminance and contrast during the patch comparison, the number of considered samples is lower for the less common patches. Moreover, this average number decreases slightly when N increases because of the introduction of subjects with less similar structural shape, contrast, or luminance in the library (i.e., with lower SSD during subject selection). However, it is interesting to note that similar patches are found in dissimilar subjects, since the average number does not drastically decrease when N increases.

- The second variable to be studied is the smoothing parameter $h(x_i)$ that represents the **minimal distance** (see Fig. 4) between $P(x_i)$ and all the patches $P(x_{s,j})$ considered during patch comparison (see Eq. 5). A high $h(x_i)$ indicates that the closest patch found in the library is not really similar to $P(x_i)$. In this case, the estimation provided by the nonlocal means estimator is less robust and leads to inaccuracies in segmentation. As shown for $N = 2$, the areas where $h(x_i)$ is high mainly correspond to false positives and false negatives (see top of Fig. 4). When $N = 20$, the higher number of considered patches enables the procedure to find more similar patches (the minimal distance $h(x_i)$ decreases). Thus, the segmentation is improved, as assessed by kappa index values (see top of Fig. 4).

- The last variable, $v(x_i)$, is the value returned by the **nonlocal means estimator** (see Eq. 2). This value can be viewed as the probability that a voxel will be included in the structure. In this case, the manual labels $y_{s,j}$ also have to be counted as probabilities. The fast decay of $v(x_i)$ shows that the nonlocal means estimator clearly distinguishes between the structure and the background. As expected, the edges obtain less discriminative values. This can be explained by the higher intra-rater variability on edges within the training database. This aspect is an inherent limitation of all methods that use expert-based manual segmentations as priors.

2.7 Implementation details

The proposed method was implemented in MATLAB 7.4.0 using C/MEX code. The experiments were conducted using a single core of an Intel Core 2 Quad Q6600 processor at 2.4 GHz with 4 GB of RAM. The different preprocessing steps needed for library construction were achieved by using **tools developed in-house in C**. The nonlocal means denoising took around 2 min, and the inhomogeneity correction, around 1 min. The linear registration required less than 2 min, and the normalization, close to 1 min. The execution times given in the results section are the times required only for segmentation, since all the compared methods required these preprocessing steps. As discussed later, many optimizations can be used because each voxel is treated independently, which allows multithreading or GPU-based computation.

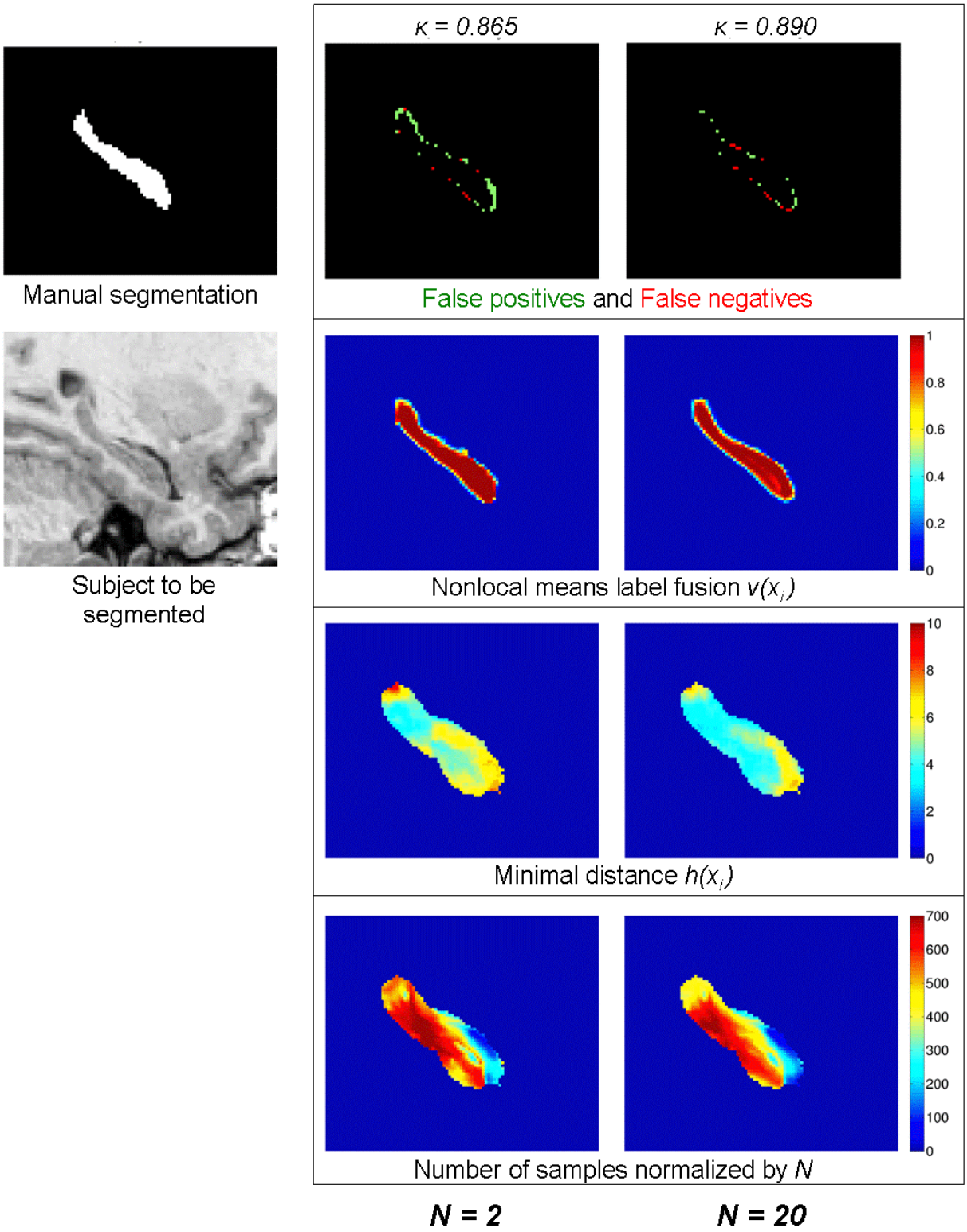

Fig. 4. Method through an example. Spatial distribution of the variables used by our method for $N = 2$ and $N = 20$. **Top. Left:** Manual segmentation by the expert and the corresponding MR image. **Right:** False positives and false negatives maps of the segmentation provided by our method. **Middle.** Spatial distribution of the nonlocal means estimator $v(x_i)$ and of the minimal distance $h(x_i)$ used as the smoothing parameter. **Bottom.** Spatial distribution of the number of samples used during the estimation normalized by the number N of training subjects. Note that the resulting number of patches evaluated for each voxel is much larger than the number of templates used for segmentation.

2.8 Validation framework

For each dataset, a leave-one-out procedure was performed for the 80 subjects. The kappa index (Dice coefficient or similarity index) (Zijdenbos et al., 1994) was then computed by comparing the expert-based segmentations with those obtained with our method. For two binary segmentations *A* and *B*, the kappa index was computed as:

$$\kappa(A,B) = \frac{2|A \cap B|}{|A| + |B|}. \tag{6}$$

As usual in quantitative MR analysis, manual segmentation is considered the gold standard (Pruessner et al., 2000). For both datasets, the impact of the patch size, search volume size, and number of training subjects was studied. Moreover, the proposed patch-based method was compared for both datasets with an appearance-based approach using level-set shape constraints (Hu and Collins, 2007) and a template-based technique inspired by Barnes et al. (2008) that uses ANIMAL (Collins et al., 1995) for the nonlinear registration of the best subject.

In the appearance-based method, only one modality was used during the processing. We used the 79 remaining subjects to construct the training dataset involved in the principal component analysis (PCA) computation. Although this number is higher than those proposed by Hu and Collins (2007) (20, 30, 40, and 60 subjects), we wanted to conduct a fair comparison with our method, since the selection of the *N* closest subjects in our patch-based method is done within the 79 remaining subjects.

For the template-based method inspired by Barnes et al. (2008), the best subject was selected using the normalized mutual information, as suggested by Aljabar et al. (2009). This subject was then nonlinearly warped to the subject under study with ANIMAL (Collins et al., 1995) within a multiresolution framework until a resolution of 2 mm. In our validation, the best subject was selected from the 79 remaining subjects during a leave-one-out procedure.

3. Results

The kappa index values obtained with the initial mask are presented in Fig. 5. The median kappa index value was 0.44 for the HC dataset and 0.41 for the ventricle dataset, which corresponds to an average percentage of false negatives (i.e., the mean number of voxels included in the mask but not in the manual segmentations) of 71% for the HC dataset and 73% for the ventricle dataset. Note that these results only give a baseline to show that the initial mask does not achieve an accurate segmentation.

3.1 Impact of the 3D patch size

First, we studied the impact of patch size on segmentation accuracy. The kappa index results are presented in Fig. 6 for both datasets. The best median kappa index value was obtained with a patch size of 7×7×7 voxels for the HC dataset ($\kappa = 0.882$) and 5×5×5 voxels for the ventricle dataset ($\kappa = 0.957$). The optimal patch size seems to reflect the complexity of the anatomical structure. The patch size needs to be larger for the HC than for the ventricle, since the intensities of the HC are less discriminative. Figure 7 shows the HC segmentation results for the best, one median, and the worst subject for the different patch sizes studied. These results indicate that the patch size needs to be large enough to capture the local geometry (holes and discontinuities in HC segmentation for a patch size of 3×3×3 voxels). Because of the high contrast between tissues for ventricle segmentation, the size of the patch can be smaller.

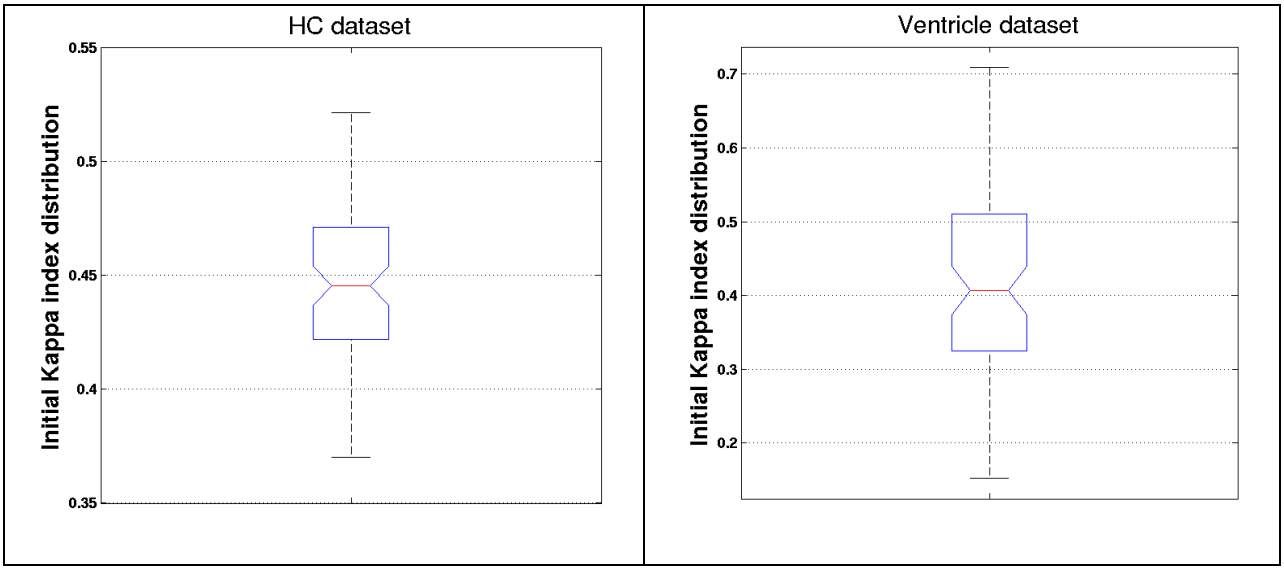

Fig. 5. Initial kappa index distribution. Box plot of the kappa index distribution of the initialization mask for the hippocampus (HC) and ventricle datasets. Boxes represent the lower quartile, the median (red line), and the upper quartile of the kappa index distribution. Moreover, whiskers indicate the most extreme values within 1.5 times the interquartile range. Finally, outliers (red +) are data with values beyond the ends of the whiskers. The size of the notches indicates the significance interval at 5%. If the notches of two distributions do not overlap, these distributions have different medians at the 5% significance level. The median kappa value was 0.44 for the HC dataset and 0.41 for the ventricle dataset, which correspond to an average percentage of false negatives of 71% for the HC dataset and 73% for the ventricle dataset.

3.2 Impact of the search volume size

We also studied the impact of the size of the search volume on segmentation accuracy. The kappa index results are presented in Fig. 8. The best median kappa index was obtained with a search volume of 9×9×9 voxels for the HC dataset ($\kappa = 0.882$) and 15×15×15 voxels for the ventricle dataset ($\kappa = 0.958$). The optimal search volume size is related to the anatomical variability of the structure within stereotaxic space. Since the HC is smaller and was segmented in healthy subjects, the variability of this structure is less than that of the ventricles of the subjects with AD; thus, a search volume of 7×7×7 voxels or 9×9×9 voxels provides good results. For the ventricle, the size of the structure and the presence of pathology mean that larger search volumes will give better results. However, the small improvement seen when the search volume increased from 11×11×11 voxels to 15×15×15 voxels may not justify the increase in computational time. Figure 9 presents the ventricle segmentation results for the best, one median, and the worst subject for different search volume sizes. As expected, when the search volume is too small, the anatomical variability of the structure of interest within stereotaxic space can lead to the selection of a subpart of the library that contains insufficient information for finding similar patches. For instance, in Fig. 9, the holes in the segmentations indicate that no similar patches were found for a search volume of 3×3×3 voxels.

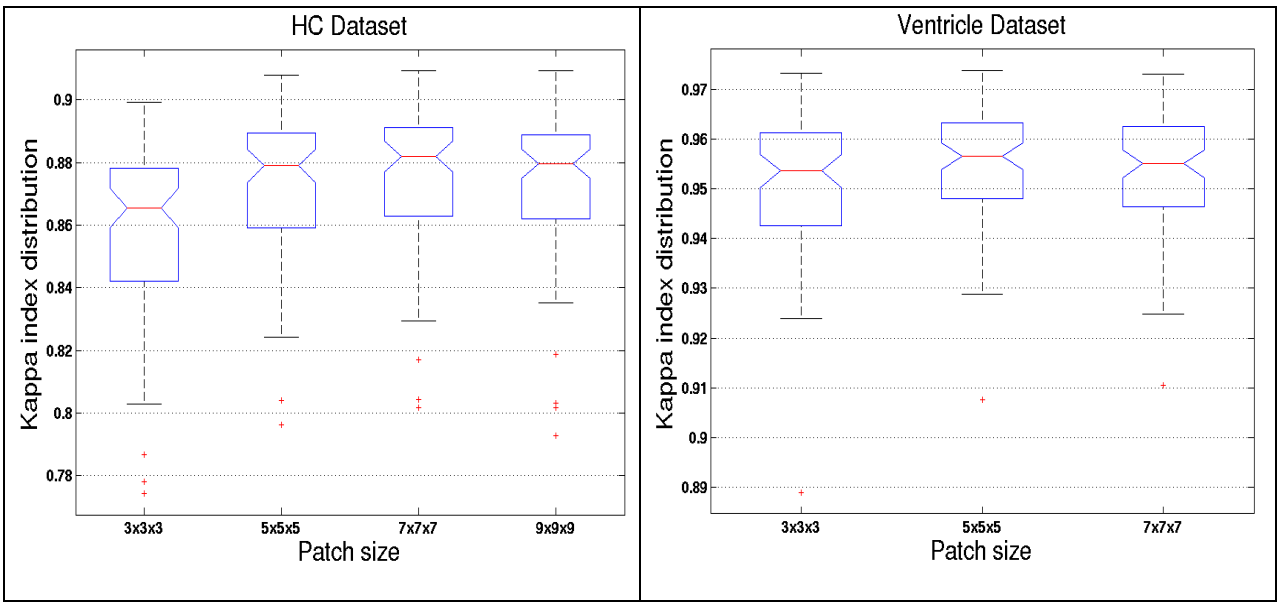

Fig. 6. Impact of the patch size. Kappa index distribution for different patch sizes for both datasets. For the HC dataset, the results were obtained using 20 training subjects and a search volume of 9×9×9 voxels. For the ventricle dataset, the results were obtained using 20 training subjects and a search volume of 11×11×11 voxels.

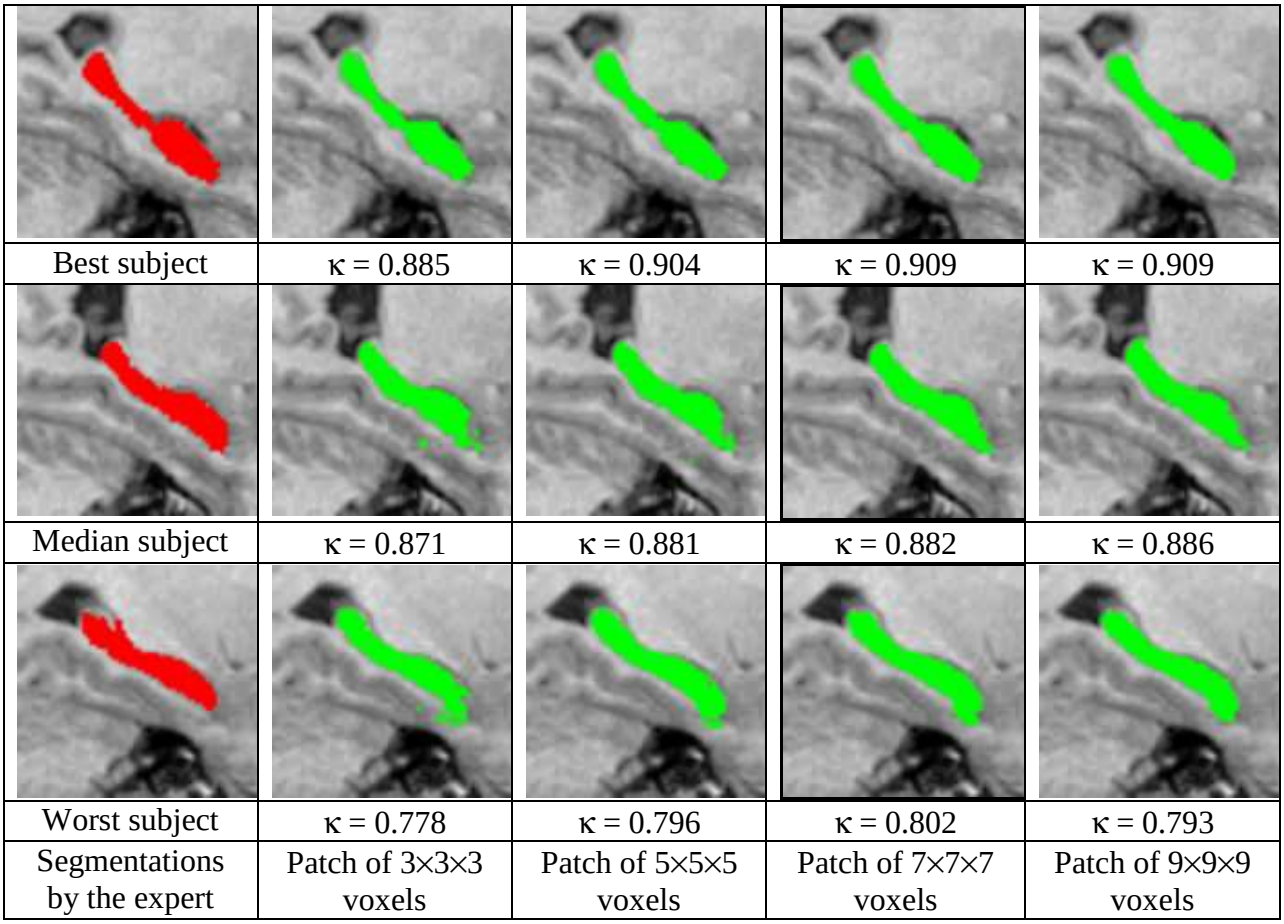

Fig. 7. Impact of the patch size. Hippocampus segmentation for the subjects with the best kappa index (top), a median kappa index (middle), and the worst kappa index (bottom) obtained by our method. These results were obtained using 20 training subjects and a search volume of 9×9×9 voxels. The expert-based segmentations are shown in red, and the segmentations obtained with our method, in green.

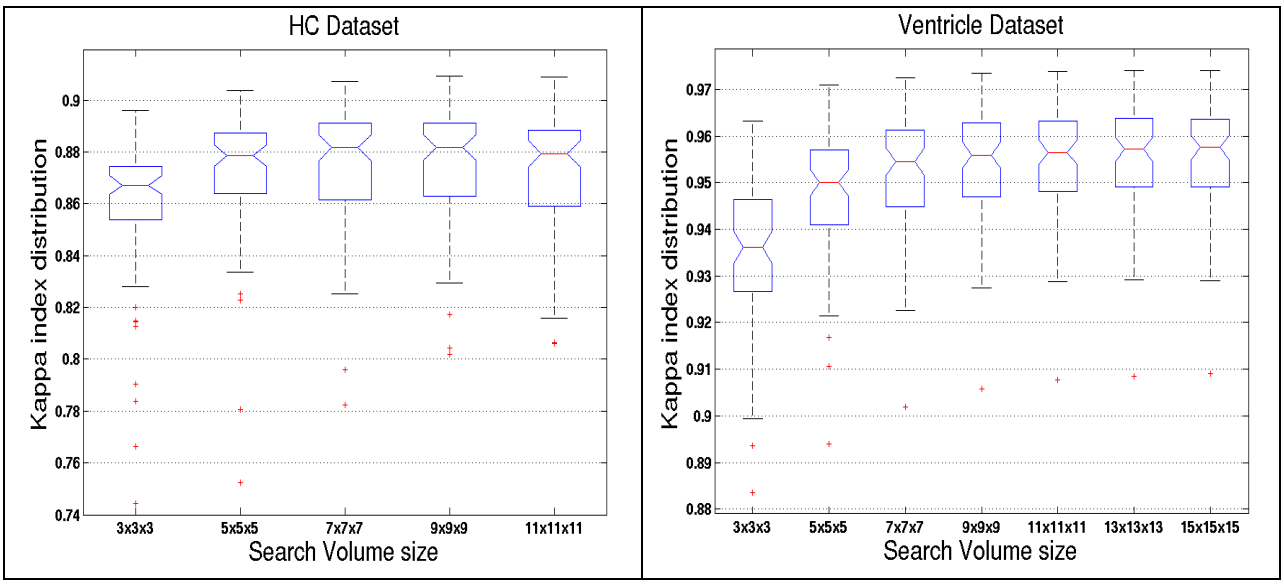

Fig. 8. Impact of search volume size. Kappa index distribution for different search volume sizes for both datasets. For the HC dataset, the results were obtained using 20 training subjects and a 3D patch size of 7×7×7 voxels. For the ventricle dataset, the results were obtained using 20 training subjects and a 3D patch size of 5×5×5 voxels.

3.3 Impact of the number of subjects

The last important parameter of the proposed method is the number of selected training subjects. During this experiment, segmentation accuracy was studied for 2 to 30 selected training subjects. The results are presented in Figs. 10 and 11. For the HC dataset, the median kappa index value was 0.848 for 2 subjects and 0.884 for 30 subjects. For the ventricle dataset, the median kappa index value was 0.942 for 2 subjects and 0.959 for 30 subjects. As expected, increasing the number of selected training subjects increased the accuracy of the segmentation. Figure 12 presents the HC segmentation results. Holes appeared in the segmentations when data from only two training subjects were used, indicating that no similar patches have been found. This aspect can be moderated by decreasing the threshold of preselection in order to increase the number of patches used during estimation. However, this strategy cannot be more efficient than increasing the number of subjects, since the final decision will be based on patches that are not very similar.

3.4 Comparison with appearance-based and template-based methods

Finally, the proposed patch-based method was compared with two other methods. Figure 13 presents the kappa index values obtained for each method applied to both datasets. The results presented for our method were obtained with $N = 20$. For HC segmentation, the appearance-based method obtained a median kappa index value of 0.800, the best template approach obtained 0.837, and the proposed method obtained 0.882. For ventricle segmentation, the appearance-based method obtained a median kappa index value of 0.788, the best template approach obtained 0.909, and the proposed approach obtained 0.957. The patch-based approach obtained significantly better results compared to the two others methods with a p-value $<< 0.001$ in both cases using Kruskal-Wallis tests. In addition, the appearance-based method was not able to capture the variability of the lateral ventricles in patients with AD. Figures 14 and 15 show 3D representations of the HC and lateral ventricle segmentations obtained by each method.

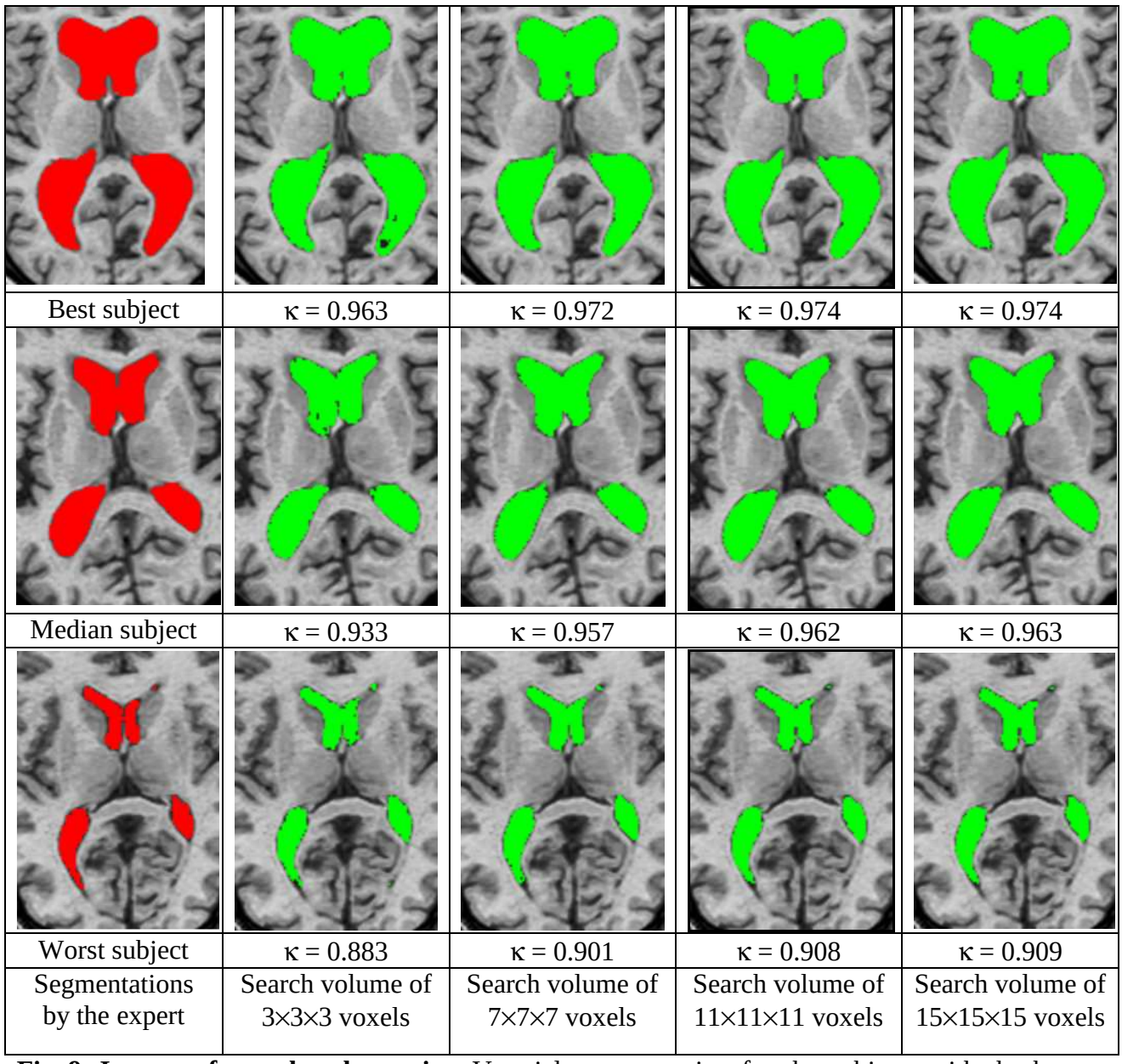

Fig. 9. Impact of search volume size. Ventricle segmentation for the subjects with the best kappa index (top), a median kappa index (middle), and the worst kappa index (bottom) obtained by our method. These results were obtained using 20 training subjects and a 3D patch size of 5×5×5 voxels. The expert segmentations are shown in red, and the segmentations obtained with our method in green.

3.5 Computational time

The computational time required by the proposed method was proportional to the number of subjects; each subject required around 40 s. This time could be easily reduced with a better initialization mask. Compared with other approaches, the appearance-based method (Hu and Collins, 2007) took around 45 s to provide the segmentation of the HC. By contrast, the best template-based approach inspired by Barnes et al. (2008) required around 6 min to achieve the nonlinear registration of the cropped images already linearly registered into stereotaxic space. Although comparing these approaches was difficult because our method was coded in C-MEX for MATLAB and not in C like the other two, these results show that the PCA-based approach was quite fast, though at the expense of accuracy. Finally, methods using nonlinear registration can become quite computationally intensive when several subjects are involved, as noticed in (Aljabar et al., 2009).

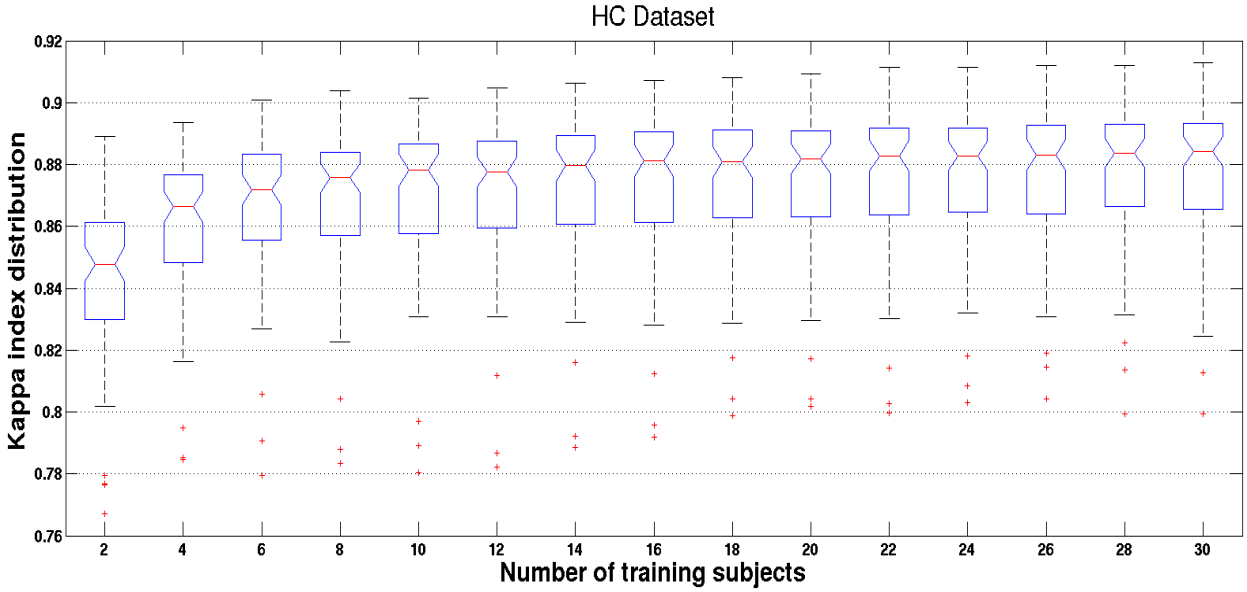

Fig. 10. Impact of the number of training subjects. Kappa index distribution according to the number of selected training subjects for the HC dataset with a patch size of 7×7×7 voxels and search volume of 9×9×9 voxels.

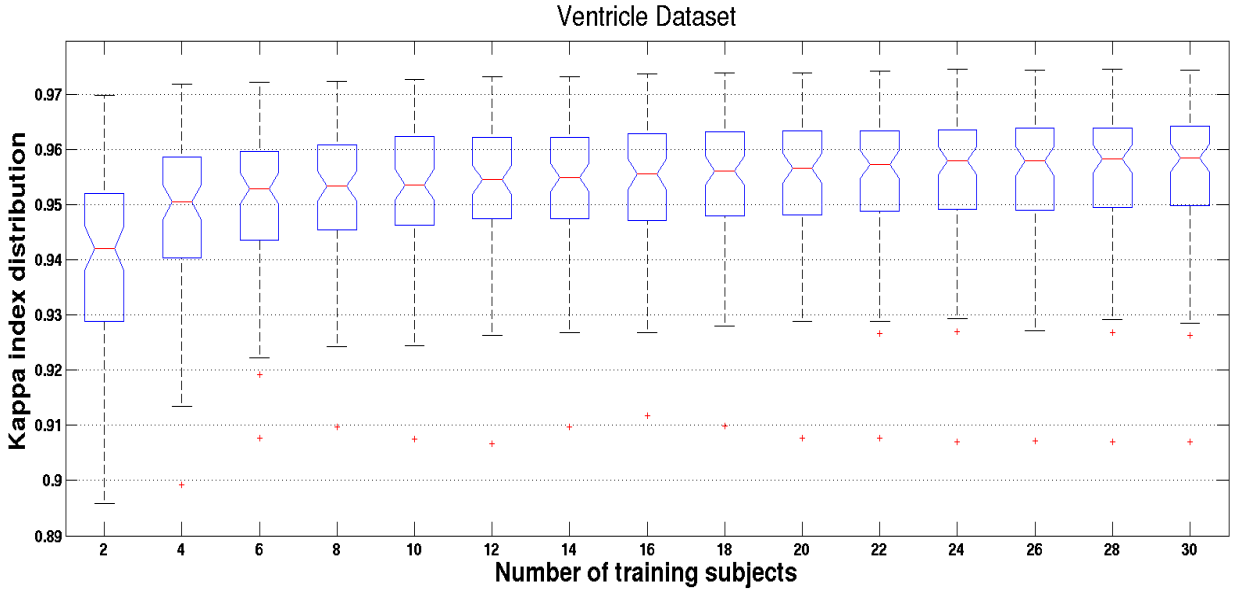

Fig. 11. Impact of the number of training subjects. Kappa index distribution according to the number of selected training subjects for the ventricle dataset with a patch size of 5×5×5 voxels and a search volume of 11×11×11 voxels.

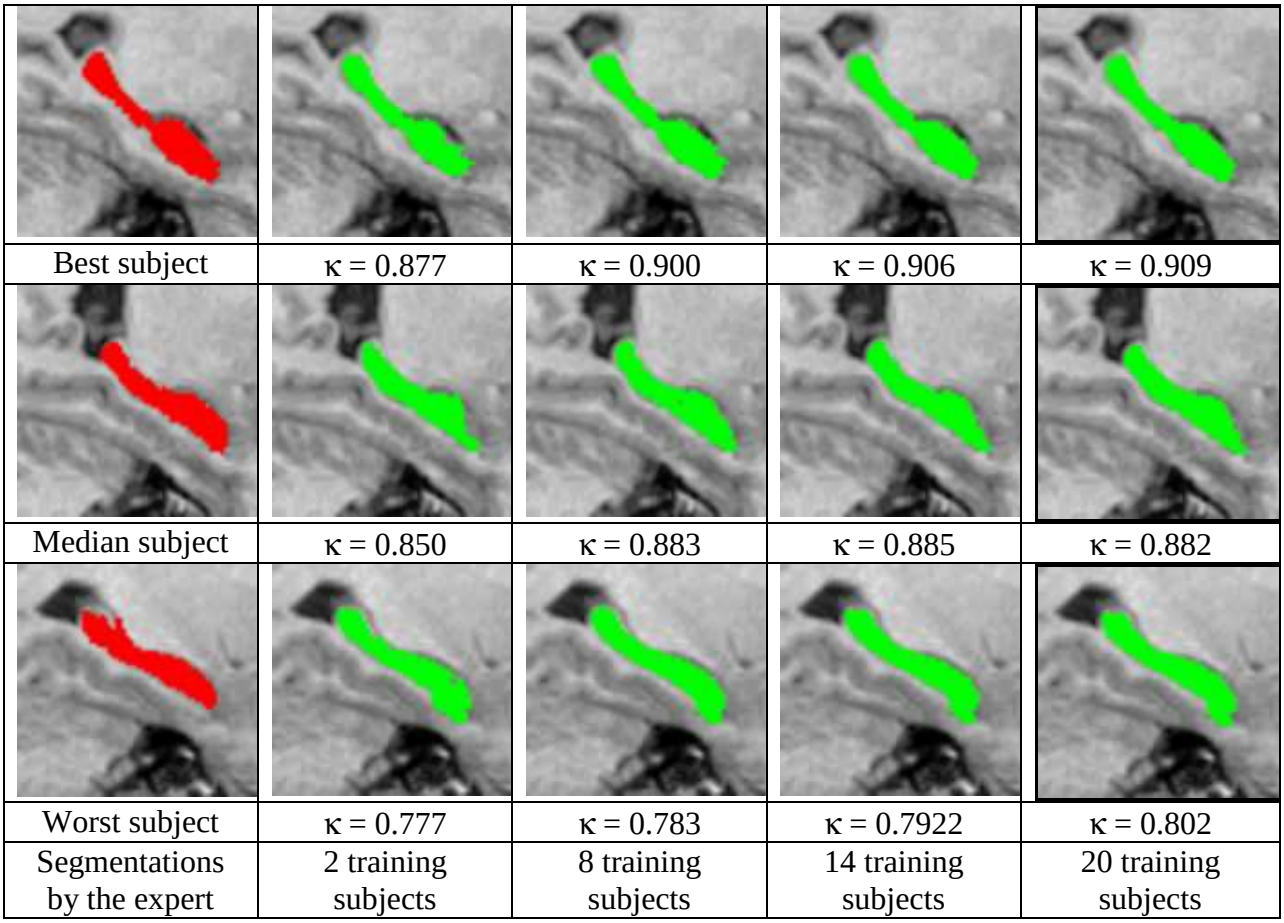

Fig. 12. Impact of the number of training subjects. Hippocampus segmentations for the subjects with the best kappa index (top), a median kappa index (middle), and the worst kappa index (bottom) obtained by our method. These results were obtained with a patch size of 7×7×7 voxels and a search volume of 9×9×9 voxels. The expert-based segmentations are shown in red, and the segmentations obtained with our method in green.

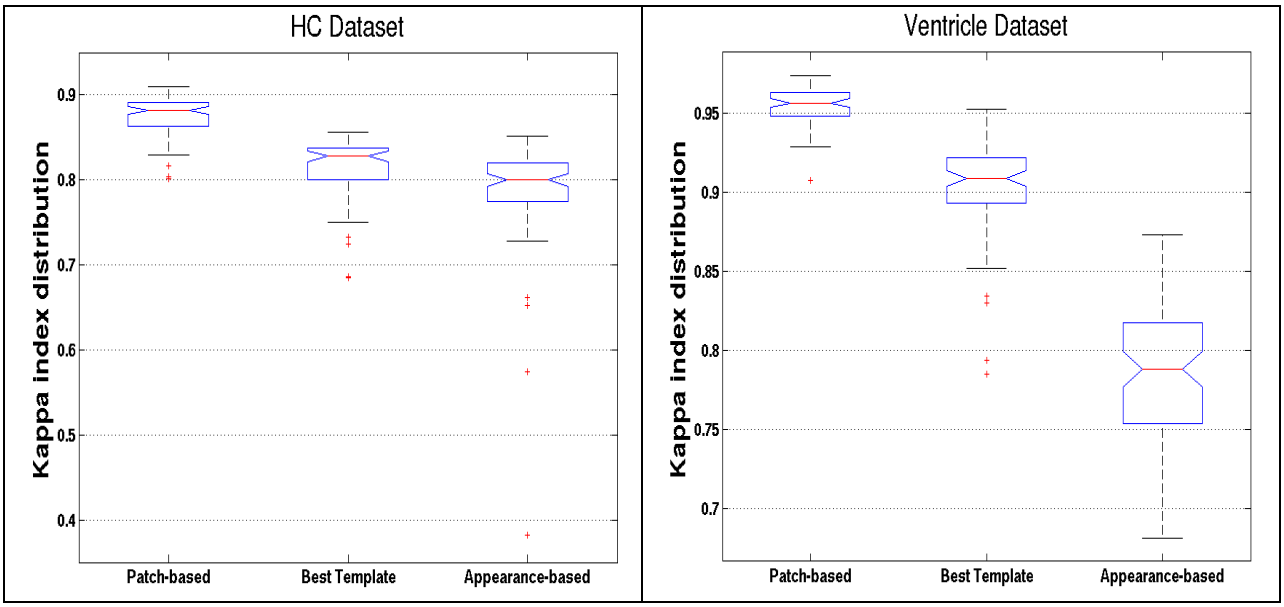

Fig. 13. Method comparison. Kappa index distribution for the three methods compared on both datasets.

4. Discussion

We propose a novel patch-based approach to automatically segment anatomical structures using the manual segmentations done by experts as priors. Despite its simplicity, the accuracy of the proposed method has been demonstrated within our validation framework for HC and lateral ventricle segmentation. The highest median kappa index values obtained during experiments were 0.884 for the HC dataset and 0.959 for the ventricle dataset, for $N = 30$ training subjects. In terms of a two-digit mean kappa index value as is widely used in the literature, our method obtained 0.88 for the HC dataset and 0.95 for the ventricle dataset. Moreover, comparison with an appearance-based (Hu and Collins, 2007) and a template-based method (Barnes et al., 2008) highlighted the competitive results obtained by the proposed nonlocal patch-based approach.

Comparing published methods is always difficult because of differences between the databases used for validation, the populations studied, the quality of expert segmentations, and the reported quality metrics. Moreover, the number of labeled samples defining the segmentation (which depends on the ratio between the volume of the structure and the voxel size) (Rohlfing et al., 2004) can impact the similarity measure. Nonetheless, interesting tendencies in method evolution can be extracted by studying published results.

For HC segmentation, recently published results (Barnes et al., 2008; Chupin et al., 2007; Morey et al., 2009; Morra et al., 2008; Pohl et al., 2007; van der Lijn et al., 2008) indicated high kappa index values greater than 0.80. The latest published methods based on the nonlinear warping of the best templates and involving a label fusion step (Collins and Pruessner, 2010; Gousias et al., 2008; Lotjonen et al., 2010) obtained kappa index values equal to or greater than 0.88. As discussed by Aljabar et al. (2009), the accuracy obtained with these techniques reach the limit of the variability of expert human raters. Gousias et al. (2008) reported a mean kappa index of 0.88 with the use of a B-spline-based nonlinear registration on the brain of a 2-year-old. Lotjonen et al. (2010) proposed two intensity-based models to improve label fusion: an extension of the graph-cut-based method described by van der Lijn et al. (2008) and an expectation-maximization (EM) approach. Using nonlinear deformations of the $N = 13$ closest templates, their graph-cut-based label fusion obtained a kappa index of 0.880 and their EM-based label fusion, a kappa index of 0.885. Obtained by using the ADNI database of healthy subjects and patients with AD, these kappa index values indicate the high performance of these approaches. In our case, only the method proposed by Collins and Pruessner (2010) can be directly compared with our proposed one as they used the same database and the same validation framework. Collins and Pruessner (2010) obtained a median kappa index of 0.886 by nonlinearly registering the $N = 11$ closest subjects with ANIMAL (Collins et al., 1995) and by fusing the resulting label with a classical majority voting scheme. By comparison, the proposed method offers the main advantages of its simplicity (no nonrigid registration required) and its computational time (40 s vs. 6 min per training subject) for a similar segmentation accuracy ($\kappa = 0.884$). As a result of the proposed automatic adaptation of the robust function parameter, our approach can be implemented simply in a fully automatic manner.

For ventricle segmentation, the large variety of databases makes comparison with the literature rather difficult. Hu and Collins (2007) used the proposed appearance-based method with a level-set constraint and obtained a mean kappa index of 0.83 for patients with multiple sclerosis. The same method obtained a median kappa index of 0.788 during our comparison using AD patients. This low kappa index might result from the higher variability of lateral ventricles in patients with AD. Schonmeyer (2006) obtained a mean kappa index of 0.90 for subjects with AD by using an object-oriented method, while Aljabar et al. (2009) reported a mean kappa index of 0.912 across a database of 275 healthy subjects. During our method comparison, we obtained a slightly lower kappa index of 0.909 with the best template approach. As previously mentioned for HC segmentation, recent template-warping approaches (Aljabar et al., 2009) with a selection strategy for the best subjects have obtained very good results in the literature.

The new approach to the label fusion problem introduced by using a patch-based method reveals several questions. First, in this proof of concept, we used linear registration of subjects to save computational time and demonstrate the robustness of the proposed method. However, the complementarity of patch-based weighted label fusion with approaches using nonlinear registration appears to be a natural extension. In this way, the spatial distance between the patches' locations could be used as a shape prior, and the initialization mask could be greatly improved, reducing computational time. This tendency toward using local intensity-based refinement after nonlinear registration seems promising, as shown by van Rikxoort et al. (2010) with local piece-wise atlas fusion and by van der Lijn (2008) and Lotjonen et al. (2010) with graph-cut-based and EM algorithms. Moreover, experiments on a larger diversity of pathologies and anatomical structures should be studied in future applications. The robustness of the proposed parameters in these situations should be also tested.

Finally, implementation optimization should be investigated. In the literature on nonlocal means denoising, many papers have been proposed on reducing computational time. In our method, a new patch preselection has been proposed and is already included. However, prototype-based (Tibell et al., 2009) or cluster tree-based (Brox et al., 2008) approaches could be faster. In addition, the noniterative nature of the nonlocal means approach is perfectly suited to parallel implementation. Work on parallelization (Coupe et al., 2008) or GPU implementations (Huang et al., 2009; Palhano Xavier de Fontes et al., 2010) has shown a significant reduction in computational time close to real-time processing.

5. Conclusion

In this paper, we propose a novel patch-based method using expert segmentations as priors to segment anatomical structures. Based on the similarity of intensity content between patches, the new label fusion is achieved by using a nonlocal means estimator. Validation of hippocampus segmentation in healthy subjects and of ventricle segmentation in patients with Alzheimer's disease was performed. In addition, comparison with an appearance-based and a template-based method demonstrated the high performance of our method. During validation, the proposed method obtained a median kappa index value of 0.884 for the HC and 0.959 for the ventricles. The use of a nonlocal means scheme in combination with a method involving nonlinear registration will be the subject of further investigation.

Acknowledgments

We thank the reviewers for their helpful comments. This study was supported by the Canadian Institutes of Health Research (CIHR, MOP-84360) and Cda (CECR)-Gevas-OE016. This work was also partially supported by the Spanish Health Institute Carlos III through the RETICS Combiomed, RD07/0067/2001 and by Elan Pharmaceuticals, Inc., and Transition Therapeutics, Inc. This work benefited from the use of ITK-SNAP from the Insight Segmentation and Registration Toolkit (ITK) for 3D rendering.

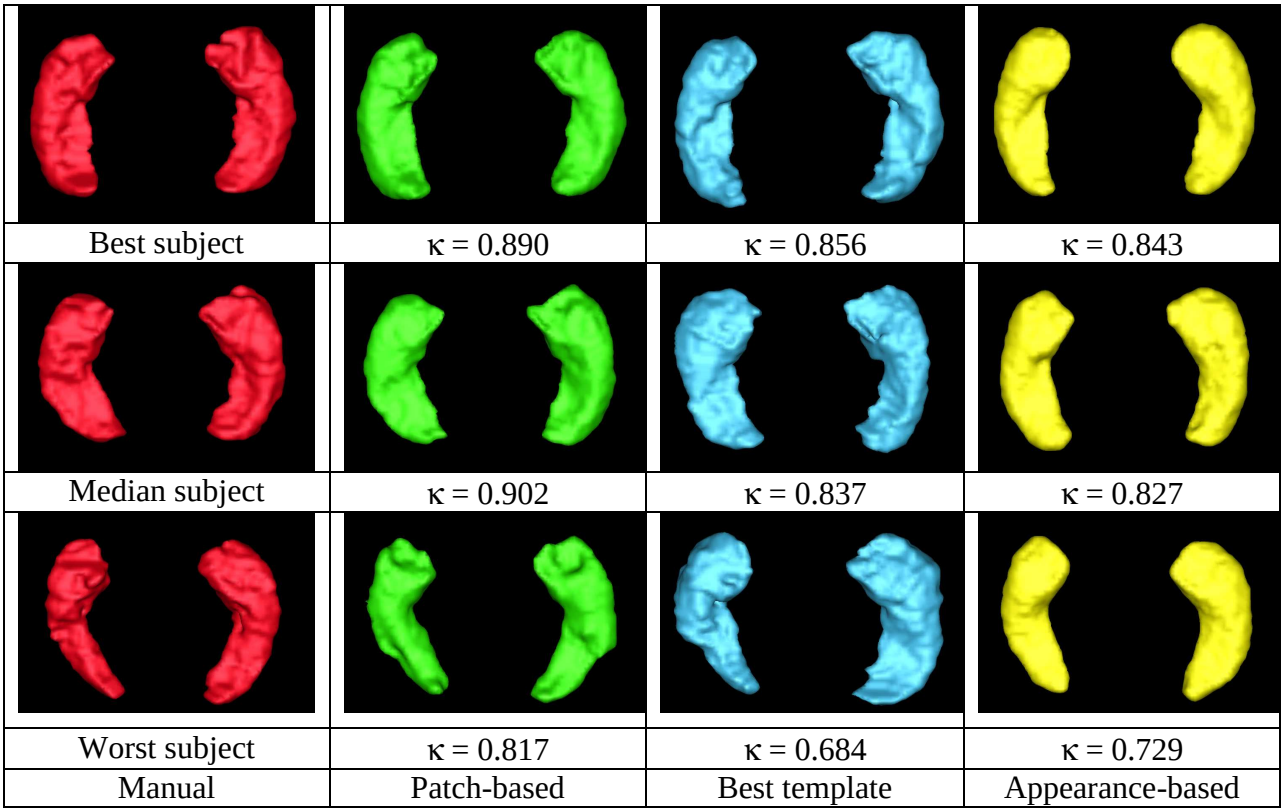

Fig. 14. Method comparison. Three-dimensional HC segmentations obtained by the three methods for the subjects with the best kappa index (top), a median kappa index (middle), and the worst kappa index (bottom) obtained by the best-template method. The expert-based segmentation is shown in red, the proposed patch-based method in green, the best template method in blue, and the appearance-based method in yellow. Note how the appearance-based result is much smoother than the other techniques.

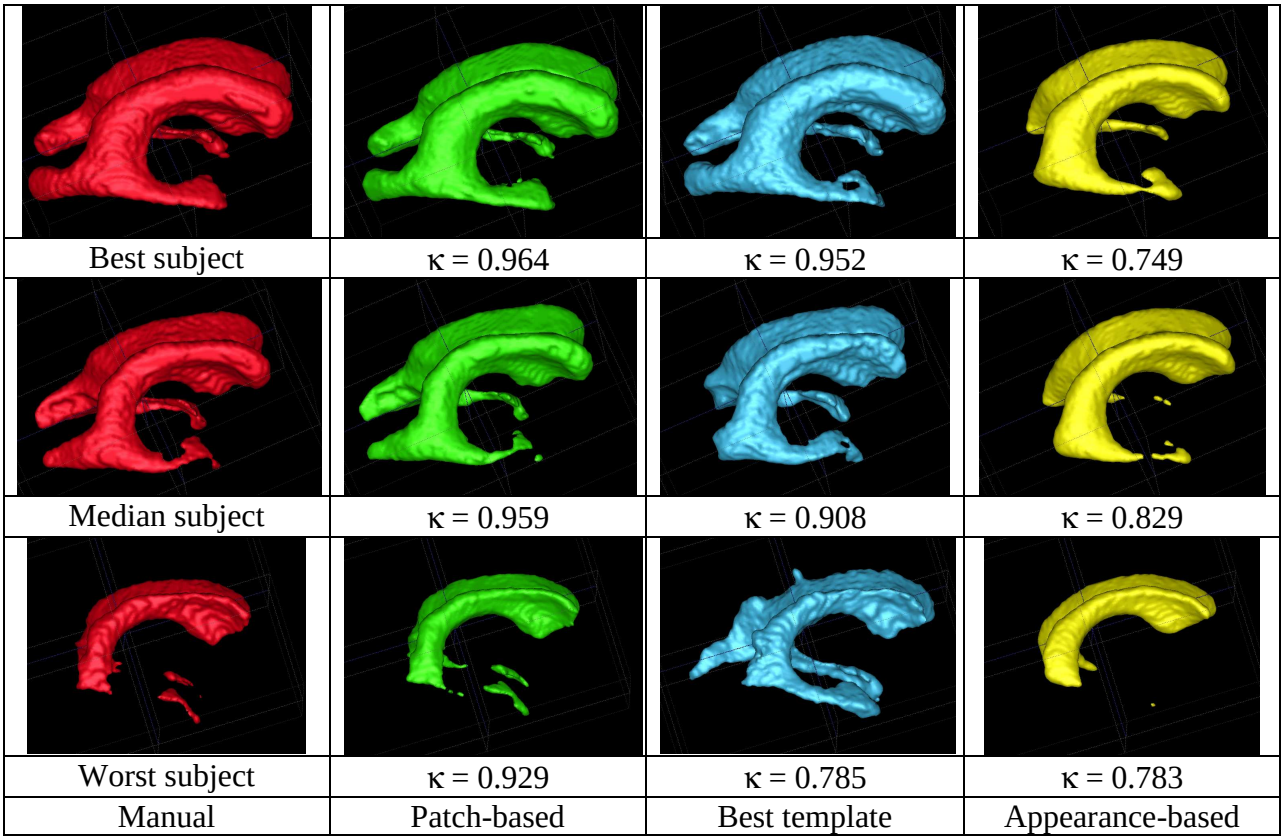

Fig. 15. Method comparison. Three-dimensional ventricle segmentations obtained by the three methods for the subjects with the best kappa index (top), a median kappa index (middle), and the worst kappa index (bottom) obtained by the best template method. The expert-based segmentation is shown in red, the proposed patch-based method in green, the best template method in blue, and the appearance-based method in yellow. Note how the both the appearance-based method and the best template method can cut off the occipital pole of the lateral ventricle. The appearance-based method also cuts off the temporal poles of the lateral ventricle.

References

Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46, 726-738.

Barnes, J., Foster, J., Boyes, R.G., Pepple, T., Moore, E.K., Schott, J.M., Frost, C., Scahill, R.I., Fox, N.C., 2008. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Neuroimage 40, 1655-1671.

Bernasconi, N., Bernasconi, A., Caramanos, Z., Antel, S.B., Andermann, F., Arnold, D.L., 2003. Mesial temporal damage in temporal lobe epilepsy: a volumetric MRI study of the hippocampus, amygdala and parahippocampal region. Brain 126, 462-469.

Bremner, J.D., Narayan, M., Anderson, E.R., Staib, L.H., Miller, H.L., Charney, D.S., 2000. Hippocampal volume reduction in major depression. Am J Psychiatry 157, 115-118.

Bremner, J.D., Randall, P., Scott, T.M., Bronen, R.A., Seibyl, J.P., Southwick, S.M., Delaney, R.C., McCarthy, G., Charney, D.S., Innis, R.B., 1995. MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. Am J Psychiatry 152, 973-981.

Brox, T., Kleinschmidt, O., Cremers, D., 2008. Efficient nonlocal means for denoising of textural patterns. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 17, 1083-1092.

Buades, A., Coll, B., Morel, J.M., 2005. A non-local algorithm for image denoising. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, Proceedings, 60-65.

Buades, A., Coll, B., Morel, J.M., 2010. Image Denoising Methods. A New Nonlocal Principle. SIAM Review 52, 113-147.

Buss, C., Lord, C., Wadiwalla, M., Hellhammer, D.H., Lupien, S.J., Meaney, M.J., Pruessner, J.C., 2007. Maternal care modulates the relationship between prenatal risk and hippocampal volume in women but not in men. J Neurosci 27, 2592-2595.

Chupin, M., Mukuna-Bantumbakulu, A.R., Hasboun, D., Bardinet, E., Baillet, S., Kinkingnehun, S., Lemieux, L., Dubois, B., Garnero, L., 2007. Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease. Neuroimage 34, 996-1019.

Collins, D.L., Holmes, C.J., Peters, T.M., Evans, A.C., 1995. Automatic 3-D model-based neuroanatomical segmentation. Human Brain Mapping 3, 190-208.

Collins, D.L., Pruessner, J.C., 2010. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Neuroimage.

Coupe, P., Manjon, J.V., Gedamu, E., Arnold, D., Robles, M., Collins, D.L., 2010. Robust Rician noise estimation for MR images. Med Image Anal 14, 483-493.

Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans Med Imaging 27, 425-441.

Criminisi, A., Perez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13, 1200-1212.

Duchesne, S., Pruessner, J., Collins, D.L., 2002. Appearance-based segmentation of medial temporal lobe structures. Neuroimage 17, 515-531.

Efros, A.A., Freeman, W.T., 2001. Image quilting for texture synthesis and transfer. Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, pp. 341-346.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341-355.

Ghanei, A., Soltanian-Zadeh, H., Windham, J.P., 1998. Segmentation of the hippocampus from brain MRI using deformable contours. Comput Med Imaging Graph 22, 203-216.

Gousias, I.S., Rueckert, D., Heckemann, R.A., Dyet, L.E., Boardman, J.P., Edwards, A.D., Hammers, A., 2008. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. Neuroimage 40, 672-684.

Hammers, A., Heckemann, R., Koepp, M.J., Duncan, J.S., Hajnal, J.V., Rueckert, D., Aljabar, P., 2007. Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof-of-principle study. Neuroimage 36, 38-47.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115-126.

Hu, S., Collins, D.L., 2007. Joint level-set shape modeling and appearance modeling for brain structure segmentation. Neuroimage 36, 672-683.

Huang, K., Zhang, D., Wang, K., 2009. Non-local means denoising algorithm accelerated by GPU. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.

Jack, C.R., Jr., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, E.G., Kokmen, E., 2000. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. Neurology 55, 484-489.

Kervrann, C., Boulanger, J., 2008. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. International Journal of Computer Vision 79, 45-69.

Lotjonen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 49, 2352-2365.

Manjon, J.V., Coupe, P., Marti-Bonmati, L., Collins, D.L., Robles, M., 2010. Adaptive non-local means denoising of MR images with spatially varying noise levels. Journal of Magnetic Resonance Imaging 31, 192-203.

Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). Neuroimage 2, 89-101.

Morey, R.A., Petty, C.M., Xu, Y., Hayes, J.P., Wagner, H.R., 2nd, Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. Neuroimage 45, 855-866.

Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Toga, A.W., Jack, C.R., Jr., Weiner, M.W., Thompson, P.M., 2008. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. Neuroimage 43, 59-68.

Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., 2008. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. Brain 131, 2443-2454.

Nyul, L.G., Udupa, J.K., 2000. Standardizing the MR image intensity scales: making MR intensities have tissue specific meaning. Medical Imaging 2000: Image Display and Visualization 1, 496-504.

Palhano Xavier de Fontes, F., Andrade Barroso, G., Coupé, P., Hellier, P., 2010. Real time ultrasound image denoising. Journal of Real-Time Image Processing.

Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E., Shenton, M.E., Wells, W.M., 2007. A hierarchical algorithm for MR brain image parcellation. IEEE Trans Med Imaging 26, 1201-1212.

Protter, M., Elad, M., Takeda, H., Milanfar, P., 2009. Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Trans Image Process 18, 36-51.

Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. Cereb Cortex 10, 433-442.

Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R., Jr., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 21, 1428-1442.

Schonmeyer, R., Prvulovic, D., Rotarska-Jagiela, A., Haenschel, C., Linden, D.E., 2006. Automated segmentation of lateral ventricles from human and primate magnetic resonance images using cognition network technology. Magn Reson Imaging 24, 1377-1387.

Shen, D., Moffat, S., Resnick, S.M., Davatzikos, C., 2002. Measuring size and shape of the hippocampus in MR images using a deformable shape model. Neuroimage 15, 422-434.

Siadat, M.R., Soltanian-Zadeh, H., Elisevich, K.V., 2007. Knowledge-based localization of hippocampus in human brain MRI. Comput Biol Med 37, 1342-1360.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. Ieee Transactions on Medical Imaging 17, 87-97.

Tanskanen, P., Veijola, J.M., Piippo, U.K., Haapea, M., Miettunen, J.A., Pyhtinen, J., Bullmore, E.T., Jones, P.B., Isohanni, M.K., 2005. Hippocampus and amygdala volumes in schizophrenia and other psychoses in the Northern Finland 1966 birth cohort. Schizophr Res 75, 283-294.

Tibell, K., Spies, H., Borga, M., 2009. Fast Prototype Based Noise Reduction. Proceedings of the 16th Scandinavian Conference on Image Analysis. Springer-Verlag, Oslo, Norway, pp. 159-168.

van der Lijn, F., den Heijer, T., Breteler, M.M., Niessen, W.J., 2008. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. Neuroimage 43, 708-720.

van Rikxoort, E.M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M.A., Pluim, J.P., van Ginneken, B., 2010. Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus. Med Image Anal 14, 39-49.

Wang, D., Doddrell, D.M., 2001. A segmentation-based and partial-volume-compensated method for an accurate measurement of lateral ventricular volumes on T(1)-weighted magnetic resonance images. Magn Reson Imaging 19, 267-273.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: From error visibility to structural similarity. Ieee Transactions on Image Processing 13, 600-612.

Wiest-Daessle, N., Prima, S., Coupe, P., Morrissey, S.P., Barillot, C., 2008. Rician noise removal by non-Local Means filtering for low signal-to-noise ratio MRI: applications to DT-MRI. Med Image Comput Comput Assist Interv 11, 171-179.

Zhou, J., Rajapakse, J.C., 2005. Segmentation of subcortical brain structures using fuzzy templates. Neuroimage 28, 915-924.

Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging 13, 716-724.