

Combining a patch-based approach with a non-rigid registration-based label fusion method for the hippocampal segmentation in Alzheimer's disease

Carlos Platero, M.Carmen Tobar

Received: date / Accepted: date

Abstract *We provide and evaluate an open-source software solution for automatically hippocampal segmentation from T1-weighted (T1w) magnetic resonance imaging (MRI). The method is applied for measuring the hippocampal volume, which allows discriminate between patients with Alzheimer's disease (AD) or mild cognitive impairment (MCI) and elderly controls (NC). The method is based on a fast patch-based label fusion method, whose selected patches and their weights are calculated from a combination of similarity measures between patches using intensity-based distances and labeling-based distances. These combined similarity measures produces better selection of the patches, and their weights are more robust.*

The algorithm is trained with the Harmonized Hippocampal Protocol (HarP). The proposal is compared with FreeSurfer and other label fusion methods. To evaluate the performance and the robustness of the proposed label fusion method, we employ two databases of T1w MRI of human brains. For AD vs NC, we obtain a high degree of accuracy, approximately 90%. For MCI vs NC, we obtain accuracies around 75%. The average time for the hippocampal segmentation from a T1w MRI is less than 17 minutes.

Keywords Atlas-based segmentation · Image registration · Patch-based label fusion · Hippocampal segmentation · Magnetic resonance imaging

Carlos Platero

Health Science Technology Group, Technical University of Madrid, Ronda de Valencia 3, 28012, Madrid, Spain.

Tel.: +34-91-3366878

Fax: +34-91-3367709

E-mail: carlos.platero@upm.es

M.Carmen Tobar

Health Science Technology Group, Technical University of Madrid, Ronda de Valencia 3, 28012, Madrid, Spain.

1 Introduction

The analysis of the hippocampus is a very active field of research because it is one of the first structures where early Alzheimer’s disease (AD) pathology is observed [1]. Patients suffering from the initial stages of AD are mostly clinically classified as amnesic mild cognitive impairment (MCI), but not all patients with amnesic MCI will develop AD [2]. It has been shown that measurements of the hippocampus from T1-weighted (T1) MRI are useful markers of AD and progression of the clinical decline at a MCI stage [2,3]. A challenge of the neuroimaging is to help in the diagnosis of early AD, particularly in amnesic MCI patients or prodromal AD.

Several approaches have been proposed for automatically classifying patients with AD and/or MCI from hippocampal segmentations extracted in brain MRI [4]. The manual segmentation of the hippocampus is considered the gold standard but is time consuming [5]. Consequently, many automatic approaches have been proposed for extracting the hippocampal structures from brain MRI [6,7,8,9,10]. Among such approaches, atlas-based methods have been demonstrated to outperform other algorithms [11] that rely on manual segmentations. However, the hippocampus is a complex anatomical structure, and different manual segmentation protocols have been proposed. In fact, a review by Konrad et al [12] identified 71 hippocampal tracing methods. The absolute volume differences between certain protocols may vary by 30%. An international effort to harmonize existing protocols has defined the Harmonized Hippocampal Protocol (HarP) [13,14]. This protocol proved to be very reliable and to provide a hippocampal segmentation estimate that can be considered as a standard measure, enabling the use of the hippocampal measures as proper bio-markers for AD and MCI.

Multi-atlas segmentation is a popular approach for labeling anatomical structures from medical images. A subset of atlases within a database are registered to a target image, and their segmentations can be transformed and subsequently fused to provide a consensus segmentation.

This paper is focused on the label fusion problem. Most existing label fusion methods are based on global weighted voting, such as majority voting [15] (MV), STAPLE [16] or weighted voting (WV) [17], which are widely used in medical image segmentation. In these approaches, each atlas contributes to the resulting segmentation with the same weight for all of its voxels. It is very sensitive to registration errors because it does not take into account the relevance of each sample.

Moreover, the errors differ if the registration is affine or non-rigid. In affine transformations, the structures of the target image and the atlases are only aligned in position, orientation and scale, but the context information of the aligned atlas is preserved in the domain of the target image. However, the atlases are typically warped to the target image using non-rigid registration techniques. This approach presents the advantage of forcing the resulting segmentation to have a similar global shape to those of expert-labeled structures

in the atlases. The major drawback of applying non-rigid registration to the atlases is the smoothing of the transferred labels.

Patch-based label fusion methods have shown great potential using only affine registrations [9, 18, 10, 19]. Rather than fusing transferred atlases into the target image, this framework is based on the non-local mean principle [20]. These methods exhibit two interesting properties: (i) this approach drastically increases the number of samples considered during the labeling estimation, and (ii) the local intensity context (i.e., patch) can be used to produce a robust comparison among samples [9]. Although these methods are powerful, they highly depend on the intensity-based similarity measures between patches. For selected patches with similar appearance according to the similarity measures used, their corresponding labelings may be very different. Image similarities over small image patches may not be an optimal estimator [10]. Moreover, the labeling is local and independent, without global constraints. To overcome these drawbacks, Asman and Landman [21] have proposed an iterative algorithm between the estimation of the labeling and the weights of the patches that maximizes the expected value of a conditional likelihood function. *Wang et al. [22] have proposed a framework that adopts sparse representation and incorporate an anatomical constrain for the brain segmentations. An initial segmentation is obtained from a subset of aligned atlases by using sparse representation in a patch-based label fusion. Then, the segmentation is iteratively refined by integration a label-based distance for obtaining the weights of the patches, which is called the anatomical constraint.*

Following this line and alternatively, we propose an algorithm without iterations. We consider the following two assumptions: (a) the label fusion methods using non-rigid registrations lead to segmentations with shape prior constraints and (b) the local context is well represented by the patches using affine transformations. The core of the proposed algorithm is to combine information from atlases using affine transformations and non-rigid registrations to increase the reliability of the automatic segmentation. Our propose generates an estimate of the labeling with constraints in the shape priors by a label fusion method based on non-rigid registrations. Then, a patch-based labeling method is applied, which selects the patches from aligned atlases, and their weights are calculated using similarity measures of the intensities and binary labelings, where the labeling distances are calculated using the output of the previous step. These combined similarity measures produce better selection of the patches, and their weights are more robust, which improves the segmentation results compared to other existing approaches.

The remainder of this paper is organized as follows. In Section 2, the proposed method of combining a patch-based labeling method with an atlas-warping using non-rigid registrations is presented. The experiments and results for the hippocampal segmentation are described in Section 3. The discussion and conclusions are presented in Section 4.

Table 1 Demographic details of the subset of the ADNI database used as the atlases with the HarP annotations. Mean (standard deviation, SD) unless specified otherwise. ANOVA with Bonferroni post hoc test is used for age and MMSE scores, except for gender where the chi-square test is used. ^a Significant compared to NC. ^b Significant compared to MCI.

	NC	MCI	AD
Number of subjects	44	45	45
Age, years	76.1 (7.4)	74.4 (8.0)	74.5 (8.1)
Gender male (%)	21 (52%)	26 (58%)	21 (47%)
MMSE	29.0 (1.0)	26.1 (2.9) ^a	21.9 (3.9) ^{a,b}

2 Materials and Methods

In this study, two databases were used to validate the proposed approach and compare it to FreeSurfer and other label fusion methods: (i) a subset of the ADNI database

(<http://www.loni.ucla.edu/ADNI>) and (ii) a private database from the Laboratory of Cognitive and Computational Neuroscience (LCCN) of the Center of Biomedical Technology (Technical University of Madrid) [23,24]. We denote this set as LCCN.

For ADNI images, MRI acquisition was performed according to the ADNI acquisition protocol [25]. The subset of the ADNI database was also used to construct the atlas set. HarP was the manual protocol for labeling the selected ADNI images. The definition of HarP has been designed to reduce inconsistencies of manual hippocampal segmentation protocols, as detailed in [26]. The mean Dice value for repeated manual segmentations between experts has been estimated to be 89% ([88%; 92%]) according to [27]. The identification numbers of the ADNI-HarP images used in this study are reported in (<http://www.hippocampal-protocol.net>) [28]. The subjects included in this study are 44 controls, 45 patients diagnosed with MCI and 45 patients with AD.

LCCN images are acquired using a Fast Spoiled Gradient Echo sequence with the following parameters: TR/TE/TI = 11.2/4.2/450 ms; flip angle of 12°; 1 mm slice thickness; a 256 × 256 matrix; and FOV of 25 cm. The diagnosis of subjects was based on a neuropsychological examination performed at the *Hospital Clínico de Madrid* and the *UPDC del Ayuntamiento de Madrid* [23, 24]. Images are collected from 52 control subjects, 97 with MCI, and 13 with AD.

Clinical and demographic data are presented in Tables 1 and 2.

2.1 Overview

The proposed segmentation scheme involves four principal steps, namely, (1) MRI pre-processing, (2) spatial normalization and defining the regions of interest (ROIs), (3) the first labeling based on atlas-warping using non-rigid

Table 2 Demographic details of the LCCN database used to compute the proposed approach. Mean (SD) unless specified otherwise. ANOVA with Bonferroni post hoc test is used for age and MMSE scores, except for gender where the chi-square test is used. ^a Significant compared to NC. ^b Significant compared to MCI.

	NC	MCI	AD
Number of subjects	52	97	13
Age, years	69.9 (4.4)	74.2 (5.2) ^a	75.6 (5.7) ^a
Gender male (%)	15 (29%)	34 (35%)	7 (58%) ^{a,b}
MMSE	29.3 (0.8)	26.6 (2.6) ^a	22.6 (5.5) ^{a,b}

registrations and (4) the final labeling by patches based on similarity measures in intensity and labeling.

During pre-processing of the images, non-brain regions are removed from all structural images and intracranial volumes are estimated. Removing non-brain tissue prior to registration is generally accepted as a means to simplify the inter-subject registration problem and thus increase the quality of the registrations [29,30]. The images are skull-stripped using BET [31]. To improve the results of BET, we follow the protocol proposed by Stein et al [32]. First, a coarse skull-stripping is obtained. Then, the bias field correction is calculated and applied to the image. Finally, the corrected image is reapplied with BET. Then, all images are spatially normalized to the same stereotactic space [33]. The MNI-152 template [33] is selected as a reference to which all images are then co-registered with an affine transformation using FLIRT with 12 degrees of freedom [34]. The intracranial volume (ICV) is calculated through registration of each skull-stripped MRI to the MNI-152 template using the same affine transformation [35].

After spatial normalization, a region of interest is defined for each structure studied (left and right hippocampus) as the minimum bounding box containing the structure for all of the training atlases expanded by three voxels along each dimension. The patient image is also processed using a skull-stripping filter, and an affine transformation is applied to the common reference space. Then, the normalized patient image is cropped around the structures of interest.

For each ROI, the normalized atlas images are first ranked based on their similarity according to the normalized target image using the mutual information (MI) measure [36]. Then, the N_R -selected atlases are registered non-rigidly into the ROI of the normalized target image. Next, the registered atlases are fused, and the labeling is calculated using graph cuts based on minimizing an energy function.

Next, an intensity normalization is applied to the above normalized atlas images using the histogram matching algorithm [37] with the ROI of the normalized target image as a reference. Now, the sum of the squared difference (SSD) measure between each atlas image and the target image is used to rank and select the first N_A atlases. This measure is chosen because SSD is related to the similarity between patches in intensities. Finally, from the previous labeling of the target image and the N_A -selected atlases, the segmentation result

random field model [38] (CRF) is defined by the following pseudo-Boolean function:

$$E(S) = \sum_{x \in \Omega} \psi_x(S(x); \theta_1(I, \mathbb{A})) + \lambda \sum_{x, y \in \mathcal{E}} \psi_{xy}(S(x), S(y); \theta_2(I)), \quad (1)$$

where $\Theta = \{\theta_1, \theta_2\}$ are the model parameters for this ROI and λ is a tunable parameter that determines the trade-off between the unary and pairwise potentials.

2.2.1 The unary potentials

The unary potentials $\psi_x(S(x); \theta_1(I, \mathbb{A}))$ use the Bayesian formulation, which allows prior information about the shape and appearance of structures to be segmented to be incorporated in the model.

Image likelihood

We assume that the observed intensities of I are independent random variables. The image likelihood $p(I|S; \mathbb{A})$ can then be written as a product of the likelihoods of the individual voxels:

$$p(I|S; \mathbb{A}) = \prod_{x \in \Omega} p(I(x)|S(x); \mathbb{A}).$$

A discriminative appearance model with low computational effort is selected. The registered atlas images are convolved with a filter-bank. A set of feature extraction kernels $\alpha_j(x)$ are used to produce different feature maps:

$$F_{\tilde{I}_i}(x) = \{\tilde{I}_i(x) * \alpha_j(x)\}_{j=1, \dots, d},$$

where d is the dimension of the feature vector and $F_{\tilde{I}_i}(x)$ denotes the resulting feature vector of $\tilde{I}_i$ at voxel x associated with the filter $\{\alpha_j(x)\}_{j=1, \dots, d}$. In this paper, 3D steerable filters [39] are adopted to extract the features. The responses for all registered atlas image voxels are whitened separately (to provide zero mean and unit covariance). These feature vectors are used to train a k -nearest neighbor (k -NN) appearance model. For computational efficiency, we use the kd -tree algorithm [40] to perform the nearest neighbor search. A kd -tree model is constructed with $\{F_{\tilde{I}_i}(x), \tilde{S}_i(x)\}_{i=1, \dots, N_R, x \in \Omega}$. The target image is also convolved and whitened.

For each label $l \in \{0, 1\}$, we consider $\mathcal{F}_l = \{F_{\tilde{I}_i}(x) / \tilde{S}_i(x) = l\}$ as the set of feature vectors extracted from the voxels that belong to the registered atlas images and whose labels are l . Let $\{F_r\}_{r \in R_l(x)} \subset \mathcal{F}_l$ be the set whose elements are nearest neighbors to $F_I(x)$ and $R_l(x)$ be the set of the indices of the feature vectors with label l that are nearest neighbors to $F_I(x)$. The image likelihoods of the individual voxels that belong to the target image are calculated using the following formula:

$$p(I(x)|l; \mathbb{A}) \propto \sum_{r \in R_l(x)} \exp(-\|F_I(x) - F_r\|_2^2). \quad (2)$$

Label prior

The label prior probability $p(S; \mathbb{A}, I)$ models the joint probability of all voxels that belong to the ROI in a particular label configuration. However, we assume that the prior probability that voxel x has label l only depends on its position, the similarity between I and $\tilde{I}_i$ and the transferred atlas-labeled images:

$$p(S; I, \mathbb{A}) = \prod_{x \in \Omega} p(S(x); I, \mathbb{A}).$$

This assumption is not realistic, but we encode the correlations of the labels using pairwise potentials. For each voxel x and each label $l \in \{0, 1\}$, we define:

$$u(S(x) = l; I, \mathbb{A}) = \sum_{i \in Q_l(x)} m(I, \tilde{I}_i, x)^q,$$

where $Q_l(x) = \{i | \tilde{S}_i(x) = l\}$, $m(I, \tilde{I}_i, x)$ is a local or global similarity measure between the target image and the registered atlas image at x , and q is an associated gain exponent [17]. The prior probability is defined as

$$p(S(x) = l; I, \mathbb{A}) = \frac{u(S(x) = l; I, \mathbb{A})}{\sum_{j \in \{0, 1\}} u(S(x) = j; I, \mathbb{A})}. \quad (3)$$

The image likelihood and label prior terms are combined to define the unary potentials $\psi_x(S(x); \theta_1(I, \mathbb{A}))$:

$$\psi_x(S(x); \theta_1(I, \mathbb{A})) = -\log \left(\frac{p(I(x)|S(x); \mathbb{A})p(S(x); I, \mathbb{A})}{p(I(x); \mathbb{A})} \right).$$

2.2.2 Spatial regularization

Following the work of Song et al. [41], a smoothness term is added to the energy function. These authors combined intensity and local boundary information into the pairwise potentials, which have been successfully applied for brain segmentation. These pairwise potentials take the form of a contrast-sensitive Potts model:

$$\psi_{xy}^R(S(x), S(y); \theta_2(I)) = \begin{cases} 0 & \text{if } S(x) = S(y), \\ \beta(S(x), S(y); \theta_2(I)) & \text{otherwise.} \end{cases}$$

where

$$\begin{aligned} \beta(S(x), S(y); \theta_2(I)) = & c \left(1 + \ln \left(1 + \frac{\|I(x) - I(y)\|^2}{2\sigma^2} \right) \right)^{-1} \\ & + (1 - c) \left(1 - \max_{r \in M_{xy}} g(\|\nabla I(r)\|) \right). \end{aligned} \quad (4)$$

where $g(\|\nabla I(x)\|) = 1 - \exp\left(-\frac{\|\nabla I(x)\|}{\sigma_G}\right)$ with a normalization factor σ_G . M_{xy} is a line that joins x and y , and σ is the robust scale of image I . The parameter $0 \leq c \leq 1$ controls the influence of the boundary-based and intensity-based parts.

2.3 Patch-based labeling method

Given a subset of the aligned atlases in a ROI of the target image using affine transformations, each voxel x in I is correlated to each voxel y in I_i with a weight of $\omega_i(x, y)$. In the conventional method, the weights are defined by intensity-based similarity measures between patches [9, 18]. A patch-based representation is extracted using a sub-volume centered at x or y from the target image or the aligned atlas. These signatures are denoted by $P_I(x)$ or $P_{I_i}(y)$, respectively. The signature difference between a reference voxel x and a voxel in an atlas image y is used to define $\omega_i(x, y)$:

$$\omega_i(x, y) = \begin{cases} e^{-\frac{\|P_I(x) - P_{I_i}(y)\|_2^2}{h^2(x)}} & \text{if } y \in \mathcal{N}(x), \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where $\mathcal{N}(x)$ denotes the neighborhood of voxel x in the atlas images. The parameter h determines how this method assigns the weights for each patch $P_{I_i}(y)$. We use h to show the minimal distance between the target patch $P_I(x)$ and its neighboring patches $P_{I_i}(y)$ [9].

Using these weights, we can use a single-point model for the label fusion from the atlases, as follows:

$$S(x) = \frac{\sum_{i=1}^{N_A} \sum_{y \in \Omega} \omega_i(x, y) \tilde{S}_i(y)}{\sum_{i=1}^{N_A} \sum_{y \in \Omega} \omega_i(x, y)},$$

where $S(x)$ is an estimation of label '1' for every voxel x and N_A is the number of the selected aligned atlases used to fuse. However, because the similarity measure is based on patches, one can obtain a multi-point model for the label fusion [18]:

$$P_S(x) = \frac{\sum_{i=1}^{N_A} \sum_{y \in \Omega} \omega_i(x, y) P_{\tilde{S}_i}(y)}{\sum_{i=1}^{N_A} \sum_{y \in \Omega} \omega_i(x, y)},$$

where $P_{\tilde{S}_i}(y)$ is a label patch that belongs to the i -th atlas and $P_S(x)$ is a label patch estimation of the target image at x , i.e., for each voxel, a vector of the likelihood for label '1' is calculated. These estimates are then aggregated using a combination classifier. In this work, we used the majority voting rule to fuse these estimates. The multi-point estimation provides better results than the single estimation [18].

2.4 Combining patch-based and registration-based label fusion

The similarity measure is the core of the patch-based labeling methods, and this measure is only calculated using intensity-based distances. The selected patches from the atlases can have similar distances to the patch of the reference voxel, but nevertheless, their label patches may be very different. In addition, these approaches estimate the labels in a local manner, without global constraints such as the shape priors. To overcome this drawback, a new similarity measure based on the labeling distance is added.

We consider the label patch as a clique, i.e., a connected set of voxels, whose labels are conditionally dependent on each other. A label patch estimation can be inferred by minimizing (1) using global shape constraints. Next, we add a measure using label-based distances:

$$\omega_i(x, y) = \begin{cases} e^{-\frac{\|P_I(x) - P_{\tilde{I}_i}(y)\|_2^2}{h_I^2(x)}} e^{-\frac{\|P_S(x) - P_{\tilde{S}_i}(y)\|_2^2}{h_S^2(x)}} & \text{if } y \in \mathcal{N}(x), \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

where $h_I(x)$ and $h_S(x)$ are defined as the minimal distances between the target patch and its neighboring patches according to the intensities and labelings, respectively (for further details, see section 2.4.2). Now, the weights of the patches are calculated considering both the local appearance and the global shape. Moreover, this proposal does not require the addition of more tunable parameters compared to the conventional path-based labeling method.

2.4.1 Setting the parameters of the label fusion method using non-rigid registrations

The non-rigid registrations, the atlas selection and the label fusion method have to be investigated to improve the performance of the multi-atlas segmentation approach. For each ROI, the atlases are ranked based on their similarity according to the target image using MI [42]. Once the atlases are ranked, we employ a leave-one-out validation strategy to determinate the number of atlases N_R that are fused to the target image. In our experiments, the first 15 atlases more similar to the target image were registered non-rigidly ($N_R = 15$) [43]. All non-rigid registrations are computed using *Elastix* [44], a publicly available package for medical image registration. Non-rigid registration of images is based on the maximization of MI, in combination with a deformation field parameterized by cubic B-splines [45]. The MI is implemented according to [46], using a joint histogram size of 32 x 32 and cubic B-spline Parzen windows. A unique resolution is employed using a B-spline control point spacing of 3.0 mm in all directions. To optimize the cost function, an iterative stochastic gradient descent optimizer is used [47]. In each iteration, 2000 random samples are used to calculate the derivative of the cost function. A maximum of 500 iterations of the stochastic optimization procedure is used. The above-described settings were determined through trial-and-error experiments on two image pairs.

The atlas-labeled images are modeled using the logarithm of odds (LogOdds) formulation, which is based on the signed distance transform [48]. This representation replaces the labels by the signed distances, which are assumed to be positive inside the structure of interest. We find that the LogOdds model produces more accurate results compared with trilinear interpolation or nearest-neighbor interpolation for transferring the atlas-labeled images [49].

Regarding the discriminative appearance model, the images are convolved with a 16-dimensional steerable filter bank [39]. From the proposed k-NN appearance model and given the feature vector of a voxel belonging to the target image, the training vectors in the kd -tree that are nearest are found. These vectors are used to calculate the distances to each label, and then equation (2) is applied for obtaining $p(I(x)|l; \mathbb{A})$ [50].

For estimating the label prior probabilities, similarity measures are required between regions of the target image and each registered atlas image, i.e., $m(I, \tilde{I}_i, x)$ in (3). A semi-global strategy is used to calculate the weight for each registered atlas. This strategy is most appropriate when the contrast between neighboring structures is low, as in the case of the hippocampus [17]. A binary mask is used for measuring this similarity between the target image and the registered atlases. This mask is constructed by joining all transferred labeled images. A voxel is considered in the binary mask if at least one vote of the foreground class is received. Because a statistical relationship is assumed among the intensities of these images, MI is used as a similarity measure. The gain exponent is set to $q = 4$ [17].

Finally, we consider a 3D grid-graph with 26 neighborhood systems in $\mathcal{E}$, and $\nabla I(x)$ is calculated from the derivatives of Gaussians at scale 1 for obtaining the pairwise potentials in (4). The final trained parameters were $\lambda = 0.2$, $c = 0.8$.

2.4.2 Setting the parameters of the patch-based labeling method

A subset of the target image domain $\Omega^* \subset \Omega$ is considered to accelerate the patch-based labeling method [9]. A binary mask is given with the union of all transferred labeled atlas images using non-rigid registrations, $\Omega^* = \bigcup \text{supp}(\tilde{S}_i)$. Figure 2 shows the first advantage of the collaboration between the atlas-warping by non-rigid registrations and the patch-based labeling method. When the domains Ω^* are calculated using non-rigid registrations, rather than using affine transformations as [9], the overlap between $\bigcup \tilde{S}_i$ and the ground-truth segmentation (S_R) is increased (i.e., the Dice coefficient [51] is higher, $DICE(\bigcup \tilde{S}_i \cap S_R, S_R)$) and the volume of $|\bigcup \tilde{S}_i|$ decreases with respect to the manually labeled hippocampal volumes (i.e., $\frac{|\bigcup \tilde{S}_i|}{|S_R|} 100[\%]$ is lower). In this experiment, two conclusions were drawn: a) because Ω^* obtained by the non-rigid registrations are more robust than those obtained by affine transformations, the success rates of the labelings are increased, and b) the computational time of the patch-based labeling method is reduced by having fewer voxels with uncertainty in their labels.

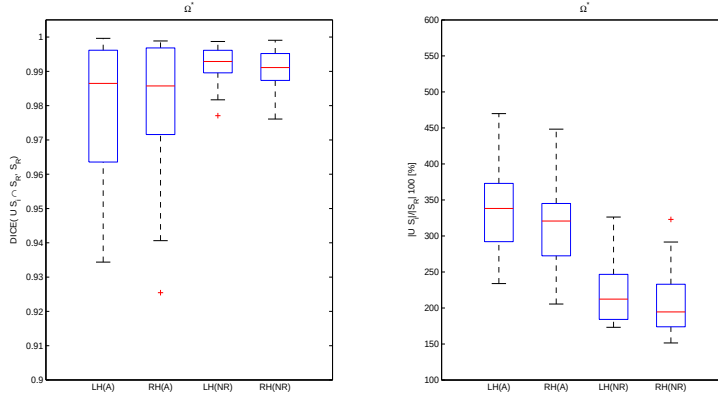


Fig. 2 Comparison of Ω^* between affine transformations (A) and non-rigid registrations (NR). For the left (LH) and right (RH) hippocampus, the first column plots the distribution of $DICE(\bigcup \tilde{S}_i \cap S_R, S_R)$, where S_R is the ground-truth segmentation. The second column plots the distribution of $\frac{|\bigcup \tilde{S}_i|}{|S_R|} 100[\%]$.

An atlas selection is also used to identify a subset of the normalized atlases to the normalized target image by affine transformations. First, the image intensities are normalized through histogram matching. Then, the SSD measure in Ω^* between each atlas image and the target image is used to rank the atlases. A compromise between the performance of the patch-based labeling method and computational efficiency is to choose the first 10 atlases that most resemble the target image [18, 19] ($N_A = 10$).

Next, for each voxel in Ω^* and whose label is uncertain, a patch pre-selection is used to accelerate the label fusion procedure and to improve the robustness of label fusion by excluding the unrelated patches [9, 52]. The pre-selection process uses a modified version of the well-know structural similarity measure [53]. We consider a similarity measure that takes both the intensity and the labeling of the candidate patch into account:

$$ss(x, y, i) = \frac{4\mu_I(x)\mu_{\tilde{I}_i}(y)\sigma_I(x)\sigma_{\tilde{I}_i}(y)}{\left(\mu_I^2(x) + \mu_{\tilde{I}_i}^2(y)\right) \left(\sigma_I^2(x) + \sigma_{\tilde{I}_i}^2(y)\right)} \cdot \frac{4\mu_S(x)\mu_{\tilde{S}_i}(y)\sigma_S(x)\sigma_{\tilde{S}_i}(y)}{\left(\mu_S^2(x) + \mu_{\tilde{S}_i}^2(y)\right) \left(\sigma_S^2(x) + \sigma_{\tilde{S}_i}^2(y)\right)} \quad (7)$$

where $(\mu_I(x), \sigma_I(x))$ and $(\mu_S(x), \sigma_S(x))$ are the mean and standard deviation of the target patches $P_I(x)$ and $P_S(x)$, respectively. Similarly, $(\mu_{\tilde{I}_i}(y), \sigma_{\tilde{I}_i}(y))$ and $(\mu_{\tilde{S}_i}(y), \sigma_{\tilde{S}_i}(y))$ are the mean and standard deviation of the candidate patch $P_{\tilde{I}_i}(y)$ and $P_{\tilde{S}_i}(y)$ belonging to the i -th selected atlas.

In the conventional patch-based labeling method, if the value of a similarity measure is larger than a given threshold ϵ , the candidate patch is selected [9]. To avoid adding more tuning parameters in our proposed approach, we consider a candidate patch if our similarity measure is greater than ϵ^2 , $ss(x, y, i) > \epsilon^2$. We evaluated the label fusion accuracy when applying different thresholds during the pre-selection. The value of ϵ was varied from 0.8 to 0.95, showing no significant improvement in the label fusion accuracy. In all experiments, we set the similarity threshold as $\epsilon^2 = 0.85^2$. On average, hundreds of patches of the selected aligned atlases are used to calculate the non-local mean label fusion in every voxel. Figure 3 shows the distributions of the number of selected patches according to the ranking of the aligned atlases using the conventional or proposed patch pre-selection procedure. The number of selected patches with our similarity measure is approximately one-third less than that of the conventional method. Our proposal selects less patches than the conventional pre-selection process and also improves the robustness of label fusion by excluding the patches whose labelings are not similar to the labeling of the target voxel. Our approach allows geometrical constraints, such as shape priors, to be imposed due to the labeling obtained by the atlas warping using non-rigid registrations. Indeed, the selected patches show similarity in appearance and labeling according to the target voxel. This is another advantage of the collaboration between the atlas-warping by non-rigid registrations and the patch-based labeling method.

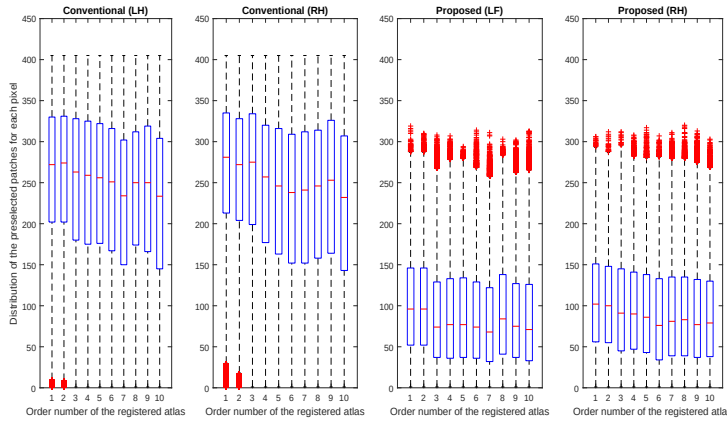


Fig. 3 Comparison between the conventional and proposed patch pre-selection procedures. The figure shows the distributions of the number of the selected patches for each voxel according to the ranking of the aligned atlases. We use $\epsilon = 0.85$.

The impacts of the patch size and the search volume size must be investigated. The patch size is related to the complexity of the anatomical structure, and the search volume size reflects the anatomy variability of the structure of interest [9]. These parameters are normally defined with a $(2r + 1) \times (2r +$

$1) \times (2r + 1)$ cube-shaped neighborhood by the radius r . In the case of the hippocampal segmentation, Coupé et al. [9] have reported a patch size of $7 \times 7 \times 7$ voxels and a search volume of $9 \times 9 \times 9$. Wang et al. [10] and Tong et al. [19] have proposed a patch size of $5 \times 5 \times 5$ and a search volume size of $7 \times 7 \times 7$, whereas Rousseau et al. [18] have used a patch size of $3 \times 3 \times 3$ and a search volume size of $11 \times 11 \times 11$. In our experiments, the patch size was tuned to $5 \times 5 \times 5$ and the search volume size was set to $7 \times 7 \times 7$.

The parameter h of (5) plays a crucial role in the weighting of the selected patches. When h is low, only a few samples are taken into account. If h is high, all selected patches tend to have the same weight, and the estimation is similar to a classical average. We use a modified version of h [9], which estimates h based on the minimal distance between the target patch $P_I(x)$ and its neighboring patches $P_{\tilde{I}_i}(y)$ [9]. A smoothing parameter β is added to the estimation of h , which depends on the noise level. The parameter β is common in image denoising tasks with the non-local mean principle [54] or in the kernel used by Rousseau et al. [18] in their patch-based labeling method. For low levels of noise in images, the best value of β is close to 0.5. For high levels of noise, this value is 1 [54]. Considering the above, we define h_I and h_S of (6) as:

$$\begin{aligned} h_I^2(x) &= \beta_I \cdot \min_{y \in \mathcal{N}(x), i=1, \dots, N_A} (\|P_I(x) - P_{\tilde{I}_i}(y)\|_2^2) + \varepsilon_I \\ h_S^2(x) &= \beta_S \cdot \min_{y \in \mathcal{N}(x), i=1, \dots, N_A} (\|P_S(x) - P_{\tilde{S}_i}(y)\|_2^2) + \varepsilon_S, \end{aligned} \quad (8)$$

where ε_I and ε_S are small constants to ensure numerical stability in case the patch under consideration is contained in its neighboring patches. The value of β_I is set to 0.5 and 1 for β_S . The aligned atlas images are considered to have low noise, but not the results of the label fusion method using non-rigid registrations, i.e., the similarity measures between labelings are considered to have a high noise level to relax restrictions imposed by the labeling obtained by the atlas warping.

2.5 Statistical analysis

Given the Dice coefficient distributions of the hippocampal segmentation algorithms, statistical significance is evaluated using the Wilcoxon signed-rank test, which does not require normality of the data. We consider that a p -value < 0.05 represents a significant improvement.

The effects of demographics and clinical measurements are tested using ANOVA with Bonferroni post hoc comparisons in continuous variables, and categorical variables are inspected using the chi-squared test. The hippocampal markers are first analyzed using MANCOVA with Bonferroni corrections by adopting a general linear model procedure, adjusting for age as a covariate. Bonferroni pairwise comparisons are performed to inspect hippocampal marker

differences between the groups. In all analysis of variance tests, homoscedasticity, normality and independence have been verified. Statistical analysis is performed using SPSS v. 22.

The hippocampal markers were subsequently analyzed using linear discrimination, which fits a normal density to each group, with a pooled estimate of covariance. The use of a simple, linear classifier ensured that the classification accuracy was primarily determined by the quality of the input data rather than by stochastic variations in the classifier. For each comparison between diagnostic groups, we computed the sensitivity (SEN), the specificity (SPE) and the accuracy value (ACC) [4]. Note that the number of subjects in each group was not the same. The classification accuracy did not enable to compare the performances between the different classification experiments. Instead, we considered both the specificity and the sensitivity.

Receiver operating characteristic (ROC) curves were calculated for each hippocampal marker and each pairwise group comparison. The discriminant value of the corresponding ROC curve was estimated using the area under curve (UAC). *DeLong's test was applied to compare AUCs between methods [55].*

A bootstrapping approach was used to evaluate the classification rate between pairs of clinical groups: for each group 75% of the subjects were randomly selected for training. The remaining 25% were then classified according to the linear discrimination model estimated in the training sets. The classification scores and interval confidences for different groups are displayed in tables after 5000 runs.

3 Results

Two types of experiments are selected for evaluating the performance of our proposal with respect to other hippocampal segmentation algorithms. First, the quality of the automatic segmentations from the compared methods is measured using the Dice coefficient. Then, the diagnostic group separation capabilities given by the methods are evaluated using a hippocampal volume marker.

In this comparative, we use Freesurfer v5.3 (FS) and four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb)[18], c) the label fusion method using non-rigid registrations described in subsection 2.2 (DS) and d) the proposed combined patch-based labeling method (PT). Freesurfer is chosen for being an application widely used in the neuroimaging community. In order to validate the proposed method, several label fusion methods are used, which share the same steps of image processing and registration. This benchmark allows us to observe the performance of our proposal regarding a conventional path-based labeling method and the label fusion method using non-rigid registrations.

Table 3 Average values and standard deviations of the Dice coefficient for all 134 images belonging to the ADNI-HarP database using FreeSurfer v5.3 (FS), and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT). Statistical significance is evaluated using the Wilcoxon signed-rank test. ^a Significant compared to FS. ^b Significant compared to MV. ^c Significant compared to Pb. ^d Significant compared to DS.

Status	Dice coefficient				
	FS	MV	Pb	DS	PT
NC	0.801(0.022)	0.830(0.024) ^a	0.850(0.016) ^{a,b}	0.853(0.029) ^{a,b}	0.862(0.017) ^{a,b,c}
MCI	0.774(0.044)	0.815(0.030) ^a	0.836(0.028) ^{a,b}	0.841(0.021) ^{a,b}	0.854(0.020) ^{a,b,c,d}
AD	0.770(0.023)	0.808(0.028) ^a	0.821(0.028) ^{a,b}	0.832(0.044) ^{a,b,c}	0.841(0.024) ^{a,b,c}
Global	0.782(0.041)	0.818(0.033) ^a	0.835(0.031) ^{a,b}	0.844(0.034) ^{a,b,c}	0.850(0.024) ^{a,b,c,d}

3.1 Accuracy of the automated segmentations

We compare the hippocampal segmentation algorithms using the 134 images belonging to ADNI-HarP. For each method in this experiment, we report the Dice coefficient. This measure is calculated using leave-one-out cross-validation. Table 3 presents the quantitative segmentation results for each method in the ADNI-HarP images (see Fig. 4 for the Dice distributions).

Our approach provides the best results with a mean Dice value of 0.850 and with significant differences in relation to FreeSurfer and the other label fusion methods. FreeSurfer returns average values of 0.782. These results are not surprising given that the conventions of manual protocols of the hippocampus differ between FreeSurfer and the rest of the methods. In this experiment, the HarP has been used as the ground-truth segmentations. It is also observed that our labeling proposal shows significant improvements over the other label fusion methods, included the conventional patch-based labeling method and the label fusion using nonrigid registrations.

Figure 5 shows the segmentation results of the FreeSurfer and our approach for one typical patient for each group (NC, MCI and AD). The subject identifications of the ADNI data are: 011_S_0021 (NC), 100_S_0892 (MCI) and 003_S_1059 (AD). The first row shows the image belonging to the ROI in the left hippocampus. In the second row, the image is overlapped with the manual protocol (HarP). The following rows show the segmentations of the FreeSurfer (FS) and our approach (PT). The results show that the proposed segmentations preserve globally the hippocampal shape.

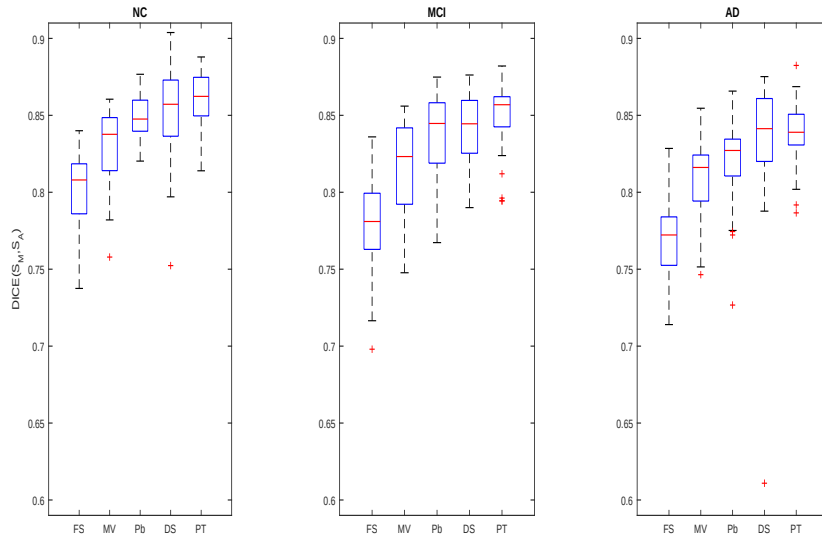


Fig. 4 The Dice distribution for 134 images belonging to the ADNI-HarP database using FreeSurfer v5.3 (FS) and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT). The row shows the Dice (S_M, S_A) distributions, where S_M is the manual hippocampal segmentation and S_A is the automated hippocampal segmentation.

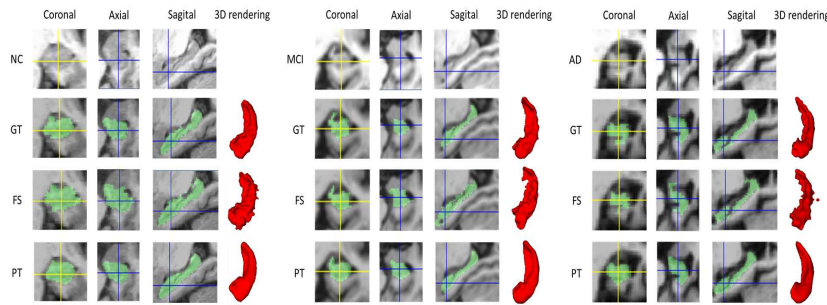


Fig. 5 Comparison between FreeSurfer and the proposed method. The first row shows the image belonging to the ROI in the left hippocampus. In the second row, the image is overlapped with the manual protocol (HarP). This is considered the Ground Truth (GT). The following rows show the segmentations of the FreeSurfer (FS) and our approach (PT).

3.2 Classification results

The normalized hippocampal volume marker is used to determine the pattern of atrophy in the progression of AD. For each subject, we compute the volume of the hippocampus. Volumes are normalized by the total intracranial volume.

Table 4 Method comparison in terms of normalized hippocampal volume marker for the ADNI-HarP and LCCN images by means FreeSurfer v5.3 (FS), and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT). The manual segmentations (HarP) are considered the Ground Truth (GT). The normalized volume is scaled using ICV in cm^3 . Data as represented as mean and standard deviation (SD). MANCOVA with Bonferroni pairwise comparisons is used for analysing the marker. Age is used as covariate. ^a Significant compared to NC. ^b Significant compared to MCI.

Data	Method	NC	MCI	AD	F	p
HarP	FS	4.78(0.58)	4.06(0.75) ^a	3.51(0.82) ^{a,b}	35.44	< 0.001
	MV	3.87(0.35)	3.42(0.49) ^a	3.15(0.50) ^{a,b}	30.13	< 0.001
	Pb	3.91(0.40)	3.41(0.58) ^a	3.08(0.52) ^{a,b}	32.77	< 0.001
	DS	4.01(0.32)	3.58(0.48) ^a	3.29(0.45) ^{a,b}	35.03	< 0.001
	PT	4.04(0.37)	3.56(0.53) ^a	3.23(0.51) ^{a,b}	35.73	< 0.001
	GT	4.19(0.55)	3.53(0.63) ^a	3.14(0.66) ^{a,b}	34.06	< 0.001
LCCN	FS	5.03(0.58)	4.44(0.75) ^a	3.88(0.82) ^{a,b}	15.89	< 0.001
	MV	3.60(0.28)	3.32(0.37) ^a	3.08(0.25) ^{a,b}	11.19	< 0.001
	Pb	3.91(0.37)	3.41(0.51) ^a	3.32(0.37) ^{a,b}	16.21	< 0.001
	DS	4.04(0.32)	3.70(0.48) ^a	3.39(0.45) ^{a,b}	16.76	< 0.001
	PT	4.07(0.37)	3.71(0.53) ^a	3.37(0.51) ^{a,b}	17.79	< 0.001

For more robustness with respect to segmentation errors, the left and right hippocampal volumes are averaged.

Table 4 and Fig. 6 present the statistical analysis results and the distributions of the normalized hippocampal volume among the diagnostic groups for both databases using the segmentation methods. MANCOVA with Bonferroni pairwise comparisons is used for analyzing the markers. Significant differences are found between NC vs MCI, NC vs AD, and MCI vs AD.

The most discriminative marker is the normalized hippocampal volume using the proposed approach (ADNI-HarP: $F = 35.73$, $p < 0.001$; LCCN: $F = 17.79$, $p < 0.001$). We also note that the marker distributions using our approach are similar between the two databases and even with the manual segmentations of the HarP annotations. By contrast, in FreeSurfer, the marker distributions are not as similar between the two databases as in our approach. Furthermore, as an effect of smoothing the hippocampal surface given by our approach, the dispersion of the markers is lower than in the manual segmentations.

We test the classification accuracy using the normalized hippocampal volume. The results of the classification experiments are summarized in Tables 5 and 6, for both databases. The FreeSurfer results for ADNI images are similar to others published previously [56,57].

In general, our approach presents slightly better classification results between diagnostic groups than FreeSurfer and the rest of label fusion methods in terms of the classification accuracy. According to the Delong's test, the classification improvement is only significant when NC vs AD are compared on the

LCCN database. Although the trends of the classification results are similar between ADNI-HarP and LCCN in all pairwise group comparisons, the improvements of our approach vary according to the database. Comparing NC vs MCI between FreeSurfer y our approach, the accuracy score increases by 8 absolute points and the AUC value increases from 0.759 to 0.792 in ADNI-HarP, whereas in the LCCN database, the improvement is only 1 absolute point and the AUC values are practically identical (from 0.794 to 0.797). Regarding the comparison between patients NC vs AD, the improvements are 2 and 4 absolute points of the accuracy scores, and the AUC values increase from 0.898 to 0.914 and from 0.915 to 0.960 in the ADNI-HarP and LCCN databases, respectively. In MCI vs AD, the improvements are minor, with 0 and 3 absolute points of the accuracy scores in the two databases, respectively. Even the AUC value decreases from 0.720 to 0.689 in ADNI, but with a positive increase from 0.685 to 0.715 in LCCN. In relation to the comparison between the fusion methods, MV gives the worst classification results. Our approach returns better scores of classification relative to the conventional path-based labeling method and the label fusion method using non-rigid registrations and graph cuts.

A specific analysis of the results obtained using our method is presented. The classification results for NC vs AD show the highest accuracy with respect to other comparative groups. The scores provide a high sensibility (over 87.8%), specificity (over 82.5%) and AUC (over 0.914). In the comparison between NC and patients with MCI, the scores on both databases are similar, and the bio-marker yields similar results with 70.2% sensibility, 75.6% specificity and an AUC in the range of 0.792 to 0.797. The classification results are more mixed in the comparison between MCI and AD. Whereas there are slight differences between the results of the ADNI-HarP and LCCN images in the other comparative groups, the classification results between MCI vs AD show some discrepancies. In ADNI-HarP images, the normalized volume provides scores with 73.9% sensibility, 60.4% specificity and 0.689 AUC, whereas in the LCCN images, the bio-marker returns scores with 55.8% sensibility, 67.7% specificity and 0.715 AUC.

3.3 Computational time

The highest consumption of computational time in our approach occurs in the skull-stripping step. This filter consume an average of 10 minutes per patient. The normalization of the brain into the MNI152 template with FLIRT requires less than 3 minutes. Regarding the hippocampal segmentation algorithm, the computational complexity is primarily due to the non-rigid registrations of the selected atlases into the target image. The computational time for segmentation increases linearly with the number of atlases that have to be registered. However, due to the availability and low cost of multi-core processors, this approach is becoming more feasible. The task of non-rigid registrations has been parallelized. The registration of the first 15 atlases in a ROI requires less than 45 seconds. The label fusion method using non-rigid registrations calculates

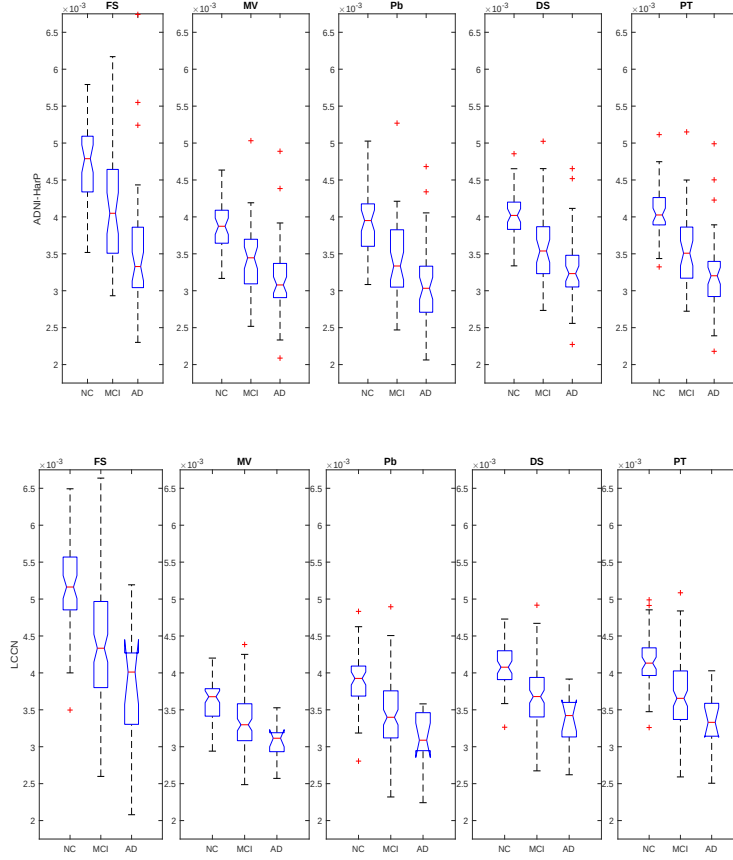


Fig. 6 The normalized hippocampal volume distributions for 134 and 162 images belonging to the ADNI-HarP and LCCN databases using FreeSurfer v5.3 (FS), and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT).

the unary and pairwise potentials of the CRF model in (1), and the labeling is found by applying the min-cut/max-flow algorithm of [58]. The CRF model is only trained for voxels whose labels have uncertainty such that the computational burden is reduced. A voxel is uncertain in its label when the atlas-labeled images are transferred and this voxel receives votes from different classes. Using the k-NN discriminative model with *kd*-tree, the semi-global strategy to infer the label prior probabilities and the graph cut techniques applied only in voxels with uncertainty makes the implementation very efficient, even without the optimized code. In the patch-based labeling method, many optimizations can be used because each voxel is treated independently, which allows multithreading. The optimized implementation should also be investigated [54].

Table 5 Results of the classifications between patients with different statuses using the normalized hippocampal volume as a bio-marker in the ADNI-HarP database by means FreeSurfer v5.3 (FS), and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT). The p -value of paired DeLong’s test are shown comparing our proposal in relation to the different methods.

Subjects	Method	SEN (%)	SPE (%)	ACC (%)	AUC	p
NC & MCI	FS	64.1(63.7 – 64.5)	65.9(65.5 – 66.3)	65.0(64.8 – 65.3)	0.759(0.756 – 0.762)	0.21
	MV	66.4(65.9 – 66.8)	72.8(72.4 – 73.1)	69.6(69.3 – 69.8)	0.765(0.761 – 0.768)	0.48
	Pb	68.7(68.3 – 69.1)	74.5(74.1 – 74.8)	71.6(71.3 – 71.8)	0.776(0.773 – 0.778)	0.86
	DS	69.9(69.5 – 70.2)	77.1(76.8 – 77.4)	73.4(73.2 – 73.7)	0.778(0.775 – 0.781)	0.91
	PT	70.2(69.9 – 70.6)	75.6(75.2 – 75.9)	72.9(72.7 – 73.1)	0.792(0.790 – 0.795)	1
NC & AD	FS	83.8(83.6 – 84.1)	83.2(82.8 – 83.5)	83.5(83.3 – 83.7)	0.898(0.897 – 0.900)	0.34
	MV	84.6(84.3 – 84.9)	82.5(82.2 – 82.8)	83.6(83.5 – 83.8)	0.898(0.896 – 0.898)	0.36
	Pb	85.3(85.0 – 85.6)	85.9(85.6 – 86.2)	84.5(84.3 – 84.7)	0.900(0.898 – 0.902)	0.45
	DS	88.6(88.4 – 88.8)	79.0(78.7 – 79.4)	84.0(83.8 – 84.2)	0.910(0.908 – 0.912)	0.67
	PT	87.8(87.5 – 85.0)	82.5(82.2 – 82.8)	85.2(85.0 – 85.4)	0.914(0.912 – 0.915)	1
MCI & AD	FS	72.4(72.0 – 72.7)	62.0(61.5 – 62.4)	67.3(67.1 – 67.6)	0.720(0.717 – 0.723)	0.23
	MV	64.7(64.3 – 65.1)	60.1(59.7 – 60.6)	62.4(62.2 – 62.7)	0.671(0.668 – 0.674)	0.28
	Pb	67.2(66.8 – 67.6)	58.2(57.8 – 58.6)	62.8(62.5 – 63.0)	0.680(0.678 – 0.682)	0.84
	DS	69.6(69.2 – 70.0)	58.5(58.1 – 58.9)	64.1(63.9 – 64.4)	0.684(0.681 – 0.687)	0.94
	PT	73.9(73.5 – 74.3)	60.4(60.0 – 60.8)	67.3(67.1 – 67.6)	0.689(0.686 – 0.692)	1

In our first version, we parallelize the labeling for each voxel of the target image with uncertainty. Our approach takes an average of approximately 180 seconds for fusing labels (included non-rigid registrations) of both the left and right hippocampus on images belonging to ADNI-HarP or LCCN data. The average time for the hippocampal segmentation from a T1-MRI is less than 17 minutes. Our experiments are carried out using a server of 16 cores at 2.27 GHz with 24 GB of RAM. The code of the label fusion methods is not yet optimized; thus, the computing times can be easily reduced. The scripts used in this study are available at https://www.nitrc.org/projects/lf_patches/.

4 Discussion and conclusion

In this paper, we proposed a fast hippocampal segmentation algorithm to discriminate patients with AD and MCI from normal aging based on classification of the normalized hippocampal volume. We developed a patch-based labeling method that cooperates with atlas-warping using non-rigid registrations. First, a labeling of the target image is inferred with atlas-warping by non-rigid registrations. Then, a patch-based label fusion method is applied, whose patches

Table 6 Results of the classifications between patients with different statuses using the normalized hippocampal volume as a bio-marker in the LCCN database by means FreeSurfer v5.3 (FS), and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT). The p-value of paired DeLong’s test are shown comparing our proposal in relation to the different methods.

Subjects	Method	SEN (%)	SPE (%)	ACC (%)	AUC	p
NC & MCI	FS	70.6(70.3 – 70.8)	79.7(79.4 – 80.0)	73.8(73.6 – 74.0)	0.794(0.792 – 0.796)	0.99
	MV	69.8(69.5 – 70.0)	69.7(69.4 – 70.0)	69.7(69.6 – 69.9)	0.744(0.742 – 0.746)	0.03
	Pb	68.6(68.4 – 68.8)	75.7(75.4 – 76.0)	71.8(71.6 – 72.0)	0.788(0.786 – 0.789)	0.78
	DS	70.2(70.0 – 70.5)	78.1(77.8 – 78.4)	73.0(72.8 – 73.2)	0.793(0.791 – 0.794)	0.94
	PT	71.6(71.4 – 71.9)	80.3(80.0 – 80.6)	74.7(74.5 – 74.9)	0.797(0.795 – 0.798)	1
NC & AD	FS	83.9(83.2 – 84.5)	87.2(83.2 – 84.5)	86.6(83.2 – 84.5)	0.915(0.912 – 0.917)	0.04
	MV	77.7(77.0 – 78.3)	78.5(78.2 – 78.8)	78.4(78.2 – 78.7)	0.884(0.882 – 0.887)	< 0.001
	Pb	79.2(78.6 – 79.9)	86.2(85.9 – 86.5)	84.6(84.2 – 84.8)	0.923(0.920 – 0.926)	0.29
	DS	91.7(91.3 – 92.2)	88.1(87.9 – 88.3)	88.7(88.6 – 88.9)	0.946(0.944 – 0.948)	0.75
	PT	90.6(90.1 – 91.1)	90.7(90.5 – 91.0)	90.7(90.5 – 90.8)	0.960(0.958 – 0.961)	1
MCI & AD	FS	52.2(51.3 – 53.0)	65.1(64.8 – 65.4)	63.6(63.4 – 63.9)	0.685(0.681 – 0.690)	0.48
	MV	59.0(58.8 – 59.3)	62.1(61.8 – 62.4)	60.6(60.4 – 60.8)	0.650(0.646 – 0.654)	0.21
	Pb	61.5(61.2 – 61.8)	62.3(61.5 – 63.2)	61.6(63.4 – 63.9)	0.664(0.660 – 0.668)	0.86
	DS	58.7(57.8 – 59.5)	65.2(65.0 – 65.5)	64.5(64.2 – 64.7)	0.715(0.711 – 0.719)	0.94
	PT	55.8(54.9 – 56.7)	67.7(67.4 – 67.9)	66.4(66.1 – 66.6)	0.715(0.711 – 0.718)	1

and weights are computed from a combination of similarity measures between patches using intensity-based distances and labeling-based distances, where the labeling distances are calculated from the previous binary labeling of the target image by atlas-warping using non-rigid registrations.

The patch-based labeling methods have the advantages of considering multiple samples during the labeling estimation and the local context is well represented by the patches, particularly with affine transformations. In contrast, the label fusion methods using non-rigid registrations lead to segmentations with shape prior constraints. When the delineation of the anatomical structures do not rely on intensity contrast, as in the case of hippocampal segmentation, the conventional patch-based labeling is not sufficient for obtaining good results. We have experimentally observed that the collaboration between these two approaches through the addition of a similarity measure based on the distance between binary labeling produces higher quality segmentations.

The collaboration between the two methods is given in the following levels: (1) The pre-selection of the patches in the atlases are improved by adding similarity measures based on both the intensity and labeling of the candidate patches. In our experiments, we observe that the numbers of selected patches with our similarity measure are approximately one-third less than that with the conventional method without adding any additional parameter to tune. (2) The weights of our selected patches are also more robust through the addition of label-based distances. Moreover, our proposed method does not require further tuning parameters compared to the conventional patch-based methods, neither in the selected patches nor in calculating the weights. (3)

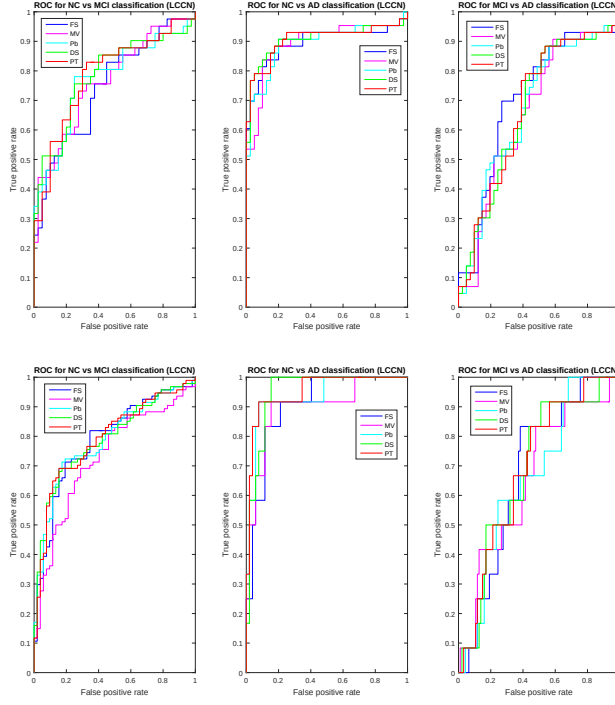


Fig. 7 The first column shows the ROC curves for the ADNI-HarP database using the normalized hippocampal volume and the second column illustrates the ROC curves for the LCCN database. Each ROC curve is obtained from Freesurfer (FS) and the four label fusion methods: a) Majority Voting (MV), b) a conventional path-based labeling method (Pb), c) the label fusion method using non-rigid registrations (DS) and d) the proposed combined patch-based labeling method (PT).

Both in the pre-selection patch process and in calculating their weights, our approach imposes geometrical restrictions, such as shape priors, using similarity measures based on binary labeling. In the conventional patch-based labeling method, there are no global constraints. (4) the sub-domain considered to accelerate the algorithm, Ω^* , generated by non-rigid registrations is smaller and has more overlap with the manual segmentation than that obtained by affine registrations. The consequence is a higher computational efficiency due to the smaller size of Ω^* and improved segmentation results by increasing the overlapping between the union of the transferred labeled atlases and the ground-truth segmentations. And (5), the spatial normalization of the atlases and the target images into a reference makes the work-flow very efficient. The ranking of the atlases for both fusion methods and the library of the patches are obtained from the spatial normalization. Indeed, the proposed approach is a multi-feature algorithm. The first three collaborations could be obtained from any initial segmentation, but the last two advantages are inherent in the

multi-atlas label fusion method. The algorithm combines the information from the registered atlases by means of affine as non-rigid registrations.

MANCOVA with Bonferroni pairwise comparisons was used for analyzing the normalized hippocampal volume markers extracted from several approaches. Significant differences were found between NC vs MCI, NC vs AD, and MCI vs AD.

The most discriminative marker was the maker using the proposed approach (ADNI-HarP: $F = 35.73$, $p < 0.001$; LCCN: $F = 17.79$, $p < 0.001$). We also observed that the marker distributions using our approach were similar between the two databases and even with the manual segmentations of the HarP annotations. By contrast, in FreeSurfer, the marker distributions were not as similar between the two databases as in our approach.

For our approach with HarP annotations, for all compared groups and the two image databases used, the classification results provide slightly better results than those obtained with the other methods. Our approach achieved classification accuracies of over 85.2% and 0.914 in AUC for AD vs NC and 72.9% and 0.792 in AUC for MCI vs NC. The scores are similar in both databases for these compared groups. The worst results were obtained in the separation between MCI vs AD, and there were score discrepancies between the two databases. In the ADNI-HarP data, the classification results were 73.9% sensibility, 60.4% specificity and 0.689 AUC, whereas in the LCCN images, the bio-marker provided scores with 55.8% sensibility, 67.7% specificity and 0.715 AUC.

Given the ADNI-HarP images, Roche et al. [59] and Giraud et al. [60] have reported mean and standard deviation values of 0.866 (0.018) and 0.898 (0.015) for the Dice distributions, respectively, whereas with our approach, the mean and standard deviation values are 0.850 (0.024). However, based on the estimations of the AUC values between pairwise group comparisons, our approach achieves better classification results between NC vs AD groups and MCI vs AD groups with 0.914 and 0.689 for the AUC estimations, respectively (see Table 4).

The reported results indicate that there is not a high correlation between the Dice distributions and estimations of AUC between pairwise patient group comparisons. We consider that the similarity with the manual segmentations is a goal to achieve, but the main objective of a hippocampal segmentation algorithm is to obtain markers that are capable of improving the discrimination between patient groups in the progression of AD. In this line, our approach preserves the shape learned from HarP and the hippocampal surfaces are smoothed relative to the manual segmentations, which produces less variability of the normalized hippocampal volumes in each diagnostic group (see Table 4), improving the classification results between NC vs AD and MCI vs AD.

Finally, in the first version of our sources, without the optimized code and only parallelization of the non-rigid registration task and the patch-based labeling, the average time for the hippocampal segmentation from a T1-MRI

Table 7 Method comparison in terms of AUC on normalized hippocampal volume marker for the ADHI-HarP images.

AUC	Manual Segmentations	Roche et al [59]	Giraud et al [60]	Our proposal
NC vs MCI	0.785	0.76	0.821	0.792
NC vs AD	0.894	0.84	0.898	0.914
MCI vs AD	0.682	0.61	0.634	0.689

is less than 17 minutes. The scripts used in this study are available at https://www.nitrc.org/projects/lf_patches/.

Future work will follow three directions. First, a new strategy of patch-based labeling has been proposed using the minimal reconstruction error [19]. The patch library of the aligned atlases is considered as a dictionary, and the target patch is modeled as a sparse linear combination of the atoms in the dictionary. An extension of the patch reconstruction can be achieved by adding the inferred labeling from the label fusion using non-rigid registrations, in the same manner as performed on the similarity measures. Second, another limitation in this work is that the weights of the patches have been computed independently for each aligned atlas, without taking into account the fact that the selected atlases may produce similar label errors due to a high correlation between them. Wang et al. [10] proposed that the weighted voting is formulated in terms of minimizing the total expectation of labeling error and in which pairwise dependency between atlases is explicitly modeled as the joint probability of two atlases, creating a segmentation error at a voxel. Third, a new approach based on PatchMatch [61], called Optimized PATCHMatch Label Fusion(OPAL), reduces the computation time required for the search of similar patches [62,60]. Furthermore, OPAL is a multi-feature framework, which allows the combination of the patches based on intensity and labeling distances. Therefore, the possibility of improving the selection of patches under a multi-feature framework using the core of this contribution, i.e. the combination of information from the atlases with affine transformations and nonrigid registrations, could improve the quality of the automatic segmentations. We leave these three themes (i.e., the minimal reconstruction error, the pairwise dependency between the patches of the atlases and OPAL) for future work.

5 Information Sharing Statement

Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:

http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

All source used in this manuscript are freely available to the public. These data are available from the website cited in Section 3.

6 Acknowledgments

Data collection for this study was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and the Laboratory of Cognitive and Computational Neuroscience (LCCN) of the Center of Biomedical Technology (Technical University of Madrid). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: Alzheimers Association; Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company.

References

1. Jack, C., Shiung, M., Gunter, J., O'Brien, P., Weigand, S., Knopman, D., Boeve, B., Ivnik, R., Smith, G., Cha, R., et al.: Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* **62** (2004) 591–600
2. Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., et al.: Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *The Lancet Neurology* **6** (2007) 734–746
3. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* **6** (2010) 67–77
4. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., et al.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* **56** (2011) 766–781
5. Clerx, L., van Rossum, I.A., Burns, L., Knol, D.L., Scheltens, P., Verhey, F., Aalten, P., Lapuerta, P., van de Pol, L., van Schijndel, R., et al.: Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neurobiology of Aging* **34** (2013) 2003–2013
6. Collins, D.L., Pruessner, J.C.: Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage* **52** (2010) 1355–1366
7. Lotjonen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* **49** (2010) 2352–2365
8. Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schuff, N., Fox, N.C., Ourselin, S., et al.: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *Neuroimage* **51** (2010) 1345–1359

9. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* **54** (2011) 940–954
10. Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (2013) 611–623
11. Nestor, S.M., Gibson, E., Gao, F.Q., Kiss, A., Black, S.E.: A direct morphometric comparison of five labeling protocols for multi-atlas driven automatic segmentation of the hippocampus in Alzheimer’s disease. *Neuroimage* **66** (2013) 50–70
12. Konrad, C., Ukas, T., Nebel, C., Arolt, V., Toga, A.W., Narr, K.: Defining the human hippocampus in cerebral magnetic resonance images: an overview of current segmentation protocols. *Neuroimage* **47** (2009) 1185–1195
13. Boccardi, M., Ganzola, R., Bocchetta, M., Pievani, M., Redolfi, A., Bartzokis, G., Camicioli, R., Csernansky, J.G., de Leon, M.J., de Toledo-Morrell, L., et al.: Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. *Journal of Alzheimer’s Disease* **26** (2011) 61–75
14. Frisoni, G.B., Jack, C.R., Bocchetta, M., Bauer, C., Frederiksen, K.S., Liu, Y., Preboske, G., Swihart, T., Blair, M., Cavedo, E., et al.: The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: Evidence of validity. *Alzheimer’s & Dementia* **11** (2015) 111–125
15. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, C.: Quo vadis, atlas-based segmentation? *Handbook of Biomedical Image Analysis* (2005) 435–486
16. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging* **23** (2004) 903–921
17. Artaechevarria, X., Muñoz-Barrutia, A., Ortiz-de Solorzano, C.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Transactions on Medical Imaging* **28** (2009) 1266–1277
18. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. *IEEE Transactions on Medical Imaging* **30** (2011) 1852–1862
19. Tong, T., Wolz, R., Coupé, P., Hajnal, J.V., Rueckert, D.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *Neuroimage* **76** (2013) 11–23
20. Buades, A., Coll, B., Morel, J.: A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation* **4** (2006) 490–530
21. Asman, A.J., Landman, B.A.: Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis* **17** (2013) 194–208
22. Wang, L., Shi, F., Gao, Y., Li, G., Gilmore, J.H., Lin, W., Shen, D.: Integration of sparse multi-modality representation and anatomical constraint for iso-intense infant brain MR image segmentation. *NeuroImage* **89** (2014) 152–164
23. Pineda-Pardo, J.A., Garcés, P., López, M.E., Aurteneixe, S., Cuesta, P., Marcos, A., Montejo, P., Yus, M., Hernández-Tamames, J.A., del Pozo, F., et al.: White matter damage disorganizes brain functional networks in amnesic mild cognitive impairment. *Brain connectivity* **4** (2014) 312–322
24. Pineda-Pardo, J.A., Bruña, R., Woolrich, M., Marcos, A., Nobre, A.C., Maestú, F., Vidaurre, D.: Guiding functional connectivity estimation by structural connectivity in meg: an application to discrimination of conditions of mild cognitive impairment. *NeuroImage* **101** (2014) 765–777
25. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* **27** (2008) 685–691
26. Boccardi, M., Bocchetta, M., Apostolova, L.G., Barnes, J., Bartzokis, G., Corbetta, G., DeCarli, C., Firbank, M., Ganzola, R., Gerritsen, L., et al.: Delphi definition of the eadc-adni harmonized protocol for hippocampal segmentation on magnetic resonance. *Alzheimer’s & Dementia* **11** (2015) 126–138

27. Tangaro, S., Amoroso, N., Boccardi, M., Bruno, S., Chincarini, A., Ferraro, G., Frisoni, G., Maglietta, R., Redolfi, A., Rei, L., et al.: Automated voxel-by-voxel tissue classification for hippocampal segmentation: methods and validation. *Physica Medica* **30** (2014) 878–887
28. Boccardi, M., Bocchetta, M., Morency, F.C., Collins, D.L., Nishikawa, M., Ganzola, R., Grothe, M.J., Wolf, D., Redolfi, A., Pievani, M., et al.: Training labels for hippocampal segmentation based on the EADC-ADNI harmonized hippocampal protocol. *Alzheimer's & Dementia* **11** (2015) 175–183
29. Battaglini, M., Smith, S.M., Brogi, S., De Stefano, N.: Enhanced brain extraction improves the accuracy of brain atrophy estimation. *Neuroimage* **40** (2008) 583–589
30. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46** (2009) 786–802
31. Smith, S.M.: Fast robust automated brain extraction. *Human Brain Mapping* **17** (2002) 143–155
32. Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Senstad, R.E., Winkler, A.M., Toro, R., Appel, K., Bartecek, R., Bergmann, Ø., et al.: Identification of common variants associated with human hippocampal and intracranial volumes. *Nature Genetics* **44** (2012) 552–561
33. Evans, A.C., Janke, A.L., Collins, D.L., Baillet, S.: Brain templates and atlases. *Neuroimage* **62** (2012) 911–922
34. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17** (2002) 825–841
35. Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z.: A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* **23** (2004) 724–738
36. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. *International Journal of Computer Vision* **24** (1997) 137–154
37. Gonzales, R.C., Woods, R.E.: *Digital image processing*. New Jersey: Prentice Hall **6** (2002) 1–689
38. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *International Conference on Machine Learning*. (2001) 282–289
39. Derpanis, K.G., Gryn, J.M.: Three-dimensional nth derivative of Gaussian separable steerable filters. In: *IEEE International Conference on Image Processing*. Volume 3., IEEE (2005) III553–6
40. Mount, D.M., Arya, S.: Ann: A library for approximate nearest neighbor searching. <http://www.cs.umd.edu/~mount/ANN/> (2010 (Accessed: 20 January 2015)) version 1.1.2.
41. Song, Z., Tustison, N., Avants, B., Gee, J.: Integrated graph cuts for brain MRI segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI* **4191** (2006) 831–838
42. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage* **46** (2009) 726–738
43. Platero, C., Tobar, M.C.: A label fusion method using conditional random fields with higher-order potentials: Application to hippocampal segmentation. *Artificial Intelligence in Medicine* **64** (2015) 117–129
44. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging* **29** (2010) 196–205
45. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* **18** (1999) 712–721
46. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing* **9** (2000) 2083–2099

47. Klein, S., Staring, M., Pluim, J.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Transactions on Image Processing* **16** (2007) 2879–2890
48. Pohl, K.M., Fisher, J., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Logarithm odds maps for shape representation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI*. Springer (2006) 955–963
49. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging* **29** (2010) 1714–1729
50. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2008) 1–8
51. Dice, L.: Measures of the amount of ecologic association between species. *Ecology* **26** (1945) 297–302
52. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., Shen, D.: A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Medical Image Analysis* **18** (2014) 881–890
53. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13** (2004) 600–612
54. Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging* **27** (2008) 425–441
55. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* (1988) 837–845
56. Khan, W., Westman, E., Jones, N., Wahlund, L.O., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Spenger, C., et al.: Automated hippocampal subfield measures as predictors of conversion from mild cognitive impairment to Alzheimer’s disease in two independent cohorts. *Brain topography* **28** (2015) 746–759
57. Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., et al.: A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *NeuroImage* **115** (2015) 117–137
58. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** (2004) 1124–1137
59. Roche, F., Schaerer, J., Gouttard, S., Istace, A., Belaroussi, B., Yu, H.J., Bracoud, L., Pachai, C., DeCarli, C.: Accuracy of bmas hippocampus segmentation using the harmonized hippocampal protocol. *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association* **10** (2014) P56
60. Giraud, R., Ta, V.T., Papadakis, N., Manjón, J.V., Collins, D.L., Coupé, P., Initiative, A.D.N., et al.: An optimized patchmatch for multi-scale and multi-feature label fusion. *NeuroImage* **124** (2016) 770–782
61. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.: Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics-TOG* **28** (2009) 24
62. Ta, V.T., Giraud, R., Collins, D.L., Coupé, P.: Optimized patchmatch for near real time and accurate label fusion. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer (2014) 105–112