

FULLY AUTOMATIC HIPPOCAMPUS SEGMENTATION DISCRIMINATES BETWEEN EARLY ALZHEIMER'S DISEASE AND NORMAL AGING

M. Chupin^{1,7}, G. Chételat², L. Lemieux³, B. Dubois^{4,7}, L. Garnero^{1,7}, H. Benali^{5,7}, F. Eustache², S. Lehericy^{4,6,7}, B. Desgranges², O. Colliot^{1,7}

¹Cognitive Neuroscience & Brain Imaging Laboratory, CNRS UPR 640-LENA, UPMC - Paris 6, Paris, France

²Inserm-EPHE-Université de Caen Basse-Normandie, Unité U923, GIP Cyceron, CHU Côte de Nacre, Caen, France

³Department of Clinical & Experimental Epilepsy, Institute of Neurology, UCL, London, UK

⁴INSERM U610, UPMC - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France

⁵INSERM UMR_S 678, UPMC - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France

⁶Center for Neuroimaging Research-CENIR, UPMC - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France

⁷IFR49, Orsay, France

ABSTRACT

The hippocampus is among the first structures affected in Alzheimer's disease (AD); hippocampal MRI volumetry is a potential biomarker for AD but is hindered by the limitations of manual segmentation. We propose a fully automatic method using probabilistic and anatomical priors for hippocampus segmentation. Probabilistic information is derived from 16 young controls and anatomical knowledge is modeled with automatically detected landmarks. The results were compared with manual segmentation on data from 16 young healthy controls, with a leave-one-out strategy, and 8 AD patients. High accuracy was found for both groups (volume error 6% and 7%, overlap 87% and 86%, respectively). The resulting volumes were used to discriminate between 25 elderly subjects, 25 early AD patients and 24 patients with amnesic mild cognitive impairment (MCI). The classification proved accurate with 87% of the AD patients and 74% of the MCI patients correctly classified with respect to the elderly controls.

Index Terms— Segmentation, Hippocampus, Alzheimer's disease, Classification

1. INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia; its early and accurate diagnosis is challenging. The hippocampus (*Hc*) is a gray matter structure of the temporal lobe known to be affected at the earliest stage of AD [1]. Hippocampal volumetry on magnetic resonance images (MRI) can thus constitute a useful diagnostic tool [2]. Up to now, hippocampal volumetry mostly relies on highly time-consuming manual segmentation, which is rater-dependent, and not feasible in clinical routine.

Automatic segmentation of the hippocampus would overcome these limitations and provide a useful biomarker of AD. The incomplete definition of *Hc* boundaries on MRI scans makes the use of prior information necessary for

ensuring accurate and robust automatic segmentation. Prior knowledge can come from statistical information on shape [3,4], deformations [5] or from registering a single subject atlas template [6]; nevertheless, these methods may be unsuitable for diseased structures. We previously proposed a semi-automatic method [7] for the segmentation of *Hc* and the amygdala (*Am*), based on simultaneous region deformation driven by anatomical priors, through formalized anatomical relationships derived from stable anatomical patterns [8]. Nevertheless, the method required manual initialization (bounding box and two voxel seeds).

Global information given by an atlas could allow automatic initialization; furthermore, segmentation using probabilistic information [9,10] offers more thorough global spatial knowledge compared to single object atlas.

To eliminate any user intervention from the procedure, we introduce a hybrid method driven by anatomical and probabilistic prior knowledge [11]. Anatomical information is derived from local anatomical patterns that are stable in controls and patients, around landmarks automatically detected during the deformation. Probabilistic information is derived from an atlas built from the registration manually segmented *Hc* and *Am* for 16 young healthy subjects. Initialization is obtained from global information and deformation is constrained by local information (both anatomical and probabilistic).

In this paper, we validate this segmentation procedure in patients with AD and assess the ability of resulting *Hc* volumes to classify subjects with normal aging, MCI (patients with impaired memory but without dementia) and early AD with no prior knowledge on their diagnosis.

2. FULLY AUTOMATIC SEGMENTATION

The segmentation is based on the alternate deformation of two objects, one for *Hc* and one for *Am*, from two initial objects, through homotopic region deformation. It is modeled in a Bayesian framework, the deformation being

driven by an iterative energy minimization. This energy is defined with a functional made of five terms: global and local data attachment, regularization and volume and surface terms [7]. The initial objects are determined from the probabilistic atlases, inside an automatically extracted bounding box. The energy functional is then iteratively minimized for Hc and Am , with additional constraints derived from the anatomical and probabilistic priors.

2.1. Probabilistic Atlas

The datasets from N (here $N=16$) young healthy subjects were manually segmented by an expert following a protocol ensuring coherence in the three planes [7]. For each of the atlas subject, $S_i \{i=1\dots N\}$, the transformation T_i to the MNI standard space is then obtained through the unified registration and segmentation module of SPM5 [12] using the native data. The transformation (expressed on a basis of ~ 1000 cosine functions) is then used to propagate the manually labeled binary masks (Hc^i and Am^i) to the MNI space. The atlases PA_{Hc} and PA_{Am} are created as follows:

$$\forall v \in \Omega, PA_{Hc}(v) = \frac{1}{N} \sum_{i=1}^N T_i(Hc_i)(v) \text{ and } PA_{Am}(v) = \frac{1}{N} \sum_{i=1}^N T_i(Am_i)(v),$$

where v is any voxel in the MRI space Ω , $PA_{Hc}(v)$ and $PA_{Am}(v)$ are the probabilities that v belongs to Hc and Am . The atlas creation step is done only once.

2.2. Initialization

The first step is to compute forward and backward transformations T and T^{-1} between native and MNI spaces. Individual atlases IPA_{Hc} and IPA_{Am} are created by back-registering the atlases PA_{Hc} and PA_{Am} using T^{-1} .

IPA_{Hc} and IPA_{Am} are used to automatically define left and right bounding boxes around the structures of interest, as the smallest boxes embedding the non-null probability values in both atlases, with an extra one-voxel margin, for the left and right hemispheres, as illustrated in figure 1 a.

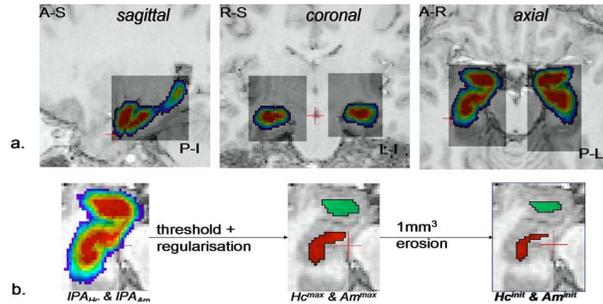


Figure 1: initialization: a. bounding boxes, b. initial objects

Atlas mismatch in the bounding box is automatically detected and corrected when necessary. For this, it is assumed that if IPA_{Hc} is locally mis-registered, the 0.5-probability object will cover a wider intensity range than if IPA_{Hc} is correctly registered. The 0.5-probability object for Hc is defined as $\{v, IPA_{Hc}(v) \geq 0.5\}$. Correction is done by

moving IPA_{Hc} in its 6-neighborhood to minimize the difference in intensity ranges.

The last step of the initialization procedure is to identify initial objects for Hc and Am , defined as the eroded regularized maximal probability objects (figure 1 b).

2.3. Deformation

The deformation is then driven by the iterative minimization of the energy functional. At each iteration, voxel candidates are determined at the border of the deforming objects, for which re-classification will be considered; meta-regions are automatically detected (interface between Hc and Am , 11 types of anatomical landmarks at the border of Hc and Am) [7]. The energy is then minimized on the voxel candidates through an ICM procedure; it is locally defined for each structure, with specific features for meta-regions: the interface, to optimize the competitive behavior, low and high likelihood zones defined around the anatomical landmarks from intensity and spatial local relationships, and three zones derived from the probability maps: $PZ^0 = \{v, IPA(v)=0\}$, $PZ^1 = \{v, IPA(v)=1\}$, $PZ^{0.75} = \{v, 0.75 \leq IPA(v) < 1\}$. These specific features are included in the anisotropic non-stationary regularization, comparing $N_o(v)$ number of O-labeled neighbors of v with a standard value \tilde{N} , a tolerance σ_I to prevent holes and wires:

$$E_o^I(v) = \left(\frac{\tilde{N} - \gamma_o^{PZ}(v) \gamma_o^{AZ}(v) \cdot (N_o(v) + \alpha_o^I(v))}{\sigma^I} \right)^5.$$

$\alpha^I = 0$, except for voxels detected as "tail of Hc " given by a local pattern (α^I then increases from 0 to 16 in the bounding box posterior half). $\gamma^{AZ} = 1$, except for voxels in low and high likelihood zones ($\gamma^{AZ} = 0.5$ in O -unlikely and 2 in O -likely zones). $\gamma^{PZ} = 1$, except for voxels in the three probability zones ($\gamma^{PZ}(v) = 0.75$ in PZ^0 , $\gamma^{PZ}(v) = 2$ in PZ^1 , $\gamma^{PZ}(v) = 1.5$ in $PZ^{0.75}$). These parameters constrain the deformation by decreasing the regularization energy in O -likely zones and vice-versa. They are chosen empirically so as to be consistent with the α^I and γ^{AZ} values given in [7].

3. EVALUATION OF SEGMENTATION ACCURACY

3.1. Evaluation Data

The segmentation was validated with a leave-one-out strategy on the data from 16 young controls used to form the atlases, and on data from 8 patients with AD, using the 16-young-healthy-subjects atlas. All data were acquired on a GE 1.5T scanner with an IR-FSPGR sequence [7]. The average hippocampal volume derived from manual segmentation was 1.9cm^3 (1.4-2.6) in AD patients, compared to 2.9cm^3 (2.1-3.6) for the young controls. Performance evaluation was done by comparing automatic and manual (reference) segmentation results. Four quantitative indices are used, as given in table 1: error in

volume (RV), Dice (DO) and Jaccard (JO) overlaps and Hausdorff symmetric distance (DM). We compared the results of the fully automatic approach with our previous semi-automatic method [7] and with an "atlas-based" segmentation given by the 0.5-level probability object.

$RV(O^S, O^R)$	$DO(O^S, O^R)$	$JO(O^S, O^R)$	$DM(O^S, O^R)$
$2 \frac{ V_{O^S} - V_{O^R} }{V_{O^S} + V_{O^R}}$	$2 \frac{V_{O^S \cap O^R}}{V_{O^S} + V_{O^R}}$	$\frac{V_{O^S \cap O^R}}{V_{O^S \cup O^R}}$	$\max_{v \in O^S} \{ \max_{v \in O^R} (d(v, O^R)) \}, \max_{v \in O^R} \{ \max_{v \in O^S} (d(v, O^S)) \}$

Table 1: quantitative indices comparing a segmentation S with a reference R (see text for details)

3.2. Validation in Young Controls

The results from the automatic segmentation were first qualitatively analyzed revealing no major discrepancy. Quantitative results are summarized in table 2, top rows.

	Index	semi-automatic [7]	0.5-level object	automatic
16 young controls				
H_c	RV (%)	7 ± 4 (0 - 14)	10 ± 6 (1 - 24)	6 ± 4 (0 - 13)
	DO (%)	84 ± 3 (78 - 89)	78 ± 5 (64 - 85)	87 ± 3 (82 - 93)
	JO (%)	72 ± 4 (64 - 79)	64 ± 6 (47 - 73)	77 ± 4 (69 - 88)
	DM (mm)	4.5 ± 1.5 (2.5 - 9)	4 ± 1.3 (2.6 - 8.5)	3.6 ± 1 (2.3 - 5.9)
Am	RV (%)	12 ± 7 (1 - 27)	10 ± 8 (0 - 33)	9 ± 6 (0 - 26)
	DO (%)	81 ± 4 (69 - 88)	83 ± 4 (70 - 89)	84 ± 4 (75 - 91.1)
	JO (%)	69 ± 6 (53 - 78)	70 ± 6 (54 - 80)	73 ± 6 (60 - 84)
	DM (mm)	3.9 ± 0.9 (2.8 - 6)	2.8 ± 0.5 (1.9 - 4)	3.2 ± 0.6 (1.9 - 5)
8 patients with AD				
H_c	RV (%)	9 ± 7 (0 - 21)	27 ± 16 (0 - 56)	7 ± 4 (1 - 13)
	DO (%)	84 ± 3 (78 - 88)	68 ± 8 (55 - 80)	86 ± 3 (81 - 91)
	JO (%)	72 ± 4 (64 - 79)	52 ± 9 (38 - 67)	75 ± 4 (68 - 83)
	DM (mm)	6.5 ± 2.4 (4.1 - 14)	4.9 ± 1.8 (3.2 - 9.7)	4.8 ± 1.8 (2.8 - 9.7)
Am	RV (%)	15 ± 13 (1 - 42)	22 ± 16 (2.3 - 43)	16 ± 13 (2 - 52)
	DO (%)	76 ± 7 (60 - 87)	71 ± 9 (57 - 83)	82 ± 6 (72 - 90)
	JO (%)	62 ± 9 (43 - 77)	55 ± 10 (40 - 72)	69 ± 8 (56 - 82)
	DM (mm)	4.5 ± 0.9 (3.1 - 5.7)	4.1 ± 0.9 (2.7 - 5.5)	3.2 ± 0.8 (1.9 - 4.9)

Table 2: value of the quantitative indices (mean \pm standard deviation (min-max)) for controls and patients with AD

All indices are consistent with improved segmentation using the automatic procedure.

3.3. Validation in Patients with Alzheimer's Disease

The automatic segmentation results are 3D-consistent and qualitatively correct, as shown in figure 2.

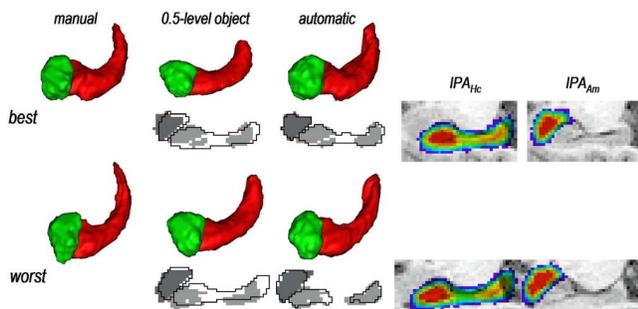


Figure 2: 3D-renderings (automatic and manual segmentation), and sagittal slices showing automatic segmentation outline overlaid on manual segmentation and IPA_{H_c} and IPA_{Am} for the best and worst results in AD patients.

All indices are consistent with improved segmentation using the automatic procedure (table 2, bottom rows).

4. APPLICATION: CLASSIFICATION OF SUBJECTS

To assess whether automatic segmentation can provide a biomarker for AD, we tested the ability of H_c volumes to discriminate between AD patients and elderly healthy subjects, and, in addition, with amnesic MCI.

4.1. Subjects

We studied 25 patients with AD (14 females, age \pm standard-deviation (SD) = 73 ± 6 years, range = 62–81 years, mini-mental score (MMS) = 24.4 ± 2.7 , range = 19–29) and 24 patients with amnesic MCI (14 females, age \pm SD = 74 ± 8 , range = 55–87, MMS = 27.2 ± 1.4 , range = 24–29) which were recruited at the Centre Hospitalo-Universitaire (CHU) of Caen. MRI data from AD and MCI patients were compared to data from 25 elderly healthy controls (12 females, age = 64 ± 8 , range = 51–84). A 3D T1-weighted MRI (voxel size $0.93 \times 0.93 \times 1.5 \text{mm}^3$) was acquired on a GE 1.5T scanner using a SPGR sequence.

4.2. Segmentation and classification

Using the fully automatic method, we segmented the hippocampus and the amygdala in all subjects using the default parameters of the algorithm. Volumes were normalized by the total intracranial volume derived from SPM5 segmentation maps; left and right volumes were averaged. Group differences were assessed using Student's t-test (a robust estimate of the p-value was computed using a bootstrap method [13]). For the classifications of AD vs controls, MCI vs controls and AD vs MCI, each participant was assigned to the closest group as follows. Robust estimates of classification rate, sensitivity and specificity were computed with a bootstrap approach for training set selection. In this procedure, we drew without replacement approximately 75% of each group to obtain a training set. On this training set, we estimated the mean normalized H_c volume for each group. Each participant in the remaining 25% was then assigned to the group which mean was closest to the volume of this participant. The procedure was repeated 5000 times.

4.3. Results

The results of group analysis and individual classification are presented in table 3 for H_c .

As for Am , we also found significant group differences between all groups of subjects (AD vs controls: 1.08 vs 1.27, $p < 0.001$; MCI vs controls: 1.18 vs 1.27, $p < 0.05$; AD vs MCI: 1.08 vs 1.18, $p < 0.05$). However, using Am volume together with H_c volume in a linear discriminant analysis procedure did not improve the classification.

	AD vs controls	MCI vs controls	AD vs MCI
Mean Hc volume (cm ³)	1.83 vs 2.59	2.14 vs 2.59	1.83 vs 2.14
Mean Vol. reduction	-29%	-17%	-14%
Statistical significance	p<0.001	p<0.001	p<0.01
Class. Rate	87%	74%	73%
Sensitivity	87%	79%	67%
Specificity	88%	69%	78%

Table 3: Upper three rows: group comparisons of Hc volumes. Lower three rows: classification rate, sensitivity and specificity for classification between AD patients, MCI patients and elderly controls.

5. DISCUSSION

The fully automatic hippocampus and amygdala segmentation method presented here has proven to be accurate for both young healthy subjects and patients with Alzheimer's disease. The process is fast (15 minutes, 10 for the registration and 5 for bilateral segmentation) and is implemented in a user-friendly environment (Brainvisa: <http://brainvisa.info>).

No algorithm or atlas modification or parameter tuning was necessary for application to AD patients. This was made possible by the hybrid anatomical and probabilistic prior. In fact, the new method is more robust to pathology and acquisition parameters than the semi-automatic method [7], in which default parameters were different for the two cohorts. Furthermore, the partial integration of probabilistic maps as a constraint in the deformation process makes it more robust to pathology than methods that rely strongly on a single atlas. In fact, it was previously demonstrated that segmentation based on the registration of a single subject atlas does not perform satisfactorily when the atlas does not belong to the same disease category as the subject [14].

Validation studies on the segmentation of Hc in AD patients are limited and difficult to compare because of different patient samples and evaluation strategies [15,16]. A comparative study evaluated seven registration methods for segmentation from the MNI, Harvard or disease-specific atlases on 20 AD, 19 MCI patients and 15 healthy subjects [14]. The fully deformable method with a disease-specific atlas gave the best result, with JO~60% for all subjects (better in healthy and MCI subjects than in AD patients) compared to 75% in our study for patients with AD.

Using fully automatic volumetry of the hippocampus, we were able to discriminate AD patients from controls with 87% accuracy. This is in line with previously published results based on manual segmentation which report accuracy between 82% and 90% for AD, e.g. [17,18]. As for automatic methods, very few studies investigated the classification of individual patients. Fischl et al. [9] detected significant group differences in hippocampal volume but did not investigate classification of individual participants. Using both volume and shape features, Csernansky et al. [6] reported a sensitivity of 83% and a specificity of 78%. The accuracy that we report for MCI (74%) is also comparable to that obtained using manual segmentation (between 60%-

74%, e.g. [18,19]). However, the MCI group is heterogeneous since not all MCI patients have incipient AD [18]. The misclassified cases may thus not develop AD. In future work, we will assess the ability of our method to predict the conversion to AD in MCI patients.

11. REFERENCES

- [1] H. Braak et al, "Staging of Alzheimer's disease-related neurofibrillary changes", *Neurobiol Aging* 16, pp 271-278, 1995
- [2] B. Dubois et al, "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria", *Lancet Neurol* 6, pp 734-746, 2007
- [3] A. Kelemen et al, "Elastic model-based segmentation of 3-D neuroradiological data sets", *IEEE TMI* 18, pp 828-839, 1999
- [4] D. Shen et al, "Measuring size and shape of the hippocampus in MR images using a deformable shape model", *Neuroimage* 15, pp 422-434, 2002
- [5] S. Duchesne et al, "Appearance-based segmentation of medial temporal lobe structures", *Neuroimage* 17, pp 515-531, 2002
- [6] J.G. Csernansky et al, "Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus", *Neurology* 55, pp 1636-1643, 2000.
- [7] M. Chupin et al, "Automated segmentation of the hippocampus and the amygdala driven by competition and anatomical priors: method and validation on healthy subjects and patients with Alzheimer's disease", *Neuroimage* 34, pp 996-1019, 2007
- [8] I. Bloch et al, "Fusion of spatial relationships for guiding recognition, example of brain structure in 3D MRI", *Pattern Recog Letters* 26, pp 449-457, 2005
- [9] B. Fischl et al, "Whole brain segmentation: automated labelling of neuroanatomical structures in the human brain", *Neuron* 33, pp 341-355, 2002
- [10] R.A. Heckemann et al, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion", *Neuroimage* 33, pp 115-126, 2006
- [11] M. Chupin et al, "Fully automatic segmentation of the hippocampus and the amygdala from MRI using hybrid prior knowledge", *Proc of MICCAI 2007 part I*, pp 875-882, 2007
- [12] J. Ashburner et al, "Unified segmentation", *Neuroimage* 26, pp 839-851, 2005
- [13] B. Efron & R.J. Tibshirani, *An introduction to the bootstrap*, New York: Chapman & Hall, 1993.
- [14] O.T. Carmichael et al, "Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment" *Neuroimage* 27, pp 979-990, 2005
- [15] Y.Y. Hsu et al, "Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia", *JMRI* 16, pp 305-310, 2002
- [16] W.R. Crum et al, "Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease", *Neuroimage* 13, pp 847-855, 2001
- [17] G.B. Frisoni et al, "Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease", *Neurology* 52, pp 91-100, 2000
- [18] Y. Xu et al, "Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD", *Neurology* 54, pp 1760-1767, 2000
- [19] C. Pennanen et al, "Hippocampus and entorhinal cortex in mild cognitive impairment and early AD", *Neurobiol Aging* 25, pp 303-310, 2004