ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 34 (2007) 996–1019

Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease

Marie Chupin,[a,*] A. Romain Mukuna-Bantumbakulu,[d] Dominique Hasboun,[b,c] Eric Bardinet,[b] Sylvain Baillet,[b] Serge Kinkingnéhun,[d] Louis Lemieux,[a] Bruno Dubois,[d] and Line Garnero[b]

[a]*Department of Clinical and Experimental Epilepsy, Institute of Neurology, UCL, UK*
[b]*Cognitive Neuroscience and Brain Imaging Laboratory, CNRS UPR640, Paris, France*
[c]*Neuroradiology Unit, Hôpital de la Salpêtrière, Paris, France*
[d]*INSERM U610 unit, Paris, France*

Received 30 June 2006; revised 25 October 2006; accepted 26 October 2006
Available online 18 December 2006

We describe a new algorithm for the automated segmentation of the hippocampus (Hc) and the amygdala (Am) in clinical Magnetic Resonance Imaging (MRI) scans. Based on homotopically deforming regions, our iterative approach allows the simultaneous extraction of both structures, by means of dual competitive growth. One of the most original features of our approach is the deformation constraint based on prior knowledge of anatomical features that are automatically retrieved from the MRI data. The only manual intervention consists of the definition of a bounding box and positioning of two seeds; total execution time for the two structures is between 5 and 7 min including initialisation. The method is evaluated on 16 young healthy subjects and 8 patients with Alzheimer's disease (AD) for whom the atrophy ranged from limited to severe. Three aspects of the performances are characterised for validating the method: accuracy (automated vs. manual segmentations), reproducibility of the automated segmentation and reproducibility of the manual segmentation. For 16 young healthy subjects, accuracy is characterised by mean relative volume error/overlap/maximal boundary distance of 7%/84%/4.5 mm for Hc and 12%/81%/3.9 mm for Am; for 8 Alzheimer's disease patients, it is 9%/84%/6.5 mm for Hc and 15%/76%/4.5 mm for Am. We conclude that the performance of this new approach in data from healthy and diseased subjects in terms of segmentation quality, reproducibility and time efficiency compares favourably with that of previously published manual and automated segmentation methods. The proposed approach provides a new framework for further developments in quantitative analyses of the pathological hippocampus and amygdala in MRI scans.

Introduction

The hippocampus (Hc) and the amygdala (Am) are medial temporal lobe grey matter structures of primary importance in fundamental cognitive processes and are involved in neurological and psychiatric diseases, including medial temporal lobe epilepsy (MTLE), Alzheimer's disease (AD) and schizophrenia. Computational studies of their anatomy are therefore crucial. These structures form a small anatomical and functional complex with anatomically ill-defined internal and external boundaries, and are subject to the partial volume effect, making their delineation in Magnetic Resonance Imaging (MRI) scans challenging.

Until recently, MRI volumetric studies of these structures have been based on careful manual segmentation (Hasboun et al., 1996; Pruessner et al., 2000) which is very time consuming (usually >1 h per structure) and suffers from considerable intra- and inter-rater variability (Bonilha et al., 2004; Free et al., 1995; Haller et al., 1997; Hogan et al., 2000; Pantel et al., 2000; Pruessner et al., 2000; Wieshmann et al., 1998). These factors limit sensitivity to time-changes, disease-specific differences and inter-site comparisons. Reducing subjectivity through automation in the segmentation process is therefore highly desirable. Several strategies have been proposed to palliate the fuzziness of some boundaries.

Purely image-based methods generally rely on a manual initialisation step. In Ashton et al. (1997), a region deformed elastically from a line of seeds according to a constraint, where both seeds and constraint were derived from manually segmented contours in three orthogonal slices. An artificial neural networks method was described by Pérez de Alejo et al. (2003), that relied on unsupervised tissue-segmentation, followed by supervised volume identification trained on a manually segmented sagittal slice. A patch-based surface model was introduced by Ghanei et al. (1998, 2001), requiring manually defined starting contours.

* Corresponding author. MRI Unit, National Society for Epilepsy, Chesham Lane, Chalfont St Peter, SL9 0RJ BUCKS, UK. Fax: +44 1494 875 666.
E-mail address: m.chupin@ion.ucl.ac.uk (M. Chupin).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.10.035

Automation was implemented for the initial step for two of these methods, with an atlas-based trained maximum likelihood classifier (Ashton et al., 2003) and landmarks derived from rules related to relative position with respect to global brain landmarks (Siadat et al., 2004).

Statistical information derived from a learning base, as in atlas-based methods or deformable templates, has led to some fully automatic methods, with built-in constraints. Shape or appearance models have been used in several methods. Statistically learned deformation modes were combined with elastic deformations of Hc, constrained by intensity profiles in Kelemen et al. (1999). A level-set method, constrained by statistical priors on shape and neighbour-relations between anatomical structures (Yang et al., 2004) or intensity (Yang and Duncan, 2004), was proposed to simultaneously segment Hc and Am. In Duchesne et al. (2002), an atlas matching method was guided by an active appearance model to segment Hc and Am, which was refined in Klemenčič et al. (2004), with another registration algorithm to build the model for Hc only. Atlas-based, fully automated methods for the segmentation of several structures (amongst which Hc and Am), have used spatial information either implicitly, as in Fischl et al. (2002) from an anatomical probabilistic atlas, or explicitly, as in Zhou and Rajapakse (2005) modelling intensity, spatial location and spatial relationships in a fuzzy template.

Some methods have combined manually and statistically derived constraints. Shen et al. (2002) deformed a shape model with geometrical and statistical priors, introducing at least 50 manually identified landmarks for segmenting Hc. A computationally expensive high-dimensional brain mapping technique (Haller et al., 1997; Hogan et al., 2000) used a fluid registration to segment Hc and required 28 landmarks to be identified manually. We note that this algorithm is unique among those with user interaction to have been subjected to a thorough evaluation of reproducibility. A semi-automated post-processing step was required in the hybrid method proposed in Pitiot et al. (2004); initialised by the registration of an hybrid MRI/structure atlas on the target image, a group of templates was then deformed according to image, shape, distance and texture constraints in a rule-controlled framework.

In this work, we present a new Markovian dual deformation method for the simultaneous segmentation of Hc and Am. The new method can be seen as an extension of the region-growing algorithm proposed in Mangin et al. (1995) and Poupon et al. (1998) for the segmentation of the cortex and the major basal ganglia. Furthermore, the method relies on a dual segmentation embedding a competition framework at the interface between Hc and Am, to ensure a correct behaviour in the unclear area between Hc and Am.

In order to be more applicable to data from both young controls and AD patients, it does not include statistical shape priors, which could bias the segmentation in atrophic cases. In fact, although there have been demonstrations of the usefulness of statistical prior knowledge in segmentation algorithms applied to data from normal subjects, such methods could be sub-optimal for the segmentation of diseased structures. Using a disease-specific atlas can lead to segmentation performance improvements compared to general atlases, as shown on data from AD or Mild Cognitive Impairment (MCI) patients (Carmichael et al., 2005); this result highlights the influence on the final segmentation of the type of subject used in the atlas.

In a preliminary study on a restricted sample of controls and patients (Chupin et al., 2006), we have observed that segmentation is rendered particularly difficult in cases with AD, due to degeneration of Hc which results in a large atrophy in patients even in early stages of the disease, and to loss of image contrast. We propose introducing a constraint in the segmentation to overcome this problem while retaining execution speed with minimal user input (to increase reproducibility) and limiting bias in relation to atrophy. In the new method, prior anatomical knowledge is explicitly considered, based on local systematic properties around automatically retrieved landmarks at the border of the structures. As opposed to statistical shape priors, anatomical knowledge gives information which is less variable and less sensitive to atrophy than shape and size (Bloch et al., 2005). These priors can use explicit knowledge, which is based on an anatomical description of the structures (topology, position, distances) and their relationships; this information is formalised by simple rules on position, geometry and intensity (Barra and Boire, 2001). In order to be introduced as a constraint in a segmentation algorithm, anatomical knowledge can be modelled in fuzzy maps (Barra and Boire, 2001; Bloch et al., 2005) or in the energy/force guiding the deformation process (Yang et al., 2004; Pitiot et al., 2004); in our method, the constraint is introduced in the energy.

Finally, we wanted to address a difficulty with regard to the evaluation of the algorithm performance in the current literature. An extended evaluation of the method is thus proposed, based on a range of commonly used measures in order to facilitate comparisons with published manual segmentation and automated methods.

Following the method's detailed description, we will present evidence of its validity and potential usefulness by comparing its outcome with manually segmented structures in MRI data from 16 representative healthy controls and 8 AD patients using both quantitative measures of agreement, with volume, region overlap and boundary distance measures, and thorough qualitative evaluation.

Algorithm

Principle

The algorithm is an iterative competitive deformation process between two objects, O_{Hc} (for Hc) and O_{Am} (for Am), in a background BG_{HcAm}. Regions deform following local topology-preserving transformations from an initial offset of seeds; as in Mangin et al. (1995) and Poupon et al. (1998), it derives from a Markovian model. We define a *voxel front* as the set of voxels at the border of a deforming object. The growth proceeds by alternate iterative deformations of O_{Hc} and O_{Am} through the re-classification of voxels in the vicinity of the current voxel front. This is achieved by minimising a global energy functional, built following a Bayesian framework in order to embody data attachment and composite priors. At each step, classification optimisation is carried out using the Iterated Conditional Modes (ICM) algorithm (Besag, 1989), a robust deterministic method for classifiers based on Markov random fields, widely used for image segmentation (Dubes et al., 1990). The segmentation process stops when stability is reached.

A competitive scheme is needed because the anatomical border between Hc and Am – both grey matter structures – is very difficult

to determine on MRI scans. This border consists of the alveus, a thin white matter structure, sometimes combined with cerebrospinal fluid (CSF) of the lateral ventricle. It is only partly visible on routine MRI scans such as those used in this work, and is therefore an unreliable feature to block the segmentation of Hc alone. In our algorithm, the current O_{Hc}–O_{Am} interface is automatically detected at the beginning of each deformation step; competition is introduced in the voxel re-classification process by considering an interface-specific optimisation and dual regularisation (interface–non-interface) for Hc. Voxels detected as part of the interface (I_{HcAm}) can be reclassified as either O_{Hc} or O_{Am}; the fact that I_{HcAm} includes only voxels previously classified either as O_{Hc} or O_{Am}, reduces the risk of including background voxels in one of the objects.

Neuroanatomical landmarks, derived from the alveus, parahippocampal gyrus, temporal horn of the lateral ventricle and isthmus are used to build priors to further constrain the segmentation process. These structures are considered for the basis of the anatomical priors, because they are commonly used by neuroanatomists during manual delineation, both at the interface and at the other borders of the two structures. During the iterative segmentation process, 11 sets of landmarks are automatically retrieved at the border of O_{Hc} and O_{Am}, corresponding to 11 patterns. These follow rules defined locally in accordance with the Markovian formalism. These rules are derived from formal empirical descriptions of patterns in brain anatomy, as seen on MRI scans, and described in Appendix A. These patterns can also be retrieved in atrophied structures. The segmentation is then constrained through regularisation: each landmark's 26-neighbourhood is divided into several likelihood zones, defined according to prior anatomical likelihood of their voxels to belong to Hc and Am. The computation of the energy is modified for these regions, to act on the deformations according to the anatomical likelihood, as detailed in Appendix B.

Outline

The three main operational steps of the segmentation process are summarised here and presented in detail in the following section. The first step is manual, its influence being discussed in Performance evaluation; the following steps are fully automatic.

Initialisation:

- Definition of a bounding box delimiting the region of interest (ROI) including Hc and Am;
- Placement of one seed in each structure (Hc and Am).

Alternate deformations:

- Resetting of landmarks to NULL set.

Homotopic deformation of Hc voxel front.

- Selection of 'candidates' to re-classification in the neighbourhood of the voxels of the Hc front;
- Detection of interface voxels, landmarks and likelihood zones, ICM initialisation;
- Voxel re-classification (ICM energy optimisation): at each iteration, for each voxel candidate, re-classification in the object leading to the smaller local energy:
  - for non-interface voxels, re-classification is restricted to either O_{Hc} or BG_{HcAm};
  - for interface voxels, re-classification is restricted to either O_{Hc} or O_{Am}.

Homotopic deformation of Am voxel front.

- Selection of 'candidates' to re-classification in the neighbourhood of the voxels of the Am front;
- Detection of interface voxels, landmarks and likelihood zones, ICM initialisation;
- Voxel re-classification (ICM energy optimisation):
  - for non-interface voxels, re-classification is restricted to either O_{Am} or BG_{HcAm};
  - for interface voxels, re-classification is restricted to either O_{Am} or O_{Hc}.

Convergence criterion and stopping.

Detailed algorithm description

Initialisation

The algorithm is implemented in the *Brainvisa* environment[1] (Cointepas et al., 2001), the complete 3D visualisation being handled by *The Anatomist* software (Rivière et al., 2000). No pre- or post-processing is used.

A region of interest (ROI) is defined manually using a mouse-driven cursor in a dedicated graphical interface for each subject and for each hemisphere. This contained Hc (about 2500 voxels) and Am (about 1250 voxels); the dimensions of the ROI are normally approximately $30 \times 50 \times 20$ voxels. The purpose of this parallelepiped ROI is to significantly decrease memory load (Duchesne et al., 2002; Hogan et al., 2000). The ROI is defined by six bounding slices according to the following rules, illustrated in Fig. 1.

1) the medial limit corresponds to the sagittal slice immediately medial to the uncus and the lateral limit to the sagittal plane containing the CSF ellipsoid of the temporal horn of the lateral ventricle;
2) the anterior limit corresponds to the coronal slice after which a grey matter mass appears in the uncal white matter, just after the isthmus and the posterior limit to the slice after the grey matter mass under the crus fornicis disappears;
3) the inferior limit corresponds to the axial slice below which grey matter disappears inside the white matter of the parahippocampal gyrus and the superior limit to the slice above which grey matter is no longer visible along the fornix.

Two seeds are positioned in the extracted ROI, according to two rules: the seeds are located near the centre of Am and the head of Hc (HHc); they are both placed approximately at the same distance from the interface. The starting point of the automated segmentation process consists of three objects; two $5 \times 5 \times 5$-voxel cubes centred on the seeds (initial O_{Hc} and O_{Am}), and all other voxels within the ROI classified as BG_{HcAm}.

Alternate deformations

The O_{Hc} and O_{Am} growth follows an alternate iterative procedure; the Hc front is considered first, followed by the Am front. Each voxel front is scanned to identify voxel re-classification candidates and only voxels identified as *simple points* in the 26-neighbour vicinity of the current voxel front are considered. A point is defined as *simple* for a given object if the topological

[1] *Brainvisa* and *The Anatomist* are available at http://brainvisa.info/, under a license *CeCILL licence version* 2.

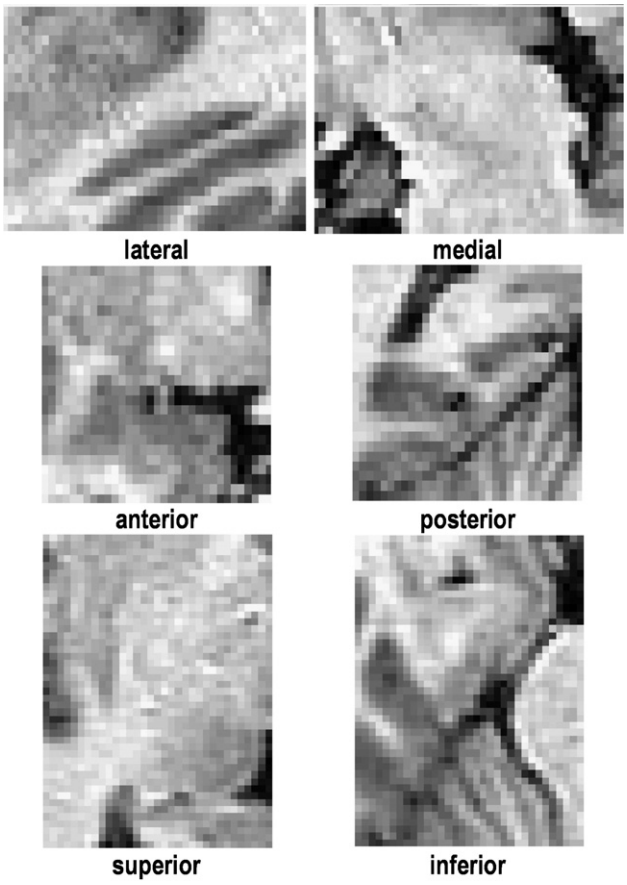

Fig. 1. Bounding box: the outermost slices defining the region of interest are shown. An extra slice is added to reduce its direct influence.

properties (one connected component, no holes) of this object do not change when the point is added to or removed from the object. It is determined by computing the number of connected components on the 26-neighbourhood of the current point for the object (in 26-connectivity) and for the background (in 6-connectivity) according to the rules defined in Malandain et al. (1993). This condition ensures that each resulting deformation step will be homotopic, and thus maintain the topology of the two objects.

The interface I_{HcAm} is considered as a meta-region during the deformation process. A meta-region is defined as a sub-region associated with a specific regularisation. It is automatically detected as the set of voxels from one object that are candidates to be re-classified as the other object.

Anatomical priors. Landmarks (segmentation priors) are detected according to intensity and relational characteristics in the 26-neighbourhood of voxel candidates. Intensity and geometric properties are modelled through invariant patterns respectively in the MRI data and the label volume. Both types of patterns model information on relative positions of structures in the vicinity of Hc and Am. Three meta-regions can then be derived with respect to the detected landmarks: low likelihood zones for Hc ($Z^{\mathrm{LL}}_{\mathrm{Hc}}$) and Am ($Z^{\mathrm{LL}}_{\mathrm{Am}}$), and high likelihood zone for Hc ($Z^{\mathrm{HL}}_{\mathrm{Hc}}$).

The landmark detection process is described in detail in Appendix A. In summary, the process is hierarchical, reflecting the relative empirical importance of each landmark as a constraint for the segmentation. Eleven sets of landmarks are detected following prior knowledge. Two are located at the interface between Hc and Am—based on the alveus and the temporal horn of the lateral ventricle. Three are defined for Hc only—two based on the alveus and one on the Hc sulcus. Five are defined for both Hc and Am – based on the parahippocampal gyrus and the temporal horn of the lateral ventricle – and used accordingly for each structure. Finally, one is defined for Am only—based on the isthmus of the temporal lobe. Each of these sets of landmarks is detected according to a set of intensity and positional rules defined with a 2D 3×3-voxels pattern (see Appendix A), empirically derived from anatomical knowledge. This pattern is composed of specific neighbours of a voxel candidate v, and rules modelling relational radiometric and geometric knowledge are derived around it. The landmarks and zones of high and low likelihood are illustrated in Fig. 2. Their influence on the results of the segmentation is studied in Performance evaluation.

Voxel re-classification. An iteration of front deformation is achieved through re-classification of the voxel candidates as belonging to either O_{Hc}, O_{Am} or $\mathrm{BG}_{\mathrm{HcAm}}$. It proceeds through the minimisation of E_{ROI}, the energy of the voxels in the ROI. E_{ROI} is defined as the sum of the local energies at every voxel, $\mathrm{mR}=\{I_{\mathrm{HcAm}}; Z^{\mathrm{LL}}_{\mathrm{Hc}}; Z^{\mathrm{LL}}_{\mathrm{Am}}; Z^{\mathrm{HL}}_{\mathrm{Hc}}\}$ being the set of meta-regions introduced in the process:

$$E_{\mathrm{ROI}} = \left[\sum_{v\in O_{\mathrm{Hc}}-\mathrm{mR}} E_{O_{\mathrm{Hc}}}(v) + \sum_{v\in O_{\mathrm{Am}}-\mathrm{mR}} E_{O_{\mathrm{Am}}}(v) + \sum_{v\in BG_{\mathrm{HcAm}}} E_{BG_{\mathrm{HcAm}}}(v)\right] + \sum_{R\in\mathrm{mR}}\left[\sum_{v\in O_{\mathrm{Hc}}\cap R} E_{O_{\mathrm{Hc}}\cap R}(v) + \sum_{v\in O_{\mathrm{Am}}\cap R} E_{O_{\mathrm{Am}}\cap R}(v)\right]. \quad (1)$$

Energy functional. The local energy $E(v)$ consists of five terms: global data attachment (E^{G}), modelling average intensity characteristics, local data attachment (E^{L}), modelling edges and context terms dedicated to Markovian regularisation (E^{I}), volume (E^{V}) and surface (E^{S}) control.

$$E_O(v) = E^{\mathrm{G}}_O(v) + E^{\mathrm{L}}_O(v) + E^{\mathrm{I}}_O(v) + E^{\mathrm{V}}_O(v) + E^{\mathrm{S}}_O(v). \quad (2)$$

Detailed descriptions of the energy terms, which are different for objects O (O_{Hc} and O_{Am}) and $\mathrm{BG}_{\mathrm{HcAm}}$, are given in Appendix B.

Optimisation. Following initialisation – re-classification of all voxel candidates according to global data attachment only – the full energy functional is minimised on the set of voxel candidates with an ICM algorithm.

At each iteration, E_{ROI} is computed for the two possible classifications for each voxel candidate. Voxels are considered sequentially, which makes re-classification equivalent to comparing local energies for each voxel and classifying the voxel in the object minimising its local energy. For example, considering an iteration of the deformation of O_{Hc}, for a voxel candidate $v_{\mathrm{C}}\notin I_{\mathrm{HcAm}}$, minimising E_{ROI} follows the scheme:

$$\text{if } \Delta E(v_{\mathrm{C}})\le 0,\ v_{\mathrm{C}}\in O_{\mathrm{Hc}};\ \text{else } v_{\mathrm{C}}\in BG_{\mathrm{HcAm}}$$
$$\text{with } \Delta E(v_{\mathrm{C}}) = E^{\text{if } v_{\mathrm{C}}\in O_{\mathrm{Hc}}}_{\mathrm{ROI}} - E^{\text{if } v_{\mathrm{C}}\in BG_{\mathrm{HcAm}}}_{\mathrm{ROI}} = E_{O_{\mathrm{Hc}}}(v_{\mathrm{C}}) - E_{BG_{\mathrm{HcAm}}}(v_{\mathrm{C}}). \quad (3)$$

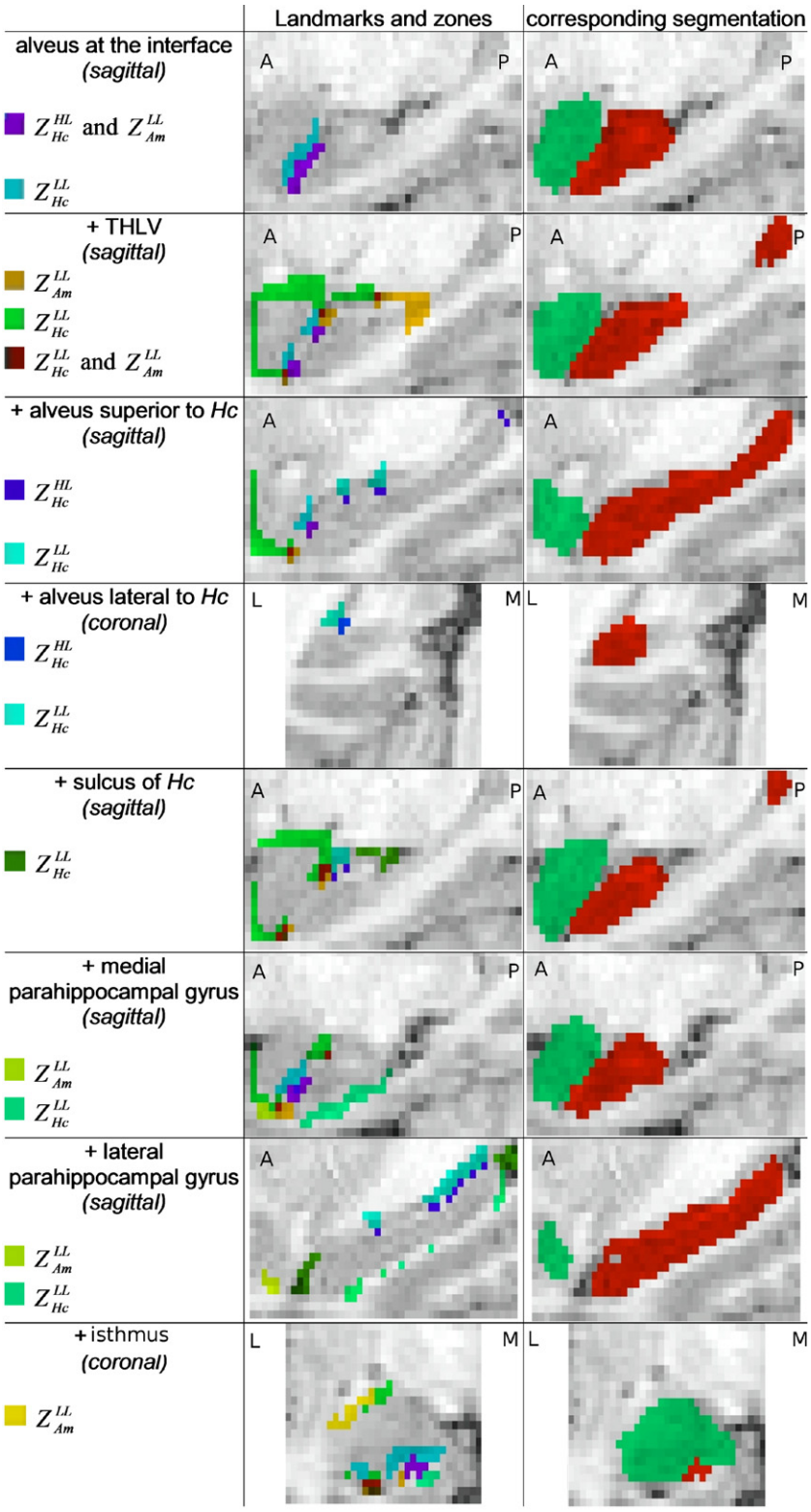

Fig. 2. Landmarks and associated likelihood zones. In each row, previously defined landmarks are kept and new ones are added. Left column shows corresponding segmentation in the same slice. THVL = Temporal Horn of the Lateral Ventricle. Landmarks were illustrated in the orientation in which they are defined, except for those derived from the parahippocampal gyrus, for which sagittal slices are more illustrative. In each case, the selected slice was chosen according to the number of voxels representing the illustrated landmark.

The ICM process stops when the deforming object is no longer modified, or when the number of changes is small enough during three consecutive ICM-iterations.

Convergence and stopping criterion

The overall deformation process ends when the number of changes in both objects is smaller than 2% of their surface for three iterations in a row.

Performance evaluation

We characterised performances by applying the algorithm in data from healthy subjects and AD patients, scanned in conditions representative of routine clinical investigations. Segmentation results were evaluated in three ways: (1) accuracy: automated results were compared visually and quantitatively to manually segmented objects; (2) reproducibility: intra-observer and inter-observer reliability was considered in relation to inter-observer reliability of manually drawn results; (3) efficacy of the anatomical priors.

Material

MRI data

Sixteen young healthy volunteers (S1–S16, age <35) and eight patients with probable AD (P1–P8, mean age 74 (range: 66–81), mean MMSE (Mini Mental State Examination) 21 (6–28)) were scanned with a 1.5 Tesla Signa scanner (GE Medical Systems, Milwaukee, WI, USA). The acquisition parameters are summarised in Table 1. Image quality was characterised by the contrast to noise ratio (CNR) and the contrast to noise and artefact ratio (CNAR), which were defined as follows: contrast being the difference between white and grey matter modes of the brain signal intensity histogram; CNR = [contrast] / [intensity standard deviation on a region of scan background without visible artefacts]; CNAR [contrast] / [intensity standard deviation on the whole background, including noise and visible artefacts (motion, wrap-around, pulsation)]. For S1–S16, CNAR = 15 ± 2 (13–19) and CNR = 20 ± 3 (16–25). For P1–P8, CNAR = 11 ± 3 (8–15) and CNR = 15 ± 4 (10–23). The slightly larger range of values for patients could reflect the different sets of acquisition parameters (Li and Mirowitz, 2004).

Gold standard: manual segmentation

The manual segmentation protocol was established by a highly trained neuroanatomist (Hasboun et al., 1996), and adapted to a voxel-based display environment. Images were manipulated with the ROI-drawing module of *The Anatomist* software (Rivière et al., 2000). The boundaries of the two structures are described in detail in Appendix C and were principally determined on coronal slices with checks on simultaneous sagittal and axial views through a linked cursor.

The degree of atrophy in P1–P8 was obtained from manual segmentations by expressing the difference between the volume of each structure and the mean volume of this structure for the young healthy controls (YHC) ($V_{Hc}^{YHC}=2.89$ cm^3, $V_{Am}^{YHC}=1.40$ cm^3), in units of standard deviation ($\sigma_{Hc}^{YHC}=0.34$ cm^3, $\sigma_{Am}^{YHC}=0.16$ cm^3). For Hc, the atrophy was $[2.9\pm1.2\ (0.9–4.4)]\cdot\sigma_{Hc}^{YHC}$ and, for Am, the atrophy was $[2.2\pm1.8\ (-1.7–4.4)]\cdot\sigma_{Am}^{YHC}$, which corresponded to limited to severe atrophy.

Based on automated segmentation results, the volume statistics for the young healthy controls were $V_{Hc}^{YHC}=2.98$ cm^3 and $V_{Am}^{YHC}=1.51$ cm^3, and the standard deviation $\sigma_{Hc}^{YHC}=0.36$ cm^3 and $\sigma_{Am}^{YHC}=0.21$ cm^3. For Hc, the atrophy was $[3.1\pm1.5\ (0.4–4.8)]\cdot\sigma_{Hc}^{YHC}$ and, for Am, the atrophy was $[2.5\pm1.1\ (0.2–4.5)]\cdot\sigma_{Am}^{YHC}$.

Design

The simultaneous segmentation of Hc and Am took less than 1 min on a 1 GHz workstation with 512 MB RAM, developed in C++, in *The Anatomist* environment. The total segmentation process with ROI determination and seed positioning for both structures in both hemispheres (4 structures) took between 10 and 15 min.

Four automated segmentations were performed: A_1^{MC} and A_2^{MC}, initialised by operator MC on two different occasions, A_1^{RM}, initialised by operator RM, and A_{noAP}^{MC} with the same ROI and seeds as A_1^{MC} but without the anatomical priors. Two manual segmentations were performed: M_1^{MC}, by operator MC and M_1^{DH}, by operator DH (S1–S6 only) or M_1^{RM}, by operator RM (P1–P8 only). The results were qualitatively evaluated through 3D renderings and visualisation of sagittal slices and quantitatively evaluated using 9 indices defined in Appendix D: relative volume error (RV), spatial overlap (K_1, K_2), False Positive rate (FP), False Negative rate (FN), Misclassified Interface Voxels (MIV), boundary distance (Dm, DM, D95).

For illustration purposes, we identified the best and worst results according to the composite index CGQ which summarises the various aspects of the quantitative indices (accuracy and reproducibility), computed for comparisons A_1^{MC} vs. M_1^{MC} and A_2^{MC} vs. A_1^{MC} (see Appendix D, Eq. (40)).

Local performance was evaluated by computing the relevant accuracy indices, RV, K_1, K_2, FP, FN, DM and D95, for three sub-regions of Hc and two of Am, defined based on the partition of the antero-posterior axis – from the first coronal slice which

Table 1
Acquisition parameters for the images used in the evaluation process

	TR	TE	TI	Flip angle	Slice thickness	Voxel size	Orientation	Matrix
S1–S8; S10–S16	14.3 ms[a]	6.3 ms[a]	600 ms	10°	1.3 mm[b]	0.9375 mm	Axial	256×192[c]
P1–P4; S9	10.3 ms	2.1 ms	600 ms	10°	1.5 mm	0.9375 mm	Axial[d]	256×192
P5–P6	12.2 ms	5.3 ms	450 ms	15°	1.5 mm	0.976562 mm	Coronal	256×192
P7–P8	9.2 ms	1.98 ms	600 ms	10°	1.3 mm	0.9375 mm	Coronal	256×256

[a] TR = 10.5 ms and TE = 2.2 ms for S15.
[b] Slice thickness of 1.5 mm for S7–S8.
[c] 256×256 matrix for S15.
[d] Coronal plane for P3.

contains the manually segmented structure to the last one – into three equal segments for Hc, approximating head, body and tail and two segments (anterior third and the remaining two thirds) for Am.

The performance of landmark detection and landmark-derived constraints was also studied. The detection performance was characterised by the stability during deformations (number of landmarks which were detected at some early iteration of the deformations but no longer in the last iteration), the stability and accuracy of the final set of landmarks (number of landmarks for each type, correct and incorrect detection of low likelihood zones outside the manually segmented objects and high likelihood zones inside the manually segmented objects). The effect of the priors on performance was assessed by comparing the performance of the algorithm for A_1^{MC} and $A_{\mathrm{noAP}}^{\mathrm{MC}}$, and analysing the improvement of CGQ.

Note that all the parameters described in Appendix B were first tuned on S1–S4, and refined on S5–S8 and P1–P4, on which only radiometric parameters could be tuned. No adjustment to the algorithm was made for application to S9–S16 and only the anisotropy parameters were modified for P5–P8.

Evaluation on young healthy controls

Qualitative analysis

Manual and automated segmentations were inspected visually for each subject and both hemispheres to assess their quality and localise any discrepancy. In general, automated segmentation tended to slightly over-estimate volume compared to manual one. Global shapes and some surface details were correctly recovered, while the main differences were in the tail and the head of Hc.

3D renderings of the best (S1 right hemisphere, S1R) and worst (S9 left hemisphere, S9L) results are displayed in Fig. 3 for M_1^{MC}, A_1^{MC}, A_2^{MC}, A_1^{RM} and $A_{\mathrm{noAP}}^{\mathrm{MC}}$. The shapes derived from manual and automated segmentations matched well for both objects. Furthermore, 3D renderings for S1R indicate that surface details extracted manually were also retrieved with the automated segmentation. Comparisons of the results with (A_2^{MC}) and without ($A_{\mathrm{noAP}}^{\mathrm{MC}}$) anatomical priors demonstrated their usefulness. Comparisons of A_1^{MC}, A_2^{MC} and A_1^{RM} revealed a reproducible behaviour of the segmentation.

To localise false positives, false negatives and overlap, a few sagittal slices are displayed for S1R and S9L (Fig. 4). For S1R,

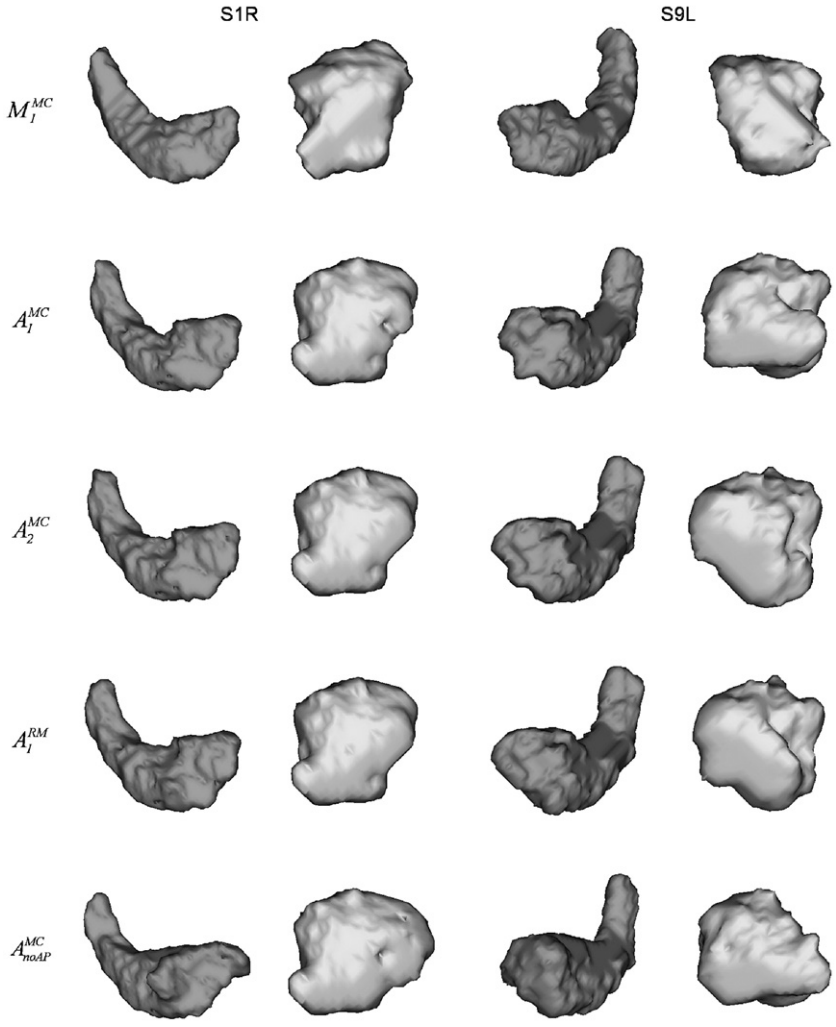

Fig. 3. 3D rendering for manual and automated segmentations for Am and Hc, for S1R and S9L. See text for details.

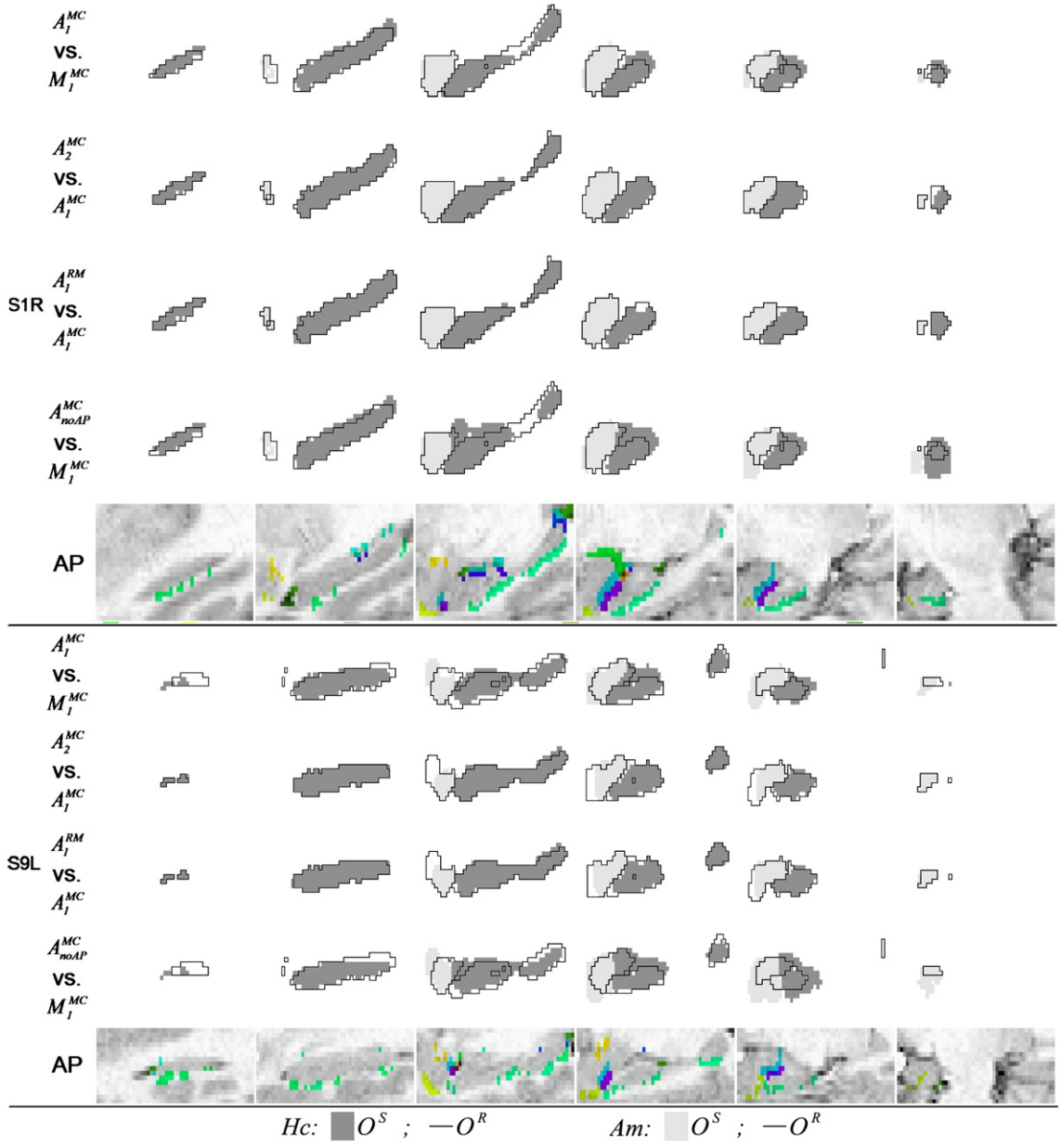

Fig. 4. Overlap between some pairs of the six segmentations for six sagittal slices (in the bounding box), for S1R and S9L for Seg vs. Ref. See text for details. Final likelihood zones for A_1^{MC} are displayed in the "AP" rows with the same colour code as Fig. 2.

automated segmentations were almost identical. For S9L, the major problems were in the Am–Hc interface region. Differences between the automated segmentations with and without anatomical priors were clear for both.

Quantitative evaluation of segmentation quality

The results are summarised in Table 2. Accuracy of the automated method A_1^{MC} vs. M_1^{MC} was close to manual reproducibility M_1^{MC} vs. M_1^{DH} as revealed by the error on volume (RV=7% (0–14) for Hc and 12% (1–27) for Am, compared to RV=7% (1–17) for Hc and 10% (1–24) for Am), the overlap (K_1=84% (78–89) for Hc and 81% (69–88) for Am, compared to K_1=90% (87–92) for Hc and 85% (81–92) for Am) and the maximal distance error (DM=4.5 mm (2.5–9) for Hc and 3.9 mm (2.8–6) for Am, compared to DM=3.7 mm (2.3–6.2) for Hc and 3.1 mm (2.1–3.9) for Am). Automated intra (A_2^{MC} vs. A_1^{MC}) and inter (A_1^{RM} vs. A_1^{MC})-rater reproducibility was better than manual reproducibility (RV=2% (*intra*) and 4% (*inter*) for Hc and 9% (*intra*) and 8% (*inter*) for Am, K_1=97% (*intra*) and 95% (*inter*) for Hc and 93% (*intra*) and 92% (*inter*) for Am and DM=2.5 mm (*intra*) and 2.9 mm (*inter*) for Hc and 2.8 mm (*intra* and *inter*) for Am). The relative segmentation error was greater for Am than for Hc. Note that the error at the interface (MIV) is negligible compared to the global overlap error; the overlap error is roughly distributed around the interface, as indicated by the average and maximal distance errors.

Segmentation quality index values for the subparts of Hc and Am are given in Table 3. RV, K_1 and Dmax values revealed that the segmentation was more accurate in the body and less accurate in

Table 2
Quantitative indices for comparison between pairs of segmentations for S1–S16

Object	Index	A_1^{MC} vs. M_1^{MC}	A_2^{MC} vs. A_1^{MC}	A_1^{RM} vs. A_1^{MC}	A_{noAP}^{MC} vs. M_1^{MC}	A_1^{DH} vs. M_1^{MC}
Hc	RV (%)	7±4 (0–14)	2±2 (0–6)	4±4 (0–15)	22±8 (4–37)	7±5 (1–17)
	K_1 (%)	84±3 (78–89)	97±1 (94–99)	95±2 (91–99)	74±4 (66–85)	90±1 (87–92)
	K_2 (%)	72±4 (64–79)	94±3 (88–99)	91±4 (83–98)	59±5 (49–74)	82±2 (76–84)
	FP (%)	15±3 (10–20)	3±2 (1–7)	5±4 (1–12)	29±4 (15–36)	6±2 (3–9)
	FN (%)	13±5 (4–21)	3±2 (0–8)	3±3 (0–15)	12±4 (6–24)	12±3 (8–19)
	MIV (%)	1.1±1 (0–3.7)	0.2±0.3 (0–1.1)	0.9±1.2 (0–4.2)	5.4±3.2 (0–12)	0.3±0.5 (0.0–1.9)
	Dm (mm)	0.5±0.1 (0.4–0.8)	0.1±0.1 (0–0.2)	0.2±0.1 (0–0.5)	1.1±0.3 (0.4–1.8)	0.3±0.1 (0.2–0.5)
	DM (mm)	4.5±1.5 (2.5–9)	2.5±0.8 (1.3–4.4)	2.9±0.9 (1.6–6)	7.4±2.6 (4–16)	3.7±1.2 (2.3–6.2)
	D95 (mm)	4±1.5 (1.9–8.5)	2.3±0.8 (1.3–4.2)	2.6±0.9 (1.6–5.5)	6.6±2.6 (3.4–15)	3.3±1.1 (1.9–5.4)
Am	RV (%)	12±7 (1–27)	9±7 (0–30)	8±5 (0–19)	16±10 (2–41)	10±7 (1–24)
	K_1 (%)	81±4 (69–88)	93±4 (78–99)	92±4 (80–99)	74±7 (53–88)	85±3 (81–92)
	K_2 (%)	69±6 (53–78)	87±7 (64–98)	85±6 (67–97)	59±9 (36–79)	74±5 (68–85)
	FP (%)	19±6 (6–32)	4±4 (0–15)	7±5 (1–22)	27±7 (14–44)	14±6 (6–26)
	FN (%)	13±5 (5–25)	9±8 (0–29)	8±5 (0–21)	14±6 (6–31)	12±6 (4–21)
	MIV (%)	1.5±1 (0.3–3.8)	0.9±1.6 (0–6.6)	0.5±0.7 (0–2.6)	3.7±3.8 (0.1–17)	2.8±2.4 (0.4–7.1)
	Dm (mm)	0.7±0.2 (0.4–1.2)	0.3±0.2 (0–0.9)	0.3±0.1 (0–0.7)	1±0.4 (0.4–2.2)	0.5±0.1 (0.2–0.7)
	DM (mm)	3.9±0.9 (2.8–6)	2.8±1.2 (0.9–5.9)	2.8±0.7 (0.9–4.3)	5.4±1.5 (2.8–9.9)	3.1±0.5 (2.1–3.9)
	D95 (mm)	3.5±0.8 (2.5–5.2)	2.6±1.2 (0.9–5.6)	2.5±0.7 (0.9–4.3)	5±1.5 (2.5–8.8)	2.9±0.6 (1.9–3.9)

Values = average ± standard deviation (minimum–maximum).

the tail for Hc (largest maximal distance error: 9 mm), whereas it was less accurate in the anterior part of Am than in any other subpart.

Finally, the usefulness of the anatomical priors can be seen by comparing the first column of Table 2 to the fourth, which reveals an important increase of the overlap between automatic and manual segmentations. The performance and influence of landmark detection are detailed in Fig. 5. The improvement brought by the anatomical priors was inversely correlated with the quality of A_{noAP}^{MC}, as shown in Fig. 5a ((Corr[CGQ(A_1^{MC}) − CGQ(A_{noPA}^{MC}),CGQ(A_{noPA}^{MC})] = −0.91). Fig. 5b shows that the landmarks tended to be robustly detected during deformations. Note that the largest changes in detected landmarks across iterations were for the temporal horn of the lateral ventricle superior–anterior to Hc and the parahippocampal gyrus medial to Am; important changes for the parahippocampal gyrus medial and lateral to Hc occurred during three times more iterations than for the other landmarks. The size of the final low and high likelihood zones appeared to be related to the size of the underlying structures they were built from (Fig. 5c); the incorrect detections were generally negligible, except for the parahippocampal gyrus medial to Hc.

Evaluation on Alzheimer's patients

Qualitative analysis

The manual and automated segmentations were again inspected visually for each subject and both hemispheres. In general, as for controls, global shapes and some surface details were correctly recovered, while the main differences were in the tail and the head of Hc. Comparison of automated segmentations with and without anatomical priors confirmed their usefulness in pathological data.

Three-dimensional renderings for the best (P4 right hemisphere, P4R) and worst (P5 right hemisphere, P5R) results for comparisons between A_1^{MC}, A_2^{MC}, A_1^{RM}, A_{noAP}^{MC} and M_1^{MC} are displayed in Fig. 6. Comparisons of A_1^{MC}, A_2^{MC} and A_1^{RM} revealed good reproducibility for Hc. For the worst result, the tail of Hc was incomplete.

To visualise false positives, false negatives and overlap for and between the two structures, a few sagittal slices are displayed for P4R and P5R (Fig. 7). False positives and false negatives tended to be homogeneously distributed, except in the anterior part of Am and the Am–Hc interface. As expected, inter-rater reproducibility was worse than intra-rater. For both subjects, problems could be observed in the Am–Hc region.

Quantitative evaluation

Segmentation results were evaluated with the nine quantitative indices as summarised in Table 4. The accuracy of the automated method was close to manual reproducibility, as revealed by the error on volume (RV = 9% (0–21) for Hc and 15% (1–42) for Am, compared to RV = 8% (2–18) for Hc and 13% (3–34) for Am), the overlap (K_1 = 84% (78–88) for Hc and 76% (60–87) for Am, compared to K_1 = 87% (83–89) for Hc and 80% (72–85) for Am) and the maximal distance error (DM = 6.5 mm (4.1–14) for Hc and 4.5 mm (3.1–5.7) for Am, compared to DM = 3.7 mm (2–6.3) for

Table 3
Quantitative indices on subparts of both structures for the accuracy for S1–S161

Index	Hc head	Hc body	Hc tail	Am anterior	Am posterior
RV (%)	13±8 (1–36)	12±7 (1–26)	13±12 (0–46)	26±15 (6–62)	11±10 (1–38)
K_1 (%)	83±5 (70–91)	85±3 (79–93)	81±6 (65–90)	78±6 (59–87)	85±4 (76–92)
K_2 (%)	71±7 (54–84)	75±5 (65–87)	69±9 (48–82)	64±8 (42–77)	74±6 (61–85)
FP (%)	14±5 (3–23)	18±4 (9–26)	15±7 (4–32)	27±11 (6–51)	10±4 (3–20)
FN (%)	15±9 (2–37)	8±4 (2–18)	16±9 (4–42)	9±6 (2–24)	17±7 (7–34)
DM (mm)	4±1.4 (2.3–7.8)	2.8±0.6 (1.5–4.5)	3.3±1.4 (1.6–9)	3.8±1 (2.6–6)	3.2±0.7 (2.1–4.7)
D95 (mm)	3.7±1.4 (1.6–7.4)	2.5±0.6 (1.5–4.1)	3.1±1.3 (1.6–8.5)	3.5±0.8 (2.1–5.4)	3±0.6 (1.9–4.6)

Values = average ± standard deviation (minimum–maximum).

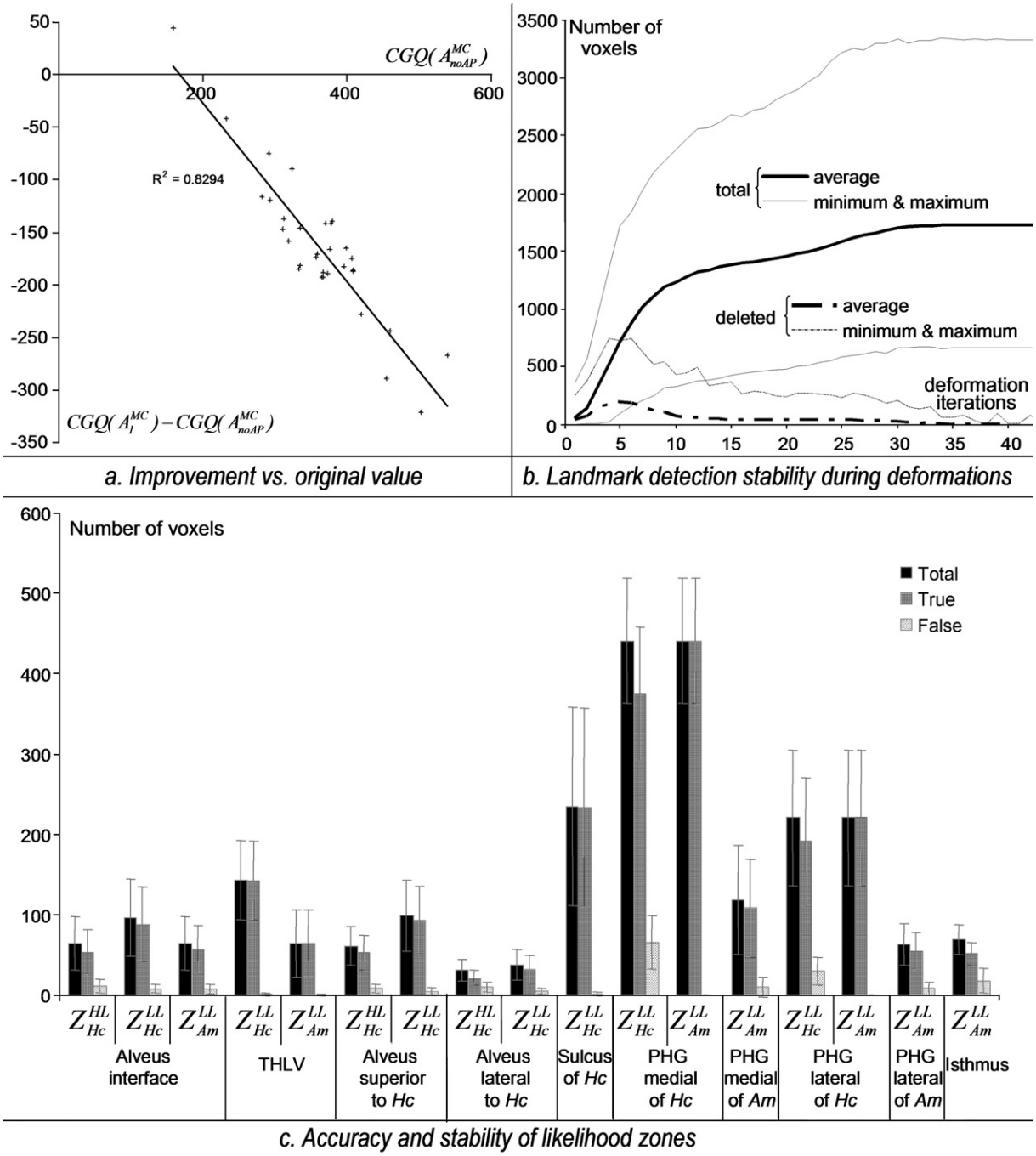

Fig. 5. Characterisation of landmark detection for S1–S16: (a) global improvement vs. global quality of A^{MC}_{noAP}; (b) total number of landmarks and number of landmarks not detected in the final iteration (minimal, average and maximal value at each iteration of the deformation of A^{MC}_{1}); (c) size of the final likelihood zones A^{MC}_{1}, with correct and incorrect classification compared to M^{MC}_{1}.

Hc and 4.3 mm (2.6–6.6) for Am). Automated intra- and inter-rater reproducibility was better than manual reproducibility (RV=6% (*intra*) and 7% (*inter*) for Hc and 12% (*intra*) and 21% (*inter*) for Am, K_1=95% (*intra*) and 94% (*inter*) for Hc and 90% (*intra*) and 82% (*inter*) for Am and DM=2.5 mm (*intra*) and 2.6 mm (*inter*) for Hc and 2.6 mm (*intra*) and 3.8 mm (*inter*) for Am). Mean values and standard deviations for volume indices remained comparable with those obtained for controls.

Segmentation quality index values for the subparts of Hc and Am are indicated in Table 5. RV, K_1 and DM values revealed that the segmentation was more accurate in the body and less in the tail, for Hc, with a largest maximal distance error of 14 mm. For Am, both parts showed similar accuracy.

Again, the values demonstrated the usefulness of the anatomical priors, and the improvement was inversely correlated with the quality of A^{MC}_{noAP} (Corr[CGQ(A^{MC}_{1})−CGQ(A^{MC}_{noAP}),CGQ(A^{MC}_{noAP})]=−0.99), as shown in Fig. 8a. The robustness of landmark detection during the deformations was comparable to that obtained for young healthy controls (Fig. 8b). Note that the largest changes in detected landmarks across iterations were for the temporal horn of the

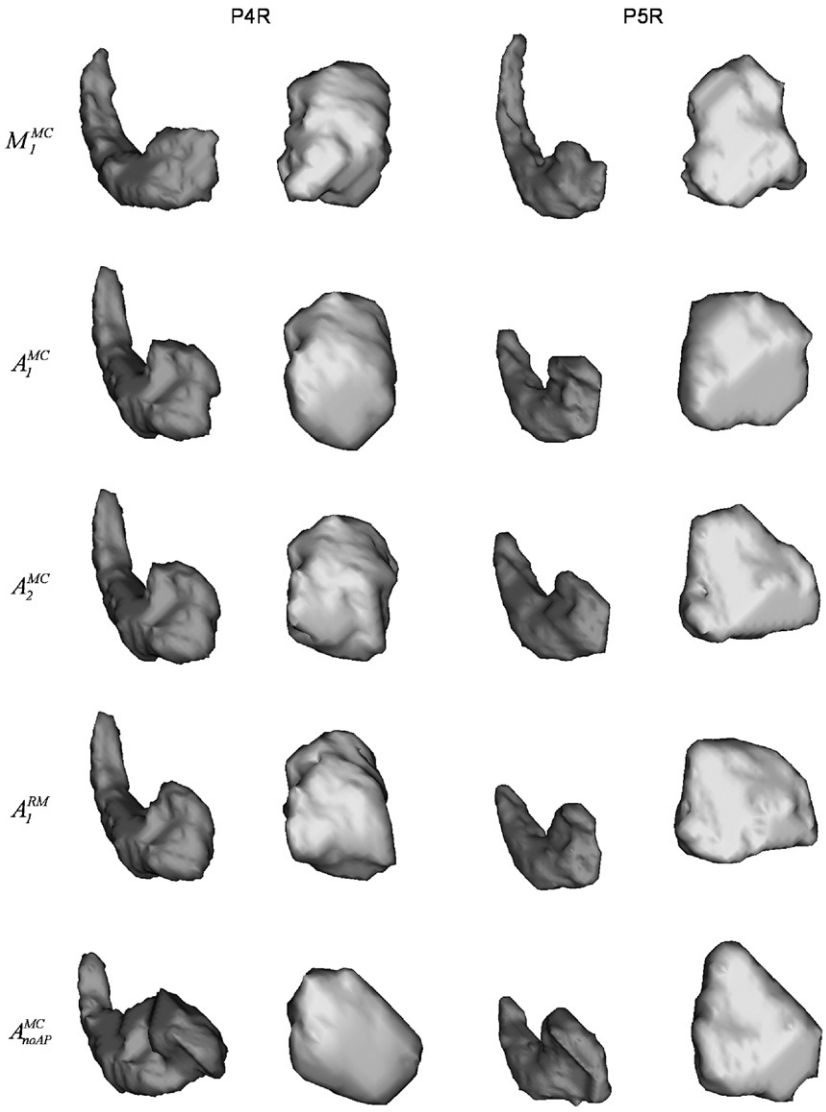

Fig. 6. 3D rendering for manual and automated segmentations for Am and Hc, for P4R and P5R. See text for details.

lateral ventricle superior–anterior to Hc and inferior–posterior to Am; the changes for the other landmarks were very small. Finally, Fig. 8c shows a different repartition of the likelihood zones compared to the young healthy controls, which reflects the effect of atrophy (the alveus at the interface was less recovered, because of large pools of CSF between Hc and Am, and the enlargement of the sulcus of Hc due to atrophy resulted in a larger detection of landmarks for the sulcus of Hc).

Discussion

We have presented a new competitive region growing method for the automated segmentation of the hippocampus and the amygdala. We introduced anatomical priors derived from systematic geometrical knowledge modelled as patterns that could be identified in both healthy subjects and patients. The method was evaluated in healthy subjects and patients with AD, qualitatively and using a set of quantitative indices characterising important aspects of the segmentation: volume, shape and position.

A previous version of the method was described in Chupin et al. (in press, in French), with limited evaluation on data from 8 young healthy controls, and its application on data from 4 patients with AD (Chupin et al., 2006). In this present work, improved anatomical priors and regularisation anisotropy were used and their full description is given for the first time; the method's performance is evaluated in a larger number of healthy controls and patients with extended evaluation design.

In the data from young healthy controls, the segmentation results were visually satisfactory with regard to both global shape and local details for both structures and their interface. The inter-rater reproducibility of the automated method surpassed that of manual segmentation, thereby justifying the proposed approach. Furthermore, the results demonstrated clearly that the anatomical priors improved the segmentation, the improvement being larger for results which were initially worse. Mean values of the quantitative indices for the sixteen young healthy controls showed a good agreement with manual segmentation, with a K_1 value of 84% for Hc and 81% for Am, and RV values of 7% for Hc and 12% for Am. The number of misclassified voxels could be directly

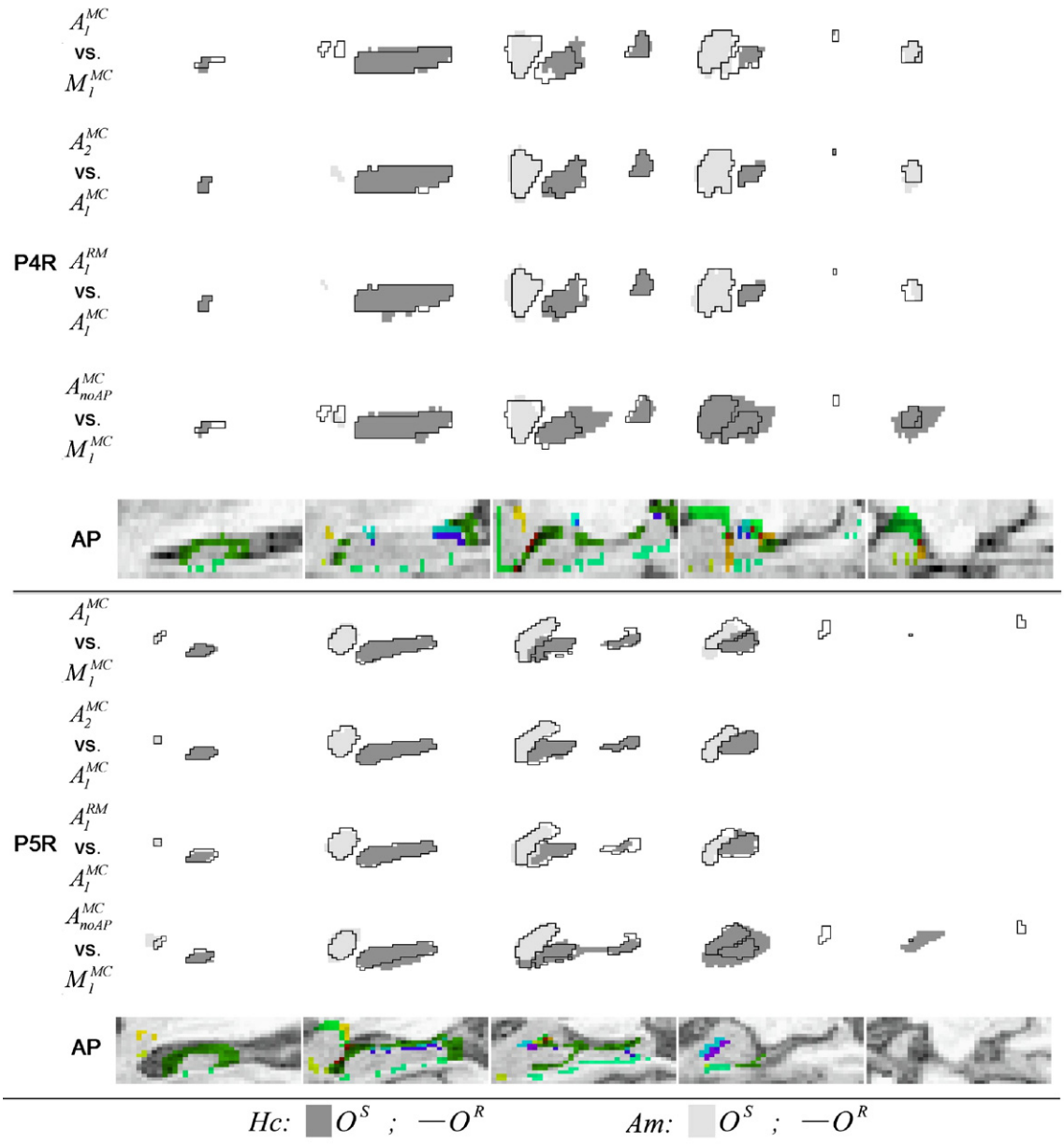

Fig. 7. Overlap between some pairs of the six segmentations for five sagittal slices (in the bounding box), for P4R and P5R for Seg vs. Ref. See text for details. Final likelihood zones for A_1^{MC} are displayed in the "AP" rows with the same colour code as Fig. 2.

characterised through the computation of $1-K_2$. Nevertheless, this index is very sensitive, as shown by the translation study in Appendix D; in summary, we found a mean of 18% of misclassified voxels between manual segmentations, 28% between automated and manual segmentations and 9% between automated segmentations, for Hc. The computation of the average distance (0.3 mm, 0.5 mm and 0.2 mm, respectively) shows that these misclassified voxels tend to be located close to the Hc–Am interface. The segmentation results were less accurate in the anterior part of Am and the tail of Hc. A possible explanation is the fuzziness of the anterior border of Am and noise blurring the transition from body to tail for Hc. This ill-definition of the anterior border of Am could also explain the higher sensitivity of the segmentation to the definition of the anterior border of the ROI, compared to the other borders.

Visually, our results for AD patients generally matched the manual segmentation, with RV = 9% and K_1 = 84% for Hc. Indices values for manual reproducibility were also less good than for controls, reflecting the difficulty of segmenting atrophied structures. Atrophy characterisation showed very similar volumes for manual and automated segmentations for controls. The same trend was also found for the atrophy of Hc in AD patients. For Am, automated volumes showed atrophy in all cases whereas manual volumes showed a hypertrophy in some cases. This emphasised the difficulty of the segmentation of Am on the available MRI scans.

Comparison with other methods

Regarding the manual gold standard, we noted that the protocol for manual segmentation was not always described in published

Table 4
Quantitative indices for comparison between pairs of segmentations for P1–P8

Object	Index	A_1^{MC} vs. M_1^{MC}	A_2^{MC} vs. A_1^{MC}	A_1^{RM} vs. A_1^{MC}	A_{noAP}^{MC} vs. M_1^{MC}	M_1^{RM} vs. M_1^{MC}
Hc	RV (%)	9±7 (0–21)	6±8 (1–31)	7±8 (0–24)	66±31 (14–142)	8±5 (2–18)
	K_1 (%)	84±3 (78–88)	95±4 (82–100)	94±4 (86–100)	55±13 (27–74)	87±2 (83–89)
	K_2 (%)	72±4 (64–79)	92±8 (70–99)	89±8 (75–100)	39±13 (15–58)	76±3 (72–80)
	FP (%)	12±4 (5–17)	5±7 (0–28)	6±7 (0–22)	47±18 (1–68)	15±3 (8–22)
	FN (%)	16±5 (8–24)	4±4 (0–17)	5±5 (0–17)	14±19 (2–84)	8±3 (5–13)
	MIV (%)	0.8±1.6 (0–6.3)	0.3±0.3 (0–1.1)	0.4±0.5 (0–1.4)	18±14 (0–36)	2.5±2.7 (0.5–10)
	Dm (mm)	0.7±0.2 (0.4–1.2)	0.2±0.3 (0–1.1)	0.2±0.2 (0–0.6)	2.9±1.8 (1.1–8.9)	0.4±0.1 (0.2–0.6)
	DM (mm)	6.5±2.4 (4.1–14)	2.5±1.6 (1–7.7)	2.6±1.2 (0–4.4)	13±4.5 (6.5–26)	3.7±1.2 (2–6.3)
	D95 (mm)	6±2.3 (3.9–13)	2.3±1.3 (1–6.7)	2.4±1.1 (0–4.1)	12±4.6 (5.2–25)	3.2±1 (1.9–5.5)
Am	RV (%)	15±13 (1–42)	12±12 (1–44)	21±18 (1–78)	97±80 (0–200)	13±8 (3–34)
	K_1 (%)	76±7 (60–87)	90±6 (76–98)	82±9 (57–90)	43±33 (0–83)	80±4 (72–85)
	K_2 (%)	62±9 (43–77)	82±9 (61–96)	70±11 (40–83)	32±27 (0–71)	67±6 (56–73)
	FP (%)	17±10 (2–36)	9±10 (0–37)	19±13 (4–57)	12±15 (0–46)	21±6 (11–35)
	FN (%)	21±8 (12–39)	9±7 (1–27)	10±10 (0–33)	55±39 (4–100)	12±5 (5–23)
	MIV (%)	4±3.3 (0–10)	0.2±0.8 (0–3.1)	0.4±1 (0–3.5)	2.9±4.3 (0–12)	1±1.1 (0–3.8)
	Dm (mm)	0.8±0.3 (0.4–1.3)	0.4±0.3 (0.1–1.1)	0.7±0.4 (0.3–2)	3.8±3.3 (0.5–10)	0.7±0.2 (0.5–1.1)
	DM (mm)	4.5±0.9 (3.1–5.7)	2.6±0.9 (1.6–4.5)	3.8±1.2 (2.6–6.7)	9.3±4.5 (3.6–18)	4.3±1.1 (2.6–6.6)
	D95 (mm)	4.2±0.8 (2.5–5.6)	2.4±0.8 (1.5–4.2)	3.6±1.1 (2.4–6)	8.6±4.2 (3.3–17)	3.7±1.1 (2.4–6.4)

Values = average ± standard deviation (minimum–maximum).

material and furthermore, in some studies, it consisted in the whole Hc–Am complex (Ashton et al., 1997; Kelemen et al., 1999) thereby making performance comparisons difficult. Previously published results are summarised in Table 6, with indications on quantitative indices and performances. Our results in young controls were superior or comparable to previously published studies. Similar values were reported for Hc in Hogan et al. (2000), but, as noted previously, this method required a preliminary manual placement of 28 landmarks and a computation time of 2 h. In Shen et al. (2002), the overlap appears to be slightly better, but the method required the preliminary manual placement of 50 landmarks. Two other methods dealt with the segmentation of both structures. The values reported in Fischl et al. (2002) were slightly inferior for Hc and substantially inferior for Am. Finally, the overlap values obtained by Duchesne et al. (2002) were inferior for both Hc and Am, but no volume error estimates were given; an overlap value of K_1 = 80% was reported for Hc (N = 28) in Klemenčič et al. (2004), with an improved version of the method reported by Duchesne. Some studies gave distance errors only (Yang et al., 2004; Pitiot et al., 2004), but these were difficult to compare with our results as they were based on surfaces and not voxel-regions, and the definition of distance differs from between surface points to between border voxel centres.

Note that, with regard to inter-rater reproducibility for manual segmentation in healthy controls, our results compare favourably to those reported by Hogan et al. (2000) and Shen et al. (2002) for Hc. Inter-rater reproducibility of the automated segmentation was only studied in Hogan et al. (2000) for Hc, and the values we report in this study were slightly better.

Of the segmentation methods constrained by shape priors, only the method of Hogan et al. (2000) was evaluated on patients (Hsu et al., 2002) based on volumetry. Comparisons between automated and manual segmentations were not reported because the protocols used to define the atlas and the manual gold standards differed. Nonetheless, the agreement between automated and manual segmentation for Hc in 5 schizophrenic patients was: RV = 9% and K_1 = 74% (Haller et al., 1997); in 5 MTLE patients with monolateral Hc sclerosis, it was: RV = 11% and K_1 = 75% (Hogan et al., 2000).

Validation data on the segmentation of Hc in AD patients are limited and difficult to compare because of the differences in atrophy. In Carmichael et al. (2005), a comparative analysis of several atlas-based segmentations was proposed on three cohorts: 20 AD patients, 19 MCI and 15 healthy subjects. The performance of one fully deformable, three semi-deformable and three affine methods, with MNI, Harvard or cohort atlases, was evaluated by comparing manual and automated segmentation results. The best performance was obtained with the fully deformable method with a cohort-based atlas: $K_2 \sim 60\%$ for all subjects (better in healthy and MCI subjects than AD patients). In Crum et al. (2001), the warping of manually segmented structures onto repeat scans (average 18 months inter-scan) was compared to manual segmentation in 15

Table 5
Quantitative indices on subparts of both structures for the accuracy for P1–P8

Index	Hc head	Hc body	Hc tail	Am anterior	Am posterior
RV (%)	19±11 (2–39)	18±10 (1–33)	22±14 (1–57)	26±17 (2–66)	20±16 (2–45)
K_1 (%)	82±7 (67–92)	88±4 (81–93)	82±6 (68–91)	77±8 (56–86)	75±8 (59–89)
K_2 (%)	69±10 (51–86)	78±6 (68–87)	69±8 (52–83)	63±10 (39–76)	61±11 (42–81)
FP (%)	10±4 (3–20)	19±7 (7–30)	7±4 (3–19)	22±14 (1–54)	12±9 (3–34)
FN (%)	21±11 (5–37)	3±2 (1–9)	23±10 (8–46)	15±11 (3–38)	27±9 (8–39)
DM (mm)	4.6±1.8 (1.9–8.5)	3.2±1.1 (1.9–6)	5.2±3 (1.9–14)	3.8±1 (2.1–5.6)	4.2±0.9 (2.5–5.7)
D95 (mm)	4.3±1.7 (1.9–8)	2.9±1 (1.9–6)	5±2.9 (1.8–13)	3.6±1 (2.1–5.6)	4±0.8 (2.5–5.3)

Values = average ± standard deviation (minimum–maximum).

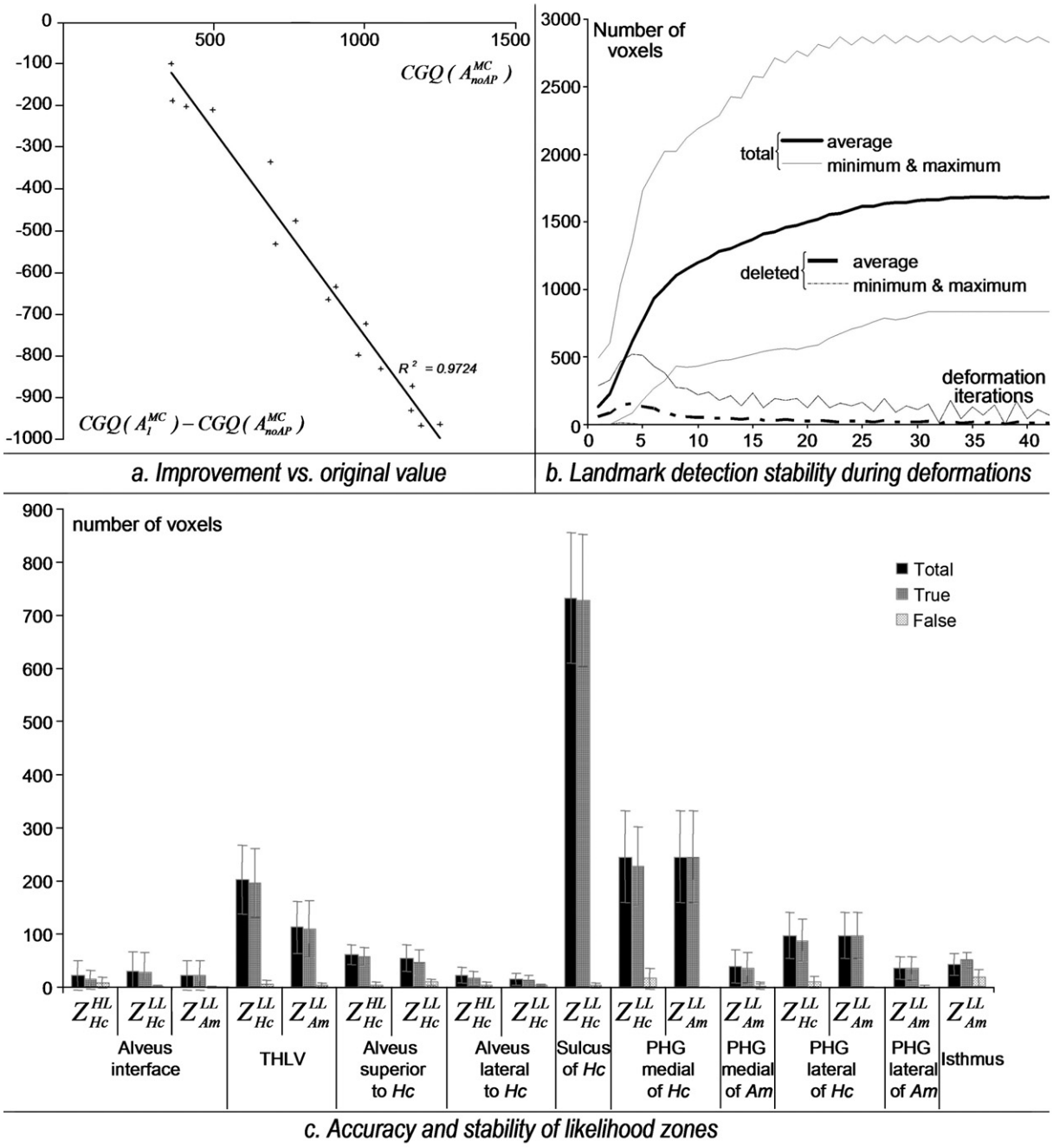

Fig. 8. Characterisation of landmark detection for P1–P8: (a) global improvement vs. global quality of A^{MC}_{noAP}; (b) total number of landmarks and number of landmarks not detected in the final iteration (minimal, average and maximal value at each iteration of the deformation of A^{MC}_{1}); (c) size of the final likelihood zones A^{MC}_{1}, with correct and incorrect classification compared to M^{MC}_{1}.

controls and 12 AD patients with the following results: RV=1%, K_2=90%,[2] and 5 subjects amongst the 27 had RV>5% and K_2<85%.

Methodological and data issues

The use of data acquired in a purely clinical context, therefore potentially sub-optimal compared to research data, is particularly interesting to test the algorithm's robustness. The choice of the different sequences used here was mainly dictated by previously defined protocols. We also wanted to test the algorithm on data acquired according to more recent protocols, for which larger data sets were available. Some of the scans were discarded due to large motion artefacts that rendered them un-interpretable.

Coronal orientation is usually preferred when studying Hc because the internal intricate details of Hc are better seen, and scans may even be acquired already reformatted perpendicular to the long axis of Hc (Hasboun et al., 1996). Scans in our data set were mainly acquired in the axial plane commonly used in non-Hc-

[2] The definition of K_2 utilised is unconventional.

Table 6
Summary comparison of our findings with previously published methods for automated segmentation of Hc and Am

		A/M			A/A			M/M			Sample size	Cpu time	Manual interaction	Qualitative evaluation	Manual protocol
		RV	K_1	Hm	RV	K_1	Hm	RV	K_1	Hm					
Chupin	H	7	84	0.5	4	95	0.2	7	90	0.3	32	<1 min Hc+Am	Low	Correct	Precise
	A	12	81	0.7	8	92	0.3	10	85	0.5					
Hogan	H	6	83		5	94		10	81		5	2 h Hc	Medium	Correct	Precise
	A														
Schen	H	6	88					3	86		20		High	Smooth	Precise
	A														
Fischl	H	10	80					15	80		14	30 min	Unclear		
	A	15	65					25	75						
Duchesne	H		68								60	20 min	Low	Unclear	Precise
	A		63												
Yang	H			1.8							12	1 h per object	Unclear	Unclear	
	A			1.6											
Pitiot	H			2.1							20	6 min, 4 objects	High	Unclear	Precise
	A														

A/M: accuracy.
A/A: inter-observer automated reproducibility.
M/M: inter-observer manual reproducibility.
Sample size: number of structures of each kind in the evaluation.
Cpu time: computation time.
Manual interaction: importance of the manual input in the method.
Qualitative evaluation: aspect of the 3D renderings or overlap slices.
Manual protocol: description of the protocol used in the manual segmentation to create the reference.

specific clinical routine, because it covers the whole brain with nearly isotropic voxels with good resolution and reduced scanning time. The alveus at the interface between Hc and Am is less blurred by partial volume effect on axial scans, which makes the automated discrimination between Hc and Am more robust in 3D.

The parameters used in the method were first tuned on S1–S4 and the radiometric parameters were refined on S5–S8 and P1–P8. In order to evaluate the influence of parameter tuning on the results and the applicability of the method on 'new' (i.e. non tuning) data sets, accuracy results are given in Table 7 for S1–S8 and S9–S16 only and for P1–P4 and P5–P8, showing that the values obtained for new data sets are similar to those obtained for the tuning data sets.

For controls, the worst result was obtained for S9, which was acquired with different acquisition parameters, with 1.5 mm axial slices and had the lowest CNR in the control group. Note that this acquisition was also used for P1–P4. In quantitative results, the systematic difference between Hc and Am could be explained partly by the fact that Am, which is roughly half the size of Hc, has blurred boundaries, and there was a significant pulsation artefact across its anterior part in 10 subjects (Fig. 9).

For AD patients, the two largest RV values for Hc for A_1^{MC} vs. M_1^{MC} were obtained for P2, for which the mean ROI grey matter intensity was consistently higher than the mean brain grey matter intensity, which was not the case for the other scans (except P7 left hemisphere). This difference could be explained by the high level of atrophy of the whole region and the under-representation of Hc grey matter in the ROI histogram grey matter mode, resulting in an incorrect estimation of the average intensity of the hippocampus. Note that the largest RV values for Hc for A_2^{MC} vs. A_1^{MC} and A_1^{RM} vs. A_1^{MC} were obtained for highly atrophic Hc (manual atrophy: 4.4 σ_{Hc}^{YHC} and 4.2 σ_{Hc}^{YHC}, respectively). The worst result, according to the combination of indices, was obtained for P5, for which the CNAR is the largest with a large amount of artefact. It can explain propagation problems towards the tail of Hc and the high variability of the segmentation for Am. Segmentation problems at the interface region and distance errors larger than for controls can be explained by the level of atrophy: when atrophy was severe, the grey matter in the head of Hc was reduced to a one-voxel ribbon against the white matter of the alveus and the parahippocampal gyrus (Fig. 10). Partial volume effect and noise made it

Table 7
Influence of the training set for parameter setting: statistics on S1–S8 ("tuning set"), S9–S16 ("test set"), P1–P4 ("tuning set") and P5–P8 ("test set") for quantitative indices for the accuracy

	Index	S1–S8	S9–S16	P1–P4	P5–P8
Hc	RV (%)	7±4 (1–13)	7±5 (0–14)	10±8 (0–21)	8±6 (0–15)
	K_1 (%)	84±3 (80–89)	83±2 (78–88)	83±2 (78–84)	85±3 (80–88)
	MIV (%)	1.1±0.9 (0–3.2)	1.2±1.2 (0–3.7)	0±0.1 (0–0.3)	1.5±2 (0–6.3)
	DM (mm)	4.2±1.6 (2.5–9)	4.8±1.5 (3–7.8)	6±1.2 (4.9–8.5)	7±3.2 (4.1–14)
Am	RV (%)	11±6 (2–21)	13±7 (1–27)	13±14 (1–42)	18±12 (1–35)
	K_1 (%)	83±3 (77–88)	80±5 (69–85)	77±5 (71–87)	76±9 (60–86)
	MIV (%)	1.5±1.1 (0.4–3.7)	1.5±0.9 (0.3–3.8)	5.1±2.6 (0–8.9)	2.9±3.8 (0–10)
	DM (mm)	3.5±0.5 (2.8–4.6)	4.3±1 (2.9–6)	4.6±0.9 (3.1–5.7)	4.4±0.9 (3.1–5.4)

Values = average ± standard deviation (minimum–maximum).

Fig. 9. Pulsation artefact in the anterior part of the amygdala (arrow) on an axial slice for S11.

appear as a white matter thread on coronal slices, and the propagation was disturbed. These errors were not reflected in the volume measurement, since they concerned only a small number of voxels, but were reflected in the distance measurements and in the indices computed on the head of Hc.

The landmarks were correctly detected, except the parahippocampal gyrus points in the head of Hc; this was due to the internal structure of Hc being occasionally visible, resulting in the detection of white matter voxels inside Hc as landmarks. About 25% of the results showed notable false detections of these landmarks in control scans, whereas only limited false detections were present in patient scans. False detection did not occur for the lower quality scans, as internal details were blurred. The impact of detection errors on the final segmentation results was small. A multi-scale approach could reduce sensitivity to spurious internal details. More generally, the large variance observed for some landmarks may be due to noise and partial volume effect. Differences between controls and patients may come from differences in voxel sizes, acquisition parameters and orientation, and the level of atrophy (modification of the shape of the sulcus of Hc). Regarding detection robustness, some landmarks detected at the first stages of the growth could be detected otherwise or rejected in later stages; this can be explained by spurious signal inside Hc and Am due to internal sub-structures and noise. Note that, due to the aspect of the Hc–Am interface, voxels erroneously detected as landmarks for the parahippocampal gyrus medial to Am at the beginning of the growth can be correctly detected later as landmarks for the alveus at the interface; this process is less likely to happen for AD patients because of the larger pool of CSF at the interface.

A possible limitation of our method is the deterministic sub-optimal algorithm (ICM) used to minimise the global energy, as there is no certainty that a local minimum will not be encountered during the deformations. Nevertheless, we found the search of the solution space with an ICM algorithm to be efficient and it was previously shown to be robust (Dubes et al., 1990). A stochastic algorithm, such as the Simulated Annealing (SA) algorithm (Geman and Geman, 1984) which would reduce the likelihood of convergence to local minima, would be computationally expensive (for an easier problem, namely a 64×64 2D image, the computation time was 6 s with ICM and several hours with SA on a Sun 4/280, as indicated in Dubes et al., 1990).

Another problem related to the ICM algorithm could arise from the predefined sequential order in which voxel candidates were considered for each ICM iteration. However, the results were on average equivalent to those with the fixed sequential order when voxel candidates were selected at random (see Table 8). Furthermore, to preserve the topological characteristics of the object, re-classification could only be considered for simple points (see definition in section Alternate Deformations). This property had to be checked for each ICM-iteration, as the objects were continuously modified making it necessary to take into account the current classification of all voxels.

Future work

The above considerations suggest that the quality of the results largely reflected that of the data. The influence of MR sequence parameters (TR, TE, TI, flip angle, acquisition matrix) on segmentation performances could be investigated more systematically (Clark et al., 2006; Li and Mirowitz, 2004). Furthermore, the necessity to use specific algorithmic parameters for AD patients could be attributed to lower CNR, linked to differences in the acquisition parameters or loss of contrast due to age. This problem requires further consideration, and in particular the investigation of optimal acquisition parameters suitable for clinical use. In addition, radiometric parameter settings could be automatically associated with specific pre-defined characteristics of the brain intensity histogram.

A new way to determine intensity parameters is being considered, in order to reduce sensitivity to ROI selection. Various degrees of automation of the initialisation step are also under study. A possibility would be to introduce probabilistic atlas-derived knowledge, in order first to extract the ROI automatically and also possibly to determine seeds from this atlas, and to assess the value of this approach in a pathological context. Segmentation errors are generally greater in the head and tail of the hippocampus and at the anterior limit of the amygdala and therefore the use of more landmarks is being investigated. Introducing a multi-scale Bayesian formalism in landmark detection and constraint could allow more flexibility in the process.

Finally, the accuracy of the algorithm was evaluated by comparing the automated segmentation with the manual one. However, the use of manual segmentation as a reference would still require an evaluation

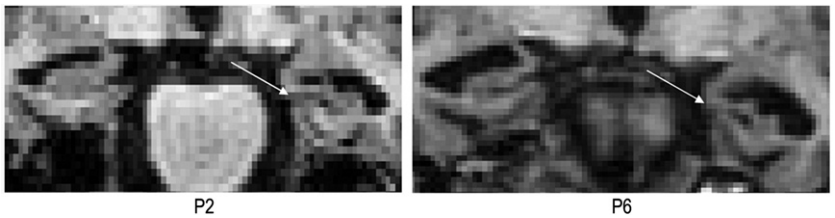

Fig. 10. Effect of the atrophy in the head of Hc for two AD patients.

Table 8
Statistics (on S1–S16) for quantitative indices for the accuracy with fixed order, average result, and best and worst indices values for random order (voxel front and voxel candidates) on 20 trials

	Index	Fixed	Average random	Best random	Worst random
Hc	RV (%)	7±4 (0–14)	8±4 (1–15)	6±4 (0–13)	11±6 (2–30)
	K_1 (%)	84±3 (78–89)	84±3 (78–89)	84±3 (80–89)	83±3 (75–88)
	MIV (%)	1.1±1 (0–3.7)	1.2±1.1 (0–3.7)	0.8±0.8 (0–3.5)	1.7±1.4 (0–4.9)
	DM (mm)	4.5±1.5 (2.5–9)	4.5±1.3 (2.4–8)	4±1 (2.1–6.5)	5.3±2.2 (3.1–14)
Am	RV (%)	12±7 (1–27)	12±7 (1–28)	10±7 (0–25)	15±7 (2–30)
	K_1 (%)	81±4 (69–88)	81±5 (69–88)	82±4 (72–89)	80±5 (65–87)
	MIV (%)	1.5±1 (0.3–3.8)	1.6±1 (0–3.8)	1.2±0.8 (0–3.6)	2.3±1.7 (0.4–6)
	DM (mm)	3.9±0.9 (2.8–6)	3.9±0.9 (2.8–6)	3.7±0.8 (2.8–6)	4.2±1 (2.8–6.4)

Values = average ± standard deviation (minimum–maximum).

with histological comparison. This difficult problem is a limitation of all MR-based segmentation of the hippocampus and amygdala and remains to be properly addressed.

Conclusion

We have presented a new automated hippocampus and amygdala MRI segmentation method. The competitive deforming region algorithm was constrained by priors derived from anatomical knowledge in the vicinity of landmarks that are automatically detected in healthy and diseased subjects. The algorithm's performance in terms of quality, reproducibility and computation time compared favourably to that of manual segmentation and other automated methods.

Acknowledgments

The authors would like to thank Fabrice Poupon and Denis Rivière for their help with the *Anatomist* software and Manik Bhattacharjee for the transfer to Brainvisa. Marie Chupin was funded for this work by a grant from the France Alzheimer Association and a post-doctoral Marie Curie fellowship. The authors would also like to thank Professor John Duncan and the National Society for Epilepsy for their support. Finally, the authors would like to thank Drs. Robert Powell and Mahinda Yogarajah for their helpful comments on English language regarding the manuscript.

Appendix A. Landmark definitions and detection

The detection of the landmarks follows a hierarchical order, defined in Fig. 11. The detection rules are detailed in the following paragraphs. Each is given in the form: *if* {set of conditions} *then* {set of actions}, in which all the conditions must be fulfilled. The six intensity thresholds used in the rules are derived from the automatically estimated average intensities and tolerances for Hc, Am and the composite object HcAm. Their setting is given in Appendix B, section B.5.

A.1. Alveus at the interface between Hc and Am

The alveus is a thin (~1 mm) white matter structure. Its anterior part is between Hc and Am. To detect the location of the alveus at the interface and use the complementary information that Am is superior

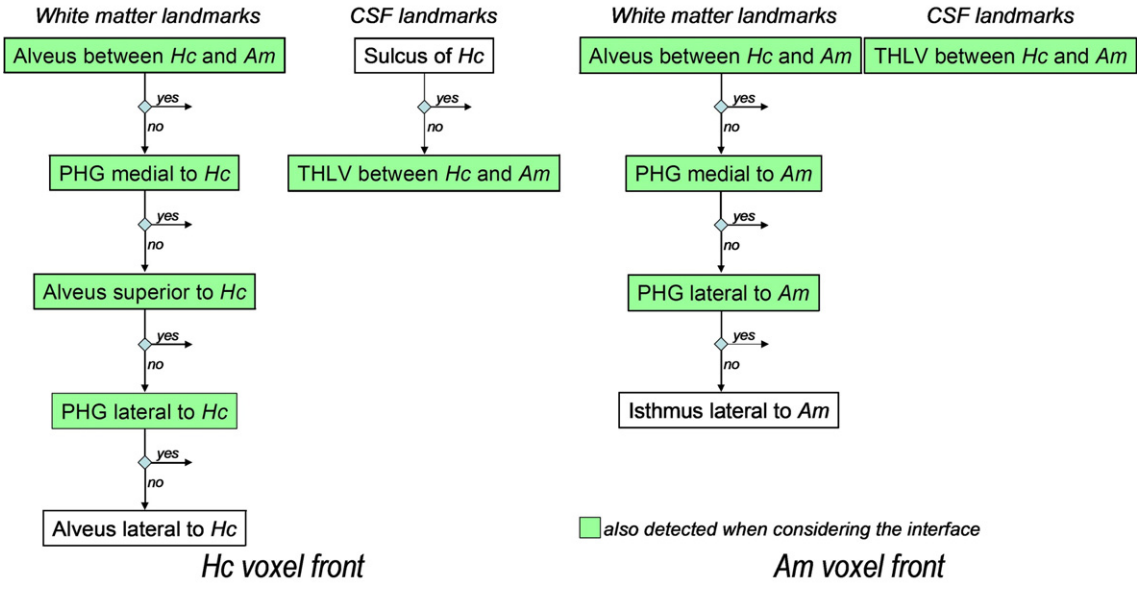

Fig. 11. Hierarchical order when detecting the landmarks.

and anterior to Hc, we compare the 2D neighbourhood pattern of the voxel in the sagittal slice with the following pattern:

$$P^{\text{sagittal\_interface}}(v) = \begin{bmatrix} v_{\text{Am}} & v_{\text{Am}} & \\ v_{\text{Am}} & v & v_{\text{Hc}} \\ & v_{\text{Hc}} & v_{\text{Hc}} \end{bmatrix}. \quad (4)$$

For the determination of the unlikely or likely zones for Hc and Am, the detection rules on the sagittal slice are:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v_{\text{Hc}} \in O_{\text{Hc}}\} \geq 2\ \&\ \text{card}\ \{v_{\text{Am}} \in O_{\text{Am}}\} \geq 2 \\ i(v) \geq i_{\text{alv}} \\ \text{card}\ \{v_{\text{Am}}\ ,\ \text{with}\ i(v_{\text{Am}}) \leq i_{\text{alv}}\} = 3 \end{array} \right\}, \quad (5)$$

$$\text{then} \left\{ \begin{array}{c} v_{\text{Hc}} \in Z_{\text{Am}}^{\text{LL}}\ \&\ v_{\text{Hc}} \in Z_{\text{Hc}}^{\text{HL}} \\ v_{\text{Am}} \in Z_{\text{Hc}}^{\text{LL}} \\ v \in Z_{Hc}^{\text{HL}} \end{array} \right\}.$$

The intensity threshold i_{alv} is derived from the characteristics for Hc. Voxels v_{Am} are not likely to belong to Am because of a possible presence of CSF between the alveus and Am.

A.2. Temporal horn of the lateral ventricle between Hc and Am

The temporal horn of the lateral ventricle (THLV) at the interface appears as a dark CSF zone between Hc and Am; again, complementary information is deduced from the anatomical position of Am, superior and anterior to Hc. $P^{\text{sagittal\_interface}}$ is used. When considering this interface, the corresponding detection rules become:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v_{\text{Hc}} \in O_{\text{Hc}}\} \geq 2\ \&\ \text{card}\ \{v_{\text{Am}} \in O_{\text{Am}}\} \geq 2 \\ i(v) \leq i_{\text{THLV}} \end{array} \right\}, \quad (6)$$

$$\text{then} \left\{ \begin{array}{c} v_{\text{Hc}} \in Z_{\text{Am}}^{\text{LL}} \\ v_{\text{Am}} \in Z_{\text{Hc}}^{\text{LL}} \\ v \in Z_{\text{Hc}}^{\text{LL}}\ \&\ v \in Z_{\text{Am}}^{\text{LL}} \end{array} \right\}.$$

When considering Hc, the detection rules are:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v_{\text{Hc}} \in O_{\text{Hc}}\} \geq 2\ \&\ \text{card}\ \{v_{\text{Am}} \in O_{\text{Hc}}\} \leq 1 \\ i(v) \leq i_{\text{THLV}} \end{array} \right\}, \quad (7)$$

$$\text{then} \left\{ \begin{array}{c} v_{\text{Hc}} \in Z_{\text{Am}}^{\text{LL}} \\ v_{\text{Am}} \in Z_{\text{Hc}}^{\text{LL}} \\ v \in Z_{\text{Hc}}^{\text{LL}}\ \&\ v \in Z_{\text{Am}}^{\text{LL}} \end{array} \right\}.$$

When considering Am, the detection rules are:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v_{\text{Am}} \in O_{\text{Am}}\} \geq 2\ \&\ \text{card}\ \{v_{\text{Hc}} \in O_{\text{Am}}\} \leq 1 \\ i(v) \leq i_{\text{THLV}} \end{array} \right\} \quad (8)$$

$$\text{then} \left\{ \begin{array}{c} v_{\text{Hc}} \in Z_{\text{Am}}^{\text{LL}} \\ v_{\text{Am}} \in Z_{\text{Hc}}^{\text{LL}} \\ v_{\text{Hc}} \in Z_{\text{Hc}}^{\text{LL}}\ \&\ v \in Z_{\text{Am}}^{\text{LL}} \end{array} \right\}.$$

The intensity threshold i_{THLV} is derived from the characteristics for HcAm.

A.3. Alveus superior to Hc

The alveus also appears on the superior border of Hc. The 2D pattern is defined on sagittal slices:

$$P^{\text{sagittal\_superior}}(v) = \begin{bmatrix} v^- & v^- & \\ & v & \\ v* & v^+ & v^+ \end{bmatrix}, \quad (9)$$

and the detection rules are:

$$\text{if} \left\{ \begin{array}{c} \text{card}\{v^+ \in O_{\text{Hc}}, v* \in O_{\text{Hc}}\} = 3\ \&\ \text{card}\{v^- \in O_{\text{Hc}}\} = 0 \\ i(v) \geq i_{\text{alv}} \\ i(v^-) \leq i_{\text{alv}}\ \&\ |i(v^-) - i(v)| \geq \sigma_{\text{alv}} \end{array} \right\}, \quad (10)$$

$$\text{then} \left\{ \begin{array}{c} v^- \in Z_{\text{Hc}}^{\text{LL}} \\ v^+ \in Z_{\text{Hc}}^{\text{HL}} \\ v \in Z_{\text{Hc}}^{\text{HL}} \end{array} \right\}.$$

The intensity tolerance σ_{alv} is derived from the characteristics for Hc.

A.4. Alveus lateral to Hc

The alveus also appears on the lateral-superior border of Hc. The 2D pattern is defined on coronal slices (displayed for the right side of the brain, in radiological convention; left side defined symmetrically):

$$P^{\text{coronal\_lateral-superior}}(v) = \begin{bmatrix} v^- & v^- & \\ v^- & v & v^+ \\ & v^+ & v^+ \end{bmatrix}, \quad (11)$$

and the detection rules are:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v^+ \in O_{\text{Hc}}\} \geq 2\ \&\ \text{card}\ \{v^- \in O_{\text{Hc}}\} = 0 \\ i(v) \geq i_{\text{alv}} \\ |i(v^-) - i(v)| \geq \sigma_{\text{alv}} \end{array} \right\}, \quad (12)$$

$$\text{then} \left\{ \begin{array}{c} v^- \in Z_{\text{Hc}}^{\text{LL}} \\ v^+ \in Z_{\text{Hc}}^{\text{HL}} \\ v \in Z_{\text{Hc}}^{\text{HL}} \end{array} \right\}.$$

A.5. Sulcus of Hc

The sulcus of Hc has variable size and shape in normal and diseased population. It appears on sagittal slices as a thin CSF space in the head of Hc, roughly oriented in an anterio-posterior direction. The 2D pattern is defined on sagittal slices:

$$P^{\text{sagittal\_middle}}(v) = \begin{bmatrix} & & v^P \\ v^A & v & v^P \\ v^A & & \end{bmatrix}, \quad (13)$$

and the detection rules are:

$$\text{if} \left\{ \begin{array}{c} i(v) \leq i_{\text{sulcus}} \\ \text{card}\{v^{\text{A}},\ \text{with}\ v^A \notin O_{\text{Hc}}\ \&\ i(v^{\text{A}}) \leq i_{\text{sulcus}}\} \geq 1 \\ \text{card}\{v^{\text{P}},\ \text{with}\ v^{\text{P}} \notin O_{\text{Hc}}\ \&\ i(v^{\text{P}}) \leq i_{\text{sulcus}}\} \geq 1 \end{array} \right\}, \quad (14)$$

$$\text{then} \left\{ \begin{array}{c} v \in Z_{\text{Hc}}^{\text{LL}} \\ \forall v^{\text{P}},\ \text{if}\ \{i(v^{\text{P}}) \leq i_{\text{sulcus}}\},\ \text{then}\ \{v^{\text{P}} \in Z_{\text{Hc}}^{\text{LL}}\} \\ \forall v^{\text{A}},\ \text{if}\{i(v^{\text{A}}) \leq i_{\text{sulcus}}\},\ \text{then}\ \{v^{\text{A}} \in Z_{\text{Hc}}^{\text{LL}}\} \end{array} \right\}.$$

The intensity threshold i_{sulcus} is derived from the characteristics for HcAm.

A.6. Parahippocampal gyrus medial to Hc or Am

The parahippocampal gyrus white matter appears partly as a white structure on the medial-inferior border of Hc and Am. The 2D pattern is defined on coronal slices (displayed for the right side of the brain, in radiological convention; left side defined symmetrically):

$$P^{\text{coronal\_medial-inferior}}(v) = \begin{bmatrix} v^+ & v^+ & \\ v^+ & v & v^- \\ v^- & v^- & v^- \end{bmatrix}, \tag{15}$$

and the detection rules are, for Hc:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v^+ \in O_{\text{Hc}}\} = 3\ \&\ \text{card}\ \{v^- \in O_{\text{Hc}}\} = 0 \\ i(v) \geq i_{\text{GPH}} \\ \text{card}\ \{i(v^-) \geq i_{\text{GPH}}\} \geq 1 \end{array} \right\},$$
$$\text{then} \left\{ \begin{array}{c} v^- \in Z^{\text{LL}}_{\text{Hc}} \\ v^- \in Z^{\text{LL}}_{\text{Am}} \\ v \in Z^{\text{LL}}_{\text{Hc}} \end{array} \right\}. \tag{16}$$

and for Am:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v^+ \in O_{\text{Am}}\} = 3\ \&\ \text{card}\ \{v^- \in O_{\text{Am}}\} = 0 \\ i(v) \geq i_{\text{GPH}} \\ \text{card}\ \{i(v^-) \geq i_{\text{GPH}}\} \geq 1 \end{array} \right\},$$
$$\text{then} \left\{ \begin{array}{c} v^- \in Z^{\text{LL}}_{\text{Am}} \\ v^- \in Z^{\text{LL}}_{\text{Am}} \end{array} \right\}. \tag{17}$$

The intensity threshold i_{GPH} is derived from the characteristics for HcAm.

A.7. Parahippocampal gyrus lateral to Hc or Am

The parahippocampal gyrus white matter also appears as a white structure on the lateral-inferior border of Hc and Am. The 2D pattern is defined on coronal slices (displayed for the right side of the brain, in radiological convention; left side defined symmetrically):

$$P^{\text{coronal\_lateral-inferior}}(v) = \begin{bmatrix} & v^+ & v^+ \\ v^- & v & v^+ \\ v^- & v^- & v^- \end{bmatrix}, \tag{18}$$

and the detection rules are, for Hc:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v^+ \in O_{\text{Hc}}\} = 3\ \&\ \text{card}\ \{v^- \in O_{\text{Hc}}\} = 0 \\ i(v) \geq i_{\text{GPH}} \\ \text{card}\ \{i(v^-) \geq i(v_{\text{GPH}}\} \geq 1 \end{array} \right\},$$
$$\text{then} \left\{ \begin{array}{c} v^- \in Z^{\text{LL}}_{\text{Hc}} \\ v^- \in Z^{\text{LL}}_{\text{Am}} \\ v \in Z^{\text{LL}}_{\text{Hc}} \end{array} \right\}. \tag{19}$$

and for Am:

$$\text{if} \left\{ \begin{array}{c} \text{card}\ \{v^+ \in O_{\text{Am}}\} = 3\ \&\ \text{card}\ \{v^- \in O_{\text{Am}}\} = 0 \\ i(v) \geq i_{\text{GPH}} \\ \text{card}\ \{i(v^-) \geq i_{\text{GPH}}\} \geq 1 \end{array} \right\},$$
$$\text{then} \left\{ \begin{array}{c} v^- \in Z^{\text{LL}}_{\text{Am}} \\ v^- \in Z^{\text{LL}}_{\text{Am}} \end{array} \right\}. \tag{20}$$

A.8. Temporal lobe isthmus for Am

The isthmus is a white matter bridge on the lateral-superior border of Am. $P^{\text{coronal\_lateral-superior}}$ is used. The detection rules are:

$$\text{if} \left\{ \begin{array}{c} \text{card}\{v^+ \in O_{\text{Hc}}\} = 3\ \&\ \text{card}\{v^- \in O_{\text{Hc}}\} = 0 \\ i(v) \geq i_{\text{isthmus}} \end{array} \right\},$$
$$\text{then} \left\{ \begin{array}{c} v^- \in Z^{\text{LL}}_{Am} \\ v \in Z^{\text{LL}}_{Am} \end{array} \right\}. \tag{21}$$

The intensity threshold i_{isthmus} is derived from the characteristics for HcAm.

A.9. Low likelihood zones propagation

$Z^{\text{LL}}_{\text{Hc}}$ and $Z^{\text{LL}}_{\text{Am}}$ are grown amongst candidate voxels following rules that were used to create them, except for the sulcus of Hc, for which no propagation was considered. $Z^{\text{HL}}_{\text{Hc}}$ could not be propagated in the same way, due to the convex shape of Hc. The propagation followed the following hierarchical order: alveus at the interface, alveus superior to Hc, the four parahippocampal gyrus landmarks together, alveus lateral to Hc (when considering the Hc front) or isthmus (when considering the Am front), and THLV.

Appendix B. Energy functional

B.1. Global data attachment term

The term for the objects O, $E^{\text{G}}_O(v)$, is based on global characteristics of the structure's intensity distribution. The intensity value at voxel v, $i(v)$, is compared with $\tilde{i}_O$, the average intensity value of O, with σ^{G}_O a standard deviation characterising the intensity range around $\tilde{i}_O$. The energy term for BG_{HcAm} is defined to exclude the intensities of both Hc and Am, defined as a single object O_{HcAm}; intensity and tolerance for O_{HcAm} are the average of those for O_{Hc} and O_{Am}. Therefore:

$$\left\{ \begin{array}{l} E^{\text{G}}_O(v) = \left(\dfrac{i(v) - \tilde{i}_O}{\sigma^{\text{G}}_O} \right)^2 \\ E^{\text{G}}_{\text{BG}_{\text{HcAm}}}(v) = \dfrac{1}{E^{G}_{O_{\text{HcAm}}} + \varepsilon} \end{array} \right., \tag{22}$$

with $\varepsilon << \left(\frac{1}{\sigma^{\text{G}}_O} \right)^2$ introduced to ensure numerical stability when $i(v) = \tilde{i}_O$.

Fig. 12 illustrates the effect of data attachment terms on a schematic drawing. The desired segmentation is delineated on the

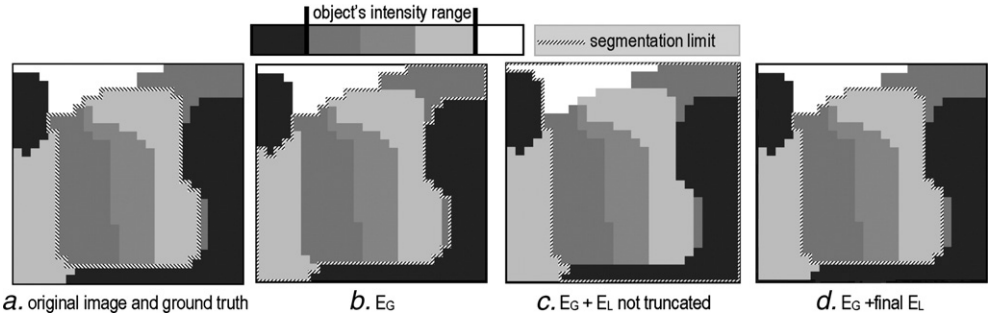

Fig. 12. Segmentation of a 2D simulation image, showing the influence of the data attachment terms on the result of the segmentation: (a) ground truth, (b) segmentation with E^G alone, (c) segmentation with E^G and E^L without truncation, and (d) segmentation with E^G and E^L with truncation.

sample image Fig. 12a. In Fig. 12b, the segmentation using E^G only is shown.

B.2. Local data attachment term

Using the local intensity properties of the 26-neighbourhood of the current voxel, E^L enables the detection of voxels creating high intensity gradients with O and avoids their addition to O. Low and high gradient patterns are defined with respect to σ^L_O, a local intensity range around $\hat{\imath}_O(v)$, the local average intensity value of O (average of the intensity values of O-labelled neighbours of v).

The result of incorporating E^G and E^L into the global energy function is illustrated in Fig. 12. To reduce the bias towards lower gradients, E^L_O is truncated to restrict its effect to intensity values within the local intensity range, $\hat{\imath}_O(v)\pm\sigma^L_O$. The effect of the truncated E^L_O is illustrated in Fig. 12d. The local data attachment term are thus defined as:

$$\begin{cases} \text{if } |i(v)-\hat{\imath}_O(v)| \geq \sigma^L_O, E^L_O(v) = \left(\dfrac{i(v)-\hat{\imath}_O(v)}{\sigma^L_O}\right)^2 ; \text{ else, } E^L_O(v)=1 \\ E^L_{\mathrm{BG}_{\mathrm{HcAm}}}(v) = \dfrac{1}{E^L_{O_{\mathrm{HcAm}}}+\varepsilon} \end{cases}. \tag{23}$$

The difference $i(v)-\hat{\imath}_O(v)$ is a first order approximation of the local intensity gradient at v.

B.3. Regularisation term

The third term is a Markovian regularisation term, modelling local classification continuity between the voxel and its neighbours in 26-connectivity, to minimise voxel wires and holes in the segmented object. It emerges from the Ising regularisation scheme (Geman and Geman, 1984; Dubes et al., 1990). It is locally expressed as the comparison of the number of O-labelled neighbours of v, $N_O(v)$, and a standard number of neighbours $\tilde{N}$ characterising regularity, with σ^I a standard deviation around $\tilde{N}$. To penalise the most irregular configurations, a non-linear function is used:

$$E^I_O(v) = \left(\frac{\tilde{N}-\gamma_O(v)\cdot N_O(v)}{\sigma^I}\right)^5. \tag{24}$$

γ_O is introduced to influence the classification process in the likely ($\gamma_O=2$) and unlikely ($\gamma_O=0.5$) meta-regions defined with respect to the sets of anatomical priors (otherwise, $\gamma_O=1$).

$N_{O_{Hc}}(v)$ is computed in an anisotropic way, in order to model the curved shape of Hc. As the seed is placed in the middle of the head of Hc (HHc), the deformations roughly follows a rostro-caudal propagation towards the tail of Hc (THc) and caudo-rostral propagation towards the anterior tip of HHc. It has been observed that, during deformations, the current voxel v should be classified in O_{Hc} to ensure a correct propagation when some of its voxel neighbours are already classified in O_{Hc}. This observation allows the definition of systematic patterns in coronal orientation (symmetrically for right and left hemispheres) characterising correct propagation in THc and HHc. For propagation towards THc, the coronal neighbourhood pattern for the right-hand side of the brain (in radiological conventions) is made of four points in the coronal slice anterior to the current voxel v:

$$P^4_{\mathrm{THc}}(v) = \underset{\text{anterior}}{\begin{bmatrix} v_{\mathrm{Hc}} & v_{\mathrm{Hc}} \\ v_{\mathrm{Hc}} & v_{\mathrm{Hc}} \end{bmatrix}} \underset{\text{current}}{\begin{bmatrix} v \end{bmatrix}} \underset{\text{posterior}}{\begin{bmatrix} \end{bmatrix}}. \tag{25}$$

For propagation towards HHc, the coronal neighbourhood pattern for the right-hand side of the brain (in radiological conventions) is made of three points (two in the coronal slice posterior to v, and one in the same coronal slice as v):

$$P^3_{\mathrm{HHc}}(v) = \underset{\text{anterior}}{\begin{bmatrix} \end{bmatrix}} \underset{\text{current}}{\begin{bmatrix} v & v_{\mathrm{Hc}} \end{bmatrix}} \underset{\text{posterior}}{\begin{bmatrix} v_{\mathrm{Hc}} & v_{\mathrm{Hc}} \end{bmatrix}}. \tag{26}$$

When one of these patterns is detected in the 26-neighbourhood of the current voxel v, v is more likely to belong to O_{Hc}. $N_{O_{Hc}}(v)$ is modified to increase the weight of the neighbours in the patterns:

$$\begin{aligned} N^{\text{anisotropic}}_{O_{\mathrm{Hc}}}(v) = N^{\text{isotropic}}_{O_{\mathrm{Hc}}}(v) &+ \alpha_{\mathrm{THc}}\cdot\delta_{O_{\mathrm{Hc}}}(P^4_{\mathrm{THc}}(v)) \\ &+ \alpha_{\mathrm{HHc}}\cdot\delta_{O_{\mathrm{Hc}}}(P^3_{\mathrm{HHc}}(v)), \end{aligned} \tag{27}$$

where δ_O is the characteristic function of O for each pattern ($\delta_O(P)=\mathrm{card}(P)$ when all the voxels in the set P are classified in O and 0 otherwise), and α_{THc} and α_{HHc}, two parameters adjusting the anisotropic effect.

B.4. Surface and volume terms

These two terms are introduced to prevent O from growing when the volume (number of voxels in O, $V_O(v)$) and surface (number of surface voxels of O, $S_O(v)$) reach unrealistically large values above given thresholds (V_O^* and S_O^*). These terms should not influence the deformation process while volume and surface are under their associated thresholds. Since these geometric characteristics are meaningless for BG_{HcAm}, a pressure term is used instead (θ positive value) to favour growth; this pressure term influences re-classification of voxel candidates otherwise undecided and prevents unsatisfactory behaviours when the objects O stop growing because of a slight local minimum.

$$\begin{cases} \text{if } V_O(v) < V_O^*,\ E_O^{\mathrm{V}}(v) = 0;\ \text{else, } E_O^{\mathrm{V}}(v) = \left(\dfrac{V_O(v) - V_O^*}{\sigma_O^{\mathrm{V}}}\right)^2, \\ E_{\mathrm{BG_{HcAm}}}^{\mathrm{V}}(v) = \theta \end{cases} \tag{28}$$

and

$$\begin{cases} \text{if } S_O(v) < S_O^*, E_O^{\mathrm{S}}(v) = 0;\ \text{else, } E_O^{\mathrm{S}}(v) = \left(\dfrac{S_O(v) - S_O^*}{\sigma_O^{\mathrm{S}}}\right)^2, \\ E_{\mathrm{BG_{HcAm}}}^{\mathrm{S}}(v) = \theta \end{cases} \tag{29}$$

with σ_O^{V} (or σ_O^{S}) a standard deviation characterising volume (or surface) range.

B.5. Parameter setting

The value of the 13 parameters used in the energy functional was either calculated automatically or fixed in the program as described in Table 9.

B.5.1. Radiometric parameters

Mean intensity values, $\tilde{i}_O$, and intensity ranges, σ_O^{G} and σ_O^{L}, are related to the scan properties and therefore computed from the data using sequence dependant factors. They were derived from the mean intensity and standard deviation of grey matter on the ROI ($_{\mathrm{GM}}$ and σ_{GM}); these were automatically retrieved from a histogram analysis on the ROI. For Hc, the mean intensity was empirically set at $\tilde{i}_{\mathrm{Hc}} = 0.95.\tilde{i}_{\mathrm{GM}}$. Our observation showed that Hc was to be considered slightly brighter than Am to ensure correct competition. To force this difference between the two average intensity values, $\tilde{i}_{\mathrm{Am}} = 0.9.\tilde{i}_{\mathrm{GM}}$.

For Hc, the global intensity range was empirically set at $\sigma_{\mathrm{Hc}}^{\mathrm{G}} = 2\sigma_{\mathrm{GM}}$ for controls' scans and $\sigma_{\mathrm{Hc}}^{G} = 2.5\sigma_{\mathrm{GM}}$ for patients' scans. For Am, it was empirically set at $\sigma_{\mathrm{Am}}^{\mathrm{G}} = 1.4\sigma_{\mathrm{GM}}$. The local intensity range, σ^{L}, was set to $\sigma_O^{\mathrm{L}} = 0.4\sigma_O^{\mathrm{G}}$ for the controls' scans and $\sigma_O^{\mathrm{L}} = 0.5\sigma_O^{\mathrm{G}}$ for the patients' scans; this ratio was inferred from the fact that the local intensity characterisation had to be more acute than the global one. The difference between sequence-dependant factors came from contrast decrease in patients' scans, which were of lesser quality.

Intensity thresholds for the detection of landmarks were empirically derived from $\tilde{i}_{\mathrm{Hc}}$ and $\tilde{i}_{\mathrm{Am}}$: $i_{\mathrm{alv}} = \tilde{i}_{\mathrm{Hc}} + 0.4\sigma_{\mathrm{Hc}}^{\mathrm{G}}$, $\sigma_{\mathrm{alv}} = 0.4\sigma_{\mathrm{Hc}}^{\mathrm{G}}$, $i_{\mathrm{THLV}} = \tilde{i}_{\mathrm{HcAm}} - 1.5\sigma_{\mathrm{HcAm}}^{\mathrm{G}}$, $i_{\mathrm{sulcus}} = \tilde{i}_{\mathrm{HcAm}} - \sigma_{\mathrm{HcAm}}^{\mathrm{G}}$, $i_{\mathrm{GPH}} = \tilde{i}_{\mathrm{HcAm}} + 0.7\sigma_{\mathrm{HcAm}}^{\mathrm{G}}$ and $i_{\mathrm{isthmus}} = \tilde{i}_{\mathrm{HcAm}} + \sigma_{\mathrm{HcAm}}^{\mathrm{G}}$.

B.5.2. Geometric parameters

Ising parameters were derived from geometrical considerations on the 26-neighbourhood ($\tilde{N}$ and σ^{I}). $\tilde{N}$ was set as half the neighbourhood size. The standard deviation σ^{I} was set to 2 to avoid irregularities such as wires (high risk if v was included but had less than 10 neighbours) or holes (high risk if v was excluded but had more than 16 neighbours). The two parameters adjusting anisotropic regularisation for Hc appeared to be sensitive to noise, and thus to be sequence-dependent and were tuned accordingly. Note that for P7 the anisotropy parameters were adapted on the left-hand side to compensate for low contrast and high noise.

Surface and volume terms were necessary to prevent the objects from growing further than reasonable sizes. Thresholds, V_O^* and S_O^*, and ranges, σ_O^{V} and σ_O^{S}, were thus set to control the relative influence of E^{V} and E^{S} in the energy functional. V_O^* and S_O^* values were derived from average values obtained from manual segmentations of a sample of young controls; we checked that volume averages corresponded to the values given in the literature, with equivalent manual protocols (Pruessner et al., 2000), but there was no data available for the surface values. Ranges were set large enough to allow variations larger than the typical inter-individual variability. Thus, the constraint derived from the thresholds would influence the deformation process when $V_O \geq V_O^* + \sigma_O^{\mathrm{V}}$ and $S_O \geq S_O^* + \sigma_O^{\mathrm{S}}$, high above average values. These terms were not modified for patient data, so that the algorithm could be used without knowing the diagnostic. The enlargement of the Hc sulcus was likely to modify the ratio volume/surface, but this did not influence the segmentation performance. The parameter θ was set to a small value; the pressure force was efficient only in cases for which the classification was unresolved by the other energy terms.

Table 9
Settings of the parameters used in the energy functional

Parameter	σ^{G}	$\tilde{i}$	σ^{L}	ε	σ^{L}	α_{THc}	α_{HHc}	$\tilde{N}$	V^*	σ^{V}	S^*	σ^{S}	θ
Setting	PS_C/PS_S	PS_C	PS_C/PS_S	PS_F	PS_F	PS_F/PS_S	PS_F/PS_S	PS_F	PS_L	PS_L	PS_L	PS_L	PS_F
Hc	–	–	–	0.001	2	1, 2 or 3	0 or 2	13	2750	700	1950	500	0.1
Am	–	–	–	0.001	2			13	1250	600	900	450	0.1

PS_S = sequence dependant.
PS_C = computed in the program.
PS_F = fixed in the program.
PS_L = derived from the literature.

Appendix C. Manual segmentation protocol

The *cornu amonis*, the *dentate gyrus*, the *alveus*, the *fimbria*, and the part of the *subiculum* directly inferior to the *cornu amonis* were included in Hc. The medial limit was geometrically derived from the *fimbria*. If this structure was not visible on a slice, we used the junction between the lateral curved part and the medial flat part of the *subiculum*, as it appeared on coronal slices. The most anterior and posterior coronal slices were given by the *alveus* and the disappearance of grey matter under the *crus fornicis*, respectively.

As for Am, the superior limit was determined following the *endorhinal sulcus*. The medial limit, when not entirely defined by the white matter of the *uncal* and *entorhinal gyri*, was drawn by prolonging it. The lateral limit was given by the white matter of the *isthmus*. The posterior limit was defined by the *alveus*, and the most anterior coronal slice was found by scrolling backward until a grey matter mass appeared in the white matter of the *uncal gyrus*, with reference to sagittal slices.

Appendix D. Segmentation quality indices

They compare the segmentation result Seg with the gold standard Ref. Nine indices were used to quantify the accuracy of the method and facilitate comparison with published values (Gerig et al., 2001).

RV is the relative error on volume; it compares volumes for segmented object O^{Seg} and reference object O^{Ref}, relatively to the average of both:

$$\mathrm{RV}\left(O^{\mathrm{Seg}},O^{\mathrm{Ref}}\right)=2\frac{\left|V_{O^{\mathrm{Seg}}}-V_{O^{\mathrm{Ref}}}\right|}{V_{O^{\mathrm{Seg}}}+V_{O^{\mathrm{Ref}}}}. \tag{30}$$

The optimal value for this index is 0%. Five measures are then used to quantify the number of properly classified voxels. The first two characterise overlap between O^{Seg} and O^{Ref}; they correspond to two ways in which the similarity index can be computed. K_1 is written:

$$K_1\left(O^{\mathrm{Seg}},O^{\mathrm{Ref}}\right)=2\frac{V_{O^{\mathrm{Seg}}\cap O^{\mathrm{Ref}}}}{V_{O^{\mathrm{Seg}}}+V_{O^{\mathrm{Seg}}}}. \tag{31}$$

This index, largely used in the literature, characterises the number of properly classified voxels, without taking into account the number of ill-classified voxels. It is thus less discriminative than K_2:

$$K_2\left(O^{\mathrm{Seg}},O^{\mathrm{Ref}}\right)=\frac{V_{O^{\mathrm{Seg}}\cap O^{\mathrm{Ref}}}}{V_{O^{\mathrm{Seg}}\cup O^{\mathrm{Ref}}}}. \tag{32}$$

The optimal value for both indices is 100%. The numbers of false positives, FP, and false negatives, FN, are computed here relatively to the number of voxels labelled as O^{Seg} or O^{Ref}:

$$\mathrm{FP}\left(O^{\mathrm{Seg}},O^{\mathrm{Ref}}\right)=\frac{V_{O^{\mathrm{Seg}}}-V_{O^{\mathrm{Seg}}\cup O^{\mathrm{Ref}}}}{V_{O^{\mathrm{Seg}}\cap O^{\mathrm{Ref}}}}, \tag{33}$$

$$\mathrm{FN}\left(O^{\mathrm{Seg}},O^{\mathrm{Ref}}\right)=\frac{V_{O^{\mathrm{Ref}}}-V_{O^{\mathrm{Seg}}\cap O^{\mathrm{Ref}}}}{V_{O^{\mathrm{Seg}}\cup O^{\mathrm{Ref}}}} \tag{34}$$

Both are optimal when equal to 0%. They can be related to the overlap index through:

$$\mathrm{FP}(O^{\mathrm{Seg}},O^{\mathrm{Ref}})+FN(O^{\mathrm{Seg}},O^{\mathrm{Ref}})=1-K_2(O^{\mathrm{Seg}},O^{\mathrm{Ref}}) \tag{35}$$

Finally, we introduce a new index, called Misclassified Interface Voxels (MIV), which quantifies the number of voxels labelled as O_1 in the segmentation and as O_2 in the reference:

$$\mathrm{MIV}\left(O_1^{\mathrm{Seg}},O_1^{\mathrm{Ref}},O_2^{\mathrm{Ref}}\right)=2\frac{V_{O_1^{\mathrm{Seg}}\cap O_2^{\mathrm{Ref}}}}{V_{O_1^{\mathrm{Seg}}}+V_{O_1^{\mathrm{Ref}}}}. \tag{36}$$

Its optimal value is 0%.

In order to estimate the sensitivity of these indices, one can compute them for simple deformations of a manual segmentation of Hc and Am. Let $O^{\mathrm{Ref}+\boldsymbol{T}}$ be O^{Ref} translated by a vector $\boldsymbol{T}$. If $\boldsymbol{T}$=(1,1,1), the values of the overlap ratios were: for Hc, $K_1(O^{\mathrm{Ref}+\boldsymbol{T}}, O^{\mathrm{Ref}})$=64% and $K_2(O^{\mathrm{Ref}+\boldsymbol{T}}, O^{\mathrm{Ref}})$=47%; for Am, $K_1(O^{\mathrm{Ref}+\boldsymbol{T}}, O^{\mathrm{Ref}})$=71% and $K_2(O^{\mathrm{Ref}+\boldsymbol{T}}, O^{\mathrm{Ref}})$=55%; for the misclassified interface voxels, for Hc, MIV $(O^{\mathrm{Ref}-\boldsymbol{T}}, O^{\mathrm{Ref}})$=7.7% and for Am, MIV$(O^{\mathrm{Ref}+\boldsymbol{T}}, O^{\mathrm{Ref}})$=16.4%. Let $O^{\mathrm{Ref}.M}$ be O^{Ref} morphologically eroded (M=E) or dilated (M=D) with a 1-voxel element. The values of the error on volume were: for Hc, RV$(O^{\mathrm{Ref}.E}, O^{\mathrm{Ref}})$=36% and RV$(O^{\mathrm{Ref}.D}, O^{\mathrm{Ref}})$=29%; for Am, RV$(O^{\mathrm{Ref}.E}, O^{\mathrm{Ref}})$=44% and RV$(O^{\mathrm{Ref}.D}, O^{\mathrm{Ref}})$=34%.

The local behaviour on the boundary can be characterised by its surface voxels (defined as the voxels of O with at least one 26-neighbour outside of O). The distance between the centre of the surface voxels of O^{Seg} and those of O^{Ref} is considered in three ways. First, the average symmetric distance on the whole boundary, Dm, is computed:

$$\mathrm{Dm}(O^{\mathrm{Seg}},O^{\mathrm{Ref}})=\max[h(O^{\mathrm{Seg}},O^{\mathrm{Ref}}),h(O^{\mathrm{Ref}},O^{\mathrm{Seg}})], \tag{37}$$

with

$$h(A,B) = \frac{1}{N_A}\sum_{a\in A} d(a,B).$$

Second, the maximum of the symmetric distance (Hausdorff distance), DM, is considered:

$$\mathrm{DM}(O^{\mathrm{Seg}},O^{\mathrm{Ref}}) = \max[H(O^{\mathrm{Seg}},O^{\mathrm{Ref}}),H(O^{\mathrm{Ref}},O^{\mathrm{Seg}})], \tag{38}$$

with

$$H(A,B) = \max_{a\in A}[d(a,B)], \tag{39}$$

with d the Euclidian distance.

Finally, the last index is used to discard sporadic errors, by considering the distance which explains the 95 percentile of DM, and it was called D95. All distances are expressed in millimetres.

A composite index, GCQ, is defined which synthesises the various information brought by the above indices. GCQ is a combination of global quality indexes, GQ(O,Seg), built from four quantitative indices, for Hc and Am:

$$\begin{cases} \mathrm{GQ}(O,\mathrm{Seg}) = 100.(\mathrm{RV}(O^{\mathrm{Seg}},O^{\mathrm{Ref}}) + 1 - K_1(O^{\mathrm{Seg}},O^{\mathrm{Ref}}) + 2.\mathrm{MIV}(O^{\mathrm{Seg}},O^{\mathrm{Ref}})) + 5.\mathrm{DM}(O^{\mathrm{Seg}},O^{\mathrm{Ref}}) \\ CGQ(\mathrm{Seg}) = 3.\mathrm{GQ}(O_{\mathrm{Hc}},\mathrm{Seg}) + \mathrm{GQ}(O_{Am},\mathrm{Seg}) \end{cases}. \tag{40}$$

The optimal value for this global index is 0.

References

Ashton, E.A., Parker, K.J., Berg, M.J., Chen, C.W., 1997. A novel volumetric feature extraction technique with applications to MR images. IEEE Trans. Med. Imag. 16 (4), 365–371.

Ashton, E.A., Riek, J., Molinelli, L., Berg, M., Parker, K., 2003. A method for fully automated measurement of neurological structures in MRI. SPIE Medical Imaging-2003, vol. 5032, pp. 1125–1134.

Barra, V., Boire, J.-Y., 2001. Automatic segmentation of subcortical brain structures in MR images using information fusion. IEEE Trans. Med. Imag. 20 (7), 549–568.

Besag, J., 1989. Towards bayesian image analysis. J. Appl. Stat. 16 (3), 395–407.

Bloch, I., Colliot, O., Camara, O., Géraud, T., 2005. Fusion of spatial relationships for guiding recognition, example of brain structure recognition in 3D MRI. Pattern Recogn. Lett. 26, 449–457.

Bonilha, L., Kobayashi, E., Cendes, F., Li, L.M., 2004. Protocol for volumetric segmentation of medial temporal structures using high-resolution 3-D magnetic resonance imaging. Hum. Brain Mapp. 22, 145–154.

Carmichael, O.T., Aizenstein, H.A., Davis, S.W., Becker, J.T., Thompson, P.M., Cidid Meltzer, C., Liu, Y., 2005. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. NeuroImage 27 (4), 979–990.

Chupin, M., Hasboun, D., Mukuna Bantumbakulu, A.R., Bardinet, E., Baillet, S., Kinkingnéhun, S., Lemieux, L., Dubois, B., Garnero, L., 2006. Competitive segmentation of the hippocampus and the amygdala from MRI data: validation on young healthy controls and Alzheimer's disease patients. SPIE Medical Imaging-2006, p. 6144.

Chupin, M., Hasboun, D., Bardinet, E., Baillet, S., Lemieux, L., Garnero, L., in press. Segmentation compétitive de l'hippocampe et de l'amygdale à partir de volumes IRM. Trait. Signal.

Clark, K.A., Woods, R.P., Rottenberg, D.A., Toga, A.W., Mazziotta, J.C., 2006. Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. NeuroImage 29 (1), 185–202.

Cointepas, Y., Mangin, J.-F., Garnero, L., Poline, J.-B., and, H., 2001. BrainVISA: software platform for visualization and analysis of multi-modality brain data. HBM-2001, p. S98.

Crum, W.R., Scahill, R.I., Fox, N.C., 2001. Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease. NeuroImage 13, 847–855.

Dubes, R.C., Jain, A.K., Nabadar, S.G., Chen, C.C., 1990. MRF model-based algorithms for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 808–814.

Duchesne, S., Pruessner, J.C., Collins, D.L., 2002. Appearance-based segmentation of medial temporal lobe structures. NeuroImage 17 (2), 515–531.

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

Free, S.L., Bergin, P.S., Fish, D.R., Cook, M.J., Shorvon, S.D., Stevens, J.M., 1995. Methods for normalization of hippocampal volumes measured with MR. Am. J. Neuroradiol. 16, 637–643.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6 (6), 721–741.

Gerig, G., Jomier, M., Chakos, M., 2001. Valmet: a new validation tool for assessing and improving 3D image segmentation. MICCAI 2001, vol. 2208, pp. 516–523.

Ghanei, A., Soltanian-Zadeh, H., Windham, J.P., 1998. A 3D deformable surface model for segmentation of objects from volumetric data in medical images. Comput. Biol. Med. 28, 239–253.

Ghanei, A., Soltanian-Zadeh, H., Elisevich, K., Fessler, J.A., 2001. A knowledge-based deformable surface model with application to segmentation of brain structures in MRI. SPIE Medical Imaging-2001, vol. 4322, pp. 356–365.

Haller, J.W., Banerjee, A., Christensen, G.E., Gado, M., Joshi, S.C., Miller, M.I., Sheline, Y., Vannier, M.W., Csernansky, J.G., 1997. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. Radiology 202, 504–510.

Hasboun, D., Chantôme, M., Zouaoui, A., Sahel, M., Deladoeuille, M., Sourour, N., Duymes, M., Baulac, M., Marsault, C., Dormont, D., 1996. MR determination of hippocampal volume: comparison of three methods. Am. J. Neuroradiol. 17, 1091–1098.

Hogan, R.E., Mark, K.E., Wang, L., Joshi, S.C., Miller, M.I., Bucholz, R.D., 2000. Mesial temporal sclerosis and temporal lobe epilepsy: MR imaging deformation-based segmentation of the hippocampus in five patients. Radiology 216, 291–297.

Hsu, Y.Y., Schuff, N., Du, A.T., Mark, K., Zhu, X., Hardin, D., Weiner, M.W., 2002. Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. J. Magn. Reson. Imaging 16, 305–310.

Kelemen, A., Szekely, G., Gerig, G., 1999. Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Trans. Med. Imag. 18 (10), 828–839.

Klemenčič, J., Pluim, J.P.W., Viergever, M.A., Schnack, H.G., Valenčič, V., 2004. Non-rigid registration based active appearance models for 3D medical image segmentation. J. Imaging Sci. Technol. 48 (2), 166–171.

Li, T., Mirowitz, S.A., 2004. Fast multi-planar gradient echo MR imaging: impact of variation in pulse sequence parameters on image quality and artifacts. Magn. Reson. Imaging 22, 807–814.

Malandain, G., Bertrand, G., Ayache, N., 1993. Topological segmentation of discrete surfaces. Int. J. Comput. Vis. 10 (2), 183–197.

Mangin, J.-F., Tupin, F., Frouin, V., Bloch, I., Rougetet, R., Régis, J., López-Krahe, J., 1995. Deformable topological models for segmentation of 3D medical image. IPMI-1995, pp. 153–164.

Pantel, J., O'Leary, D.S., Cretsinger, K., Bockholt, H.J., Keefe, H., Magnotta, V.A., Andreasen, N.C., 2000. A new method for the in vivo volumetric measurement of the human hippocampous with high neuroanatomical accuracy. Hippocampus 10 (6), 752–758.

Pérez de Alejo, R., Ruiz-Cabello, J., Cortijo, M., Rodriguez, I., Echave, I., Regadera, J., Arrazola, J., Avilés, P., Barreiro, P., Gargallo, D., Grana, M., 2003. Computer-assisted enhanced volumetric segmentation magnetic resonance imaging data using a mixture of artificial neural networks. Magn. Reson. Imaging 21, 901–912.

Pitiot, A., Delingette, H., Thompson, P.M., Ayache, N., 2004. Expert knowledge-guided segmentation system for brain MRI. NeuroImage 23, S85–S96.

Poupon, F., Mangin, J.F., Hasboun, D., Poupon, C., Magnin, I., Frouin, V., 1998. Multi-object deformable templates dedicated to segmentation of brain deep structures. MICCAI-1998, vol. 1496, pp. 1134–1143.

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

Rivière, D., Papadopoulos, D., Poupon, C., Poupon, F., Coulon, O., Poline, J.-B., Frouin, V., Régis, J., Mangin, J.-F., 2000. A structural browser for human brain mapping. HBM-2000, vol. 11(5), p. 912.

Shen, D., Moffat, S., Resnick, S.M., Davatzikos, C., 2002. Measuring size and shape of the hippocampus in MR images using a deformable shape model. NeuroImage 15 (2), 422–434.

Siadat, M.R., Soltanian-Zadeh, H., Fotouhi, F., Elisevich, K., 2004. Bayesian landmark identification in medical images. SPIE Medical Imaging-2004, vol. 5370, pp. 628–639.

Wieshmann, U.C., Free, S.L., Stevens, J.M., Shorvon, S.D., 1998. Image contrast and hippocampal volumetric measurements. Magn. Reson. Imaging 16 (1), 13–172.

Yang, J., Duncan, J.S., 2004. 3D image segmentation of deformable objects with joint shape-intensity prior models using level sets. Med. Image Anal. 8, 285–294.

Yang, J., Staib, L.H., Duncan, J.S., 2004. Neighbor-constrained segmentation with 3D deformable models. IEEE Trans. Med. Imag. 23 (8), 940–948.

Zhou, J., Rajapakse, J., 2005. Segmentation of subcortical brain structures using fuzzy templates. NeuroImage 28 (4), 915–924.