

Received July 30, 2013; revised November 4, 2013; accepted December 14, 2013. Date of publication 9 January 2014;
date of current version 3 February 2014.

Digital Object Identifier 10.1109/JTEHM.2014.2297953

Accurate and Fully Automatic Hippocampus Segmentation Using Subject-Specific 3D Optimal Local Maps Into a Hybrid Active Contour Model

DIMITRIOS ZARPALAS^{1,2}, POLYXENI GKONTRA¹, PETROS DARAS¹ (Senior Member, IEEE),
AND NICOS MAGLAVERAS^{2,3} (Senior Member, IEEE)

¹Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki 57001, Greece

²Laboratory of Medical Informatics, the Medical School, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

³Centre for Research and Technology Hellas, Institute of Applied Biosciences, Thessaloniki 57001, Greece

CORRESPONDING AUTHOR: D. ZARPALAS (zarpalas@iti.gr)

ABSTRACT Assessing the structural integrity of the hippocampus (HC) is an essential step toward prevention, diagnosis, and follow-up of various brain disorders due to the implication of the structural changes of the HC in those disorders. In this respect, the development of automatic segmentation methods that can accurately, reliably, and reproducibly segment the HC has attracted considerable attention over the past decades. This paper presents an innovative 3-D fully automatic method to be used on top of the multiatlas concept for the HC segmentation. The method is based on a subject-specific set of 3-D optimal local maps (OLMs) that locally control the influence of each energy term of a hybrid active contour model (ACM). The complete set of the OLMs for a set of training images is defined simultaneously via an optimization scheme. At the same time, the optimal ACM parameters are also calculated. Therefore, heuristic parameter fine-tuning is not required. Training OLMs are subsequently combined, by applying an extended multiatlas concept, to produce the OLMs that are anatomically more suitable to the test image. The proposed algorithm was tested on three different and publicly available data sets. Its accuracy was compared with that of state-of-the-art methods demonstrating the efficacy and robustness of the proposed method.

INDEX TERMS Hippocampus segmentation, hybrid active contour model (ACM), multi-atlas, prior knowledge, local weighting scheme, optimal local maps (OLMs).

I. INTRODUCTION

Brain disorders are major contributors to morbidity, disability, and premature mortality in many developed and developing countries worldwide [1]. Every year, over one-quarter of adult Americans are diagnosed with a mental illness, such as Major Depressive Disorder (MDD), Post-Traumatic Stress Disorder (PTSD), schizophrenia, and Alzheimer's Disease [2]. Moreover, every year a third of the EU's population is diagnosed with mental disorders [3]. The number of patients becomes even higher if neurological disorders, such as epilepsy and dementia, are also taken into account. Apart from making life difficult for patients on a personal level, brain disorders have a considerable societal and financial cost. In this respect, improved prevention and treatment of brain disorders is a key-issue, and could alleviate healthcare costs. Assessing the

structural integrity of the hippocampus (HC), which is a structure of the limbic system (Fig. 1), is an essential step toward this, due to its implication in such disorders. Dysfunction and neurodegeneration of HC plays a fundamental role in the development of various brain disorders. Many studies support that altered HC volume and connectivity represents a specific endophenotype; indicatively in schizophrenia [4]–[6], in first-episode schizophrenic patients [7], in bipolar disorder [8]–[10], in epilepsy [11], in Alzheimer's [12], [13], in Mild Cognitive Impairment [14], [15], in dementia associated with multiple sclerosis [16], [17], and in Down's syndrome [18]. Hence, HC morphology alterations have shown the capacity to be potentially used as a biomarker in decision making systems regarding various brain disorders. Thus, HC morphometry is a potentially powerful tool for many diseases

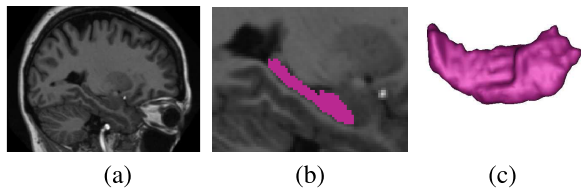


FIGURE 1. A sagittal slice of a 3D MR volume (a), zoomed version of (a) where the hippocampal region is indicated with magenta color (b), and the reconstructed 3D model of the hippocampus (c).

diagnosis, prognosis and monitoring. However, apart from the appropriate evidence and widespread agreement of the usefulness of HC volumetry, its establishment as a biomarker requires that it can be measured with appropriate accuracy and reproducibility [19].

Traditionally, HC structural assessment has been based on manual or semi-automated segmentation from MRI scans. However, time-constraints posed by those methods, largely as a result of the vast amount of data produced by MRI, rater bias and cost, constitute the major obstacles in the effective, large-scale morphological study of HC. Therefore, reliable automatic HC segmentation offers a valuable clinical tool, already showing its usefulness. Recently, in a large-scale, genome-wide association meta-analysis of hippocampal, brain, and intracranial volume [20] automated hippocampal volumetry has successfully enabled the discovery of novel genes associated with hippocampal volume in schizophrenia. In [21] it is clearly stated that the aim is fostering the use of hippocampal volumetry in routine clinical settings regarding Alzheimer's, which requires standardization firstly on the segmentation protocol (given the variety of protocols), and secondly on the automatic segmentation method. Similarly, a collaborative initiative on Alzheimer's between Europe and USA (EADC-ADNI) plans the adoption of HC volumetry as a new diagnostic criterion of Alzheimer's and in therapeutic trials [19]. Thus a roadmap has been defined in order to establish HC volumetry as an Alzheimer's biomarker, which requires reliable automatic segmentation. Once HC volumetry is established, every day clinical practice would then require subjective and highly accurate HC segmentation for proper and reliable disease diagnosis (potentially within a decision support system), monitoring, and treatment evaluation, which can even lead to drug discovery. Given the existing evidence of HC alterations in other disorders (e.g. schizophrenia, bipolar disease, epilepsy, etc.), similar actions are foreseen for these cases too.

Several methods have been proposed for automatic HC segmentation. However, it remains a very challenging task. Results in literature report that HC is among the brain structures for which the segmentation accuracy of automatic segmentation methods is lower compared to other brain structures [22]–[24]. Automatic segmentation methods of deep brain structures, such as the HC, can be broadly divided into three major categories: (1) atlas-based techniques, (2) deformable models and (3) active appearance models.

Atlas and multi-atlas based methods imply non-rigidly registering one or multiple (in the case of multi-atlas techniques) training images to the target image using some similarity measure. The labeled image(s) of the training image(s) are subsequently propagated to the space of the target image using the calculated wrapping fields, to offer the final segmentation. In the case of the multi-atlas methods an extra step is required; fusing the transformed labeled images or a selected subset of them. In literature, there exists a substantial amount of variations of the multi-atlas concept [23], [25]–[28]. These methods mainly differ in the registration method, the way for selecting a subset of the training labeled images before fusion, as well as the label fusion approach that is followed. A recent workshop offered comparative evaluation among state-of-the-art multi-atlas techniques [29]. In total, 25 algorithms entered the challenge and their performance was evaluated on a publicly available dataset (abbreviated as OASIS-MICCAI dataset in this work). Among the various multi-atlas methods that were evaluated, the ones that reached higher accuracy in terms of Dice similarity coefficient [30] are the joint label fusion technique proposed by Wang et al. [31] and the Non-Local STAPLE proposed by Asman et al. [32]. The first one is based on joint label fusion combined with the bias correction [33], and was proved to be the top performer of the challenge. The Non-Local Staple method is a statistical fusion technique using the non-local means framework.

In the second category, popular examples of deformable models are the Active Contour Models (ACM). ACMs allow a contour to deform iteratively to partition an image into distinct regions according to the image gradients (edge-based ACM), or the intensities' statistical information (region-based ACM). ACMs when used in combination with an implicit representation of the object of interest, have proved to be powerful tools in image segmentation. One popular example of edge-based methods is the Geometric Active Contour model (GAC) [34], whose evolution is terminated when "strong" edges are encountered. On the other hand, region-based ACMs use statistical information regarding the distribution of intensities, inside and outside the contour, making them less sensitive to image distortions, as well as to the "leakage" effect. The Chan-Vese model [35], that is based on the Mumford-Shah segmentation framework [36], is one of the most widely used region-based ACMs to detect objects whose boundaries are not defined by strong edges. However, the sole use of regional information can lead to erroneous segmentation results in the case of objects with well-defined boundaries due to the lack of boundary terms. To tackle the problems posed by the use of solely region or edge-based information, hybrid approaches have been proposed [37].

Nonetheless, ACMs solely depend on image information. Thus, their drawback is the lack of anatomical knowledge about structures undergoing segmentation. This limitation, which can be overcome through modeling and integration of prior knowledge of anatomical structures into the segmentation framework, has become a key-issue in medical

image analysis. One of the earliest and most influential works towards this direction was that of Cootes et al. [38] who incorporated into the ACM framework global shape constraints learned by means of Principal Component Analysis (PCA). They named their method as Active Shape Models (ASM) to avoid confusion with traditional ACM. In Leventon et al. [39], a non-parametric, intrinsic model based on the implicit representation of the shapes was constructed and incorporated into a GAC segmentation framework. The same approach was later adapted by Yang et al. [40], to build a statistical neighbor prior, able to constraint the segmentation based on neighborhood properties between adjacent structures. Yang et al. incorporated the constructed models into a region-based ACM framework to achieve simultaneous segmentation of neighboring structures. The aforementioned ASMs, however, are modeling global shape-prior knowledge through PCA, thus cannot account for local shape variations.

Incorporation of texture in the ASM framework led to Active Appearance Models (AAMs) [41]–[44]. AAMs use PCA-based linear subspaces to model variation of both shape and texture information from a training set. The initial concept of AAM required the identification of landmarks. In an effort to overcome this issue, the integration of level-sets in the AAM framework was proposed by Hu et al. [45], [46]. In Toth et al. [47], a multi-feature landmark-free AAM was presented. Other interesting approaches extending the initial idea of AAMs include combination of the AAM with patch-based label fusion [48], and the Bayesian appearance model [49]. Despite its advantages, such as fast performance, AAM is a local search technique and thus requires good initialization [50].

All the aforementioned active models focus on modeling and integrating prior information, rather than optimally balancing the degree of local influence that the prior knowledge and image information should have at a voxel level. Global weighting, which implies consistent boundary properties, has traditionally been used. However, this hypothesis is not true in some challenging cases, and thus is removed within this work. In fact, HC has spatially varying boundary properties, demonstrating clear, blurry and even missing borders. Towards this direction, we recently developed a local weighting scheme [51] to improve the weighting between the image and the prior term. A local weighting map, called Gradient Distribution on Hippocampus Boundary map (GDHB), was built based on the learned gradient values across the boundary of the HC.

Based on the same concept, in [52] we defined an Optimal Local Weighting map (OLW), via an optimization procedure. The optimization criteria are designed to generate the most accurate segmentations for a set of training images, given the corresponding ground-truth segmentations. The training OLWs are adapted on the test image and fused, to generate the OLW values of the latter. OLW is subsequently incorporated into an ACM framework defined by two energy terms, the region-based image term and the prior term. The efficacy of this concept was validated through experiments on the central sagittal slice of HC.

Hereby, the concept of OLW is further extended and a fully automatic subject-specific segmentation framework is proposed that models the local properties of HC by making use of a complete set of Optimal Local Maps (OLMs), applying it on 3D MR images. The OLMs produced are incorporated into a hybrid ACM framework, which includes three energy terms: a region-based term, an edge-based term and a prior term. The prior term, which is a label spatial distribution map, is built based on a straightforward multi-atlas framework. Thus, the proposed method is a mixture of the multi-atlas concept together with the ACM framework, in which OLMs are used to locally blend the energy terms.

OLMs refer to three different 3D maps: (i) a map that applies the local weighting between the prior energy term and the image derived terms (W_1), (ii) a map that locally balances the contribution between the region- and the edge-based term (W_2), and (iii) a map that controls locally the time step used in the evolution of the level set (S). To the best of our knowledge this is the first work to define the purpose and extraction of such three maps. All parameters included into the hybrid ACM model are calculated during training through an optimization procedure, avoiding heuristic parameter fine-tuning, ensuring optimal contour evolution based on the captured HC boundary and shape properties. The advantage of the proposed scheme, is that the ACM based on OLMs allows for capturing of fine details. Furthermore, the multi-atlas concept is utilized, since it can naturally incorporate the training OLMs, treating them as extra atlas modalities, thus constructing multi-fold atlases, i.e the atlas image, its label image and the corresponding OLMs (left part of Fig. 2).

The proposed algorithm was tested on three different datasets and demonstrated its appropriateness to be used as a supplementary technique to the multi-atlas methods.

II. METHODS

A. DESCRIPTION OF DATA

Three datasets were used in the context of this work:

1) OASIS Dataset

The OASIS database [53] consists of T1-weighted MR image volumes acquired with a 1.5T Vision scanner, produced by averaging four scans of the same individual, offering images with reduced noise levels. The MR image volumes were resampled to produce images with resolution $1.0mm \times 1.0mm \times 1.0mm$ and were spatially warped into the Talairach space. The size of the volumes is $176 \times 176 \times 208$ voxels. The database is very large (416 subjects), but no manual segmentations are available. As a result, a subset was chosen so as to cover the entire age span of the subjects and to include subjects with different degrees of dementia. A professional radiologist provided us with manual segmentations of HC, which we offer publicly to the research community.¹ The selected subset consists of 23 right-handed subjects (13 females and 10 males) with ages ranging from 18 to 96 years old.

¹<http://vcl.itl.gr/hippocampus-segmentation/>

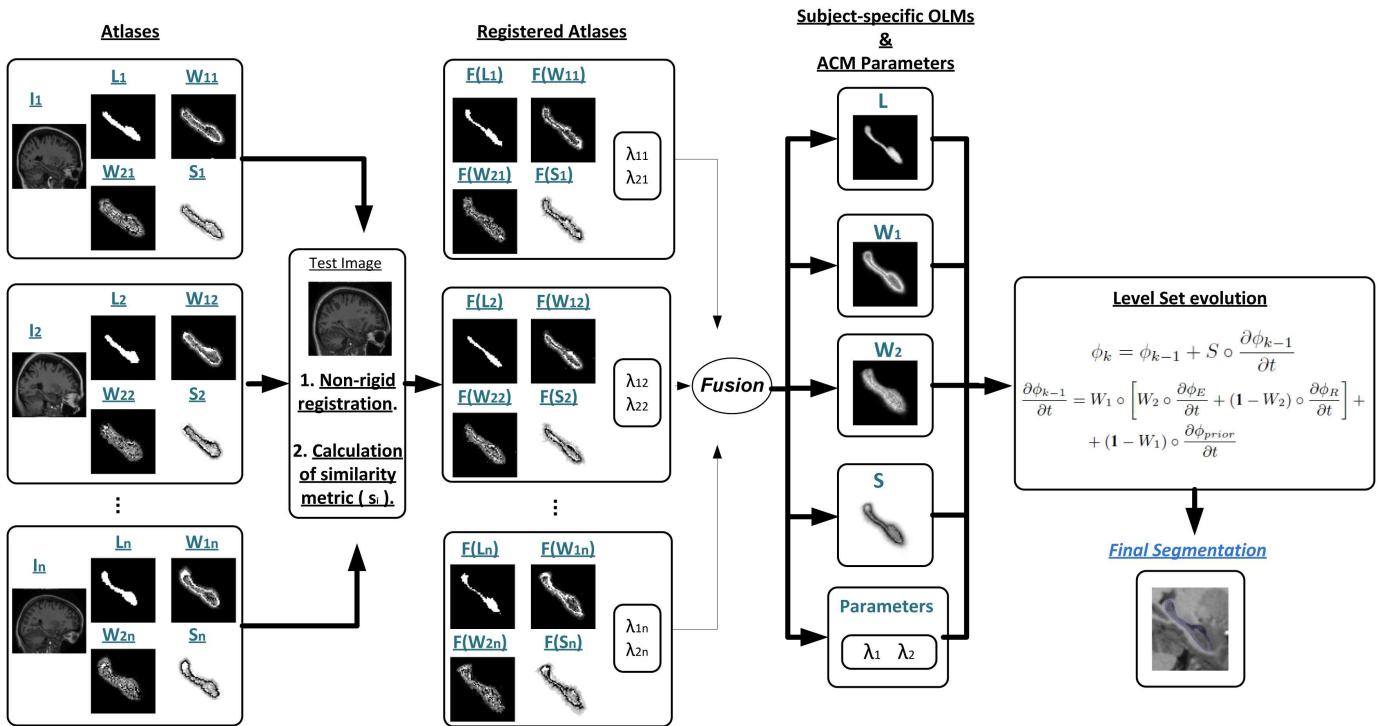


FIGURE 2. Overview of the proposed methodology for calculating the spatial distribution map L , the OLMs as well as the ACM parameters for a target image. The resulting L , OLMs and ACM parameters are incorporated into the ACM framework, to produce the final segmentation.

Among them, 2 subjects have a Clinical Dementia Rating scale (CDR) equal to 0.5, indicating very mild dementia while 2 subjects have CDR scale of 1, indicating mild dementia.

The manual protocol followed for the segmentation of the OASIS dataset is a close variant of the protocol used in the study of Narr et al. [54]. This protocol is an adaptation from existing protocols [55]–[60] and defines HC as a homogeneous gray matter structure. However, it should be noted the current discussion on whether to include or not non-gray matter parts in the hippocampal formation [61], [62]. The non-gray matter parts under discussion to be included in hippocampus are the alveus and fimbria.

2) IBSR DATA SET

The IBSR dataset is provided by the Center for Morphometric Analysis at Massachusetts General Hospital.² It contains T1-weighted MR image volumes of various image resolutions (from $0.8mm \times 0.8mm \times 1.5mm$ to $1.0mm \times 1.0mm \times 1.5mm$) from 18 subjects. The volumes have been spatially normalized into the Talairach orientation (rotation only). The subjects' age varies from youngsters, of less than 7 years of age, to older people of 71 years. Among the subjects, 4 of them were female, while the rest 14 subjects were male. Volumes' size is $256 \times 256 \times 128$ voxels.

The IBSR repository further offers the corresponding manual segmentations of HC. The manual segmentation protocol

followed is described in Makris et al. [63] and regards HC as a homogeneous gray matter structure.

3) OASIS-MICCAI DATA SET

The OASIS-MICCAI dataset is also a subset of the OASIS database, that was used in the recent evaluation workshop of MICCAI 2012 “Workshop on Multi-Atlas Labeling” [29]. Of the 35 MR image volumes in the subset, 15 were used for training and 20 for testing. The average age of the subjects in the training set is 23 years old (ages ranging from 19 to 34 years old), while in the testing set the average age is 47.5 years old (ages ranging from 18 to 90 years old). Both training and testing datasets contain female and male subjects; 10 MRIs from females and 5 from males, and 12 from females and 8 from males, are included in the training and testing sets, respectively.

Manual segmentations for the dataset are provided by Neuromorphometrics, Inc.³ under academic subscription. The manual segmentation protocol⁴ used for HC segmentation includes also white matter parts (fimbria/alveus) in the hippocampal region.

B. OVERVIEW

The key point of the proposed method is the blending of different types of information in a hybrid ACM framework through the incorporation of a set of subject-specific Optimal

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

³<http://Neuromorphometrics.com/>

⁴www.braincolor.org

Local Maps (i.e. W_1 , W_2 and S) to be used on top of the multi-atlas concept. The level set evolution depends on three energy terms: the edge-based term, the region-based term and the prior term. The latter is built by the subject-specific spatial distribution of labels map L offered through multi-atlas. In this scheme, W_2 is used to balance the contribution of each image-derived energy term. Hence, in the presence of strong edges W_2 weights more the edge-based term on that region, while in low gradient regions the region-based term is trusted more. Similarly, W_1 balances between the combined image terms and the prior term. Thus, the prior term is used in regions where the image information is not sufficient to drive the segmentation in the right direction. S aims to control the time step for the level set evolution, defining smaller time steps when the level set is close to convergence. Vice versa, S takes higher values on homogeneous regions, where the evolving contour is far from the actual boundary, to speed up evolution and convergence.

As Fig. 2 shows, once the OLMs and ACM parameters of the atlases are extracted, the atlases are registered to the test image. The fusion step follows, to extract the subject-specific OLMs, the spatial distribution label map L , as well as the ACM parameters. Then, the ACM evolution starts evolving a contour both on the MR image and on L , and optimally blending their outcomes. For the initialization of the ACM evolution, the region with the most likely voxels to belong to hippocampus (defined as the regions of L with the highest values) is used.

C. PRIOR INFORMATION

The first step towards modeling prior information is to build L , which offers information about the spatial distribution of the structure's labels. In this work, we investigated two different fusion techniques for the construction of L . The first one is a simple multi-atlas fusion concept based on a global weighted average technique. Thus, given a set of n training images I_i , $i = 1, \dots, n$ and their corresponding labeled images L_i , as well as the wrapping fields calculated by non-rigidly registering I_i to the target image I , L is provided as:

$$L = \sum_{i=1, \dots, n} s_i \cdot F(L_i) \quad (1)$$

where F represents the wrapping process of the labeled images L_i to the space of I , and s_i stands for the similarity between the registered training image I_i and the target image I expressed by means of cross-correlation. All s_i are normalized so that $\sum_{i=1, \dots, N} s_i = 1$.

The tasks of non-rigid registration and similarity calculation are performed with the ANTs toolkit. More precisely, the symmetric normalization methodology (SyN) [64] is utilized, which is based on optimizing and integrating a time-varying velocity field. The instructions in Klein et al. [65] were followed to choose the similarity metrics and the velocity field regularization.

The second fusion technique that was investigated for the construction of L is based on the recent joint label fusion technique⁵ which was proposed by Wang et al. [31]. According to this approach:

$$L_{joint}(x, y, z) = \sum_{i=1, \dots, n} w_i(x, y, z) \cdot F(L_i(x, y, z)) \quad (2)$$

where w_i refer to the voting weights (3D matrices of size equal to the size of L_i) calculated by the use of the joint label fusion technique, and are subject to $\sum_{i=1, \dots, n} w_i = 1$.

The reason for investigating two different multi-atlas fusion techniques is two-fold. Firstly, the multi-atlas concept is crucial for the performance of the proposed method as it offers L , which is used to provide the prior term. Secondly and most important, the proposed method is a locally weighted ACM on top of the multi-atlas concept and it is designed to improve the multi-atlas result no matter how accurate the latter is. By incorporating one of the most accurate multi-atlas methods according to the results from the MICCAI 2012 workshop, it is possible to demonstrate the appropriateness of the proposed method to work as a complementary method even to sophisticated and highly accurate multi-atlas methods.

D. ACM EVOLUTION

Let C denote the evolving curve, which is implicitly represented as the zero level set of a signed distance function ϕ . The evolution of the curve C is driven by the image terms and the prior term. Hence, by introducing the local weighting maps W_1 and W_2 the contour update equation is defined as:

$$\frac{\partial \phi}{\partial t} = W_1 \circ \frac{\partial \phi_{image}}{\partial t} + (1 - W_1) \circ \frac{\partial \phi_{prior}}{\partial t} \quad (3)$$

where the operation \circ denotes the Hadamard product, and

$$\frac{\partial \phi_{image}}{\partial t} = W_2 \circ \frac{\partial \phi_E}{\partial t} + (1 - W_2) \circ \frac{\partial \phi_R}{\partial t} \quad (4)$$

where ϕ_R , ϕ_E and ϕ_{prior} correspond to the evolving contours based on the region-based (E_R), the edge-based (E_E) and prior (E_{prior}) term, respectively. Finally, by introducing S in the level set framework, the evolving shape at iteration k is calculated by:

$$\phi_k = \phi_{k-1} + S \circ \frac{\partial \phi_{k-1}}{\partial t} \quad (5)$$

The derivation of ϕ_R , ϕ_E , ϕ_{prior} and based on them the final form of the evolution equation $\frac{\partial \phi}{\partial t}$ can be found in the Appendix section.

E. CALCULATION OF TRAINING OLMs AND ACM PARAMETERS THROUGH GRAPH-CUTS

Another means of capturing prior knowledge is the modeling of the varying boundary properties of HC, through the construction of local blending maps (OLMs), that define at voxel

⁵<http://www.nitrc.org/projects/picslmal/f/>

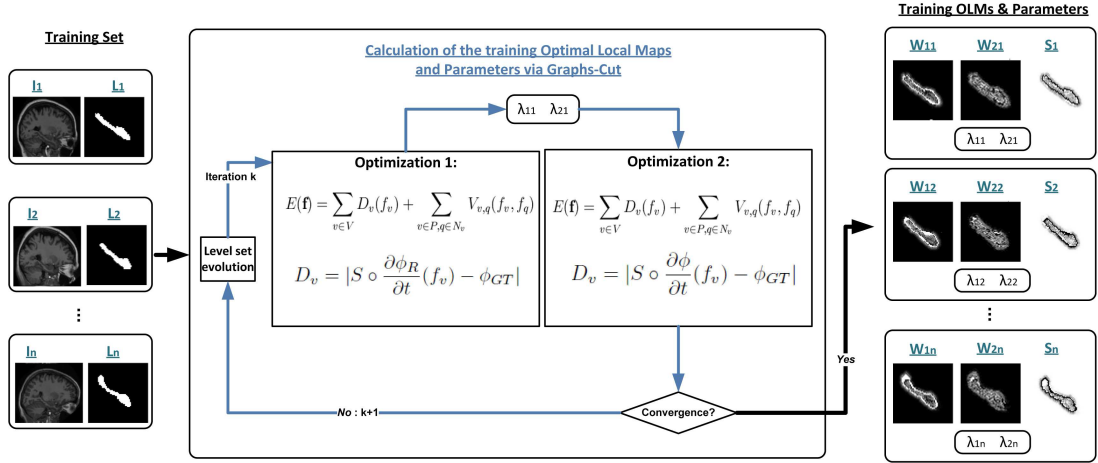


FIGURE 3. Training phase: Overview of the procedure for calculating the training OLMs and ACM parameters for the training images via an optimization scheme.

level which energy terms are to be trusted more for accurate segmentation results.

Graph cuts [66] have been widely used in computer vision in various problems, whose solution can be found through discrete pixel labeling. Graph cuts require formulating the pixel labeling in terms of energy minimization, assuming that the minimum energy solution corresponds to the maximum a posteriori estimate. Hereby, the Maxflow algorithm, introduced by Boykov and Kolmogorov [67], is used to minimize two energy functionals $E(f)$ (Fig. 3). These allow the calculation firstly of the training OLMs and secondly of the ACM parameters for the n training images, i.e. W_{1i} , W_{2i} , S_i , and λ_{1i} , λ_{2i} , $i = 1, \dots, n$.

Minimizing the two energy functionals requires that W_{1i} , W_{2i} , S_i and λ_{1i} , λ_{2i} are defined such as the image and prior terms that drive the evolving level-set will force it to move towards the corresponding ground-truth level set, despite its initial position. Thus, by imposing minimum difference between the needed move of the level set and the ground-truth zero level set extracted by the label image, and repeating the procedure until convergence of the level set, the curve will approach and finally fall onto the ground-truth contour. Once convergence is accomplished, the training OLMs and ACM parameters are defined as the average from all iterations. More details can be found in the Appendix section.

F. ADAPTING TRAINING OLMs AND ACM PARAMETERS THROUGH REGISTRATION

The focus of this section is on creating a subject-specific segmentation framework that accounts for the subject's anatomy. Towards this aim, a multi-atlas concept is used to produce subject-specific OLMs and ACM parameters for the target image. Attention should be paid to the fact that in the context of this section, as an atlas we consider the coupling of an anatomical image with its corresponding OLMs.

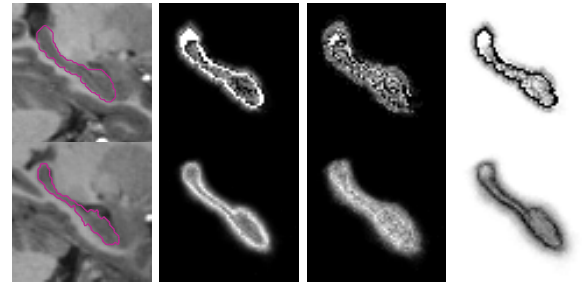


FIGURE 4. Sagittal slices of a training MR image I_i (top row) and a test MR image (bottom) of the OASIS dataset. The contours of hippocampus are represented with magenta. The corresponding slices from the 3D training W_{1j} , W_{2j} , S_j (top) and of the testing W_1 , W_2 , S (bottom) are presented in columns 2, 3 and 4 respectively.

As mentioned in subsection II-C, each anatomical image in the training set I_i , $i = 1, \dots, n$ is non-rigidly registered to the test image. The resulting transformations are used to propagate the training OLMs to the space of the target image (Fig. 4). Denoting the wrapping procedure as F , the resulting local maps and ACM parameters are combined according to the similarity s_i :

$$\begin{Bmatrix} W_1 \\ W_2 \\ S \\ \lambda_1 \\ \lambda_2 \end{Bmatrix} = \sum_{i=1, \dots, n} s_i \cdot \begin{Bmatrix} F(W_{1i}) \\ F(W_{2i}) \\ F(S_i) \\ \lambda_{1i} \\ \lambda_{2i} \end{Bmatrix} \quad (6)$$

III. EXPERIMENTS AND RESULTS

A. EVALUATION FRAMEWORK

The performance of an algorithm is potentially affected by the scanner type, imaging conditions, demographic characteristics and even by the quality of manual segmentations and the segmentation protocol used. To overcome these limitations and achieve fair comparisons, we evaluated the

TABLE 1. Comparison results for the OASIS dataset using mean dice's similarity index.

OASIS dataset	Dice $\mu \pm \sigma$	Method Description
OLM-ACM_Joint	0.86 \pm 0.05	Hybrid ACM with OLMs & multi-atlas with joint label fusion
ACM_Joint	0.84 \pm 0.08	Hybrid ACM with global weighting & multi-atlas with joint label fusion
Multi-atlas_Joint ('PICSL_Joint')	0.84 \pm 0.04	Multi-atlas & with joint label fusion
OLM-ACM	0.81 \pm 0.09	Hybrid ACM with OLMs & multi-atlas with weighted average fusion
ACM	0.79 \pm 0.02	Hybrid ACM with global weighting & multi-atlas with weighted average fusion
Multi-atlas	0.80 \pm 0.08	Multi-atlas with & weighted average fusion
Babalola et al. [50]	0.77 \pm 0.07	AAM model

proposed method (abb. OLM-ACM when weighted average fusion is used for building the spatial distribution map, and OLM-ACM_Joint when the sophisticated joint label fusion scheme is incorporated) in 3D MR images using three different datasets that vary in terms of the aforementioned characteristics.

To assess the behavior of the proposed methodology on HC segmentation, experiments were conducted to evaluate its performance through comparison with other methods. For the IBSR and OASIS datasets, the broadly-used leave-one-MRI-out procedure was followed, in order to offer fair comparison with the published results of other methods. For the OASIS-MICCAI dataset we followed the evaluation protocol of the MICCAI challenge to enable a straightforward comparison of our results, where 15 MRIs were used for training and 20 for testing. The Dice similarity coefficient (D) is used in all datasets as a segmentation performance measure due to its popularity and importance in evaluating and comparing the performance of segmentation methods. D is given by:

$$D = \frac{2|\hat{H} \cap H|}{|\hat{H}| + |H|} = \frac{2 \cdot Pr \cdot Re}{Pr + Re}, \quad D \in [0, 1] \quad (7)$$

where Pr and Re stand for Precision and Recall respectively. $D = 0$ indicates no overlap between the actual (H) and the estimated volume (\hat{H}), while $D = 1$ indicates perfect agreement.

B. RESULTS AND COMPARISONS

1) OASIS DATA SET

Due to the absence of published results in our OASIS dataset, a state-of-the-art AAM segmentation algorithm [50] offers a valuable indication on the expected Dice values.

The implementation of the latter is publicly available.⁶ In addition, the proposed method's performance is compared with that of the corresponding multi-atlas method on which the spatial distribution map relies on. More precisely, the ACM framework based on OLMs that uses either the weighted averaging for building the prior term (OLM-ACM), or the sophisticated joint label fusion (OLM-ACM_Joint), are compared with Multi-atlas and Multi-atlas_Joint methods, which are produced by applying majority voting on L and L_{joint} , respectively. This comparison actually reveals the contribution of the proposed methodology on top of the multi-atlas, regardless of the fusion technique. It should be noted that the Multi-atlas_Joint method is our reproduction (using the publicly available tools of ANTs toolkit and joint label fusion) of the method proposed by [31] and abbreviated as 'PICSL_Joint' during the MICCAI 2012 workshop. PICSL_Joint ranked 3rd in the challenge, while when combined with bias correction reached the first place. Moreover, segmentation results using a hybrid ACM based on global weighting for blending the edge, the region and the prior term were produced. As in the case of the proposed method, the performance of the hybrid ACM method using two different approaches for building the prior term is evaluated. In the first case the prior term is build using L (ACM), while in the second the L_{joint} is used instead (ACM_Joint).

The resulting mean Dice similarity coefficient and the corresponding standard deviations for all experiments are presented in Table 1. Comparing the OLM-ACM with the Multi-atlas method, and the OLM-ACM_Joint with the Multi-atlas_Joint method respectively, an improvement of 1-2% can be observed (p-values from paired t-test are 0.045 and 0.047, respectively). This demonstrates that the multi-atlas approach still leaves space for improvements that the proposed methodology takes advantage of by combining both image (edge and intensity) and prior information in an optimal way. Comparing the resulting Dice similarity coefficients by means of the ACM method and the ACM_Joint approach (0.79 and 0.84) with those achieved when using the Multi-atlas and Multi-atlas_Joint methods (0.80 and 0.84 respectively), no improvement can be seen. This suggests that the hybrid ACM with global weighting is insufficient for improving the multi-atlas result. It is important to note that in order to find the adequate ACM parameters for the hybrid ACM without OLMs, exhaustive heuristic fine-tuning was used and only the best achieved results are presented here. Therefore, the incorporation of the OLMs concept, which uses local weighting and parameters calculated with the use of an optimization procedure, is required in order for the ACM framework to be able to offer improvements to the multi-atlas methods.

A comparison plot of the Dice similarity coefficient for each subject is provided in Fig. 5 with the aim to allow comparisons among the OLM-ACM_Joint, Multi-atlas_Joint and the AAM methods. The Clinical Dementia Rating (CDR)

⁶http://www.isbe.man.ac.uk/kob/vaam_1_0/index.html

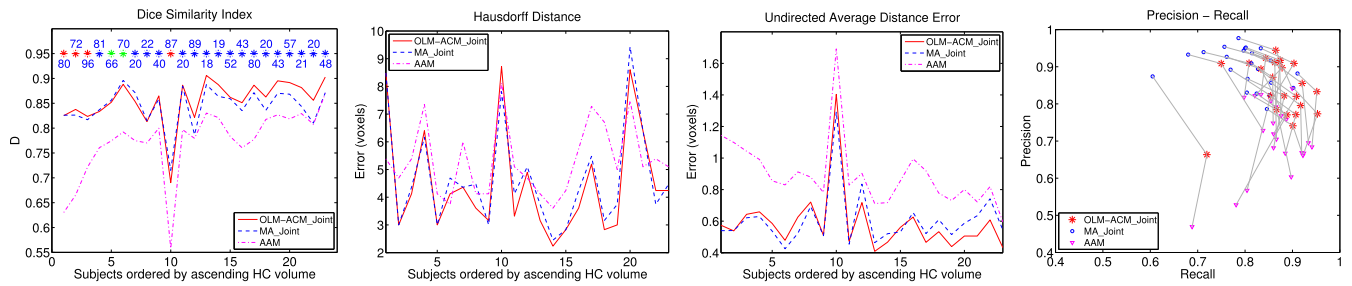


FIGURE 5. Comparison metrics for the performance of OLM-ACM_Joint, Multi-atlas_Joint and AAM methods. On the precision vs recall plot, the results of the OLM-ACM_Joint method are connected with those of each of the methods under comparison on the same subject to allow for direct comparison. The optimum position in the Precision-Recall space is the upper right corner (1, 1). On the Dice similarity index plot, colored asterisks stand for the Clinical Dementia Rating (CDR) of each subject (red asterisk stands for CDR=1 which indicates mild dementia, green for CDR=0.5 indicating very mild dementia, while blue for CDR=0 which indicates no dementia). The age of each subject appears above or below the corresponding CDR asterisk. Subjects are sorted by ascending ground-truth HC volume.

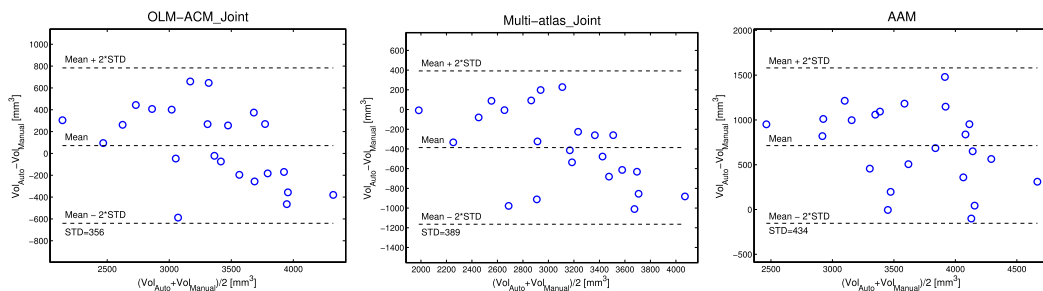


FIGURE 6. Bland-Altman plots for the OASIS dataset showing graphically the agreement between the manually segmented volumes and the volumes segmented by means of OLM-ACM_Joint, Multi-atlas_Joint and AAM method.

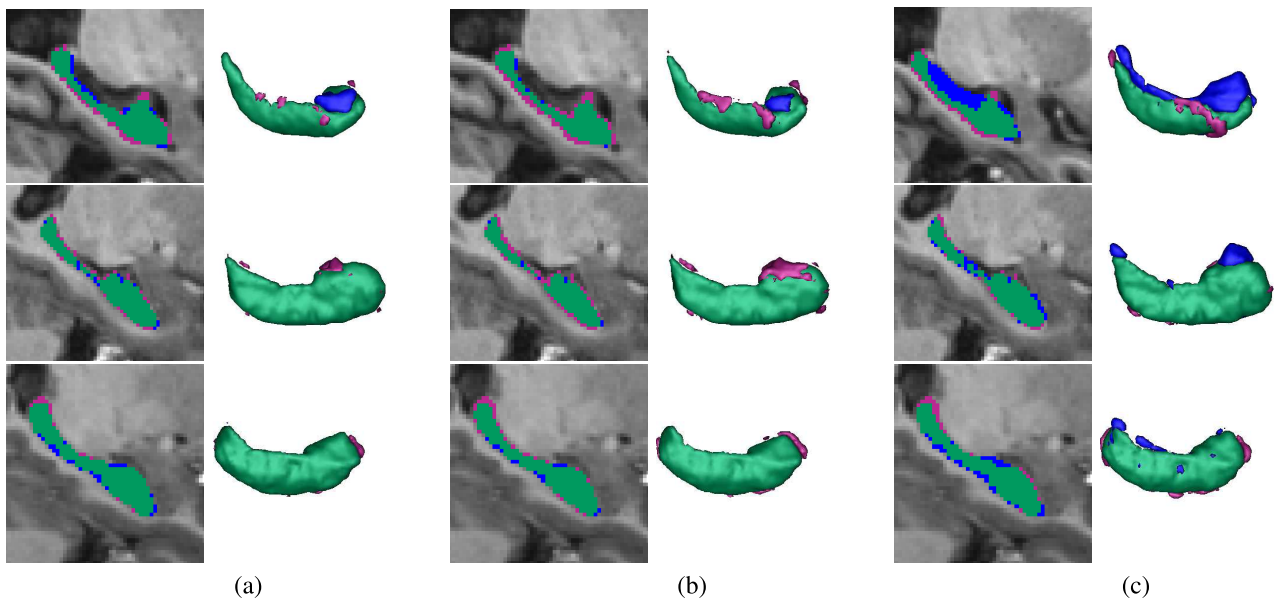


FIGURE 7. Segmentation results for subjects 2 (1st row), 12 (2nd row), 14 (3rd row) from OASIS dataset by (a) OLM-ACM_Joint, (b) Multi-atlas_Joint, and (c) the AAM method of [50]. On the both 2D slices and 3D models, blue color represents false positives, green true positives and magenta false negatives.

and the age of every subject are also provided in the plot. The subjects have been sorted according to ascending hippocampal volume in an effort to demonstrate the influence of volume in the performance of the methods. Further,

comparison plots on additional metrics are also provided, i.e. the precision and recall metrics, the Hausdorff distance and the undirected average difference. The plots clarify that in the OASIS dataset, OLM-ACM_Joint performs better than

TABLE 2. Comparison results for the IBSR dataset using mean Dice's similarity index (μ) and the corresponding standard deviation (σ) where available. Three decimal points are used when available to allow for fair ranking, as differences among methods are in some cases lower than 1%.

IBSR dataset	Dice $\mu \pm \sigma$	Method Description
OLM-ACM_Joint	0.852 \pm 0.021	Hybrid ACM with OLMs & multi-atlas with joint label fusion
Multi-atlas_Joint ('PICSL_Joint')	0.847 \pm 0.023	Multi-atlas & with joint label fusion
OLM-ACM	0.839 \pm 0.024	Hybrid ACM with OLMs & multi-atlas with weighted average fusion
Multi-atlas	0.833 \pm 0.026	Multi-atlas with & weighted average fusion
Rousseau et al. [68]	0.83	Patch-based labeling
Lötjönen et al. [28]	0.814	Multi-atlas & intensity modeling
Sdika [69]	0.81	Multi-atlas & intensity classification
Khan et al. [24]	0.76 \pm 0.03	Multi-structure registration & supervised atlas & correction
Arteachevarria et al. [70]	0.75	Multi-atlas & multiple combination strategies
Akselrod et al. [71]	0.69	Multiscale segmentation with probabilistic atlas prior atlas prior
Fischl et al. [22]	0.75 \pm 0.02	FreeSurfer

the Babalola et al. approach [50] and Multi-atlas_Joint for most subjects on every metric. It can also be observed that there is a decrease in segmentation performance on older subjects, especially the ones suffering from dementia for all aforementioned methods. However, segmentation performance in those subjects may have been affected by the lack of sufficient amount of similar cases. Apparently, experimenting with a dataset, with only a few problematic cases, is not sufficient for drawing conclusions on a method's behavior in those cases.

Furthermore, the agreement between the automatically and manually segmented volumes was studied with the use of the Bland-Altman analysis (Fig. 6). A high overestimation bias for the AAM method can be observed, while the Multi-atlas_Joint presents an underestimation bias. Furthermore, Multi-atlas_Joint shows a light tendency to overestimate small volumes and to underestimate the large ones. The same tendency can be observed for the OLM-ACM_Joint method. However, the OLM-ACM_Joint method has a much lower bias when compared to the other two methods. This indicates that the segmented volumes, calculated by means of the OLM-ACM_Joint method, are closer to the manually segmented ones. Fig. 7 illustrates segmentation results for 3 different subjects.

2) IBSR DATA SET

In order to validate the performance of the proposed method in the IBSR dataset, the segmentation results produced are compared with the results published over the years on this

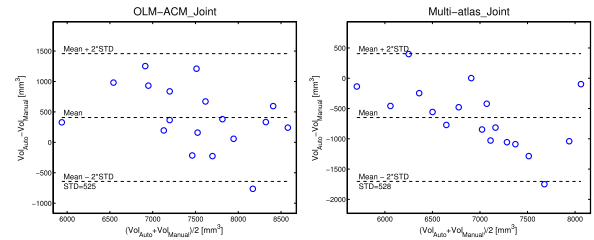


FIGURE 8. Bland-Altman plot for the IBSR dataset showing graphically the agreement between the manually segmented volumes and the volumes segmented by means of OLM-ACM_Joint.

TABLE 3. Comparison results for the OASIS-MICCAI dataset using mean Dice's similarity coefficient. Three decimal points are used to allow for fair ranking of the methods.

Method	Dice $\mu \pm \sigma$
1. 'PICSL_BC'	0.869 \pm 0.020
2. 'NonLocalSTAPLE'	0.866 \pm 0.024
3. OLM-ACM_Joint	0.865 \pm 0.026
4. 'PICSL_Joint' - Multi-atlas_Joint	0.862 \pm 0.026
5. 'MALP_EM'	0.860 \pm 0.020
7. 'CIS_JHU'	0.851 \pm 0.022
8. 'maper'	0.849 \pm 0.031
9. 'BIC-IPL-HR'	0.846 \pm 0.018
11. 'SpatialSTAPLE'	0.846 \pm 0.018
10. OLM-ACM	0.843 \pm 0.037
12. 'CRL_Weighted_STAPLE_ANTs+Baloo'	0.843 \pm 0.038
13. 'DISPATCH'	0.841 \pm 0.036
14. 'STEPS'	0.841 \pm 0.038
15. 'CRL_Weighted_STAPLE_ANTs'	0.840 \pm 0.036
16. 'SBIA_SimMSVoting'	0.840 \pm 0.043
17. 'SBIA_SimRank+NormMS'	0.839 \pm 0.038
18. 'CRL_MV_ANTs'	0.839 \pm 0.036
19. 'SBIA_BrainROIMaps_MV_IntCorr'	0.838 \pm 0.037
20. 'CRL_STAPLE_ANTs'	0.838 \pm 0.037
21. 'CRL_MV_ANTs+Baloo'	0.837 \pm 0.041
22. 'CRL_STAPLE_ANTs+Baloo'	0.837 \pm 0.042
23. 'BIC-IPL'	0.837 \pm 0.022
24. 'SBIA_BrainROIMaps_JaccDet_IntCorr'	0.837 \pm 0.037
25. 'SBIA_SimRank+NormMS+WtROI'	0.836 \pm 0.041
26. Multi-atlas	0.833 \pm 0.032
27. 'UNC_NIRAL'	0.830 \pm 0.034
28. 'CRL_Probabilistic_STAPLE_ANTs'	0.828 \pm 0.038
29. 'CRL_Probabilistic_STAPLE_ANTs+Baloo'	0.827 \pm 0.044

dataset from state-of-the-art segmentation methods, including various multi-atlas based methods. The resulting mean Dice similarity coefficient and the standard deviation of the methods are presented in Table 2 allowing a direct comparison among methods. The results indicate that OLM-ACM_Joint outperforms all previously published results. Furthermore, comparing the OLM-ACM with the Multi-atlas and the OLM-ACM_Joint with the Multi-Atlas_Joint presents a consistent improvement of 0.5-0.6% (p-values of paired t-tests 0.03 and 0.042 respectively), as well as smaller dispersion of the resulting Dice similarity coefficients (as demonstrated by the σ values). MR images from the IBSR dataset differ significantly from those of the OASIS dataset in terms of

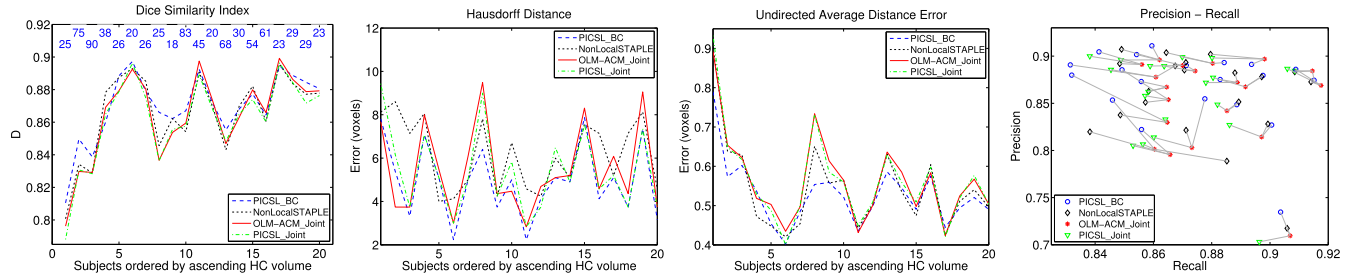


FIGURE 9. OASIS-MICCAI subset: Comparison of the four top ranked methods, as presented in Table 3, on four metrics. On the precision vs recall plot, the results of the OLM-ACM_Joint method and each of the remaining methods on the same subject are connected to allow for direct comparison. Subjects are ranked by ascending ground-truth HC volume and their ages are provided in the Dice plot. Note that in the OASIS-MICCAI subset, the Clinical Dementia Rating (CDR) for all subjects, but subject 16, is 0 or not provided by OASIS (young subjects), in contrast with OASIS subset where subjects with different CDR were included (Fig. 5).

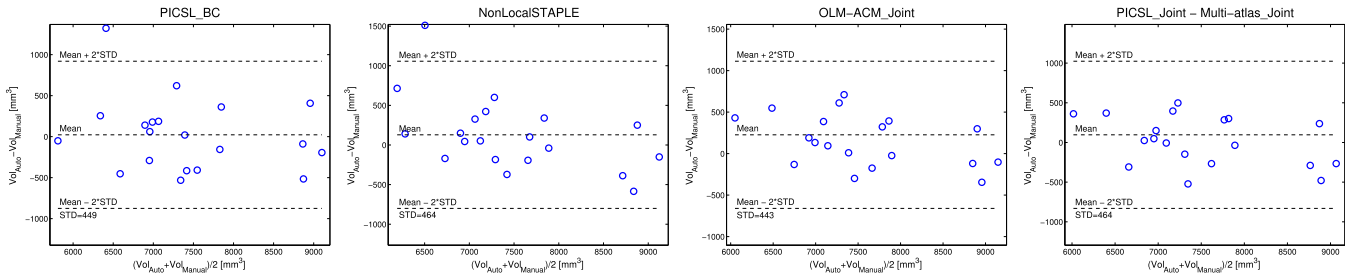


FIGURE 10. OASIS-MICCAI subset: Bland-Altman plots showing graphically the agreement between the manually segmented volumes and the volumes segmented by means of the four top ranked methods presented in Table 3.

imaging quality/varying resolution and scanner types used. Therefore, the improvement in image segmentation for both datasets suggests that the proposed method may be insensitive to differences in scanner type and image quality.

Moreover, the Bland-Altman analysis in Fig. 8 demonstrates an overestimation bias for OLM-ACM_Joint method, while the Multi-atlas_Joint method presents (similarly with the OASIS results) a larger underestimation bias. Furthermore, for the Multi-Atlas_Joint method the tendency to underestimate volumes is stronger for subjects with large HC volumes.

3) OASIS-MICCAI DATA SET

Multi-atlas labeling techniques have gained increased popularity over the past years for the segmentation of brain structures, including the hippocampus. The ‘‘Grand Challenge on Multi-Atlas Labeling’’ at the MICCAI 2012 workshop has provided the scientific community with an insight to the theory and application of current state-of-the-art multi-atlas methods, as well as with a comparative evaluation among them using the mean Dice similarity coefficient. In total, 25 different multi-atlas approaches were presented and validated, while the segmentation masks have been made publicly available. Following the same protocol as in the challenge, the proposed methodology is applied to the challenge’s dataset. Please note that around half of the methods, including the three highly ranked, have used the ANTs toolkit for the task of non-rigid registration, as the proposed method does. Thus, a fair comparison is available.

The mean Dice similarity coefficient values obtained by means of all 25 methods, as well as OLM-ACM, OLM-ACM_Joint, Multi-atlas and Multi-atlas_Joint, are provided in Table 3. It should be noted that the method ‘PICSL_Joint’ of Table 3 is actually the Multi-atlas_Joint method. The results demonstrate that the proposed methodology, when combined with the joint label fusion scheme, achieves accuracy of 0.865, with the highest accuracy achieved in this dataset being that of the ‘PICSL_BC’ method [72], which equals to 0.869. However, comparing OLM-ACM with Multi-atlas and OLM-ACM_Joint with Multi-atlas_Joint, it is clear that the application of the proposed ACM framework on top of the multi-atlas concept is beneficial also in this dataset.

Furthermore, Fig. 9 presents comparison plots for the four top ranked methods presented in Table 3 using four metrics. The Dice similarity coefficient plot shows that, except for the three smallest volumes, no bias between volume size and segmentation performance was observed for any of the methods. Furthermore, the precision-recall diagram demonstrates higher recall values for the proposed methodology, while the rest of the methods demonstrate higher precision. Moreover, the agreement between manually and automatically segmented volumes by means of the four aforementioned methods is indicated using the Bland-Altman analysis (Fig. 10). OLM-ACM_Joint presents a higher overestimation bias than the rest of the methods, while has smaller variation than them.

It is worth mentioning that the proposed concept was tested on datasets that differ in terms of the manual segmentation

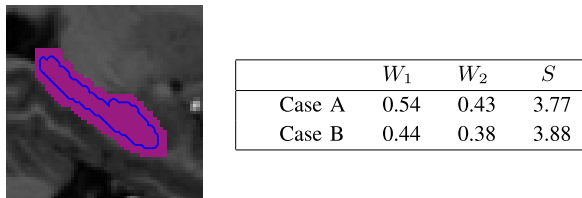


FIGURE 11. The actual boundary (blue color) and HC dilated region used for the calculation of the mean values of OLMs presented in the table on the right as case A (on the boundary) and B (purple region), respectively.

protocol. As mentioned in II-A, the manual segmentation protocol used in the OASIS-MICCAI dataset includes non gray-matter parts in the hippocampal region, while those used in IBSR and OASIS dataset do not. The proposed methodology was designed to work according to manual segmentation protocols that consider HC as a homogeneous gray matter structure. However, for the sake of completeness we wanted to show our performance also in the OASIS-MICCAI dataset, in which apparently our method can not perform in an optimum way; the region based term cannot support the inclusion of white matter, since the vast majority of HC voxels have a darker intensity. However, the proposed method ranked high in all datasets, regardless of the manual protocol used. Thus, the results suggest the potential robustness of the proposed method to the segmentation protocol.

IV. CONCLUSION

This paper advocated the incorporation of OLMs into a hybrid ACM, to be used on top of the multi-atlas concept for HC segmentation. OLM-ACM tends to improve segmentation accuracy compared to traditional prior-knowledge and data driven ACM. This is because the latter makes use of the hypothesis of consistent boundary properties and thus applies global weighting to the energy terms. On the contrary, OLM-ACM defines each term's contribution at a voxel level, taking into account the spatially varying properties of boundaries and thus allowing the optimal exploitation of the ACM energy terms. Furthermore, it consistently improves the result of the multi-atlas methods in all three datasets, which demonstrates its efficacy as a supplementary technique to the multi-atlas methods.

W_1 tends to underline image properties, either edges or statistical differences of intensities, in those regions which are located close to the boundary, by weighting them more. In this respect, W_1 makes it possible for a level set to converge at an accurate voxel point where the actual boundary is located. The table in Fig. 11 allows us to observe that W_1 does indeed take its higher values on the boundary, as desired. The sole use of W_1 is not, however, a sufficient means of achieving optimal segmentation. This is because, apart from the general knowledge where the image term should be trusted more, it is of great significance to determine at voxel level whether the edge or the region term is more trustworthy.

The concept of W_2 was introduced to tackle this issue. In regions of the boundary where gradients are high, the edge-based term is used more frequently. Vice versa, in regions characterized by a lack of strong edges the region term is preferred to allow the level set to evolve correctly. It should be noted that the use of the region-based term is generally preferred, as HC is a structure with mainly ambiguous boundaries. This is confirmed by Fig. 11 ($W_2 < 0.5$ means more weight on E_R). In addition, the level set evolution depends on the step used for its evolution. When large steps are used the evolution process is accelerated. However, when the level set is close to the real boundary, the step should be small enough to capture small deformations that are needed to achieve segmentation accuracy. This is why the use of S in ACM methods seems to be of high importance.

Regarding execution time, the major bottleneck is the task of non-rigid registration, included in the multi-atlas procedure. More precisely, the testing procedure involves registering the test image with each training image. This procedure requires $n \times 2$ hours (n is the size of the training set) with ANTs toolkit routines on an Intel Core i7 3.90Ghz computer (using 1 core). The subsequent transformation of the labels and training OLMs to the space of the target image and the calculation of the similarity metrics s_i takes on the same computer 5 min, while around an hour is required for the joint label fusion algorithm. The ACM evolution requires only 6 min on average (with the use of un-optimized Matlab code). This means that any burden regarding the computation time needed is due to the registration procedure, since during the testing phase our method increases the computational time infinitesimally comparing with the multi-atlas required time. For this reason, future work will focus on avoiding the task of non-rigid registration. Some first works towards this direction have recently been presented [68], [73], [74]. As far as the training phase is concerned, it is also computational heavy due to the sophisticated and complex nature of extracting the OLMs. On the same computer the training requires $n \times 2.6$ hours on average. However, as any other training procedure, the training is performed only once and is an offline procedure.

The proposed framework was evaluated in three publicly available datasets, none of them equipped with statistics on manual segmentation variations. These offer a good indication of the segmentation task's difficulty in a given dataset, as the goal is to offer less, or even similar, variability than the one observed between different experts. Indicatively, in two recent 3 Tesla HC studies the reported inter-rater variability was 0.91 [75], while in [11] it was 0.832 (with the intrarater being 0.891, and the automatic segmentation performance 0.844). The results presented in [76] show comparable Intra-class Correlation Coefficient between the automatic and manual volumes (0.898), compared to the inter-rater reliability (0.929). Similarly, [77] reports higher manual-manual (0.63) compared to manual-automated (0.61) HC agreement, while in [78] the difference is much higher in two datasets (inter-rater 0.80 vs automatic 0.77, and inter-rater 0.90 vs

automatic 0.75). In [79] two raters were using the same tool to enhance their HC segmentation skills. In a two series experiment both of them managed to raise their intra-rater agreement (from 0.79 to 0.94). Interestingly, once this was accomplished, their inter-rater agreement decreased from 0.68 to 0.57. This could mean that the two raters were doing excellent but different segmentations, and is further suggesting that inter-rater reliability may be a useful indication but perhaps insufficient too.

Overall, the proposed method is an ACM based extension of the multi-atlas methodology. Experimental results on three datasets, with different manual segmentation protocols, demonstrate the efficacy of the proposed method and its appropriateness to be used on top of multi-atlas methods, even the sophisticated ones. Thus, combination of the proposed method with an even better performing multi-atlas based algorithm (such as the PICSL_BC [33]) can lead to further improvements and is inline with our future work. However, results from the OASIS-MICCAI dataset in comparison with those from OASIS and IBSR datasets, show that there exists space for further improvements in datasets for which the manual segmentation protocol followed includes white matter parts in the hippocampal region. In this respect, future work will include investigating ways to assign in W_1 higher values to the prior term in the alveus/fimbria regions. Given that the multi-atlas based prior knowledge mapped in L will be voting the inclusion of alveus/fimbria, this modification will allow the proposed methodology to perform in an optimum way in such datasets too.

In conclusion, evidence favors the inclusion of HC volumetry in clinical practice, to enhance disease diagnosis, within a decision support system. Hence, actions are envisaged for establishing it as a biomarker. The above highlight the need for automatic HC segmentation methods that can offer as high accuracy as possible. Any improvement that is proved statistically significant could help identifying a more precise and reliable biomarker. The proposed framework demonstrates a supplementary technique to the multi-atlas methods, consistently improving their performance, while slightly increasing the computational cost. It ranked high in three datasets (even in one with a different definition of hippocampus), posing itself as a promising candidate for large-scale experimentation.

ACKNOWLEDGMENT

The authors are grateful to the OASIS team for providing us with their dataset as well as the IBSR repository for providing us with manually-guided expert segmentation results along with their MRI dataset. We would also like to give special thanks to Angelos Baltatzidis M.D., Radiologist for providing us with the manual segmentations for the OASIS dataset. Furthermore, for the OASIS-MICCAI dataset we would like to thank the workshop organizers, Prof. Bennett Landman and Prof. Simon Warfield, as well as Neuromorphometrics, Inc. for providing the manual segmentations.

APPENDIX

A. ENERGY MODEL

Following the level set method [80], in the image domain $\Omega \in R^3$, we define an evolving curve C implicitly represented as the zero level set of a signed distance function $\phi : R^3 \rightarrow \Omega$

$$C = \{(x, y, z) \in \Omega \mid \phi(x, y, z) = 0\} \quad (8)$$

where $\phi(x, y, z) < 0$ inside the contour C and $\phi(x, y, z) > 0$ outside the contour C .

The contour update equation based on the local weighting maps W_1 and W_2 is defined as: where the operation \circ denotes the Hadamard product and

$$\begin{aligned} \frac{\partial \phi}{\partial t} = & W_1 \circ \left[W_2 \circ \frac{\partial \phi_E}{\partial t} + (1 - W_2) \circ \frac{\partial \phi_R}{\partial t} \right] \\ & + (1 - W_1) \circ \frac{\partial \phi_{prior}}{\partial t} \end{aligned} \quad (9)$$

B. REGION-BASED TERM (E_R)

The region-based term used is the one presented by Chan-Vese in [35], where the curve is being evolved to minimize the following energy functional:

$$\begin{aligned} E_R = & \lambda_1 \int_{\Omega_1} |I(x, y, z) - c_1|^2 dx dy dz \\ & + \lambda_2 \int_{\Omega_2} |I(x, y, z) - c_2|^2 dx dy dz, \quad (x, y, z) \in \Omega \end{aligned} \quad (10)$$

where I is the target MR image, c_1 and c_2 are the average intensities of the regions inside and outside the contour, respectively and $\lambda_1, \lambda_2 \geq 0$ are balancing factors for the properties of the interior and the exterior regions of the estimated boundary. Based on E_R , the evolution equation of the contour driven by the region-based term becomes:

$$\frac{\partial \phi_R}{\partial t} = \delta_\epsilon(\phi) \left[\mu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nu - \lambda_1 (I - c_1)^2 + \lambda_2 (I - c_2)^2 \right] \quad (11)$$

where $\delta_\epsilon(\phi)$ is the Dirac function, ν controls the propagation speed, and $\mu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right)$ is a regularization term that controls the smoothness of the contour.

C. EDGE-BASED TERM (E_E)

The edge-based term is formulated by minimizing the energy functional defined in Caselles et al. [34]:

$$E_E = \int_{\Omega} g(\mathbf{v}) |\nabla \phi(\mathbf{v})| d\mathbf{v} \quad (12)$$

where g is an edge stopping function defined as in [34]:

$$g(|\nabla(I)|) = \frac{1}{1 + |\nabla G_\sigma * I|} \quad (13)$$

with G_σ standing for the Gaussian convolution kernel of size $3 \times 3 \times 3$ and standard deviation 0.5. The contour evolution

equation driven only by the edge-based term reads:

$$\frac{\partial \phi_E}{\partial t} = \left[g |\nabla(\phi)| \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \nabla g \cdot \nabla \phi \right] \quad (14)$$

where $\frac{\nabla \phi}{|\nabla \phi|}$ is the regularization term.

D. PRIOR TERM (E_{PRIOR})

The prior term is modeled by applying the region-based ACM on L . The selection of the Chan-Vese approach to model the prior term is based on the fact that L is an image with very smooth transitions. Thus, the energy functional is defined as:

$$E_{\text{prior}} = v_1 \int_{\Omega_1} |L(x, y, z) - d_1|^2 dx dy dz + v_2 \int_{\Omega_2} |L(x, y, z) - d_2|^2 dx dy dz, \quad (x, y, z) \in \Omega \quad (15)$$

where d_1 and d_2 are the mean values in the regions of L inside and outside C . Similarly to equation (10), v_1 and v_2 are balancing factors for the properties of the interior and the exterior regions, which were set equal to one, since both inside and outside regions are smooth and homogeneous. Based on E_{prior} , the evolution equation for the contour driven by the prior term is defined as:

$$\frac{\partial \phi_{\text{prior}}}{\partial t} = \delta_\epsilon(\phi) \left[\mu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - v - v_1(L - d_1)^2 + v_2(L - d_2)^2 \right] \quad (16)$$

By means of equations (9), (11), (14), (16), the overall contour update formula becomes:

$$\begin{aligned} \frac{\partial \phi}{\partial t} = & W_1 \circ W_2 \circ \left[g |\nabla(\phi)| \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \nabla g \cdot \nabla \phi \right] \\ & + \delta_\epsilon(\phi) \left[(1 - W_1 \circ W_2) \mu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right. \\ & - W_1 \circ (1 - W_2) \circ \left(\lambda_1(I - c_1)^2 - \lambda_2(I - c_2)^2 \right) \\ & \left. - (1 - W_1) \circ \left(v_1(L - d_1)^2 + v_2(L - d_2)^2 \right) \right] \quad (17) \end{aligned}$$

E. CALCULATION OF TRAINING OLMS AND ACM PARAMETERS THROUGH GRAPH-CUTS

Fig. 3 provides the overview of this procedure. Let us consider the problem of finding the optimum combination of values for W_{1i} , W_{2i} , S_i at a voxel v for a training image I_i , $i = 1, \dots, n$. Such a procedure can be handled as a graph-cut labeling problem where each label f_v is being mapped to a combination of three labels: a label $f_{v1} \in [0, 1]$ that represents the amount of contribution of the prior term (W_{1i}), a label $f_{v2} \in [0, 1]$ that represents the amount of contribution of the edge-based term (W_{2i}) as well as a label $f_{v3} \in [1, 6]$ representing the step of evolution (S_i). W_{1i} 's and W_{2i} 's values are in the interval $[0, 1]$ as they represent percentages of contribution of the various energy terms, while the values of S_i stand for the size of the

time steps and thus, can be integer numbers. The mapping function is expressed as:

$$f : f_v \in [0, P] \rightarrow f_{v1} \in [0, 1] \wedge f_{v2} \in [0, 1] \wedge f_{v3} \in [1, 6] \quad (18)$$

where P is the number of possible permutations of f_{v1}, f_{v2}, f_{v3} . Due to computational considerations, only 8 discrete values in the interval $[0, 1]$ were used for both f_{v1} and f_{v2} . For the same reason, the S 's possible values were limited to 6. It is obvious that using more values could lead to better accuracy, but this selection was done having in mind an optimal balance between accuracy and computational cost in terms of memory requirements.

In order to formulate our problem, we consider the complete set of voxels \mathbf{V} which belong to image I_i of the training set and its corresponding label image L_i that serves as the ground-truth image. The goal is to define an optimal labeling f for \mathbf{V} . Finding the optimal labeling is equivalent to minimizing an energy functional $E(\mathbf{f})$. According to graph cut theory, the energy functional can be formulated as:

$$E(\mathbf{f}) = \sum_{v \in \mathbf{V}} D_v(f_v) + \sum_{v \in \mathbf{P}, q \in N_v} \mathcal{V}_{v,q}(f_v, f_q) \quad (19)$$

where D_v is the individual voxel cost for voxel v and measures at which extent label f_v fits for voxel v given the ground-truth segmentation and the resulting one. N_v is the set of neighboring voxels of v and $\mathcal{V}_{v,q}(f_v, f_q)$ [66] is the interaction potential between voxels v, q that penalizes discontinuities between neighboring voxels and thus encourages spatial coherence and it is defined as $\mathcal{V}_{v,q}(f_v, f_q) = \min(|f_{v1} - f_{q1}| + |f_{v2} - f_{q2}| + |f_{v3} - f_{q3}|, K)$, where K is set equal to 4 based on experimentation. Within our framework, the data cost function D_v is defined as:

$$D_v = |S \circ \frac{\partial \phi}{\partial t}(f_v) - \phi_{GT}| \quad (20)$$

where $\frac{\partial \phi}{\partial t}$ is given by equation (17) and ϕ_{GT} stands for the level set formulation of the corresponding label image.

A similar formulation with that in equation (19) is also used in order to find optimal ACM parameters, $\lambda_{1i}, \lambda_{2i}$ for each image in the training set. Thus, to calculate $\lambda_{1i}, \lambda_{2i}$, which are set in $[0, 1]$, graph-cuts are used to minimize the difference between the region update term and the level set formulation of GT:

$$D_v = |S \circ \frac{\partial \phi_R}{\partial t}(f_v) - \phi_{GT}| \quad (21)$$

REFERENCES

- [1] Pan American Health Organization. (2009). *Strategy and plan of action on mental health*. Washington, DC, USA. Tech. Rep. CD 49/11, 2009 [Online]. Available: http://www2.paho.org/hq/dmdocuments/2009/MENTAL_HEALTH.pdf
- [2] N. A. DeCarolis and A. J. Eisch, "Hippocampal neurogenesis as a target for the treatment of mental illness: A critical evaluation," *Neuropharmacology*, vol. 58, no. 6, pp. 884–893, 2010.
- [3] H. U. Wittchen, F. Jacobi, J. Rehm, A. Gustavsson, M. Svensson, B. Jönsson, et al., "The size and burden of mental disorders and other disorders of the brain in Europe 2010," *Eur. Neuropsychopharmacol.*, vol. 21, pp. 655–679, 2011.

- [4] B. Bogerts, M. Ashtari, G. Degreef, J. M. Alvir, R. M. Bilder, and J. A. Lieberman, "Reduced temporal limbic structure volumes on magnetic resonance images in first episode schizophrenia," *Psychiatry Res*, vol. 35, no. 1, pp. 1–13, 1990.
- [5] A. Breier, R. W. Buchanan, A. Elkashef, R. C. Munson, B. Kirkpatrick, and F. Gellad, "Brain morphology and schizophrenia. A magnetic resonance imaging study of limbic, prefrontal cortex, and caudate structures," *Archives General Psychiatry*, vol. 49, no. 12, pp. 921–926, 1992.
- [6] D. Velakoulis, S. J. Wood, M. T. Wong, P. D. McGorry, A. Yung, L. Phillips, et al., "Hippocampal and Amygdala volumes according to Psychosis stage and diagnosis: A magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals," *Archives General Psychiatry*, vol. 63, no. 2, pp. 139–49, 2006.
- [7] A. Sumich, X. A. Chitnis, D. G. Fannon, S. O'Ceallaigh, V. C. Doku, and A. Falrowicz, "Temporal lobe abnormalities in first-episode psychosis," *Amer. J. Psychiatry*, vol. 159, no. 7, pp. 1232–1235, 2002.
- [8] P. Brambilla, J. P. Hatch, and J. C. Soares, "Limbic changes identified by imaging in bipolar patients," *Current Psychiatry Rep.*, vol. 10, no. 6, pp. 505–509, 2008.
- [9] C. Langan and C. McDonald, "Neurobiological trait abnormalities in bipolar disorder," *Mol. Psychiatry*, vol. 14, no. 9, pp. 833–846, 2009.
- [10] H. P. Blumberg, J. Kaufman, A. Martin, R. Whiteman, J. H. Zhang, J. C. Gore, et al., "Amygdala and hippocampal volumes in adolescents and adults with bipolar disorder," *Archives General Psychiatry*, vol. 60, no. 12, pp. 1201–1208, 2003.
- [11] G. P. Winston, M. J. Cardoso, E. J. Williams, J. L. Burdett, P. A. Bartlett, M. Espak, et al., "Automated hippocampal segmentation in patients with epilepsy: Available free online," *Epilepsia*, vol. 54, no. 12, pp. 2166–2173, 2013.
- [12] F. de Andrés Trelles, L. Pani, B. Jonsson, and M. Isaac, "Qualification opinion of low hippocampal volume (atrophy) by MRI for use in regulatory clinical trials in pre-dementia stage of Alzheimer's disease," European Medicines Agency, London, U.K., Tech. Rep. EMA/CHMP/SAWP/809208/2011, 2011.
- [13] R. La Joie, A. Perrotin, V. de La Sayette, S. Egret, L. Doevure, S. Belliard, et al., "Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia," *NeuroImage, Clin.*, vol. 3, pp. 155–162, Aug. 2013.
- [14] A. T. Du, N. Schuff, D. Amend, M. Laakso, Y. Hsu, W. Jagust, et al., "Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease," *J. Neurol. Neurosurgery Psychiatry*, vol. 71, no. 4, pp. 441–447, 2001.
- [15] C. R. Jack, R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, et al., "Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease," *Neurology*, vol. 49, no. 3, pp. 786–794, 1997.
- [16] B. Fontaine, D. Seilhean, A. Tourbah, C. Daumas-Duport, C. Duyckaerts, N. Benoit, et al., "Dementia in two histologically confirmed cases of multiple sclerosis: One case with isolated dementia and one case associated with psychiatric symptoms," *J. Neurol. Neurosurgery Psychiatry*, vol. 57, no. 3, pp. 353–359, 1994.
- [17] M. Tsolaki, A. Drevelegas, S. Karachristianou, K. Kapinas, D. Divanoglou, and K. Routsonis, "Correlation of dementia, neuropsychological and MRI findings in multiple sclerosis," *Dementia*, vol. 5, no. 1, pp. 48–52, 1994.
- [18] S. J. Teipel, M. B. Schapiro, G. E. Alexander, J. S. Krasuski, B. Horwitz, C. Hoehne, et al., "Relation of corpus callosum and hippocampal size to age in nondemented adults with down's syndrome," *Amer. J. Psychiatry*, vol. 160, no. 10, pp. 1870–1878, 2003.
- [19] C. R. Jack, F. Barkhof, M. A. Bernstein, M. Cantillon, P. E. Cole, C. DeCarli, et al., "Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criteria for Alzheimer's disease," *Alzheimers Dement.*, vol. 7, no. 4, pp. 474–485, 2011.
- [20] J. Hass, E. Walton, H. Kirsten, J. Liu, L. Priebe, C. Wolf, et al., "A genome-wide association study suggests novel loci associated with a schizophrenia-related brain-based phenotype," *PLoS ONE*, vol. 8, no. 6, p. e64872, 2013.
- [21] G. B. Frisoni and C. R. Jack, "Harmonization of magnetic resonance-based manual hippocampal segmentation: A mandatory step for wide clinical use," *Alzheimers Dementia*, vol. 7, no. 2, pp. 171–174, 2011.
- [22] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, et al., "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.
- [23] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, 2009.
- [24] A. R. Khan, M. K. Chung, and M. F. Beg, "Robust atlas-based brain segmentation using multi-structure confidence-weighted registration," in *Proc. Int. Conf. MICCAI*, vol. 12, 2009, pp. 549–557.
- [25] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.
- [26] A. R. Khan, N. Cherbuin, W. Wen, K. Anstey, P. S. Sachdev, and M. F. Beg, "Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuprDyn): Validation on hippocampus segmentation," *NeuroImage*, vol. 56, no. 1, pp. 126–139, 2011.
- [27] D. L. Collins and J. C. Pruessner, "Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion," *NeuroImage*, vol. 52, no. 4, pp. 1355–1366, Oct. 2010.
- [28] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, et al., "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage*, vol. 49, no. 3, pp. 2352–2365, 2010.
- [29] B. Landman and S. Warfield, "MICCAI 2012 workshop on multi-atlas labeling," in *Proc. Med. Image Comput. Comput. Assisted Intervent. Conf. Grand Challenge Workshop Multi-Atlas Labeling Challenge Result*, 2012.
- [30] L. Dice, "Measure of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [31] H. Wang, J. W. Suh, S. R. Das, J. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Mar. 2013.
- [32] A. J. Asman and B. A. Landman, "Non-local statistical label fusion for multi-atlas segmentation," *Med. Image Anal.*, vol. 17, no. 2, pp. 194–208, 2013.
- [33] H. Wang and P. A. Yushkevich, "Spatial bias in multi-atlas based segmentation," in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 909–916.
- [34] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.
- [35] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [36] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [37] Y. Zhang, B. J. Matuszewski, L. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *Proc. 5th Int. Conf. BioMed. Visualizat.*, Jul. 2008, pp. 71–76.
- [38] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and applications," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [39] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2000, pp. 316–323.
- [40] J. Yang, L. H. Staib, and J. S. Duncan, "Neighbor-constrained segmentation with level set based 3-D deformable models," *IEEE Trans. Med. Imag.*, vol. 23, no. 8, pp. 940–948, Aug. 2004.
- [41] G. J. Edwards, C. J. Taylor, and T. F. Cootes, "Interpreting face images using active appearance models," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 300–305.
- [42] T. F. Cootes, G. J. Edward, and C. J. Taylor, "Active appearance model," in *Proc. Eur. Conf. Comput. Visualizat.*, 1998, pp. 484–498.
- [43] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
- [44] M. B. Stegmann and R. Larsen, "Multi-band modelling of appearance," *Image Vis. Comput.*, vol. 21, no. 1, pp. 61–67, 2003.
- [45] S. Hu and D. L. Collins, "Joint level-set shape modeling and appearance modeling for brain structure segmentation," *NeuroImage*, vol. 36, no. 3, pp. 672–683, 2007.
- [46] S. Hu, P. Coupé, J. C. Pruessner, and D. L. Collins, "Appearance-based modeling for segmentation of hippocampus and amygdala using multi-contrast MR imaging," *NeuroImage*, vol. 58, no. 2, pp. 549–559, 2011.

- [47] R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: Application to prostate MRI segmentation," *IEEE Trans. Med. Imag.*, vol. 31, no. 8, pp. 1638–50, Aug. 2012.
- [48] S. Hu, P. Coupé, J. C. Pruessner, and D. L. Collins, "Validation of appearance-model based segmentation with patch-based refinement on medial temporal lobe structures," in *Proc. MICCAI Workshop Multi-Atlas Label. Statist. Fusion*, 2011, pp. 28–37.
- [49] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson, "A Bayesian model of shape and appearance for subcortical brain segmentation," *NeuroImage*, vol. 56, no. 3, pp. 907–922, 2011.
- [50] K. Babalola and T. Cootes, "Using parts and geometry models to initialise active appearance models for automated segmentation of 3D medical images," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2010, pp. 1069–1072.
- [51] D. Zarpalas, A. Zafeiropoulos, P. Daras, and N. Maglaveras, "Hippocampus segmentation using a local prior model on its boundary," in *Proc. Int. Conf. Mach. Vis., Image Process. Pattern Anal.*, 2011, pp. 1–6.
- [52] D. Zarpalas, P. Gkontra, P. Daras, and N. Maglaveras, "Hippocampus segmentation by optimizing the local contribution of image and prior terms, through graph cuts and multi-atlas," in *Proc. 9th IEEE ISBI*, May 2012, pp. 1168–1171.
- [53] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cognit. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [54] K. L. Narr, P. M. Thompson, P. Szaszko, D. Robinson, S. Jang, R. P. Woods, et al., "Regional specificity of hippocampal volume reductions in first-episode schizophrenia," *NeuroImage*, vol. 21, no. 4, pp. 1563–1575, 2004.
- [55] K. L. Narr, P. M. Thompson, T. Sharma, J. Moussai, R. Blanton, B. Anvar, et al., "Three-dimensional mapping of temporo-limbic regions and the lateral ventricles in schizophrenia: Gender effects," *Biol. Psychiatry*, vol. 50, no. 2, pp. 84–97, 2001.
- [56] K. L. Narr, T. G. M. van Erp, T. D. Cannon, R. P. Woods, P. M. Thompson, S. Jang, et al., "A twin study of genetic contributions to hippocampal morphology in schizophrenia," *Neurobiol. Dis.*, vol. 11, no. 1, pp. 83–95, 2002.
- [57] C. R. Jack, W. H. Theodore, M. Cook, and G. McCarthy, "MRI-based hippocampal volumetrics: Data acquisition, normal ranges, and optimal protocol," *Magn. Reson. Imag.*, vol. 13, no. 8, pp. 1057–1064, 1995.
- [58] J. G. Levitt, R. E. Blanton, R. Caplan, R. Asarnow, D. Guthrie, A. W. Toga, et al., "Medial temporal lobe in childhood-onset schizophrenia," *Psychiatry Res.*, vol. 108, no. 1, pp. 17–27, 2001.
- [59] J. Pantel, D. S. O'Leary, K. Cretsinger, H. J. Bockholt, H. Keefe, V. A. Magnozza, et al., "A new method for the in vivo volumetric measurement of the human hippocampus with high neuroanatomical accuracy," *Hippocampus*, vol. 10, no. 6, pp. 752–758, 2000.
- [60] J. C. Pruessner, L. M. Li, W. Serles, M. Pruessner, D. L. Collins, N. Kabani, et al., "Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories," *Cereb. Cortex*, vol. 10, no. 4, pp. 433–442, 2000.
- [61] M. Boccardi, M. Bocchetta, R. Ganzola, N. Robitaille, A. Redolfi, S. Duchesne, et al. (2013, May). Operationalizing protocol differences for EADCADNI manual hippocampal segmentation. *Alzheimers Dementia* [Online]. Available: <http://www.hippocampalprotocol.net/SOPs/publications.php>
- [62] C. Konrad, T. Ukas, C. Nebel, V. Arolt, A. W. Toga, and K. L. Narr, "Defining the human hippocampus in cerebral magnetic resonance images—An overview of current segmentation protocols," *NeuroImage*, vol. 47, no. 4, pp. 1185–1195, 2009.
- [63] N. Makris, D. N. Kennedy, J. Meyer, A. Worth, V. S. Caviness, L. Seidman, et al. (2004). *General brain segmentation: Method and utilization. Center for morphometric analysis*. Massachusetts General Hospital. Boston, MA, USA. Tech. Rep. [Online]. Available: <http://www.cma.mgh.harvard.edu/manuals/>
- [64] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.
- [65] A. Klein, S. S. Ghosh, B. Avants, B. T. T. Yeo, B. Fischl, B. Ardekani, et al., "Evaluation of volume-based and surface-based brain image registration methods," *NeuroImage*, vol. 51, no. 1, pp. 214–220, 2010.
- [66] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [67] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [68] F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE Trans. Med. Imag.*, vol. 30, no. 10, pp. 1852–1862, Oct. 2011.
- [69] M. Sdika, "Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote," *Med. Image Anal.*, vol. 14, no. 2, pp. 219–226, 2010.
- [70] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1266–1277, Aug. 2009.
- [71] A. Akselrod-Ballin, M. Galun, J. M. Gomori, A. Brandt, and R. Basri, "Prior knowledge driven multiscale segmentation of brain MRI," in *Proc. MICCAI*, vol. 10, 2007, pp. 118–126.
- [72] H. Wang, B. Avants, and P. A. Yushkevich, "Grand challenge on multi-atlas segmentation: A combined joint label fusion and corrective learning approach," in *Proc. MICCAI Workshop Multi-Atlas Label.*, 2012.
- [73] P. Coupé, J. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Nonlocal patch-based label fusion for hippocampus segmentation," in *Proc. Int. Conf. MICCAI*, Beijing, China, 2010, pp. 129–136.
- [74] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [75] J. L. Winterburn, J. C. Pruessner, S. Chavez, M. M. Schira, N. J. Lobaugh, A. N. Voineskos, et al. "A novel in vivo atlas of human hippocampal subfields using high-resolution 3T magnetic resonance imaging," *NeuroImage*, vol. 74, pp. 254–265, Jul. 2013.
- [76] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and the Alzheimer's Disease Neuroimaging Initiative, "LEAP: Learning embeddings for atlas propagation," *NeuroImage*, vol. 49, no. 2, pp. 1316–1325, 2010.
- [77] O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, et al., "Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 27, no. 4, pp. 979–990, 2005.
- [78] C. A. Bishop, M. Jenkinson, J. Declerck, and D. Merhof, "Evaluation of hippocampal segmentation methods for healthy and pathological subjects," in *Proc. Eurograph. Workshop Vis. Comput. Biol. Med.*, 2010, pp. 17–24.
- [79] G. Gerig, M. Jomier, and M. Chakos, "Valmet: A new validation tool for assessing and improving 3D object segmentation," in *Proc. Int. Conf. MICCAI*, 2001, pp. 1–8.
- [80] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. New York, NY, USA: Springer-Verlag, 2002.



DIMITRIOS ZARPALAS received the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Greece, in 2003, and the M.Sc. degree in computer vision from the Department of Electrical Engineering, The Pennsylvania State University, USA, in 2006. Currently, he is a Ph.D. candidate with the Laboratory of Medical Informatics, the Medical School, Aristotle University of Thessaloniki, and an Associate Researcher with the Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki. His main research interests include 3-D medical image processing, 3-D shape analysis, real time 3-D reconstruction from multiple passive/active sensors, dynamic mesh coding, 3-D motion analysis, 3-D object recognition, search and retrieval, and classification of 3-D objects.



POLYXENI GKONTRA received the Diploma degree in electrical and computer engineering and the M.Sc. degree in medical informatics from the Aristotle University of Thessaloniki, Greece, in 2008 and 2013, respectively. Since 2011, she has been a Research Assistant with the Information Technologies Institute, Centre for Research and Technology Hellas. Her research interests lie in the area of biomedical engineering, including biomedical image and signal processing and analysis, with a special focus on processing of brain MRI, prostate ultrasound, multiphoton microscopy, IVUS, neurocognitive testing, and vital signs.



PETROS DARAS is a Researcher Grade B with the Information Technologies Institute, Centre for Research and Technology Hellas. His main research interests include multimedia processing, multimedia and multimodal search engines, 3-D reconstruction from multiple sensors, dynamic mesh coding, medical image processing, and bioinformatics. He has co-authored more than 30 papers in refereed journals, 25 book chapters, and more than 70 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. He has been the Chair of the IEEE Interest Group (IG) on Image, Video and Mesh Coding since 2012 - <http://committees.comsoc.org/mmc/IG/IMVIMECIG.asp> and a Key Member of the IEEE IG on 3-D Rendering, Processing and Communications since 2010 - <http://committees.comsoc.org/mmc/IG/3DRPCIG.asp>. He regularly acts as a reviewer/evaluator for the EC.



NICOS MAGLAVERAS received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki (A.U.Th.), Greece, in 1982, and the M.Sc. and Ph.D. degrees in electrical engineering and computer science from Northwestern University, Evanston, IL, USA, in 1985 and 1988, respectively. He is currently a Professor and Director with the Laboratory of Medical Informatics, A.U.Th. He is the Head of the graduate program in medical informatics at A.U.Th. and is a Collaborating Researcher with the Center of Research and Technology Hellas, Institute of Applied Biosciences. His current research interests include nonlinear biological systems simulation, cardiovascular engineering, biomedical informatics, ehealth, AAL, personalized health, biosignal analysis, medical imaging, CDSS, and neurosciences. He has published more than 300 papers in peer reviewed international journals, books, and conference proceedings. He served as the President of the EANBES from 2008 to 2010, and a member of the editorial board or invited editor in a number of international journals and participated as a core partner in over 35 National and EU funded projects.