Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Adaptive Bayesian label fusion using kernel-based similarity metrics in hippocampus segmentation

David Cárdenas-Peña
Andres Tobar-Rodríguez
German Castellanos-Dominguez
Alzheimer's Disease Neuroimaging Initiative

SPIE.

David Cárdenas-Peña, Andres Tobar-Rodríguez, German Castellanos-Dominguez, Alzheimer's Disease Neuroimaging Initiative, "Adaptive Bayesian label fusion using kernel-based similarity metrics in hippocampus segmentation," *J. Med. Imag.* **6**(1), 014003 (2019), doi: 10.1117/1.JMI.6.1.014003.

Adaptive Bayesian label fusion using kernel-based similarity metrics in hippocampus segmentation

David Cárdenas-Peña,* Andres Tobar-Rodríguez, German Castellanos-Dominguez, and Alzheimer's Disease Neuroimaging Initiative†
Universidad Nacional de Colombia, Signal Processing and Recognition Group, Manizales, Colombia

Abstract. The effectiveness of brain magnetic resonance imaging (MRI) as a useful evaluation tool strongly depends on the performed segmentation of associated tissues or anatomical structures. We introduce an enhanced brain segmentation approach of Bayesian label fusion that includes the construction of adaptive target-specific probabilistic priors using atlases ranked by kernel-based similarity metrics to deal with the anatomical variability of collected MRI data. In particular, the developed segmentation approach appraises patch-based voxel representation to enhance the voxel embedding in spaces with increased tissue discrimination, as well as the construction of a neighborhood-dependent model that addresses the label assignment of each region with a different patch complexity. To measure the similarity between the target and training atlases, we propose a tensor-based kernel metric that also includes the training labeling set. We evaluate the proposed approach, adaptive Bayesian label fusion using kernel-based similarity metrics, in the specific case of hippocampus segmentation of five benchmark MRI collections, including ADNI dataset, resulting in an increased performance (assessed through the Dice index) as compared to other recent works. [DOI: 10.1117/1.JMI.6.1.014003]

Keywords: brain tissue segmentation; Bayesian segmentation; interslice kernel.

Paper 18158RR received Jul. 25, 2018; accepted for publication Dec. 27, 2018; published online Feb. 4, 2019.

1 Introduction

Magnetic resonance imaging (MRI) is a dynamic and fast-growing imaging modality that has been increasingly admitted in research and clinical applications such as detection of subtle or small brain abnormalities and identification of pathological structure conditions,[1,2] monitoring of disease progression,[3] and diagnosis of soft-tissue lesions and inflammatory foci.[4,5] Nonetheless, in cases of neuroimaging, the effectiveness of MRI tools strongly depends on the performed segmentation of associated brain tissues or anatomical structures, for which qualitative expert assessments may become impractical because of increased time and human resource consumption. To name a few applications: the detection of subtle or small brain abnormalities and identification of pathological structure conditions,[1,2] monitoring of disease progression,[3] and diagnosis of soft-tissue lesions and inflammatory foci.[4,5] Nonetheless, the effectiveness of MRI tools for neuroimaging applications strongly depends on the performed segmentation of associated tissues or anatomical structures.

In a first approach, brain tissue intensities can be assumed to be linearly separable, supporting the classification of tissues, having very different diamagnetic properties such as white matter, gray matter, and cerebrospinal fluid. Nevertheless, this assumption does not hold in the volume compartments with very similar magnetic susceptibility (as the hippocampus, basal ganglia, and putamen), producing a histogram overlapping that degrades automatic segmentation.[6] An effective way to overcome this issue is to incorporate the prior information about the expected location and structure of volume shapes using image templates or atlases, which are MRI pairs (intensity images joined to their label maps) that must be spatially warped to the target spatial coordinates. To encompass the inherent ambiguity of volume inhomogeneities, Bayesian algorithms are employed together with local intensity models,[7] resulting in a single spatial model prior. For instance, the unified segmentation approach simultaneously performs registration, filtering, and delineation of brain structures using a single probabilistic template together with Gaussian mixtures to model space-varying intensity.[8–10] To extract a broader anatomical heterogeneity, however, priors are to be collected from large human groups, which may incorporate high inter- and intrasubject variability of intensity distributions.[11]

Modern segmentation approaches consider subject-specific atlas strategies to tackle the intrasubject variability, which demand the registration of available training MRI sets to a unique image. Hence, the full processing stage is carried out over the target coordinate space, but weighting the contribution of each training image. Thus, the most similar templates to the target image are engaged, applying different measures of proximity: active shape patterns,[12] manifold learning,[13] appearance models,[14] and probabilistic atlases.[15] All these metrics highly rely on the accuracy supplied by each applied pairwise image alignment that, in practice, is deteriorated by the misregistration artifact, increasing the computational cost that becomes huge for large training datasets.

*Address all correspondence to David Cárdenas-Peña, E-mail: dcardenasp@unal.edu.co

†Part of the data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.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.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

2329-4302/2019/$25.00 © 2019 SPIE

Aiming at dealing with atlas misalignments, the pairwise agreement is computed between each atlas to the target using a template search strategy (nonlocal methods), for which the voxel labeling depends on the intensity similarity over a predefined neighborhood. To infer the tissue class, the similarity is computed between intensity patches centered at target and atlas voxels. Therefore, the patch-based approaches differ in the selected label fusion strategy. Thus, basic methods assign the labels using a weighted voting strategy ruled by patch similarity metrics measured over neighboring patches.[16,17] Also, the hierarchical patch-based label fusion involves multiscale feature representation to characterize the local anatomical information, dynamically adjusting the patch size during the label procedure.[18] More elaborate approaches are spatially localized random forests,[19,20] patchwise metric learning with multiscale features,[21,22] probabilistic labeling following confidence maps,[23] and progressive labeling by multilayer dictionaries.[24] In attempting to undertake the space-varying labeling difficulty, however, these approaches demand a significant number of atlas voxels, therefore reducing the benefit of the voting strategies.[25] Although the above approaches attempt to undertake the space-varying labeling difficulty by increasing the number of atlas voxels, still many atlas voxels may cause a segmentation loss if they are neither representative for the query tissue nor suitably aligned with the target voxel. In either case, instead of reinforcing the labeling decision, the arisen uncertainty decreases the achieved performance.

In this paper, for the concrete case of hippocampus, we propose a brain tissue segmentation methodology that aims at selecting the most related atlases with a query, using more elaborate kernel-based similarity metrics and adapting the Bayesian model parameters to the spatial labeling complexity. The methodology, termed adaptive Bayesian label fusion using kernel-based similarity metrics (ABKS), comprises two stages: (i) an atlas selection strategy using a tensor-based similarity between the target and atlases at regions of interest (ROI) in an attempt to better the anatomical variability extraction from a pool of image collections, accurately fitting the target brain structure and enhancing the tissue delineation. (ii) Estimation of adaptive Bayesian models according to the neighborhood labeling complexity. In the cases of labels with low uncertainty, estimation relies on a maximum *a posteriori* classifier that is supplied by spatially varying priors at neighborhoods. Otherwise, a nonlocal patch scheme is applied with conditional distributions learned from structural information and uniform priors to tackle regions with high uncertainty. Further, the incorporated collectionwise agreement threshold delimits the neighborhood labeling complexity, ruling more accurately the segmentation model. To assess the benefit of the proposed methodology, we consider the popular segmentation of labeling the hippocampus from public MRI collections, showing that proposed ABKS approach achieves a Dice index similarity than the compared state-of-the-art approaches. The remainder of the paper is organized as follows: Sec. 2 introduces the proposed tissue segmentation approach, ABKS, explaining the ranking procedure of the training atlases in the template search strategy. Then, Sec. 3 describes the experimental setup validated on four publicly available datasets, providing results that are discussed in Sec. 4. Last, conclusions and future work are outlined in Sec. 5.

2 Methods

2.1 *Bayesian Segmentation Using Atlas-Based Priors*

Let $\mathcal{X} = \{x_r : r \in \Omega\}$ be a raw intensity image set, holding a measurement $x_r \in, \mathbb{R}$ at location r defined over the spatial domain Ω. Provided a tissue class set $\mathcal{C}, = \{1, \ldots, C\}$, the image segmentation task intends to find the optimal partition, that is, it builds the label image set $\mathcal{L} = \{l_r \in \mathcal{C}\}$ by assigning a label l_r to r'th spatial element, relying on its measured intensity x_r.

In image processing, Bayesian methods perform the optimal segmentation task determined as the partition that has the highest *a posteriori* probability, incorporating the anatomical information extracted from an available atlas (or prior image). Thus, the posterior conditional probability $P(\mathcal{L}|\mathcal{X})$ (or segmentation probability) is expressed in terms of $P(\mathcal{X}|\mathcal{L})$, the conditional probability computed from the image $\mathcal{X}$, given a segmentation set $\mathcal{L}$ as well as the corresponding prior probability $P(\mathcal{L})$. So, the segmentation probability reads as follows:

$$P(\mathcal{L}|\mathcal{X}) \propto P(\mathcal{X}|\mathcal{L})P(\mathcal{L}). \tag{1}$$

Under the assumption that all spatial elements are i.i.d., the segmentation probability of a given image can be expressed as $P(\mathcal{L}|\mathcal{X}) = \prod_{r\in\Omega} P(l_r = c|x_r)$, so that the maximum a posteriori criterion assigns a label $l_r^\star$ as

$$\begin{aligned} l_r^\star &= \arg\max_{c\in\mathcal{C}} P(l_r = c|x_r) \\ &= \arg\max_{c\in\mathcal{C}} P(x_r|l_r = c)P(l_r = c), \end{aligned} \tag{2}$$

where the conditional probability $P(x_r|l_r = c)$ describes the intensity distribution for class c at position r, and $P(l_r = c)$ is the label prior probability regardless of its intensity. Considering that the tissue conditional probability locally follows a Gaussian distribution $\mathcal{N}(x_r|\mu_{rc}, \sigma_{rc}^2)$, both probability parameters are derived from an available atlas set, $\mathcal{A} = \{\mathcal{X}^n, \mathcal{L}^n : n \in [1, \ldots, N]\}$, as[7]

$$P(l_r = c) = \mathbb{E}_n\{P(l_r^n = c)\}, \tag{3}$$

$$\mu_{rc} = \mathbb{E}_n\{x_r^n P(l_r^n = c)\}, \tag{4}$$

$$\sigma_{rc}^2 = \mathbb{E}_n\{(x_r^n - \mu_{rc})^2 P(l_r^n = c)\}, \tag{5}$$

where $\mu_{rc} \in \mathbb{R}$ and $\sigma_{rc}^2 \in \mathbb{R}^+$ are the estimated mean and variance of tissue c at location r, and $P(l_r^n = c)$ is the probability that n'th atlas has the label c at r. Notation $\mathbb{E}_n\{\cdot\}$ stands for the expectation operator over variable n.

2.2 *Atlas Selection by Similarity using Tensor-Product Kernels*

Due to the anatomical variability between subjects, Eqs. (3)–(5) are biased toward the population average, hindering the segmentation performance. To overcome this issue, we introduce an atlas selection stage that allows computing the parameters of local Gaussians from the atlas subset with the largest similarity to the target image. That is, it holds that $\{\mathcal{X}^n, \mathcal{L}^n : k(\mathcal{X}, \mathcal{X}^n) > \epsilon\}$, being $k(\mathcal{X}, \mathcal{X}^n) \in \mathbb{R}^+$ a similarity

function between the target and n'th atlas. For the implementation of this stage, we consider our previous tensor-product kernel-based (TPK) similarity to build the atlas subset.[26] To this end, we extract the set of pairwise interslice kernel (ISK) relationships along each MR image axis into a set of kernel matrices $\{\boldsymbol{S}^{\nu} \in \mathbb{R}^{L_{\nu} \times L_{\nu}} : \nu \in \{a, s, c\}\}$, where L_{v} is the number of slices across ν'th axis, namely, axial (a), sagittal (s), and coronal (c). Relying on a positive-definite kernel function $\kappa(\cdot, \cdot)$, any intensity image is mapped into the symmetric matrix $\boldsymbol{S}^{\nu}$ with elements $s^{\nu}_{ll'} = \kappa_X(\boldsymbol{X}^{\nu}_{l}, \boldsymbol{X}^{\nu}_{l'})$, where $\boldsymbol{X}^{\nu}_{l}$ is l'th slice of ν'th axis. The resulting set of ISK relationships feeds the TPK similarity, combining the pairwise image similarity along each axis as

$$k(\mathcal{X}, \mathcal{X}^n) = \prod_{v \in \{a,s,c\}} \kappa_S(\mathbf{S}^{\nu}, \mathbf{S}^{\nu}_{n}), \quad \forall\, n \in [1, N]. \tag{6}$$

In particular, both kernels follow the Gaussian function:

$$\kappa_X(\mathbf{X}^{\nu}_{l}, \mathbf{X}^{\nu}_{l'}) = \exp[-\|\mathbf{X}^{\nu}_{l} - \mathbf{X}^{\nu}_{l'}\|^2_F/(2\sigma^2_X)], \tag{7}$$

$$\kappa_S(\boldsymbol{S}^{\nu}, \boldsymbol{S}^{\nu}_{n}) = \exp(-\|\boldsymbol{S}^{\nu} - \boldsymbol{S}^{\nu}_{m}\|^2_F/(2\sigma^2_S)), \tag{8}$$

where notation $\|\cdot\|_F$ stands for matrix-based Frobenius norm.

Then, the similarity function in Eq. (6) benefits of the anatomical changes on images to be compared across the axes. As a result, the combination of axiswise similarities allows ranking the atlases according to their similarity with the target image so that the closer each atlas is—the better its contribution to building the local probability distributions in Eq. (2).

2.3 *Adaptive Fusion Labeling in Neighboring Patches*

Commonly, the estimation in Eq. (3) implies the existence of one-to-one relationships between the training dataset images. Nonetheless, the voxels, neighboring a boundary region $\mathcal{B}$, hold mostly more uncertainty than the far voxels do, making the segmentation of voxels with less uncertainty (e.g., deep inside the structures) have lower computational effort. This means that the assumption of a one-to-one relationship does not hold entirely for the edges that belong to the tissues of interest, introducing several errors into the label fusion procedure.

To address above issue, we define the probability distribution models in Eqs. (3)–(5) for boundary and nonboundary regions. The nonboundary region is composed by voxels r where the maximum *a priori* probability [Eq. (3)] is larger than or equal to an introduced threshold $\alpha \in (1/C, 100\%]$., i.e., $\max_c \mathbb{E}_n\{P(l^n_r = c)\} \geq \alpha$. On the contrary, all voxels that fail this condition (that is, having the largest label uncertainty) are assumed to belong to the boundary region. In either case of regions, estimation of the prior and likelihood function parameters are as follows:

$$\text{boundary regions}: \max_c \mathbb{E}_n\{P(l^n_r = c)\} < \alpha$$
$$P(l_r = c) = 1/C, \tag{9}$$

$$\mu_{rc} = \mathbb{E}_{n,s}\{x^s_r P(l^n_s = c) P(\mathrm{d}_{\mathcal{X}} \mathcal{X}^q_r \mathcal{X}^n_s | \epsilon_{\varphi})\}, \quad \forall\, s \in \mathcal{N}(r), \tag{10}$$

$$\sigma^2_{rc} = \mathbb{E}_{n,s}\{(x^s_r - \mu_{rc})^2 P(l^n_s = c) P(\mathrm{d}_{\mathcal{X}} \mathcal{X}^q_r \mathcal{X}^n_s | \epsilon_{\varphi})\}, \tag{11}$$

$$\text{nonboundary regions}: \max_c \mathbb{E}_n\{P(l^n_r = c)\} \geq \alpha$$
$$P(l_r = c) = \delta(l_r - c) \tag{12}$$

$$\mu_{rc} = \mathbb{E}_n\{x^n_r P(l^n_r = c)\}, \tag{13}$$

$$\sigma^2_{rc} = \mathbb{E}_n\{(x^n_r - \mu_{rc})^2 P(l^n_r = c)\}, \tag{14}$$

where $\mathcal{N}(r)$ describes a cubic searching volume (or neighborhood) that is centered at r and limited in size by an included radius parameter $\epsilon_{\Omega} \in \mathbb{R}^+$ (neighborhood radius). The local intensity set $\mathcal{X}^n_r$ builds a patch of radius $\epsilon_{\beta} \in \mathbb{R}^+$ that represents the r'th voxel. Further, $P[d_{\mathcal{X}}(\mathcal{X}^q_r \mathcal{X}^n_s) | \epsilon_{\varphi}]$ is the probability that the patches $\mathcal{X}^q_r$ and $\mathcal{X}^n_s$ become similar within the threshold value $\epsilon_{\varphi} \in \mathbb{R}^+$, employing the distance function $\mathrm{d}_{\mathcal{X}}(\mathcal{X}^q_r \mathcal{X}^n_s) \in \mathbb{R}^+$.

As a result, the agreement threshold α selects the model according to the labeling complexity within the boundary and nonboundary regions. In the former case, we propose the uniform distribution as the priors to reduce the labeling errors related to a poor agreement between training atlases in Eq. (3). Moreover, the conditional distributions (computed from the neighborhood search and pairwise volume similarity) are incorporated to locate the most similar voxels across the atlas set, and therefore, the label of a targeted voxel is assigned more accurately. For the nonboundary regions, inside most of the atlases agree, a voxelwise probabilistic model reduces the computational cost, without compromising the tissue delineation.

3 Experimental Setup

3.1 *MRI Datasets and Image Preprocessing*

We evaluate the proposed approach of adaptive Bayesian label fusion using kernel-based similarity metrics in delineating the hippocampal tissue, appraising the following pipeline stages: (i) preprocessing of MRI scans using the ANTS tool to perform the affine-registration of all volumes into each target; (ii) subject-dependent atlas selection based on the TPK, ranking the atlases according to their similarity to the target in the RKHS space; (iii) construction of a spatial domain, aiming to gather all hippocampus-labeled voxels of the whole dataset within a region of interest and thus reducing the computational effort; (iv) tissue label estimation, relying on a nonlocal patch search over the sets of intensity and label images for boundary regions and point-to-point comparisons for nonboundary regions;

Fixing the acquisition parameters as shown in Table 1, the discussed segmentation approach is tested on the following benchmark MRI collections:

(a) SATA: This collection, publicly available,[27] had been created by the "Multi-Atlas Labeling beyond the Cranial Vault: Workshop and Challenge" for assessing the image segmentation of blind-folded data. We consider the training subset, which holds 35 T1 MR images and is equipped with the manually delineated hippocampus.

(b) LONI:[28] This database consists of 40 T1-weighted MR brain images that had been collected from healthy volunteers (20 males and 20 females), aging

Table 1 Technical specifications of the tested benchmark databases.

Database	Volumes	Volume size (voxels)	Voxel size (mm)
SATA	35 – training	$256 \times 256 \times 261$	$1.0 \times 1.0 \times 1.0$
	12 – testing	to $256 \times 256 \times 334$	
LONI	40	$220 \times 220 \times 184$	$1.0 \times 1.0 \times 1.0$
IBSR	18	$256 \times 256 \times 128$	$0.8 \times 0.8 \times 1.5$
			to $1.0 \times 1.0 \times 1.5$
HAMMERS	30	$192 \times 256 \times 124$	$0.937 \times 0.937 \times 1.5$
ADNI	101	$160 \times 192 \times 192$	$1.2 \times 1.25 \times 1.25$
		$166 \times 256 \times 256$	$1.2 \times 0.94 \times 0.94$
		$256 \times 256 \times 256$	$1.0 \times 1.0 \times 1.0$

from 20 to 40 years (29.2 average $\pm$6.4 years). Each scan is manually delineated to extract 56 partitions; most of them within the cortex. In addition to brain atlas construction, this image reservoir is widely used for evaluating segmentation algorithms of hippocampal subcortical structures.

(c) IBSR. This repository[29] assembles 18 T1-weighted MRI scans acquired from normal subjects (14 males and 4 females, aging from 7 to 71 years), and allows evaluating the sensitivity of segmentation tools to the signal-to-noise ratio, contrast-to-noise ratio, shape complexity, and degree of partial volume effect.

(d) HAMMERS. The dataset available at Ref. 30 contains 30 T1-weighted MR brain images, which have been collected from 15 women (median age 31 years and range 20 to 54) and 15 men (median age 30 years and range 20 to 53); all of them are subjects with no neurological, medical, or psychiatric conditions. Each scan has 83 manually labeled ROI sets with hippocampal segmentation.

(e) ADNI. The Alzheimer's disease neuroimaging initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease (AD). For up-to-date information, see Ref. 31, for download, see Ref. 32 gathers imaging biomarkers, biosignals, and neuropsychological tests to characterize dementia patients. We selected 101 T1 whole-brain MR images from different healthy control subjects, along with their hippocampus masks. Selected subjects average age 80-years old and 49% are men.

During validation, the leave-one-out strategy provides the segmentation performance assessed by Dice similarity index $d_S \in [0,100]$, defined as follows:

$$d_S(\Omega_H \tilde{\Omega}_H) = 200 \frac{|\Omega_H \cap \tilde{\Omega}_H|}{|\Omega_H| + |\tilde{\Omega}_H|}, \tag{15}$$

where $\Omega_H = \{r \in \Omega : l_r = H\}$ holds the ground-truth voxels that belong to the hippocampus, and $\tilde{\Omega}_H = \{r : \tilde{l}_r = H\}$—the automatically labeled voxels. Notation $\cap$ stands for the intersection set operator. The Dice similarity index measures the overlap between the estimated and manual segmentation label sets, rating 100 for fully matching regions and 0 for nonoverlap at all.

Note that for comparison purposes, the above-processing pipeline of hippocampal segmentation is also performed for the following approaches: adaptive gradient distribution on the boundary map based on active contour models and multiatlas concept (AGDB),[33] hierarchical patchbased label fusion (HPBLF),[18] patch-based label fusion with discriminative dimensionality reduction (PLFDDR),[34] patch-based label fusion via matrix completion (PLFMC),[17] spatial confidence maps label fusion (SCMLF),[23] ensembling multiview convolutional neural network (9ViewEnsem-Net1),[35] embedding patch-based label fusion (EPLF),[36] and multiscale patch-based label fusion (MSPLF).[37] The latter one was implemented setting its parameters as it is suggested in the document.

3.2 *Parameter Setting in Evaluating Segmentation Pipeline*

In the beginning, we tune the kernel bandwidth [see Eqs. (7) and (8)] in accordance with the procedure proposed by Ref. 26. That is, in the reproduced-kernel Hilbert space, all variances are maximized, enhancing the difference among slices and images. Section 3.2 displays the resulting pairwise volume, showing that the MRI dataset distribution differs from each other. At the same time, HAMMERS and IBSR evidence several image clusters, whereas the SATA, ADNI, and LONI collections lack any group structure (Fig. 1).

Then, the pairwise volume similarity values in Eq. (6) feeds the atlas selection stage, having as the main parameter the probability function that measures the voxel similarity across a patch domain as

$$P(d_{\mathcal{X}} \mathcal{X}_r^q \mathcal{X}_s^n) | \epsilon_\varphi = \begin{cases} \exp(-\|\mathcal{X}_r^q - \mathcal{X}_s^n\|_F^2 / \gamma_n) & \epsilon_\varphi \le d_{\mathcal{X}}(\mathcal{X}_r^q \mathcal{X}_s^n) \le 1 \\ 0 & \text{Otherwise} \end{cases}, \tag{16}$$

where the RBF kernel width hyperparameter γ_n is set, as suggested in Ref. 38, as the minimal distance between the target patch $\mathcal{X}_r^q$ and the neighboring patches under consideration $\mathcal{X}_s^n$, that is, $\gamma_n = \min_{\forall s} \|\mathcal{X}_r^q - \mathcal{X}_s^n\|_F^2$. The value $\mathrm{d}_{\mathcal{X}} \mathcal{X}_r^q \mathcal{X}_s^n \in \mathbb{R}^+$ is the structure similarity index that allows selecting the patches with similar contrast and luminance so that the exponential similarity calculation in $\exp(-\|\cdot\|_F^2/\gamma_n)$ is not performed over unlike patches, decreasing substantially the computational burden of the accomplished nonlocal patch search. The structure similarity index is given as[39]

$$d_{\mathcal{X}}(\mathcal{X}_r^q \mathcal{X}_s^n) = \left(\frac{2\mu_s \mu_r}{\mu_s^2 + \mu_r^2} \right) \left(\frac{2\sigma_s \sigma_r}{\sigma_s^2 + \sigma_r^2} \right), \tag{17}$$

where $\mu_z \in \mathbb{R}$ and $\sigma_z \in \mathbb{R}^+$ are the mean and standard deviation of the intensity patch $\mathcal{X}_z^q$ centered at $z \in \Omega$, respectively.

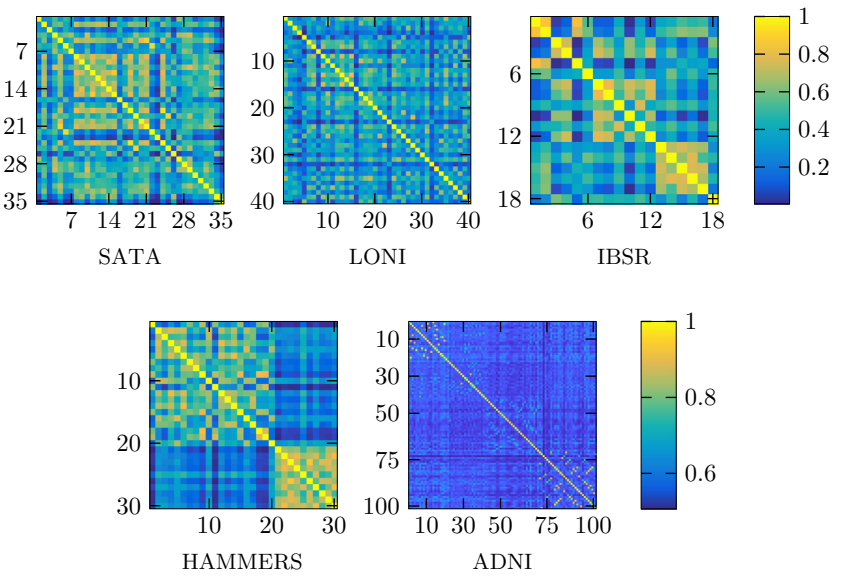

Fig. 1 Resulting matrices of pairwise volume similarity of all tested MRI dataset, using their optimal parameters.

Practically, the performance of the proposed Bayesian segmentation depends on the accurate tuning of algorithm hyperparameters: global and local. The parameter tuning follows a two-staged grid search of the maximum average Dice index similarity, employing a leave-one-out scheme. In the first stage, we optimize both global parameters (number of atlases $|\mathcal{A}|$ and agreement threshold α) by fixing the local values of $(\epsilon_\varphi, \epsilon_\beta, \epsilon_\Omega) = (0.9, 2, 3)$ as recommended in Refs. 40 and 41. In the second stage, we search the optimal local parameters within $\epsilon_\beta \in [2,3]$ (that is, $5\times5\times5$ and $7\times7\times7$ voxels) and $\epsilon_\Omega \in [2,7]$ (ranging from $5\times5\times5$ till $15\times15\times15$ voxels) by setting $|\mathcal{A}|$ and α to above values. Figure 2 shows the average Dice index that is estimated on each considered MRI collection, at different values of agreement threshold. As seen, the optimal parameters are attained at: $(|\mathcal{A}|, \alpha) = (18,65\%)$, $(13,65\%), (11,53\%), (9,65\%), (8,53\%)$ for SATA, LONI, IBSR, HAMMERS, and ADNI subjects, respectively. Note that the performed Dice index values prove that the use of all available atlases hampers the segmentation performance.

For the second tuning stage, we searched the optimal local parameters within $\epsilon_\beta \in [2,3]$ (that is, $5\times5\times5$ and $7\times7\times7$ voxels) and $\epsilon_\Omega \in [2,7]$ (ranging from $5\times5\times5$ till $15\times15\times15$ voxels) by fixing $|\mathcal{A}|$ and α to above values. Figure 3 shows that the pairs achieving the best performance

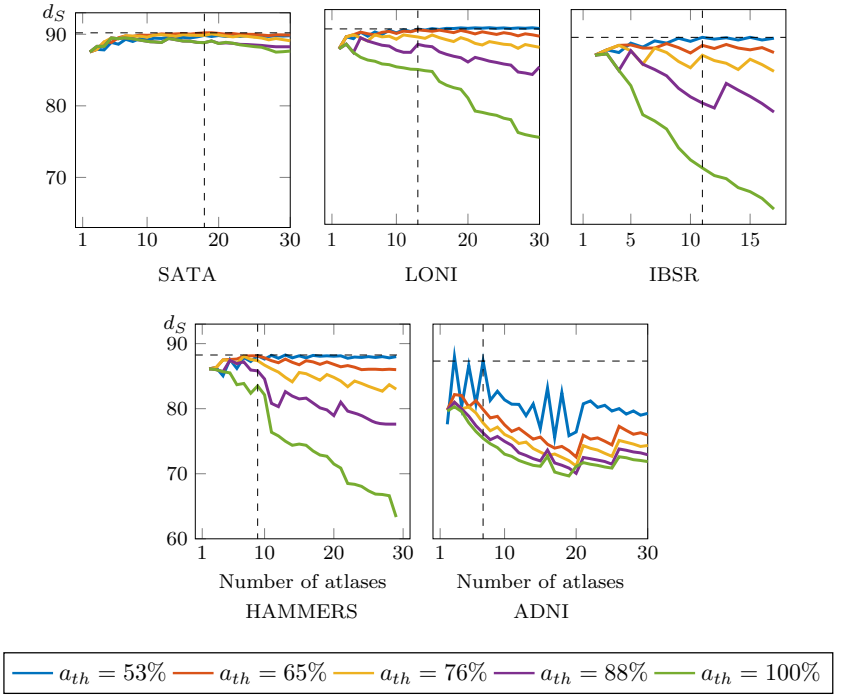

Fig. 2 Global parameters tuning for the different datasets over our method is tested.

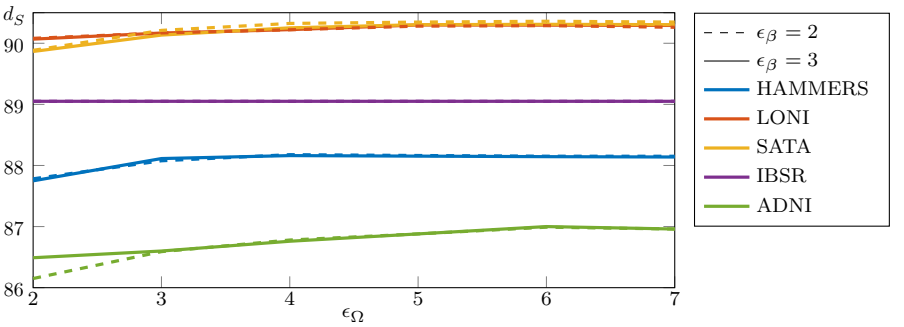

Fig. 3 Tuning of neighborhood and patch radii using the average Dice index for evaluation segmentation quality over 10 subjects for each dataset.

are $(\epsilon_\beta, \epsilon_\Omega) = (2,6), (2,5), (2,3), (2,4), (2,6)$ for SATA, LONI, IBSR, HAMMERS, and ADNI, respectively. In almost all cases, the Dice index increases as the neighborhood radius also rises, meaning that the misalignment produced by the affine registration demands the presence of a larger number of regions to achieve a set of similar patches. Nevertheless, d_S decreases as ϵ_Ω becomes too large since more similar patches may be discovered across the adjacent structures. Regarding ϵ_β, though the computational cost rises, there is a very slight improvement in some cases as the patch radius increases because the large patches more adequately reflect the tissue properties.

Summarizing, the resulting values of the calculated optimal parameters are as follows: $(|\mathcal{A}|_{\text{op}}, \alpha, \epsilon_\Omega, \epsilon_\beta, \epsilon_\varphi) = (18{,}65\%, 6{,}2{,}0.9), (13{,}65\%{,}5{,}2{,}0.9), (11{,}53\%{,}2{,}3{,}0.9), (9{,}65\%{,}4{,}2{,}0.9), (8{,}53\%{,}6{,}2{,}0.9)$ for SATA, LONI, IBSR, HAMMERS, and ADNI subjects, respectively. Note that the values of threshold and patch radius tend together, respectively, to $\alpha = 65\%$ and $\epsilon_\beta = 2$ in most of the tested databases. In turn, in each one of the databases, diverse tuned values are accomplished by the number of atlases $|\mathcal{A}|_{\text{op}}$ and neighborhood radius ϵ_Ω, which varies within a broad range (from 2 to 6) across databases in an attempt to compensate for the misalignment effect as reported in Ref. 41.

For visual quality inspection of the segmented hippocampus, Fig. 4 shows the best (top row) and worst (bottom row) subjects of each database performed by the proposed method, marking in a green, blue, and red dots the places denote that are correctly labeled (true positives), undersegmented (false negatives), and oversegmented (false positives) voxels, respectively. As expected, most of the mistakes arise in those locations placed near to the tissue boundaries. Thus, for the best-segmented subjects, the proposed approach performs similarly in the four dataset sets in terms of true positives, false negatives, and false positives. In addition, the number of false positives is smaller than the false negatives implying that most of the errors are due to undersegmented voxels. On the contrary, errors in subjects with the lowest Dice lack consistency: the subject is oversegmented for SATA, IBSR, and ADNI; undersegmented for HAMMERS; and the errors are evenly distributed in the case of LONI.

Overall, the reached results show that ABKS outperforms other approaches in almost all evaluated databases, as shown in Table 2 that displays the averaged Dice index of segmenting the left and right hippocampus, using the tuned parameter set. Hence, we hypothesize that combination of the atlas selection and adaptive fusion labeling enhances the hippocampal segmentation.

4 Discussion

To overcome the morphological variations and template-to-target misalignment, this work introduces a brain tissue segmentation approach, adaptive Bayesian label fusion using kernel-based similarity metrics – ABKS, that constructs adaptive target-specific probabilistic atlases to be ranked in the template search strategy, appraising the following stages: Atlas selection for dealing with segmentation of MRI data with high anatomical variability; patch-based voxel representation to

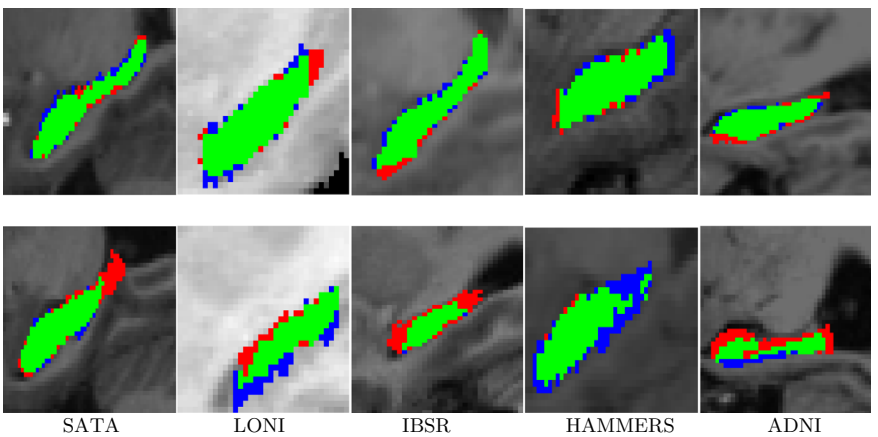

Fig. 4 Examples of segmented hippocampus performed by the proposed method in each database. Best (top) and worst (bottom) subjects. Green, blue, and red regions mean correctly labeled, undersegmented, and oversegmented voxels, respectively.

Table 2 Mean Dice index score for hippocampus segmentation accomplished by the compared methods. Notation "—" stands for nonreported values.

Approach	SATA	LONI	IBSR	HAMMERS	ADNI
AGDB[33]	—	—	84.0	—	—
HPBLF[18]	86.5	81.4	—	84.6	88.5
PLFDDR[34]	83.8	—	—	—	81.1
PLFMC[17]	—	83.9	—	—	84.1
SCMLF[23]	87.7	—	—	—	86.6
ABKS	**90.3 ± 1.66**	**90.3 ± 1.58**	**89.0 ± 1.43**	**88.1 ± 1.37**	86.6
MSPLF[37]	85.5	86.1	83.3	82.0	—
EPLF[36]	—	—	—	—	85.4
9ViewEnsem-Net1[35]	—	—	—	—	**89.5 ± 1.45**

Note: Boldface stands for the best performing approach.

enhance the voxel embedding in spaces with increased tissue discrimination, and construction of a neighborhood-dependent model that addresses the label assignment of each region with a diverse patch complexity. From the experiments performed on four popular MRI collections, the following remarks are worth being mentioned:

Atlas selection through similarity measures enables segmentation with wide anatomical variability. Expressed in terms of pairwise TPK relationships, the morphological variability within each collection allows distinguishing cases of similarity, ranging from very related images to remarkably different volumes. Moreover, several distribution modes of atlases can be inferred relying on the presence of diagonal blocks with separate pairwise volume similarities in the kernel matrices. In the specific case of HAMMERS and IBSR collections, the modes may point to demographic particularities of imaged subjects as seen in Sec. 3.2. Still, the quality of the coarse registration constraints the introduced representation. Figure 2 evidences that the use of TPK takes advantage of this fact, ranking the atlases that contribute the most to the image labeling. As a result, the introduced atlas selection undertakes the anatomical variability of MRI collections by selecting the image subset that better matches to the morphological similarity of each target image, enhancing the labeling efficiency.

Within the Bayesian segmentation framework, the spatial priors of tissue modeling can be enhanced by the involved adaptive label fusion of neighboring patches, adjusting the patch and neighborhood radii properly. Using a baseline heuristic grid search, both ruling parameters are tuned in accordance to the voxel uncertainty that highly depends on its boundary proximity. Therefore, the segmentation performance tends to increase as the neighborhood radius also rises as clearly shown in Fig. 3 for LONI and SATA. This result suggests that the misalignment produced by the affine registration demands the presence of a more significant number of regions to achieve a set of similar patches. However, the segmentation performance tends to decrease as the neighborhood became too large due to more similar patches may be found within adjacent structures (such as in LONI and HAMMERS when the parameter is larger than 11 voxels). Note that the application of TPK-based similarity implies a tuned parameter set, which is specific for each database, making this issue the most challenging shortcoming for generalizing the obtained results over other MRI collections. As a result, the better performance at larger neighborhood radius implies that anatomical relationships must be searched in a broader region, and the patch size defines the shape variability to be encoded at each voxel.

The segmentation framework enhances the shape variability representation at separate levels of patch dimension, employing two complementary strategies: atlas selection (collecting global information) together with adaptive modeling (accumulating local information). Results in Table 2 prove that the atlas selection stage improves the multiatlas concept (AGDB) at determining the most relevant templates for labeling a query image. Regarding the patch-based approaches, proposed adaptive Bayesian model better decodes the local labeling complexity than SCMLF; and the conditional probability estimator for boundary regions enhances the local tissue modeling of HPBLF, SCMLF, and MSPLF. Therefore, we conclude that the introduced atlas selection together with adaptive fusion labeling enhances the hippocampal segmentation with performance values comparable with convolutional neural network approaches for filtering (EPLF) and labeling (9ViewEnsem-Net1).

5 Conclusion

With the purpose of enhancing the brain tissue segmentation, this work introduces the ABKS approach that includes the construction of adaptive target-specific probabilistic priors using similarity-ranked atlases for dealing anatomical variability of collected MRI data. In particular, the segmentation approach appraises patch-based voxel representation to enhance the voxel embedding in spaces with increased tissue discrimination, as well as the construction of a neighborhood-dependent model that addresses the label assignment of each region with diverse patch complexity. To measure the similarity between the target and atlases, we propose a tensor-based kernel metric that also includes the labeling set. We evaluate the proposed approach on four benchmark MRI collections used for hippocampus segmentation, resulting in an increased segmentation

performance (assessed through the Dice index) as compared to other recent works.

Nonetheless, some remarks remain to be improved in the proposed segmentation approach. Thus, the incorporated atlas selection stage may not extract accurately enough a very wide heterogeneity from image collections as it is the case for the IBSR and HAMMERS databases, yielding over- or undersegmentation flaws as seen in Fig. 4. Therefore, as a future work, the authors plan to extend the proposed segmentation approach to a wider class of brain structural variations, including abnormal masses of tissue, and MRI data with high demographical disparity (age, ethnicity, etc.). Another factor to be reflected is the effect of varying acquisition conditions.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This work was developed within the framework of the research project 1110-744-55778, funded by COLCIENCIAS. Part of the data used for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

References

1. R. Sutter et al., "Significance of parenchymal brain damage in patients with critical illness," *Neurocrit. Care* **23**(2), 243–252 (2015).
2. L. Homeida et al., "Synchronous presentation of trigeminal, glossopharyngeal and geniculate neuralgias in a single patient," *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **121**(6), 626–628 (2016).
3. A. S. Chua et al., "Handling changes in {MRI} acquisition parameters in modeling whole brain lesion volume and atrophy data in multiple sclerosis subjects: comparison of linear mixed-effect models," *NeuroImage Clin.* **8**, 606–610 (2015).
4. A. Agarwal et al., "Soft tissue masses of hand: a radio-pathological correlation," *Radiol. Res. Pract.* **2015**, 1–10 (2015).
5. L. Alfonso et al., "A comparison of magnetization transfer methods to assess brain and cervical cord microstructure in multiple sclerosis," *J. Neuroimaging* **27**(2), 221–226 (2016).
6. M. A. Balafar et al., "Review of brain MRI image segmentation methods," *Artif. Intell. Rev.* **33**(3), 261–274 (2010).
7. B. Fischl et al., "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron* **33**(3), 341–355 (2002).
8. J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage* **26**(3), 839–851 (2005).
9. R. J. Beare et al., "Neonatal brain tissue classification with morphological adaptation and unified segmentation," *Front. Neuroinf.* **10**, 12 (2016).
10. C. R. Hernandez-Castillo, C. Limperopoulos, and J. Diedrichsen, "A representative template of the neonatal cerebellum," *NeuroImage* **184**, 450–454 (2019).
11. R. Wolz et al., "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Trans. Med. Imaging* **32**(9), 1723–1730 (2013).
12. F. M. Sukno et al., "Active shape models with invariant optimal features: application to facial analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(7), 1105–1117 (2007).
13. R. Wolz et al., "Leap: learning embeddings for atlas propagation," *NeuroImage* **49**(2), 1316–1325 (2010).
14. B. Patenaude et al., "A Bayesian model of shape and appearance for subcortical brain segmentation," *NeuroImage* **56**(3), 907–922 (2011).
15. C. Chu et al., "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images," *Lect. Notes Comput. Sci.* **8150**, 165–172 (2013).
16. J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Med. Image Anal.* **24**(1), 205–219 (2015).
17. G. Sanroma et al., "A transversal approach for patch-based label fusion via matrix completion," *Med. Image Anal.* **24**(1), 135–148 (2015).
18. G. Wu et al., "Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition," *NeuroImage* **106**, 34–46 (2015).
19. Q. Zheng and Y. Fan, "Integrating semi-supervised label propagation and random forests for multi-atlas based hippocampus segmentation," in *IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018)*, IEEE, pp. 154–157 (2018).
20. L. Zhang et al., "Concatenated spatially-localized random forests for hippocampus labeling in adult and infant MR brain images," *Neurocomputing* **229**, 3–12 (2017).
21. Y. Wang et al., "Patch-based label fusion with structured discriminant embedding for hippocampus segmentation," *Neuroinformatics* **16**, 411–423 (2018).
22. H. Zhu et al., "Metric learning for multi-atlas based segmentation of hippocampus," *Neuroinformatics* **15**(1), 41–50 (2017).
23. O. M. Benkarim et al., "Discriminative confidence estimation for probabilistic multi-atlas label fusion," *Med. Image Anal.* **42**, 274–287 (2017).
24. Y. Song , G. Wu, and K. Bahrami et al., "Progressive multi-atlas label fusion by dictionary evolution," *Med. Image Anal.* **36**, 162–171 (2017).
25. O. V. Senyukova and A. Y. Zubov, "Full anatomical labeling of magnetic resonance images of human brain by registration with multiple atlases," *Programm. Comput. Softw.* **42**(6), 356–360 (2016).
26. A. Álvarez-Meza et al., "Tensor-product kernel-based representation encoding joint MRI view similarity," in *36th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc. (EMBC)*, IEEE, pp. 3897–3900 (2014).
27. "Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge," 2013, https://www.synapse.org/#!Synapse:syn3193805/wiki/217780.
28. D. W. Shattuck et al., "Construction of a 3D probabilistic atlas of human cortical structures," *NeuroImage* **39**(3), 1064–1080 (2008).
29. National Institute of Neurological Disorders and Stroke (NINDS), "Internet brain segmentation repository," 2007, http://www.cma.mgh.harvard.edu/ibsr.
30. I. S. Gousias et al., "Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest," 2008, http://brain-development.org/brain-atlases/adult-brain-atlases/individual-adult-brain-atlases/ (February 2018).
31. M. W. Weiner, "Alzheimer's Disease Neuroimaging Initiative," 2003, http://www.adni-info.org (December 2018).
32. M. W. Weiner, "Alzheimer's Disease Neuroimaging Initiative," 2003, http://adni.loni.usc.edu/ (September 2018).
33. D. Zarpalas et al., "Hippocampus segmentation through gradient based reliability maps for local blending of ACM energy terms," in *IEEE 10th Int. Symp. Biomed. Imaging (ISBI)*, IEEE, pp. 53–56 (2013).
34. G. Sanroma et al., "Discriminative dimensionality reduction for patch-based label fusion," *Lect. Notes Comput. Sci.* **9487**, 94–103 (2015).
35. Y. Chen et al., "Hippocampus segmentation through multi-view ensemble Convnets," in *IEEE 14th Int. Symp. Biomed. Imaging (ISBI 2017)*, IEEE, pp. 192–196 (2017).
36. G. Sanroma et al., "Learning non-linear patch embeddings with neural networks for label fusion," *Med. Image Anal.* **44**, 143–155 (2018).
37. X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Med. Image Anal.* **31**, 77–87 (2016).
38. P. Coupé et al., "Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation," *NeuroImage* **54**(2), 940–954 (2011).
39. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
40. G. Wu et al., "A generative probability model of joint label fusion for multi-atlas based brain segmentation," *Med. Image Anal.* **18**(6), 881–890 (2014).
41. T. Tong et al., "Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling," *NeuroImage* **76**, 11–23 (2013).

Biographies of the authors are not available.