Computer Vision and Image Understanding 116 (2012) 146–165

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations

Geoffroy Fouquier [a,*], Jamal Atif [b], Isabelle Bloch [a]

[a] Institut Telecom, Telecom ParisTech, CNRS LTCI, 46 rue Barrault, 75013 Paris, France
[b] TAO INRIA, CNRS, LRI – Paris-Sud University, 91405 Orsay Cedex, France

ARTICLE INFO

Article history:
Received 13 October 2010
Accepted 14 September 2011
Available online 2 October 2011

Keywords:
Segmentation
Knowledge-based system
Spatial relations
Graph representations
Fuzzy sets
Medical images
MRI

ABSTRACT

A sequential segmentation framework, where objects in an image are successively segmented, generally raises some questions about the "best" segmentation sequence to follow and/or how to avoid error propagation. In this work, we propose original approaches to answer these questions in the case where the objects to segment are represented by a model describing the spatial relations between objects. The process is guided by a criterion derived from visual attention, and more precisely from a saliency map, along with some spatial information to focus the attention. This criterion is used to optimize the segmentation sequence. Spatial knowledge is also used to ensure the consistency of the results and to allow backtracking on the segmentation order if needed. The proposed approach was applied for the segmentation of internal brain structures in magnetic resonance images. The results show the relevance of the optimization criteria and the interest of the backtracking procedure to guarantee good and consistent results.

1. Introduction

In this paper, we deal with segmentation and recognition of objects or structures in an image, based on a generic model of the scene. As a typical example, we focus on the recognition of internal brain structures in 3D magnetic resonance images (MRI), based on an anatomical model. More specifically, we address two important problems occurring in sequential approaches, as detailed below.

In Refs. [1,2], the authors introduced a new paradigm combining segmentation and recognition tasks. We will refer to this paradigm in the remainder of this paper as sequential segmentation and interpretation. It is defined as a knowledge-based object recognition approach where objects are segmented in a predefined order, starting from the simplest object to segment to the most difficult one. The segmentation and recognition of each object are then based on a generic model of the scene and rely on the previously recognized objects. This approach uses a graph which models the generic spatial information about the scene in an intuitive and explicit way (presented in [3]). This sequential segmentation framework allows decomposing the initial problem into several sub-problems easier to solve, using the generic knowledge about the scene. This approach differs from a regular divide-and-conquer approach since each sub-problem contributes to improve the resolution of the next sub-problems. It also avoids relying on an initial segmentation of the whole image.

This approach, as pointed out in Ref. [2], requires to define the order according to which the objects have to be recognized and the choice of the most appropriate order is one of the problems that remain open. It also lacks a step which could evaluate the quality of the segmentation of a particular object and detect errors to prevent their propagation.

In this paper, we propose original methods to answer these two open questions. Our contribution is twofold: first, we extend the sequential segmentation framework by introducing a pre-attentional mechanism, which is used, in combination with spatial relations, to derive a criterion for the optimization of the segmentation order. Secondly, we introduce criteria and a data structure which allow us to detect the potential errors and control the ordering strategy.

The pre-attentional mechanisms were defined in [4–6] to guide the focus of attention in modeling the visual system such as in feature integration theory. The sequential segmentation framework may be viewed as a way to focus attention on a small part of the scene and thus limit the search domain and the computational load. Among these mechanisms, we propose to use the notion of saliency to optimize the sequence of segmentation.

Our approach is applied to the segmentation and the recognition of internal brain structures in 3D magnetic resonance images. The intrinsic variability of these structures, the lack of clear boundaries and the insufficient radiometry make this

* Corresponding author.
E-mail addresses: geoffroy.fouquier@telecom-paristech.fr (G. Fouquier), atif@lri.fr (J. Atif), isabelle.bloch@telecom-paristech.fr (I. Bloch).

1077-3142/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2011.09.004

segmentation problem a difficult one. Some of the difficulties can be overcome by relying on generic knowledge about the human anatomy, that will be exploited to derive the model guiding the whole process.

This article is organized as follows. First we present in Section 2 a survey of knowledge based-approaches to the recognition of objects in a scene and provide an overview of the proposed approach. Section 3 presents the knowledge representation model. In Section 4 we propose to use some concepts of the visual attention to optimize the sequential segmentation framework. Then, the optimization of the sequential segmentation itself is described in Section 5 and the mechanisms for evaluating each structure segmentation in Section 6. Experiments on internal brain structure segmentation and results are presented in Section 7. Finally we draw some conclusions in Section 8.

2. Knowledge-based systems and spatial reasoning

The sequential segmentation framework of Colliot et al. [2] relies on a priori knowledge about the scene and uses intensively this knowledge at each step of the process. Thus, this framework may be described as a knowledge-based system using spatial relations. One can find a review of these systems in Refs. [7,8]. In this section, we focus on knowledge-based systems using spatial relations to describe the structure of the scene that have been applied to the recognition of brain structures in medical images.

Spatial relations play a crucial role in model-based image recognition and interpretation due to their stability compared to many other image appearance characteristics. They constitute structural information, which is particularly relevant when the intrinsic features of the objects are not sufficient to discriminate them.

2.1. Knowledge-based approaches for internal brain structures recognition

The difficulty of segmenting internal brain structures is due to the similarity between their gray levels, the lack of clear boundaries at some places and the partial volume effect. Their intrinsic features present a natural variability (in size and shape for example) between individuals, which is further increased in pathological cases. On the contrary, the spatial arrangement of these structures, i.e., their relative positions, is stable in healthy cases and even quite stable in pathological cases. For all these reasons, structural models of the internal brain structures have been used to segment and recognize the internal structures.

2.1.1. Structural model of the brain structures

One can find several anatomical descriptions of the brain, as atlas [9], nomenclature [10] or ontology [11]. These descriptions are often organized as a hierarchy of structures and provide descriptions of structures and relations between them. In Ref. [3], in collaboration with a neuro-anatomist, the internal brain structures are represented as a hierarchical graph where each vertex corresponds to an anatomical structure and each edge carries spatial relations between anatomical structures. This representation has been extended as the GRAFIP[1] [12] to include information about the structures composition, functional knowledge and about the pathologies.

2.1.2. Segmentation and recognition

Several classes of approaches for internal brain structures segmentation have been proposed in the literature. The first class of approaches uses a model graph and the image to segment is represented as a graph too. The segmentation and recognition process is then formalized as a graph matching problem [13]. The authors in Refs. [14,15] proposed to find a fuzzy morphism between a model graph built from a manual segmentation and an over-segmented image represented as a graph. Several optimization techniques have been proposed for this task [16,17]. Another approach was proposed in Ref. [18] and used an over-segmentation. The matching is viewed as a constraint satisfaction problem, with two levels of constraints and an ad-hoc algorithm. The authors recently extended their approach to cope with unexpected structures, such as tumors [19]. For this class of approaches, the initial graph is usually built from an over-segmentation of the image to segment, and the complexity of the method increases as the number of regions obtained from the over-segmentation grows.

In the second class of approaches, a sequential segmentation of the internal brain structures is performed, as proposed in Refs. [1,2]. In these approaches, the segmentation and the recognition are achieved at the same time. Each segmentation uses the spatial information encoded in the model, and more specifically the spatial relations to the already segmented structures. This information allows restricting the search domain around the structure. In these approaches, there is no initial segmentation of the image, but it raises some questions like the order of segmentation of the different objects or how to avoid the propagation of potential errors. Our approach belongs to this class and our contribution is an original answer to both questions.

The authors in Refs. [20,21] proposed a different type of approach, which is global and uses a constraint network. They proposed to link each anatomical structure with a region of space which satisfies all constraints in the network. Since it is hard to solve this problem directly, only the bounds of the domain of each variable (i.e. structure to be segmented) are modified by the process and sequentially reduced using specifically designed propagators derived from the spatial constraints. Finally, a segmentation is extracted using a minimal surface algorithm. This approach provides good results and does not need an initial segmentation either. However, due to the number of constraints, it is quite complex and the computation time is high, especially in 3D.

2.2. Proposed framework

We propose to extend the sequential segmentation framework proposed in Ref. [2], where structures are sequentially segmented from the easiest to segment to the most difficult ones. Each structure segmentation uses the information provided by the previous segmentations. Our extension aims at answering the following questions raised by this framework: "in which order should the objects of the scene be segmented?" and "how to assess the segmentation result in order to detect potential errors and avoid their propagation?".

The proposed framework has two levels, as depicted in Fig. 1. The first level is a generic bottom-up module which allows selecting the next structure to segment. This level does not rely on an initial segmentation or classification, but instead on a focus of attention and a map of generic features described in Section 4. The sequential approach allows this level to use two types of knowledge: generic and domain independent features in unexplored area of the image to segment, and high-level knowledge such as spatial relations linked to the already recognized structures. We propose to answer the first question by deriving a selection criterion from a pre-attentional mechanism: a saliency map. This criterion is used to optimize the segmentation order and to select the next structure to segment at each step.

The second level achieves recognition and segmentation of the selected structure, as well as the evaluation of the segmentation. The recognition of the structure is achieved at the same time as

[1] For "Graph of Representation of Anatomical and Functional data for Individual patients including Pathologies".

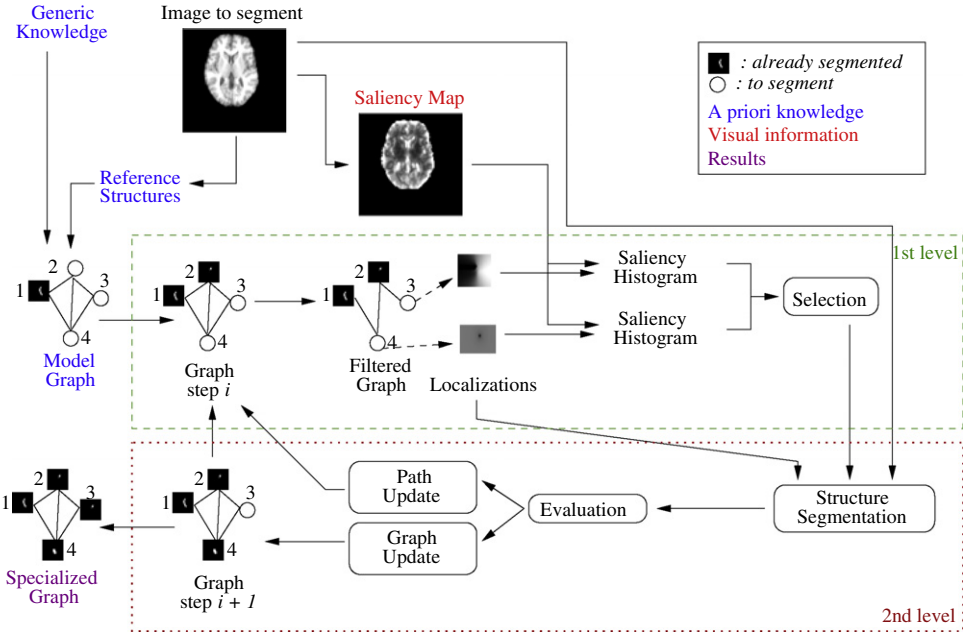


Fig. 1. General scheme of the sequential segmentation framework. The graph initially represents only the generic knowledge and the reference structures. At each step, a structure is selected according to the saliency of its localization and to the presented criterion. This structure is then segmented and the result is evaluated. In case of success, the graph is updated and the process is iterated until the graph is completely specialized or no more structure can be segmented. In case of failure, the system is constrained to select another path to segment and the process is iterated.

the segmentation. This level is composed by the segmentation method defined in Ref. [2] and an original evaluation method. It uses two types of a priori information: the spatial information which allows us to reduce the search area, and a radiometric estimation of the intensity of the structure. Therefore, the radiometric estimation needs to discriminate the intensity of the structure only in the search area and not in the whole image. Once a structure is segmented and recognized, this level also evaluates the quality of the result and proposes a strategy to guarantee the spatial consistency of the result and to potentially backtrack on the segmentation order. This allows answering the second question.

The two levels rely on graph representations described in the next section.

3. Knowledge representation

Graphs are well adapted to represent generic knowledge, such as spatial relations between the objects of a scene. In the sequential segmentation framework, the generic model of the scene is modeled as a graph where each vertex represents an object and each edge represents one or more spatial relations between two objects. We introduce the following notations: Let Σ_V,Σ_E be the sets of vertex labels and edge labels, respectively. Let V be a finite nonempty set of vertices, L_v be a vertex interpreter $L_v{:}V \to \Sigma_V$, E be a set of ordered pairs of vertices called edges, and L_e be an edge interpreter $L_e{:}E \to \Sigma_E$. Then $G = (V,L_v,E,L_e)$ is a labeled graph with directed edges. For $v \in V$ and $e \in V \times V$, $\delta(v,e)$ is a transition function that returns the vertex v' such that $e = (v,v')$. For $v \in V$, $A(v)$ returns the set of edges adjacent to v. Finally, $p = (v_1, v_2,\ldots,v_n)$ is a path of length n labeled as $l_p = (v_1, e(v_1,v_2), v_2,\ldots,v_n)$.

A knowledge base KB defines all the spatial relations existing between vertices in the graph:

$$KB = \{v_i R v_j, v_i, v_j \in V, R \in \mathcal{R}\} \text{ and}$$
$$e = (v_1, v_2) \in E \iff \exists R \in \mathcal{R}, (v_1 R v_2) \in KB,$$

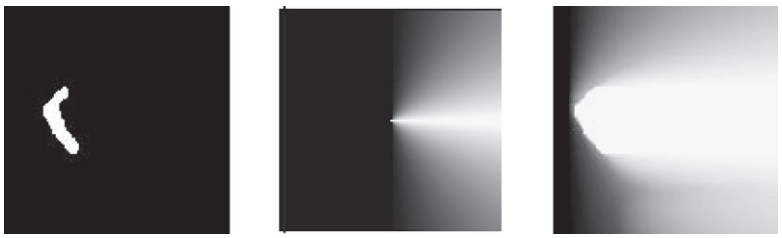
where $\mathcal{R}$ is the set of relations. In the following, we use fuzzy representations of the spatial relations, since they are appropriate to model the intrinsic imprecision of several relations (such as "close to" and "behind"), their potential variability (even if it is reduced in normal cases) and the necessary flexibility for spatial reasoning [22]. Here, the representation of a spatial relation is computed as the region of space in which the relation R to an object A is satisfied. The membership degree of each point corresponds to the satisfaction degree of the relation at this point. Fig. 2 presents an example of a structure and the region of space corresponding to the region "to the right of" this structure.

A directed edge between two vertices v_1 and v_2 carries at least one spatial relation between these objects. An edge interpretor associates to each edge a fuzzy set μ_{Rel}, defined in the spatial domain $\mathcal{S}$, representing the conjunctive merging of all the representations of the spatial relations carried by this edge to a reference structure. Each fuzzy set gives an estimation of the localization of an object. By localization, we mean an approximate region containing the object. A conjunction of all these fuzzy sets gives the most precise estimation of the localization. Since there is at least one spatial relation carried by an edge, μ_{Rel} cannot be empty. Let $\mu^e_{R_i}, i = 1,\ldots,n_e$ the n_e relations carried by an edge e. Then μ^e_{Rel} is expressed as: $\mu^e_{Rel} = \tau_{i=1..n_e}\left(\mu^e_{R_i}\right)$ with τ a t-norm (fuzzy conjunction) [23].

We now briefly describe the modeling of the main relations that we use: distances and directional relative positions. More details can be found in Ref. [22]:

A **distance** relation can be defined as a fuzzy interval f of trapezoidal shape on $\mathbb{R}^+$. A fuzzy subset μ_d of the image space $\mathcal{S}$ can then be derived by combining f with a distance map d_A to the reference object $A: \forall x \in \mathcal{S}, \mu_d(x) = f(d_A(x))$, where $d_A(x) = \inf_{y \in A} d(x,y)$.

The relation **"close to"** can be defined as a function of the distance between two sets: $\mu_{close}(A,B) = h(d(A,B))$ where $d(A,B)$ denotes the minimal distance between points of A and B: $d(A,B) = \inf_{x \in A, y \in B} d(x,y)$, and h is a decreasing function of d, from $\mathbb{R}^+$ into $[0,1]$. We assume that $A \cap B = \emptyset$. The relation of **adjacency** can be defined likewise as a "very close to" relation, leading to a degree of adjacency instead of a Boolean value, making it more robust to small errors.



(a) Lateral Ventricle (b) Structuring element (c) "Left of LVl"

Fig. 2. (a) Binary segmentation of a left lateral ventricle (a slice of a 3D volume) denoted by *LVl*. (b) Structuring element representing the semantic of the spatial relation "left of". (c) Fuzzy landscape representing the spatial relation "left of *LVl*" (note that the usual convention in medical imaging "left is right" is used here, and anatomically left means right on the displayed image).

Directional relations are represented using the "fuzzy landscape approach" [24]. A morphological dilation δ_{ν_α} by a fuzzy structuring element ν_α representing the semantics of the relation "in direction α" is applied to the reference object $A : \mu_\alpha = \delta_{\nu_\alpha}(A)$, where ν_α is defined, for x in $\mathcal{S}$ given in polar coordinates (ρ, θ), as: $\nu_\alpha(x) = g(|\theta - \alpha|)$, where g is a decreasing function from $[0, \pi]$ to $[0, 1]$, and $|\theta - \alpha|$ is defined modulo π. This definition extends to 3D by using two angles to define a direction. Fig. 2 presents an example of fuzzy landscape representing a directional relation.

Other relations can be modeled in a similar way [22]. These models are generic, but the membership functions depend on a few parameters that have to be tuned for each application domain according to the semantics of the relations in that domain. Here we propose to learn these parameters from a database of segmented images.

3.1. Images database

A database of 44 brain MRI, manually segmented, is used. This database is composed by 30 healthy images and 14 images presenting a brain tumor (with different localizations, types and sizes). The set of healthy images is composed by the IBSR database[2] and some images from the OASIS database ("Open Access Series of Imaging Studies").[3] Manual segmentations are available for the IBSR database. All other images have been manually segmented and tumor segmentations have been validated by experts. These segmentations are used for learning the parameters of the relations, and to evaluate the results.

3.2. Learning of spatial relations

The modeled spatial relations are based on fuzzy intervals that are chosen of trapezoidal shape for the sake of simplicity. They define the functions f and g introduced above. The parameters of the fuzzy intervals are learned for each triplet (A, R, B) where A and B are two objects and R a spatial relation. The learning procedure [25] basically consists in enlarging the kernel and the support of the spatial relation in a way that all the targeted structures are included in this support. Fig. 3 illustrates the effect of the learning on the fuzzy interval. For example, let us consider the relation "the putamen is on the left of the caudate nucleus". The objective of the learning procedure is to ensure that the putamen is localized in the support of the relation "on the left of the caudate nucleus".

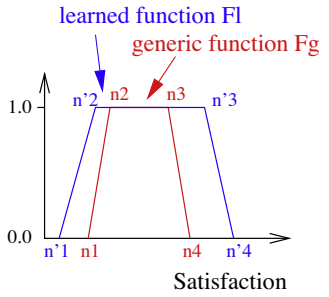


Fig. 3. Fuzzy intervals of trapezoidal shape. The learning procedure consists in defining the parameters $n_1, \ldots, n_4$ in a way that the targeted function is included in the kernel of the representation of the function. A relation R can be defined in a generic way (red interval) and then specified for two structures a and b to represent the relation aRb (blue interval).

The learning procedure consists of three steps:

- For each image of the learning database, the relation ("on the left of the caudate nucleus" in our example) is represented with a generic function F_g, i.e. with generic values for the relation "left of". Fig. 4b shows an example of a fuzzy subset obtained with such values.
- For each resulting fuzzy subset, we compute the satisfaction values at each point of the targeted structure and extremal values (minimum and maximum) are kept. If the targeted structure is included in the kernel of the relation, the satisfaction value at each point is 1.00. In our example in Fig. 4b, the putamen is not completely included in the kernel and the minimum of satisfaction is 0.37 (the maximum is 1.00).
- The mean $m_{\min}$ and standard deviation $\sigma_{\min}$ of the minimum values (respectively $m_{\max}$ and $\sigma_{\max}$ for the maximum values) are computed and a new function Fl is defined with the following parameters:

$$\begin{aligned} \text{n1} &= m_{\min} - \sigma_{\min} & \text{n3} &= m_{\max} \\ \text{n2} &= m_{\min} & \text{n4} &= m_{\max} + \sigma_{\max} \end{aligned}$$

An example of this function is given in Fig. 4c and the fuzzy subset using this function is displayed in Fig. 4d. This subset presents a larger kernel in this example.

3.3. Localization of a structure

We define the localization of a structure as the conjunctive merging of all spatial relations targeting a structure. This corresponds to a region of interest defined by the constraints on a structure. The learning step ensures that an object is localized in the

[2] Internet Brain Segmentation Repository. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/.

[3] http://www.oasis-brains.org, built thanks to Pubmed Central submissions: P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

support of all spatial relations targeting this object. Therefore, each spatial relation representation provides a rough localization which is larger than the target object and includes it. Then a conjunction of all spatial relations targeting an object allows us to get a more precise localization. Fig. 5 presents the graph used in our experiments and an example of localization.

4. Visual attention to optimize a sequence of segmentation

Visual attention is often referred to as a "spotlight" on the visual field, i.e., at a given moment, the visual attention is restricted to a spatial area (or a number of visual objects). The exploration of the visual field is thus sequential. The sequential segmentation framework may be viewed as the progressive exploration of a scene where the "spotlight" of the visual attention corresponds to the consecutive segmentation of objects of the scene.

Visual attention was first modeled as two sequential steps: the attentional step itself and a pre-attentional step dedicated to guide the "spotlight" of visual attention by selecting the area of space to visit. The relations between these two steps are in fact more complex and both steps are intertwined.

The pre-attentional mechanisms were introduced in Refs. [4–6,26] as bottom-up mechanisms, computed on the whole scene and using specific features computed simultaneously. The pre-attentional mechanisms guide the attentional step by selecting "salient" areas or objects, i.e., regions which have a quality that thrusts itself into attention. Pre-attentive features are characterized by the "pop-out" effect, i.e., the detection is fast and not

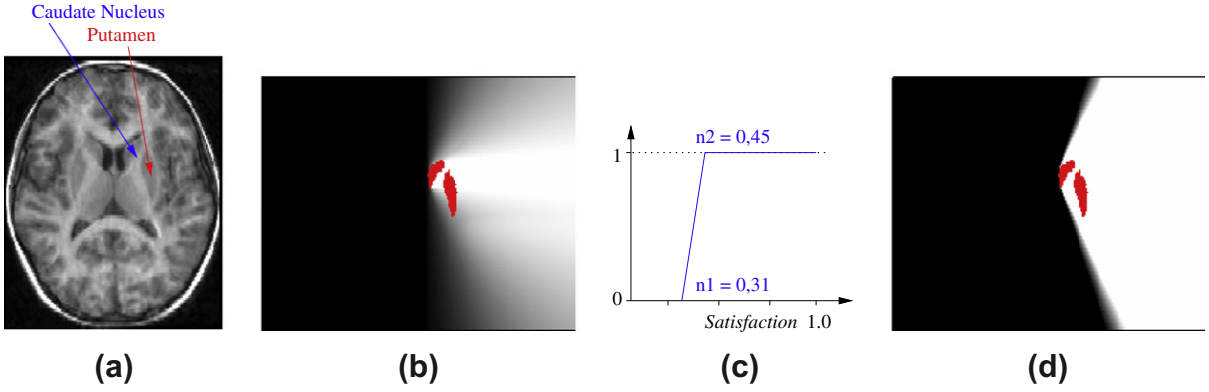


Fig. 4. Learning the parameters of the trapezoidal fuzzy set which represents the relation "the caudate nucleus is on the right of the putamen". (b) For all images, the fuzzy set representing the relation is computed with default parameters. (c) Extremal values of satisfaction at the location of the putamen are used to compute the parameters of the fuzzy numbers. (d) The relation may be computed with the new set of parameters which shows a larger kernel in this example.

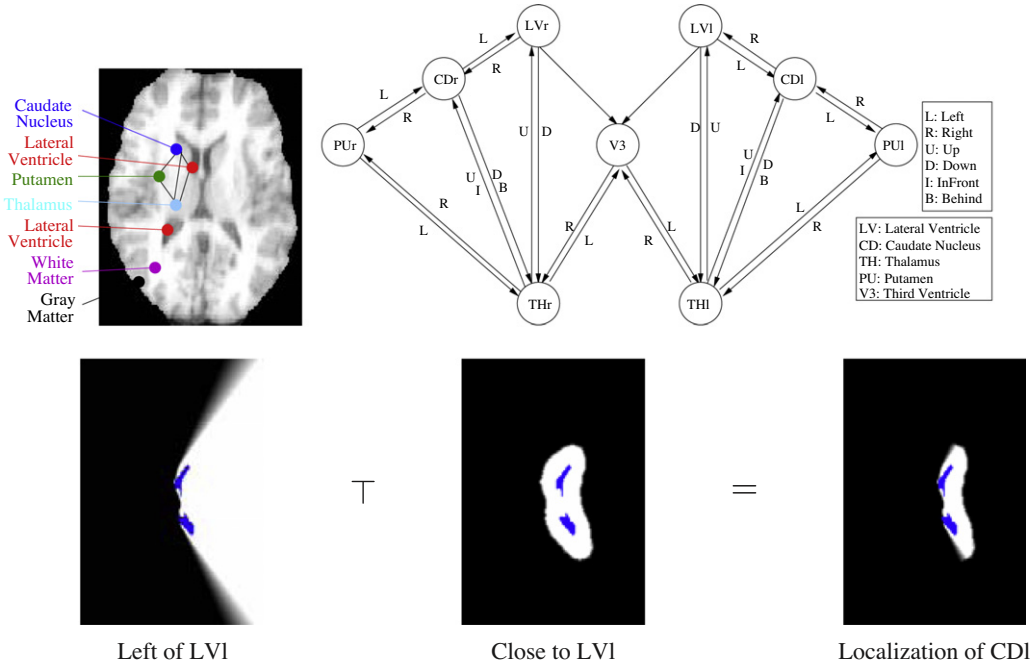


Fig. 5. The graph used in our experiments. Vertices represent anatomical structures and edges represent spatial relations. Only directional relations have been displayed on this graph, but each edge carries other relations as well. Below, the representation of two spatial relations carried by the edge between the lateral ventricle (displayed in blue) and the caudate nucleus and the resulting localization of the caudate nucleus.

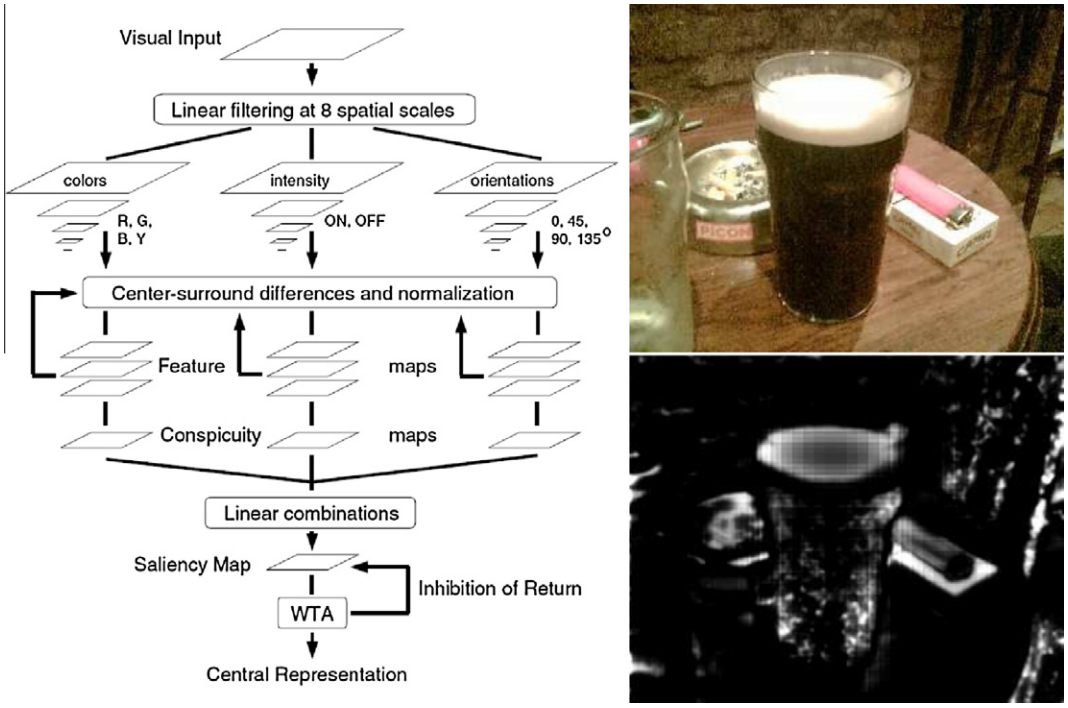


Fig. 6. The generation of a saliency map proposed in Ref. [28] on the left. An image and the corresponding saliency map on the right. The high saliency values correspond to regions with a high contrast with respect to their environment and/or geometrical structures.

correlated to the number of objects in the scene. A description and examples are presented in Ref. [27].

4.1. Saliency and saliency maps

Among the pre-attentional mechanisms, we focus on the saliency map, as defined by Itti and Koch [28,29] for 2D images. This mechanism uses three different types of pre-attentional features: opposition of colors (red[4] vs. green, blue vs. yellow), intensity and orientation (a Gabor filter with four different orientations). For each feature the original image is filtered and a Gaussian pyramid is built from the filtered image. Basically, the way of considering each feature is to look at discontinuities within each pyramid by comparing "fine" scale and "coarse" scale. A fine scale is a scale close to the original image. Each comparison generates a "feature map" reflecting discontinuities for a specific feature and with a certain scale factor. All feature maps issued from the same pyramid are merged after normalization into a conspicuity map (one per each type of features, so three maps here). Finally a weighted mean of conspicuity maps produces the saliency maps.

The full process is described in Ref. [28] and illustrated in Fig. 6. We describe now the different steps and the required adaptation to compute saliency maps on 3D brain MRI.

4.1.1. Pre-processing: brain extraction

Our application focuses on recognition of internal brain structures. Therefore only the brain is needed in the image. The skull, the eyes and other parts may be discarded. Thus, the brain is first extracted from the 3D volume using the method proposed in Ref. [30]. This allows us to reduce the search domain so as to consider only the most relevant information for our task.

[4] For interpretation of color in Figs. 1, 3–6, 9–14, 16–24, the reader is referred to the web version of this article.

4.1.2. Pre-processing: resampling

For each feature, a multi-scale analysis is performed. Since the original resolution of 3D MRI is often anisotropic, a resampling to a volume of 256 cubic voxels allows us to compute saliency maps on a volume with a fixed size and an isotropic resolution (the choice of 256 voxels is guided by the most frequent size of the images in our database described in Section 3). The chosen interpolation method is a spline resample interpolation [31], available for 3D MRI in Brainvisa.[5]

4.1.3. Features and filtering

The original method uses three different types of features: intensity, oppositions of colors and orientations. There is no color in MRI. The intensity feature is the same as in the original method.

For the orientation, a 3D Gabor filter is used as described in [32,33]. The bandwidth parameter is fixed to $B = 0.55$ in our experiments. We use the following orientations (angles θ and ϕ in spherical coordinates):

$\theta\backslash\phi$	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$
0	×							
$\frac{\pi}{4}$	×	×	×	×	×	×	×	×
$\frac{\pi}{2}$	×	×	×	×				

Each filter is symmetric thus only a half sphere is sampled. The number of orientations is limited in order to reduce memory usage and computation time.

4.1.4. Pyramids generation

A dyadic pyramid is built from each filtered image (1 for intensity and 13 orientations, so 14 pyramids). In the original method, a

[5] http://www.brainvisa.info.

Gaussian pyramid is built with 8 levels, but here, due to the size of resampled brain MRI (256), we limit our pyramid to 5 scales.

For intensity, we build a Gaussian pyramid where the initial level $i = 0$ is the original image. At each level, the size of the image remains the same, but the width of the Gaussian Filter (σ) is adapted: $\sigma_i = i + 0.5$.

For the orientations, instead of a Gaussian pyramid, we can take advantage of the parameter of the Gabor filter, to directly produce a pyramid where each level corresponds to a different filtering by a Gabor filter in the same orientation. At each level, we adapt the frequency of the Gabor filter, starting at 0.4 and adding 0.05 at each level. Each resulting image is then smoothed with a Gaussian filter ($\sigma = 0.5$) to remove noise.

4.1.5. Feature maps

Feature maps are computed between "fine" scales and "coarse" scales of a pyramid. The fine scales used to compute maps are 1 and 2. The coarse scales are the fine scales plus a step $\delta \in \{1,2\}$, i.e., 1 + 1,1 + 2,2 + 1,2 + 2. A feature map is a point-to-point difference between both scales which, in this approach, have the same size. Pyramids and feature maps are illustrated in Fig. 7 for both intensity and orientations.

4.1.6. Normalization

There are 14 pyramids with four feature maps each. The normalization step is therefore very important. The normalization operator $\mathcal{N}$ we use is the same operator as in the original method [34]. This operator is designed to promote maps where there are few high peaks, rather than maps where there is a large number of peaks but with the same values.[6] The normalization is achieved in three steps:

- normalize the map in an interval $[0,M]$ with a fixed M to remove features specific dynamics,
- compute the average value $\hat{m}$ of all local maxima lower than M,
- multiply each point of the map by $(M - \hat{m})^2$.

4.1.7. Merging

All feature maps belonging to a same pyramid are merged and produce a "conspicuity map". All the conspicuity maps belonging to a same feature are also merged in order to produce a unique conspicuity map per feature type (intensity and orientation in our case).

For intensity, only one pyramid is built. The conspicuity map is generated as:

$$C_{int} = \oplus\{\mathcal{N}(I_{ce} \ominus I_{co}), ce \in \{1,2\}, co = ce + \delta, \delta \in \{1,2\}\},$$

with $\oplus$ a point-to-point addition and $\ominus$ a point-to-point difference.

For orientations, an intermediary map is generated for each pyramid. All these maps are then normalized and merged in the same fashion:

$$C_{\theta,\phi} = \oplus\left\{\mathcal{N}\left(I^{\theta,\phi}_{ce} \ominus I^{\theta,\phi}_{co}\right), ce \in \{1,2\}, co = ce + \delta, \delta \in \{1,2\}\right\}$$
$$C_{orient} = \sum_{\theta,\phi} \mathcal{N}(C_{\theta,\phi}).$$

The saliency map is then generated as a weighted mean of conspicuity maps:

$$\text{Saliency Map} = \frac{\mathcal{N}(C_{int}) + \mathcal{N}(C_{orient})}{2}$$

Fig. 8 presents some examples of saliency maps generated from brain MRI.

[6] The normalization achieved by the operator $\mathcal{N}$ is not a normalization in the common sense.

4.2. Using focus of saliency maps as a region feature

In a sequential segmentation framework, a usual question is the order of the successive segmentations. The saliency map is a bottom-up pre-attentional mechanism designed to guide the attentional step. Therefore, considering a parallel between the attentional step and the segmentation step in sequential segmentation, we propose to use a pre-attentional mechanism to guide the segmentation process, i.e. define the best sequence of segmentation.

Thanks to the spatial information contained in the graph, we are able to compute the localizations of all structures connected to a previously segmented structure, as described in Section 3. The selection of the next structure to segment is achieved by comparing the saliency at each localization of candidate structures (these localizations may overlap each other). The histogram of the saliency map corresponding to the localization is generated. Thus, the comparison between localizations is a comparison between histograms of saliency of each region.

The computation of the saliency map described above shows that the saliency information is based on discontinuities for several pre-attentive features. Hence, since usually a structure is easier to segment if its border is well defined, we can assume that a structure is easier to segment than another one if its localization is more salient.

4.2.1. Comparison of localizations

To compare two histograms (previously normalized), we choose the Earth mover's distance (EMD) [35]. This measure gives the transportation cost between two distributions. If p and q are two probability distribution functions and N the number of bins, then the EMD measure is defined as:

$$emd(p,q) = \min_{\alpha_{i,j} \in \mathcal{M}} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{ij} c(i,j),$$

where $\mathcal{M} = \{(\alpha_{i,j}); \alpha_{i,j} \geqslant 0, \sum_j \alpha_{i,j} = p[i], \sum_i \alpha_{i,j} = q[j]\}$ and $c(.,.)$ is a distance between bins. For non-circular 1D histogram, if $c(i,j) = \frac{|i-j|}{N}$, then the EMD measure may be computed as the difference between corresponding cumulative histograms [36]:

$$emd(p,q) = \frac{\sum_{i=1}^{N} |P[i] - Q[i]|}{N}, \quad (1)$$

where p and q are two probability distributions, P and Q the two corresponding cumulative histograms and N the number of bins. The computation is then direct and very simple in this case.

In order to define an order between two distributions, we compute the following criterion:

$$s(p,q) = \sum_{i=1}^{N} P[i] - Q[i],$$

with the same notations as before. A signed EMD measure, denoted EMDS, is then defined as:

$$emds(p,q) = \begin{cases} emd(p,q) & \text{if } s(p,q) < 0, \\ -emd(p,q) & \text{if } s(p,q) \geqslant 0. \end{cases} \quad (2)$$

Fig. 9 (a) presents cumulative histograms for the localization of the caudate nucleus (CDl) and the thalamus (THl). The EMD measure between these distributions is $emd(CDl,THl) = 0.0084$. With the EMDS measure, we are able to determine which distribution presents the most salient value: $emds(CDl,THl) = 0.0084$ and $emds(THl,CDl) = -0.0084$. In this example, the localization of the caudate nucleus is preferred (but both distributions are very close to each other).

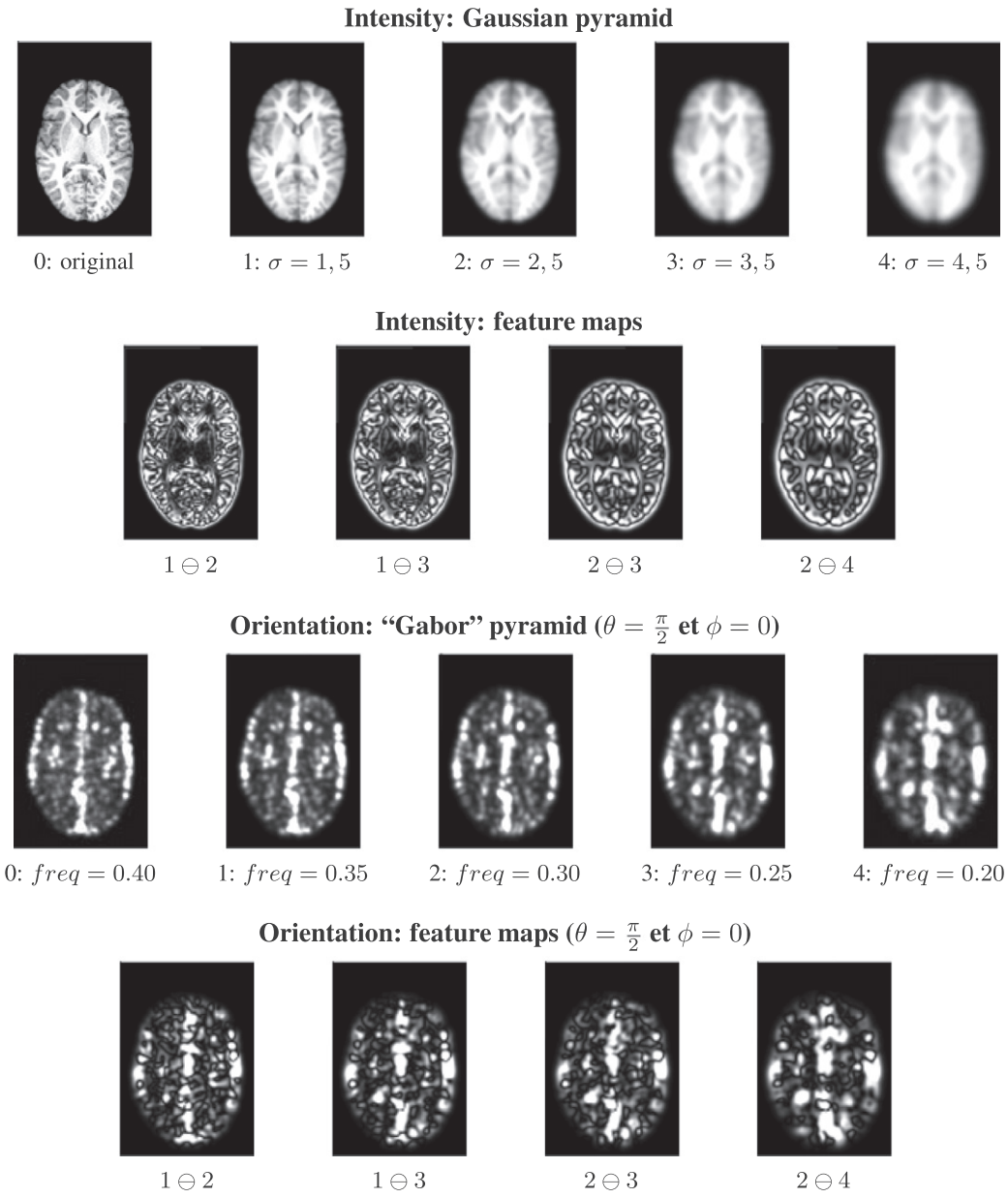


Fig. 7. Generation of a saliency map. A slice of a 3D MRI is presented at the top left of the figure, and corresponds to the initial level of the Gaussian pyramid (on top) generated for intensity features. Second line: corresponding feature maps, computed by applying the center-surround operator ($\ominus$) between fine and coarse scales of the pyramid. Third line: "Gabor" pyramid in a selected orientation and corresponding feature maps (last line) for the same slice on the brain MRI.

4.2.2. A saliency-based criterion

A criterion is defined to compare localizations and to select one of them. It aims at selecting the localization which presents the highest saliency distribution, as defined above. From experimental observations, it appears that the precision of the localization (defined below) must be taken into account too. We thus define a criterion that merges both aspects, in order to privilege regions that are more focused, i.e. with a more restricted support.

The way the localizations are generated ensures that the candidate structure is included in the localization (considering that the previous segmentations are consistent with the model). However, the support of the localization may be large (i.e. including several other objects), for example, if the only spatial relation available to define the localization of a structure is a directional relation. Therefore, the more a localization includes other parts of the image, the less the saliency of this localization provides relevant information about the targeted structure. Thus the precision aims at estimating how much the support is restricted.

Since it is difficult to estimate the precision of a given localization (before segmentation and without a priori information about the structure volume), another measure is used to evaluate the saliency of a localization based on a comparison with a learned saliency distribution mod_o, where the subscript stands for the targeted structure o. This distribution is computed on the same database as before and corresponds to the average of all the distributions obtained for the segmentations of the structure o in the database. The mean $\widehat{mod_o}$ and the standard deviation σ_{mod_o} are also computed in order to center and reduce the measure. The distance between the saliency of the localization (loc_o) and the learned saliency (mod_o) for an object o is estimated with a regular EMD as:

$$d_o(loc_o, mod_o) = \frac{emd(loc_o, mod_o) - \widehat{mod_o}}{\sigma_{mod_o}}. \quad (3)$$

Fig. 9 (b and c) shows the distribution computed from the localizations and the learned distribution for the left caudate nucleus (CDl)

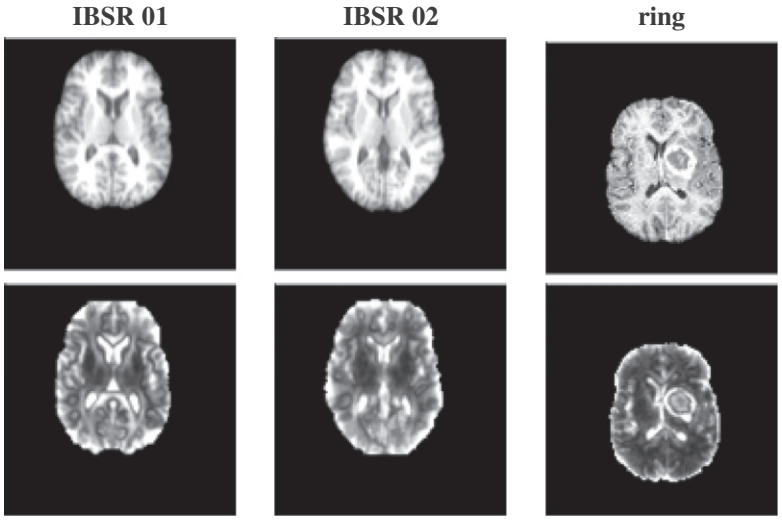


Fig. 8. Some images and their corresponding saliency maps. All computations are done in 3D but only one slice is presented here. Parts of the brain MRI presenting a high contrast (like the lateral ventricles, dark structures in the center) present high saliency values in the corresponding image. The tumor in the example on the right also presents high saliency values.

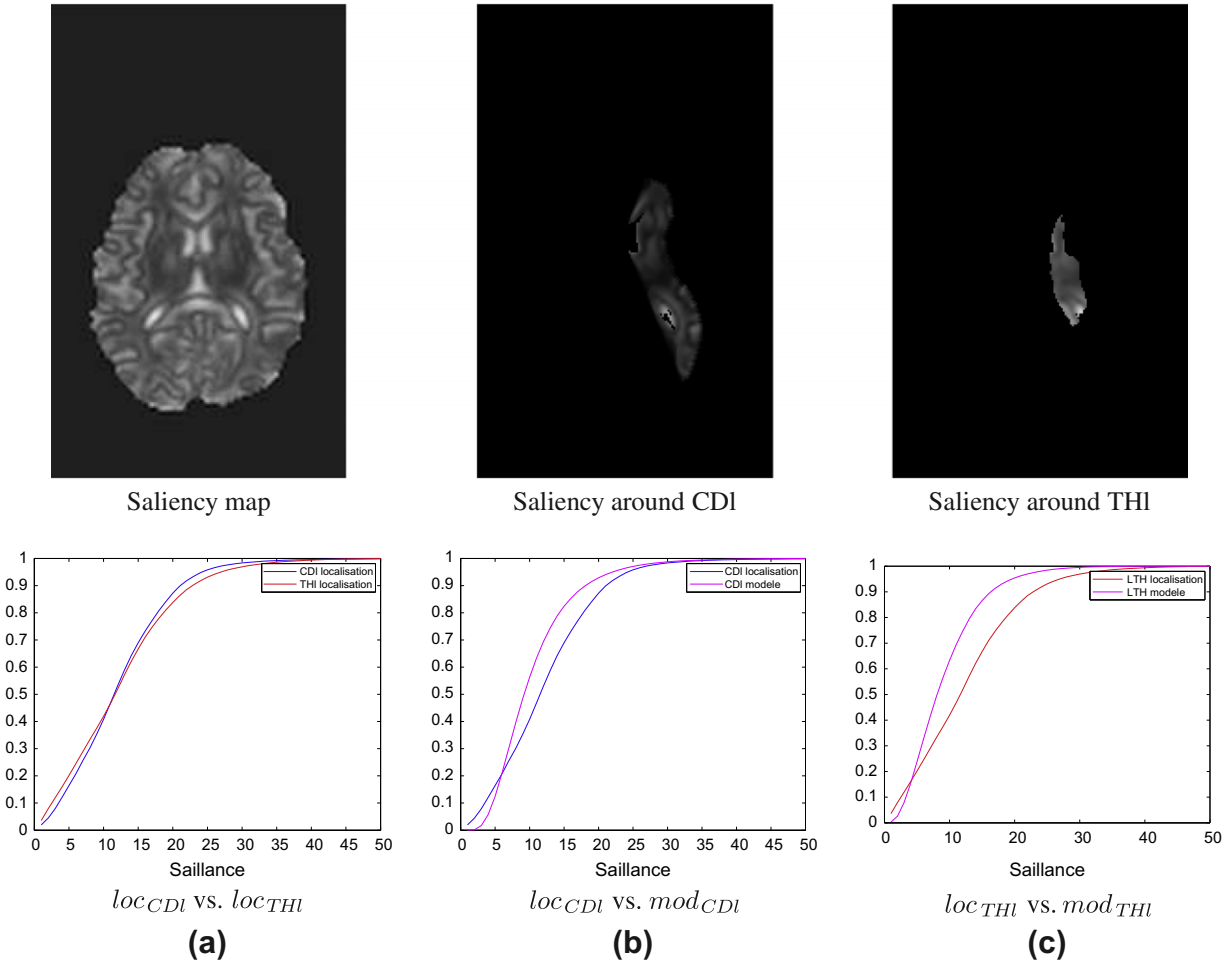


Fig. 9. The saliency map of a brain MRI and the saliency at the localization of the left caudate nucleus (CDl) and the left thalamus (THl). (a) The selection among these two structures takes into account the comparison between both saliency pdf computed from the localization (here the localization of CDl is the most salient one with $emds(CDl, THl) = 0.0084$), but also a comparison between the localization and the model for each structure: (b) the comparison for the CDl is $d_{CDl} = -0.089$, and (c) for THl: $d_{THl} = 0.791$. Finally, the CDl is selected in this example among four structures with a criterion value $c_{CDl} = 0.076$.

and the left thalamus (THl). The following comparison values are obtained: $d_{CDl} = -0.089$, and $d_{THl} = 0.791$.

The criterion to select the best localization is then defined as:

$$c_o = |d_o(loc_o, mod_o)| - \sum_{o' \in V_c \setminus \{o\}} \text{emds}(loc_o, loc_{o'}) \quad (4)$$

where V_c is the set of all candidates. The structure o minimizing c_o is selected. In the example illustrated in Fig. 9, the caudate nucleus is selected with a value $c_{CDl} = 0.076$.

4.3. Saliency in pathological cases

There are different types of tumors, with different visual appearances and thus different saliency maps. Fig. 10 presents two images with a tumor and their corresponding saliency maps. The tumor in the first case (on the left) presents a high contrast with respect to its surrounding and to the necrotic part. The saliency of this tumor is higher than in most of the other parts of the brain (the necrosis saliency however is low). On the contrary, the second type of tumor is large and homogeneous and thus does not present a high contrast with respect to its surrounding. The saliency of this tumor is lower than the one of the brain. For several other tumors, the saliency at the location of the tumor is not higher or lower than in the brain.

5. Optimization of a sequence of segmentation

The process of sequential recognition is viewed as the sequential specialization of a generic graph to a case-specific graph, i.e., where each node representing an anatomical structure has been linked with the corresponding region of the image. If the generic graph includes only a part of the object represented by the image, then the segmentation process segments only these objects and parts of the image remain unexplored.

The process is viewed as the progressive exploration of the image, starting from a reference object. For instance, the ventricles of the brain are the reference structures for the recognition of the internal brain structures. These structures present a high contrast with respect to the gray and white matter and may be easily segmented in most of the cases. Furthermore, they also present a high saliency. Their choice as a starting point for the exploration of the image is then consistent with an exploration of the image like the visual system would do.

The exploration is achieved using the available spatial information in the graph. The spatial relations representations allow us to answer the following question: "from a reference object, which are the locations in the image space where the spatial relation is satisfied to a given degree". Therefore, only the spatial relations with an available (i.e. segmented) reference object are representable, and only the objects connected by an edge to a segmented object have a localization that can be actually computed.

At each step of the process, the graph is filtered to keep the relevant information: two sets of vertices and the set of edges between these two sets are defined. The first set V_{fs} is the set of vertices which are already segmented and connected to a non-segmented vertex. The second set V_{fo} is the set of vertices which are not segmented and connected to the first set. This set includes all vertices which may be segmented at this step of the process. The set of edges E_f represents all spatial relations representable at this step of the process and which target a non-segmented structure. Fig. 11 presents the initial graph and the filtered graph at the first step of the process.

Once the graph is filtered and thus the candidate structures identified, their localization is computed as the conjunction of the representations of all spatial relations targeting this structure, as presented in Section 3. The selection of the next structure to segment is achieved according to the criterion c_o presented in Section 4.2.

5.1. Segmentation of a structure

The segmentation method we use has been proposed in Ref. [2] and is not part of our work. We briefly present this approach here to understand its influence on the segmentation results.

This segmentation approach uses two knowledge sources: a radiometric estimation of the intensity of the structure and the spatial relations targeting the structure. These two different types of information are closely intertwined in the proposed approach: the radiometric estimation for this problem is not enough to segment a structure and is necessarily combined with spatial information, which reduces the search domain. Furthermore, the spatial relations are used to guide and constrain the segmentation process. This approach is composed of two steps. The first step combines the knowledge to recognize the structure and provides a rough segmentation. In the second step, the segmentation is refined by a deformable model method which also uses spatial information as an additional energy term to guide the process.

Fig. 12 illustrates the procedure. A map corresponding to the searched structure (here a thalamus) is generated by thresholding the original image with the radiometric estimation. This estimation is composed by two parameters α and β which are used to express the average ($\dot{x}$) and the standard deviation (σ_x) of the intensity of a given structure as a function of the white matter (wm) and the gray matter (gm) of this particular image: $\dot{x} = \alpha \dot{x}_{wm} + (1-\alpha)\dot{x}_{gm}$ and $\sigma_x = \beta \frac{(\sigma_{wm}+\sigma_{gm})}{2}$, as proposed in Ref. [37]. The image is thresholded as follows: if a voxel has a radiometry between

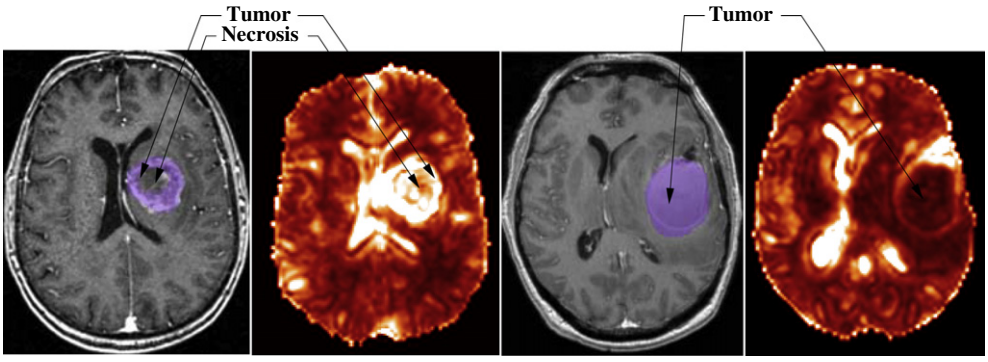


Fig. 10. Two pathological cases and the corresponding saliency maps. The saliency of the tumor highly depends on the tumor type and aspect. The tumor on the left presents a high-saliency corresponding to the high contrast between the tumor and its surrounding (outside the tumor and the necrosis). The tumor on the right is large and homogeneous. The saliency of the latter is very low.

$\hat{x} - \sigma_x$ and $\hat{x} + \sigma_x$, its value is set to '1', and '0' otherwise. The resulting map is masked by the spatial information, i.e., the localization of the structure as defined in Section 3, in order to reduce the search domain around the structure, and then filtered using morphological operations. The largest connected component is then identified as the structure and corresponds to the initial segmentation. A deformable model guided by the spatial information (fully described in Ref. [2]) produces the final segmentation.

Both sources of information, spatial and radiometric, are crucial for this approach and are mutually dependent: the restriction of the search domain thanks to the spatial information allows the radiometric information to be only used to discriminate the

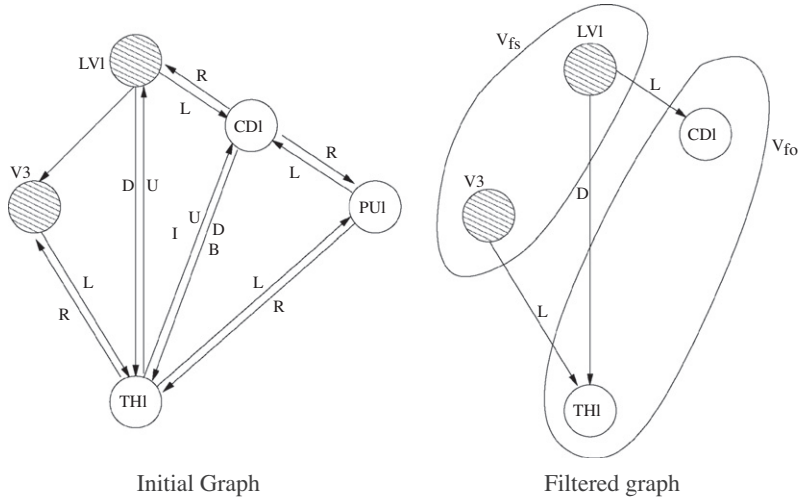


Fig. 11. Initial graph (for the left hemisphere only) and the corresponding filtered graph at the first step of the process. The reference structures (lateral ventricle and third ventricle) are in the set V_{fs} of vertices already segmented. The caudate nucleus and the thalamus are in the set of vertices not segmented V_{fo}. The putamen is not connected to V_{fs} and therefore is not in V_{fo}. The set of edges E_f includes all edges oriented from a vertex of V_{fs} to a vertex of V_{fo}.

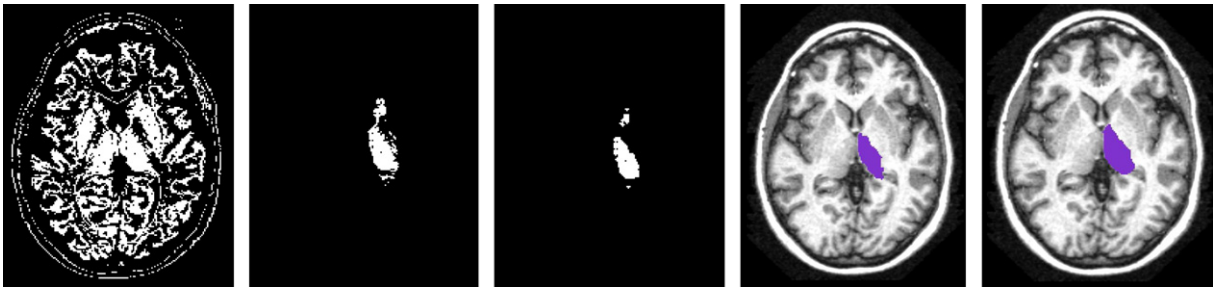
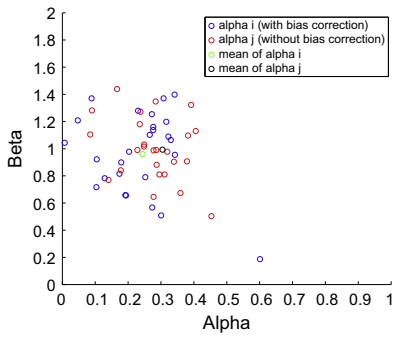


Fig. 12. The process to segment a single structure proposed in Ref. [2]. The first step combines knowledge to identify the component corresponding to the structure. The second step refines the segmentation with a deformable model approach guided by the spatial information.



	[37]		IBSR	
	α	β	α	β
CD	0.305	1.328	0.216	1.208
TH	0.633	1.374	0.557	1.586
PU	0.508	1.072	0.545	0.976
	OASIS		Patho. cases	
	α	β	α	β
CD	0.278	1.398	0.303	1.693
TH	0.606	1.152	0.592	1.483
PU	0.505	1.024	0.485	1.341

Fig. 13. Radiometric estimations for different structures expressed as a function of the intensity of white and gray matters. Left: exact (α_i, β_i) values computed for the caudate nucleus in two cases: with (in blue) or without (in red) bias correction of the MRI. The average values (resp. in red and green) are the resulting α parameter. Right: comparison between the values of Poupon et al. [37] and the learned values on different sets of our database.

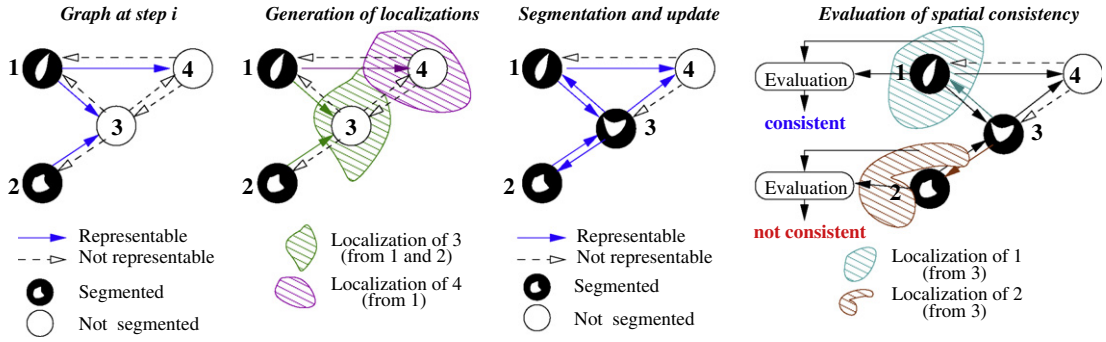


Fig. 14. Evaluation of spatial consistency. At a step i of the process, two structures (3 and 4) are candidate for segmentation. The localizations of both structures are computed using the available spatial information (edge issuing from segmented structures 1 and 2). Structure 3 is segmented and the graph is updated. The spatial relations issued from structure 3 are now representable, particularly relations targeting segmented structures 1 and 2. The fuzzy satisfiability is computed between these relations and the segmentation of the targeted structure. Structure 1 is localized in the kernel of the relation, so it is spatially consistent. Structure 2 is outside the kernel of the relation and this is not consistent.

targeted structure and its surrounding, and not the whole image. Errors may occur when the spatial information does not allow us to reduce the search domain in a way that the radiometric information is relevant, or when the radiometry of a structure does not allow discriminating it from its immediate surrounding.

We recompute the estimations on our database, using the manual segmentation of each structure. For each image of a given set, the exact parameters (α_i, β_i) are computed. The α parameter for this set is the average of the α_i values. The β parameter is the maximum of all β_i values. Fig. 13 (on the left) presents each (α_i, β_i) computed for a given structure and the corresponding α parameter. In this plot, the α_i values present a large dynamic and therefore the average is inexact for several images. In order to reduce the distance between the estimation and the exact values, three different estimations have been computed on three sets of images composing our database (IBSR, OASIS and pathological images). The obtained values for each set, used in our experiments, are presented in Fig. 13 (on the right).

When the radiometric estimation is not correct, the structure may be incomplete (a missing part), or include its surrounding or other parts of the image. In all these cases, an erroneous segmentation is produced and propagates through the representation of the spatial relations using this structure as a reference. Therefore, the segmentation of a particular structure has to be evaluated and the process may incriminate the previous steps when errors occur.

In the next section, we present how to assess the segmentation in order to detect possible errors.

6. Segmentation assessment

As mentioned, during the segmentation of a particular structure, errors may occur and propagate. Therefore, the process must be able to detect errors immediately or a posteriori and to update its strategy, i.e. backtrack and change the sequence of segmentation even if this implies to discard previous structures segmentations. To this end, two criteria are proposed here as well as a structure of control, which consists of a tree of all current and past segmentations, used to update the strategy during the process.

6.1. Criteria for segmentation evaluation

The first criterion concerns the spatial information and controls the consistency of the structural model. The parameters of each spatial relation are learned in a way that the targeted structure is included in the kernel of the relation as described in Section 3. The spatial consistency criterion evaluates if this assertion is still true once a new structure segmentation has been added into the graph. The spatial consistency is not evaluated in the whole graph at each step, but only on the spatial relations using the recently segmented structures as reference. Fig. 14 illustrates how the spatial consistency is evaluated for a small graph. A structure (3) of the graph is segmented using the spatial information from segmented structures 1 and 2. The spatial relations issued from structure 3 and targeting segmented structures are represented. A criterion (presented below) allows us to compare the resulting fuzzy subset and the segmentation, which has to be localized in the kernel of the relation.

To evaluate the spatial consistency of a spatial relation μ_{Rel} targeting a structure μ_{Obj}, we compute a fuzzy satisfiability [38] between the fuzzy subset representing the relation and the targeted structure:

$$f_s(Rel, Obj) = \frac{\sum_{x \in \mathcal{S}} \min(\mu_{Rel}(x), \mu_{Obj}(x))}{\sum_{x \in \mathcal{S}} \mu_{Obj}(x)}, \tag{5}$$

where $\mathcal{S}$ denotes the image space. The fuzzy satisfiability is maximal if the targeted structure is included in the kernel of the fuzzy subset representing the relation.

The second criterion is an intrinsic criterion which compares the segmentation result to a model. In fact, due to the intrinsic variability in shape and size of the internal brain structures, this criterion compares the learned pdf of saliency, i.e., checks if the "visual aspect" or the appearance against the surrounding of the structure is the expected one. The criterion is the EMD distance between the pdf of saliency computed with the segmentation and the pdf learned for this structure, as in Eq. (3):

$$d_o(seg_o, mod_o) = \frac{emd(seg_o, mod_o) - \widehat{mod_o}}{\sigma_{mod_o}}. \tag{6}$$

where seg_o represents the saliency pdf computed from the segmentation, mod_o the saliency pdf learned for this structure, $\widehat{mod_o}$ the mean EMD distance between each case in the database and the learned pdf and σ_{mod_o} the standard deviation of this measure.

These two criteria are used to update the strategy of choice, as described below. We first introduce the data structure used to keep information about the steps of the process.

6.2. Segmentation tree

The previous criteria allow us to detect an erroneous structure segmentation. These errors may happen because of the intrinsic difficulty of segmenting a structure or because of the radiometric estimation. The error may also be caused by the propagation of

previous (undetected) errors. Typically, a wrong segmentation propagates because the spatial relation using this structure as a reference will be wrong too. Therefore, we need to keep track of the history of the previous steps of the process to be able to backtrack.

A tree structure, which contains information about all the segmentations done by the process, is used as a journal of each realized sequence (even sequences finishing by failure). The root of the tree is composed by all the reference structures. Each node corresponds to a segmentation of a particular structure (i.e. a same structure may appear in different sequences, but only one is segmented at a given step). The success or failure of the segmentation is encoded in the node. Sequences without failure are denoted "active sequences".

For each segmented structure, its localization is generated using all spatial relations targeting this structure and these spatial relations use one or more reference structures (already segmented). Among these structures, we denote as the "parent structure" the most recently segmented structure. When a structure is segmented, it is attached in the segmentation tree to its parent structure in the active sequence. If there is no parent structure, then the node is attached to the root of the tree.

This tree structure allows us to know during the process which sequence of segmentation is already tested and therefore to avoid loops (if two sequences are alternatively tested with failure). It is also possible to easily find the untested sequences and eventually to stop the process without finishing if all sequences lead to a failure.

6.3. Backtrack and path selection

In case of an error occurring during the segmentation of a structure and detected thanks to the previous criteria, the strategy of control of the process is simple: it consists in preventing the system of trying the same sequence, which is immediate thanks to the segmentation tree.

The evaluation procedure is presented as pseudo source code in Fig. 15. When the evaluation indicates an error, the following cases are considered:

- if there is no segmentation produced (i.e. the resulting binary map is empty), the parent segmentation (if there is one) is discarded;
- if there is a segmentation, then the spatial consistency criterion is tested. The fuzzy satisfiability is a value in the interval [0,1] and a threshold is fixed at 0.8. In case of failure, the parent segmentation (if it exists) is discarded as well as the current segmentation, otherwise only the current segmentation is discarded;
- the saliency criterion is then tested. A threshold has been set to $T = 2\sigma_{mod_o}$. The current segmentation is discarded in case of failure;
- if both criteria are satisfied, then the segmentation is accepted and the graph is updated.

Fig. 16 presents an example of the segmentation tree at different steps of the process. The right caudate nucleus is segmented first, followed by the right thalamus with a failure. This failure discards the first segmentation. After successfully segmenting two structures on the left (CDl and THl), the right thalamus is segmented (but in first position this time) and then the caudate nucleus, with a failure which discards the thalamus segmentation. The segmentation tree allows us to easily find the untested configurations, and the segmentation is finally achieved by restoring the initial segmentation of the right caudate nucleus and then by segmenting the putamen before the thalamus.

In the worst case, the complexity of this procedure can be high, since potentially the whole tree of possible paths could be explored. However, in practice, we have observed that only few backtracking steps are actually performed (see Section 7), which makes the approach tractable.

7. Experiments on MRI images for internal brain structures segmentation

In this section, we present the experiments conducted on the images of the database described in Section 3. The proposed method is applied on each image by computing the parameters of the spatial relations using a leave-one-out procedure. We first

```
StructureSegmentationEvaluation(currentStructure)
 1  parent ← findParent(currentStructure)
 2  if (SegmentationExist(currentStructure)
 3     then
 4          evalSpaCons ← evaluateSpatialConsistency(currentStructure, parent)
 5          evalSaliency ← evaluateSaliencyCriteria(currentStructure, model)
 6          if (evalSpaCons > threshold_SpaCons)
 7            then
 8                 if (evalSaliency < threshold_Saliency)
 9                   then
10                        acceptSegmentation(currentStructure)
11                   else
12                        discardSegmentation(currentStructure)
13            else
14                 if (parentExist(parent))
15                   then
16                        discardSegmentation(parent)
17                   discardSegmentation(currentStructure)
18     else
19          if (parentExist(parent))
20            then
21                 discardSegmentation(parent)
```

Fig. 15. Pseudo-source code of the evaluation process. Both the current structure and its parent structure are concerned during this procedure. The values of both criteria allow to separate different cases where the segmentation is accepted or discarded, and eventually the parent segmentation is discarded too.

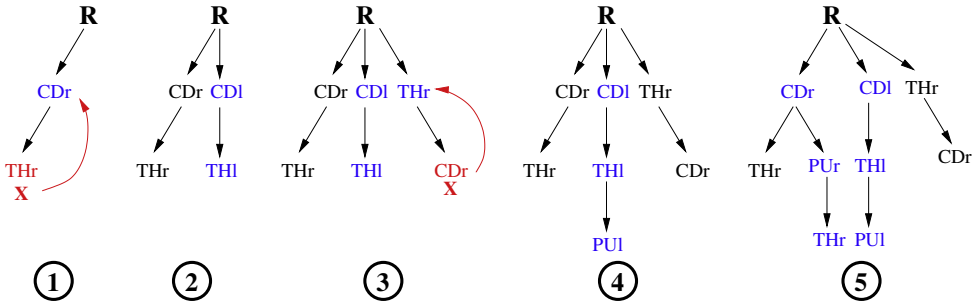


Fig. 16. Structure of control of the segmentation results and configuration of the process. This structure keeps information about past segmentations of structures with different configurations to prevent the process of trying an already known configuration and to easily find remaining not-tested configurations.

Table 1
Numerical evaluations of segmentation results. For each structure, a comparison with a manual segmentation is realized and the mean distance is computed. An average of these distances is computed for each structure over all healthy cases in the database. The segmentation scheme is the same for each structure segmentation. The differences consist of the spatial information used as an input to the segmentation method. The average mean distance (AMD) for the caudate nucleus is higher with the expert path, since it is always segmented first and thus with less spatial information. The mean distances of the other segmentations are similar for all methods.

Struct	Expert path		Optimized path		Optim. path + belief revision	
	# segm.	AMD	# segm.	AMD	# segm.	AMD
CDl	30	1.69	30	1.64	28	1.30
CDr	30	4.63	29	2.20	29	1.49
THl	27	1.90	27	2.25	28	2.39
THr	23	2.36	27	2.71	27	2.30
PUl	26	3.21	27	3.11	27	3.03
PUr	22	3.28	26	3.43	27	3.42

illustrate step by step the segmentation process on one example. Then, the results on the whole database are presented and compared to those obtained with a segmentation path defined a priori by an expert. Quantitative evaluations are provided and the influence of the radiometric estimations as well as of the parameters of the spatial relations is discussed. The test database is composed by 30 healthy cases and 14 pathological ones and comparisons are performed on the whole data, except for the comparison between structures presented in Table 1, which are evaluated only on the healthy cases (this avoids including the potential impact of a pathology on the normal structures in this evaluation).

7.1. Segmentation of an image step by step

As an example, we illustrate the segmentation process on an image of the OASIS database depicted in Fig. 17. The parameters of the spatial relations are learned on the whole database (healthy and pathological cases) following a leave-one-out procedure, i.e. without considering the processed example in the learning step. In the following figures, each illustration presents the same slice of the 3D volume (but the whole process is actually applied in 3D). The path derived from the optimization method and followed during the segmentation is the following: right caudate nucleus, right thalamus, right putamen, left thalamus, left caudate nucleus and left putamen. Note that the path is not the same on both sides of the brain.

Fig. 17 illustrates the first step of the process. For the sake of simplicity, only the structures of the right hemisphere (i.e. on the left on the displayed images) are represented. For visualization purpose, and to show the relevance of the computed localizations, the candidate structures are drawn on the localizations in green. The localizations are computed using the spatial relations to the reference structures and the selection is achieved using the localization, the saliency map and the criterion c_o defined in Eq. (4). In this example, the right caudate nucleus is selected and segmented. The graph is then updated. In the second step (Fig. 18), the putamen is now a candidate, since it is connected to a segmented structure, the caudate nucleus. The localizations of the candidates are computed. The localization of the thalamus is not the same as in the previous step since it now benefits from the spatial relations to the caudate nucleus, which leads to a more precise localization. It is selected as the next structure to be segmented. After its segmentation, the localization of the putamen is recomputed and also benefits from new spatial relations (to the thalamus). The putamen is then selected and segmented. At each step the segmented structure is labeled as such in the graph (in blue in the figures) and the graph is updated with the new candidate structures. The process then goes on with the other structures. The whole segmentation sequence is finally the following: CDr, THr, PUr, THl, CDl, PUl. The flexibility of the path optimization and its adaptive feature depending on the data are clear here, since the optimized sequence is the same as the one used in [2] in the right hemisphere, while it is different in the left one. The final results are presented in Fig. 19, showing that all structures are well recognized and segmented.

7.2. Comparison with a fixed path

In this section, we compare the results obtained with three different approaches, on the healthy cases of the database:

1. a priori defined path, called "expert path", where the caudate nucleus is segmented first, then the thalamus and finally the putamen in both hemispheres, as in Ref. [2]. This path is used for all images and not adapted to each case. Note that with this method, if an error occurs, it may prevent the correct segmentation of other structures in the same hemisphere;
2. proposed optimized path, computed for each individual case, and therefore potentially different from one image to another one. Note that the only difference with the first method is the order in which structures are segmented and the possibility to backtrack;

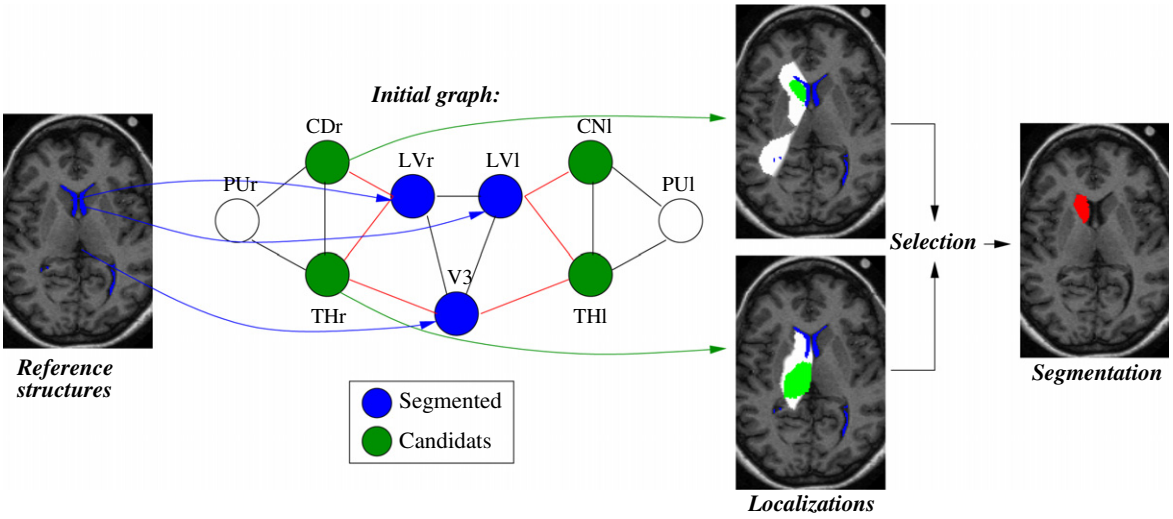


Fig. 17. Initial step of the process. Only the reference structures (lateral ventricles and third ventricle) are segmented and represented in blue in the graph. Four structures are candidates for the next segmentation step (left and right caudate nuclei and thalami, represented in green in the graph). The localizations of these structures are computed using the spatial relations to the reference structures (only two are represented here, as white regions, in the right hemisphere). A structure is then selected (here the right caudate nucleus), according to the criteria c_o, and segmented.

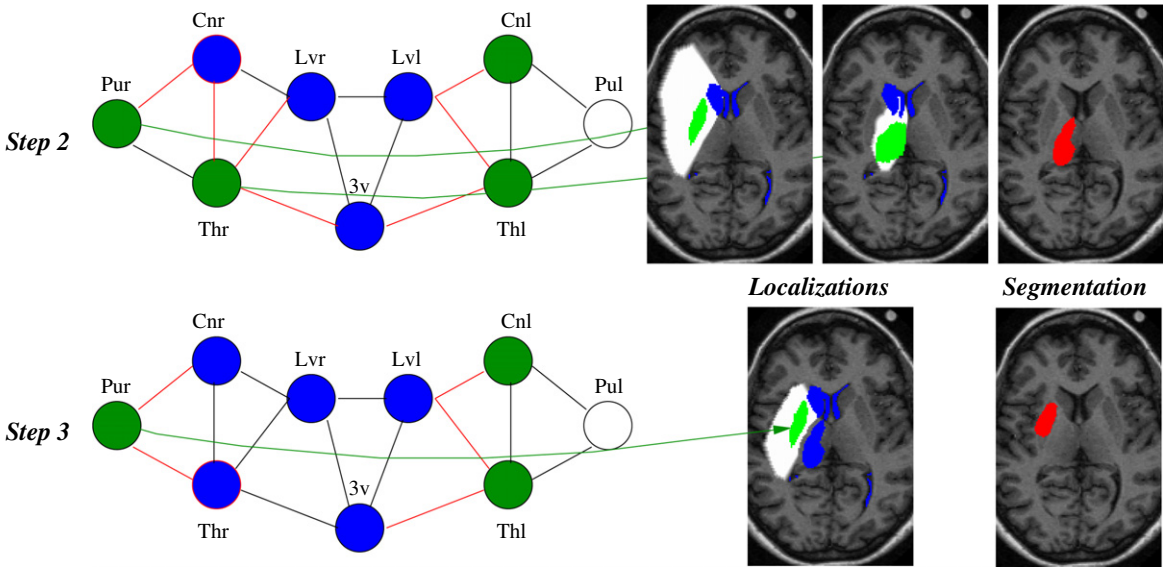


Fig. 18. Second step (top) and third step (bottom). After the segmentation of the right caudate nucleus in the first step, the putamen becomes a candidate. The localization of the thalamus now benefits from spatial information related to three structures and is more precise than in the previous step (the white region is reduced). In the third step, the right putamen is segmented.

3. proposed optimized path and modified segmentation scheme using a simple "belief revision" scheme: when a structure is segmented and accepted, the resulting segmentation is used to recompute the parameters α and β (see equations in Section 5) used for the radiometric estimations. A new segmentation of the same structure is achieved and if the new segmentation improves both evaluation criteria, it replaces the previous segmentation.

Fig. 20 presents a comparison between the three methods: sequential segmentation following an expert path, an optimized sequence and the optimized sequence with belief revision. In the left hemisphere, the path followed is the same for all approaches and results are identical for the expert path and for the optimized one. With an additional belief revision approach, the numerical

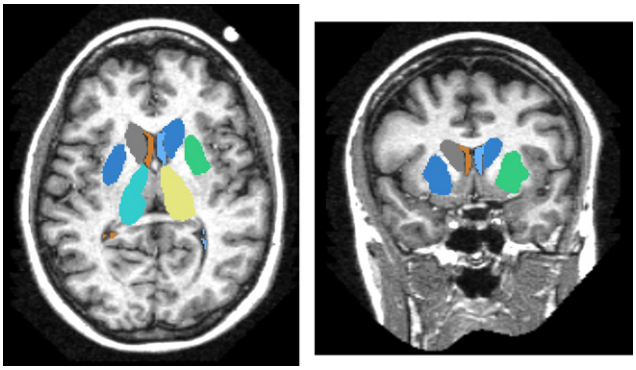
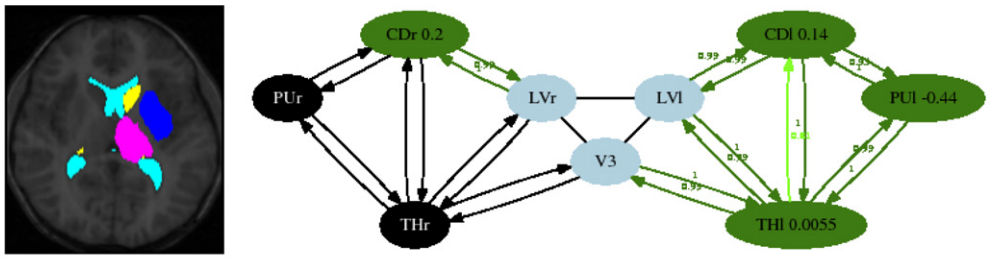


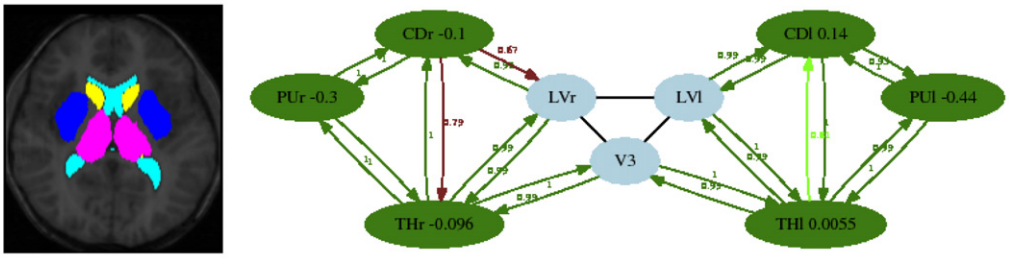
Fig. 19. Final segmentation after six steps of the process (one axial slice and one coronal slice are displayed): all structures have been successfully recognized and segmented.

(a) Expert path:



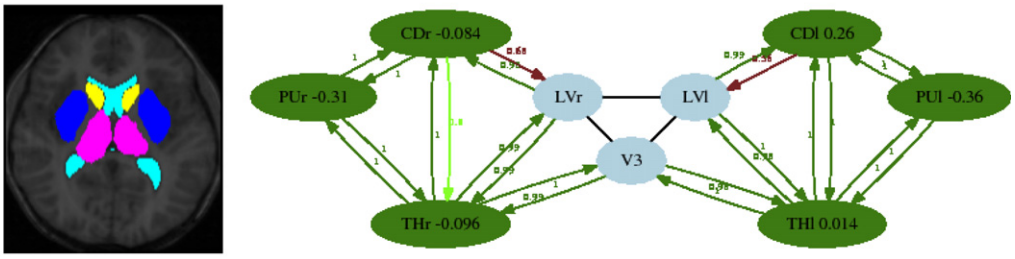
Segmentation path: CDl CDr THl PUl

(b) Optimized path:



Segmentation path: THr CDr CDl THl PUr PUl

(c) Optimized path with belief revision:



Segmentation path: THr CDr CDl THl PUl PUr

Numerical evaluation of structure segmentations (mean distance):

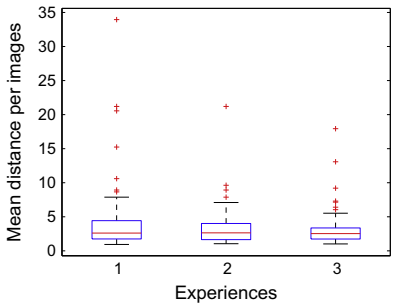
	CDl	THl	PUl	CDr	THr	PUr
expert path	4.13	0.95	2.62	26.9	NS	NS
optimized path	4.13	0.95	2.62	0.63	2.34	2.87
w. belief revision	0.94	1.01	2.49	0.64	2.34	2.91

Fig. 20. Comparison between sequential segmentations following an expert path (a), our approach (b), and our approach with belief revision (c). Values of the two criteria are given in the graph (spatial consistency on edges and saliency on vertices). In the left hemisphere, the same path is followed in (a) and (b) and thus the numerical evaluations of each structure segmentation are the same. With belief revision, the numerical evaluations of the CDl and the PUl are better, and slightly higher for the thalamus. In the right hemisphere, different paths are followed. The expert path segments the caudate nucleus (with errors) and fails to segment two structures, while the other approaches segment all structures correctly. The numerical evaluation of the caudate nucleus is significantly better.

evaluations are better for the segmentation of two structures. In the right hemisphere, the segmentation according to the expert path fails to segment the caudate nucleus: the procedure produces a segmentation, but it is not correct and the structure is mis-recognized (with a part of its tail and the back part of the thalamus) and thus the other structures cannot be segmented. With an optimized path, the thalamus is first segmented and then the caudate nucleus is segmented with success. Finally the putamen is segmented. As shown in this example, our approach allows detecting the mis-recognition of the right caudate nucleus and adapting the path by backtracking on the segmentation ordering, in order to correctly recognize and segment all structures.

7.3. Segmentation evaluations

In order to provide a quantitative evaluation of the results, the mean distance between the obtained segmentation and a manual segmentation is computed for each structure. An average value is then computed for each structure over all healthy images in the database. Table 1 presents the obtained values. The results show that the mean distance for the caudate nucleus is better with an optimized path. An explanation is that with the expert path, the caudate nucleus is always segmented in first place and thus with less spatial information than a structure segmented later on in the sequence. On the contrary, in the proposed approach, the



	expert	saliency	belief
# images	44	44	44
AMD	5.024	3.545	3.514

Fig. 21. Numerical evaluations of segmentation results (1: expert path, 2: optimized path, 3: optimized path with belief revision). The table on the right presents the average of all numerical evaluations of segmented structures in all images. The average mean distance (AMD) is lower for an optimized path than for the expert path.

Table 2
Quantitative evaluation of segmentation results on the 44 images of the database (30 healthy and 14 pathological ones). Values on top represent the final total number of segmentations realized by the process. There are 264 structures in total and our approach allows us to segment more structures than a sequential segmentation following an expert path (there are less failures). The spatial consistency criterion is more often used than the saliency criterion. In the bottom part of the table, the number of accepted segmentations against the final number of segmentations shows the number of adaptations of the path needed to achieve the segmentations.

		Expert	Saliency	Belief
# Segmentations	Correctly segmented structures	209	224	233
	Failures	55	40	31
Criteria	Saliency	2	6	13
	Spatial consistency	21	79	65
	Both	2	3	3
Segm stats	Accepted	209	309	309
	Failed (no result image produced)	10	12	16
	Discarded (itself)	2	6	13
	Discarded (as parent)	0	85	76

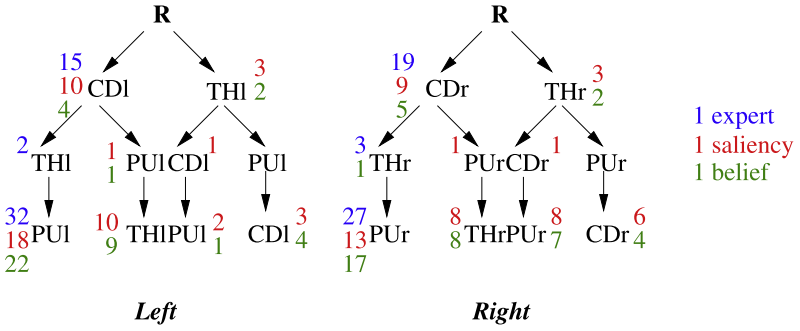


Fig. 22. Distribution of the segmentation sequences. The expert path (CD, TH, and then PU) is the most frequent path, but other paths are followed when needed. This is consistent with what is expected from the optimization procedure, which adapts the path when needed, according to the data. There are also less unfinished paths with one or two missing structures, which indicates that the followed path allows a better segmentation.

caudate nucleus may be segmented later on in the optimized sequence, and is then likely to be better segmented. The mean distances obtained for the other structures are similar for all paths, with no significant difference.

Table 1 also presents the total number of segmentations per structure over the 30 healthy cases of the database. There is a similar number of caudate nucleus segmentations for all paths. However, there are more thalamus and putamen segmentations (and particularly on the right side without any evident reason). The number of caudate nucleus segmentations is related to the order followed by the expert path, i.e. the caudate nucleus is always segmented in first position. In this case, there is no backtrack of the process, thus an erroneous segmentation of the caudate nucleus prevents the segmentation of the other structures but cannot be discarded (note that in the latter case, the erroneous segmentation appears in the number of segmentations of the caudate nucleus). Fig. 20 presents an example of an erroneous segmentation.

For each image the average value of the mean distances computed for all structure segmentations is calculated. The mean of these values over the 44 images of the database is presented in Fig. 21 for the three methods. On the left, the average mean distances are represented with box plots (1: expert path, 2: optimized path, 3: optimized path with belief revision). The upper quartile and largest observation are higher, while the lower quartile and the median values are similar. Extremal values are also higher with the expert path. This indicates that the segmentation with an optimized path allows us to correct the largest errors of segmentation but does not improve the other segmentations. The optimized path allows us to detect recognition errors and inconsistencies and to propose a strategy to avoid errors. Note that the precision of the segmentation of a structure mainly depends on the chosen approach for the final segmentation (here the deformable model proposed in Ref. [2]). On the right of Fig. 21, the results show a better average numerical evaluation when the segmentation is performed with an optimized path than with an expert path. The belief revision step further improves the results, but only slightly in average.

Considering the whole data base (healthy and pathological cases), the number of segmented structures is larger with our

sequential segmentation framework with an optimized sequence than with the same framework but following an expert path. This result shows that the dynamic path selection allows us to recognize and segment more structures. Furthermore, no false positives (misrecognized structures) appear in the results. A poor radiometric estimation may lead to segment two structures as only one and thus the recognition maybe partially wrong. But the spatial consistency criterion allows avoiding such errors, when applied with a large enough threshold value. Table 2 presents these values as well as the number of path changes needed to achieve the segmentation. The criterion evaluating the spatial consistency is more often used than the other criteria. This shows the relevance of this criterion. On the one hand, this result is consistent with the important usage of spatial information to guide the process of each structure's segmentation. On the other hand, we choose not to rely on intrinsic features of the structure to segment and thus the saliency criterion used to evaluate the segmentation result is less relevant.

Fig. 22 presents the different sequences of segmentation with the number of occurrences of each sequence, at the end of the segmentation process, i.e. the final path, after potential backtracking, on healthy and pathological cases. The repartition shows that the most frequent path is the expert path, but other paths are also used. This is an expected result for our approach.

7.4. Influence of radiometric estimation

In order to estimate the influence of the radiometric estimation on the segmentation, experiments have been conducted with parameters estimated on different sets of images. The first couple of parameters (α, β), called "exact", is computed on the same image with a manual segmentation. The second experiment is achieved with the values presented in Section 5, where the learning database is separated in three sets: IBSR, OASIS and pathological cases. The third experiment uses the parameters described in Ref. [37]. Remember that the parameters α and β allow defining the average intensity value of a structure and the standard deviation respectively, as functions of the intensity of white and gray matter.

Examples from the resulting segmentations are presented in Fig. 23. The segmentations achieved with the "exact" parameters are not the best segmentations. The parameters learned on three subsets of our database (following a leave-one-out procedure) give the best results for the four cases presented here. However, the differences between segmentations of the same image with different radiometric estimations show the influence of these parameters.

7.5. Influence of the learning of spatial relation parameters

Finally, we propose to analyze the influence of the parameters of the spatial relations by applying the proposed approach to the same image and using the same segmentation scheme (with an optimized path without belief revision), but with different parameters for the spatial relations. The purpose of this experiment is to establish whether the results are improved when the spatial relations are more precise, or if the imprecision of the spatial relation does not impact the result.

Three experiments are carried out with parameters learned on different sets of images. The default set (denoted by *all*) is the whole learning database (44 images) including both healthy and pathological cases (with a leave-one-out procedure). A smaller and more homogeneous set (denoted by *healthy*) is composed by the 30 healthy images only (with a leave-one-out procedure too). Finally, an experiment denoted by *exact* is achieved with parameters derived from the manual segmentation of the image, i.e. exact parameters for this image.

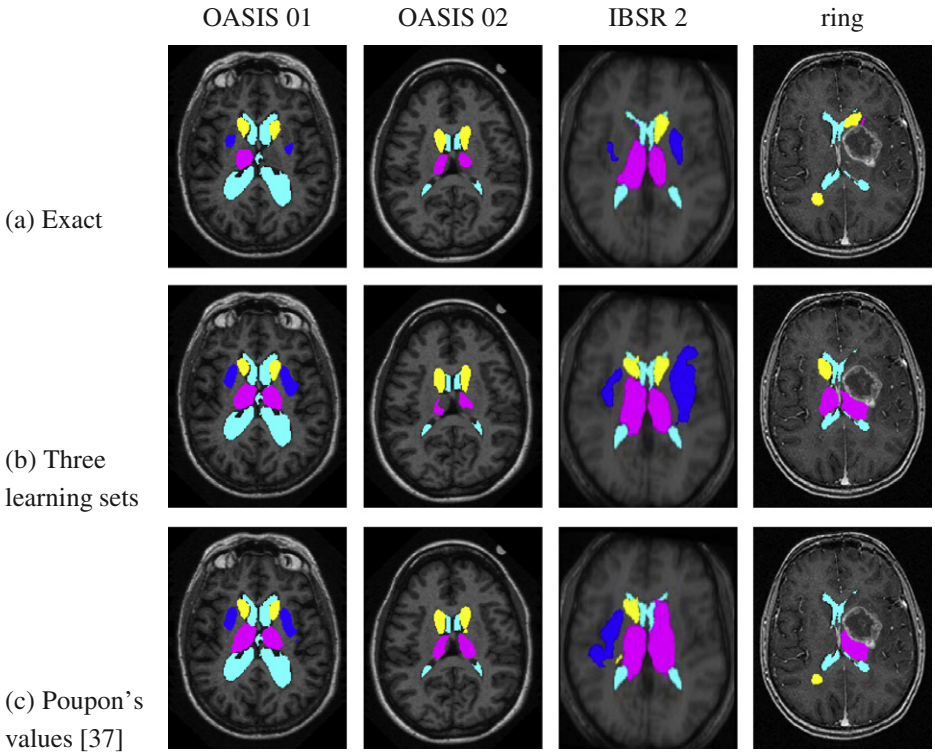


Fig. 23. Comparison of sequential segmentations with radiometric estimation learned on different sets of images (two examples of the IBSR data base, two of the OASIS database and one pathological case, with a ring-shaped tumor). (a) Segmentation results obtained with "exact" parameters. (b) Results using parameters computed on a clustered database and used in our other experiments. (c) Sequential segmentation using the parameters proposed in Ref. [37]. The "exact" parameters do not give the best results. The α parameter is related to the mean of the intensity values and it is important that this estimation reflects the intensity of the structure but also discriminates it from the other structures.

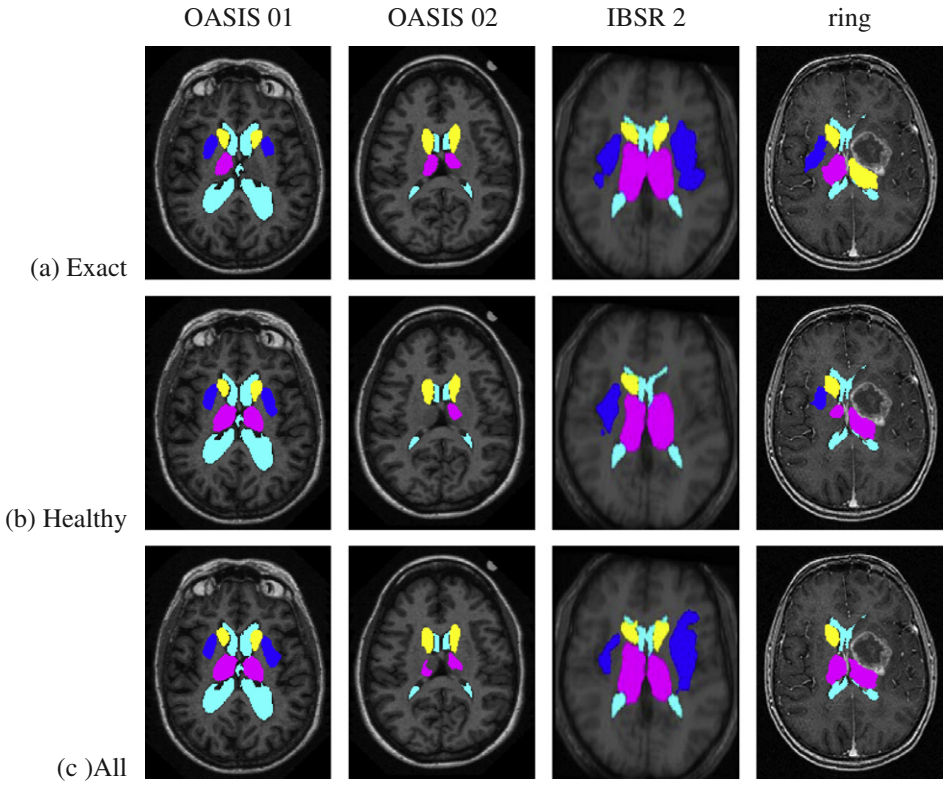


Fig. 24. Comparison of sequential segmentations with learning of the spatial relation parameters on different sets of images. (a) The parameters are "exact", i.e. learned on the segmented image only. (b) The parameters are learned on all healthy cases of the database (homogeneous set). (c) The parameters are learned on the whole database (healthy and pathological cases). This experiment shows that more precise spatial information does not necessarily provide better results, and even sometimes worse. The intrinsic imprecision of the spatial relations provides the necessary flexibility for spatial reasoning.

Slices from the resulting segmentations are presented in Fig. 24. The first row presents the segmentation obtained with *exact* parameters and these segmentations are sometimes improved in the other experiments. The segmentation results are not improved when using more precise spatial relations. On the contrary, the larger learning set, which allows a more flexible spatial reasoning, provides the best results. This is a very encouraging result, since it shows that the variability is well taken into account by the proposed approach.

8. Conclusion

In this paper we addressed two important problems in sequential segmentation. The first one is related to the choice of the order in which structures are segmented. To solve this problem, we proposed to optimize a criterion combining saliency information computed in each image to be processed and generic structural information about the spatial relations between structures, derived from an anatomical model. This contribution extends the framework developed in Ref. [2], where the segmentation order was fixed in an ad-hoc way and was the same for all processed cases. The proposed optimization procedure allows reducing the number of segmentation failures by adapting the segmentation order to the specificities of each image.

Furthermore, the proposed criterion involves a number of parameters, related to the definition of spatial relations and to radiometric information. We have proposed a learning procedure to estimate these parameters, thus avoiding tedious manual fine tuning.

The second problem related to sequential segmentation is the influence of a potential error on the subsequent steps of the process. We proposed an original method to control the result obtained at each step, and its consistency with respect to the model. Additionally, we developed a backtracking procedure, which allows, in case an error is detected, to change the segmentation order and to choose another strategy. From an algorithmic point of view, the efficiency of the proposed method is ensured by a tree structure which keeps trace of all segmentations and already explored paths. The experiments have shown that this control and backtracking process is efficient and allows segmenting more structures in a correct and consistent way.

In the proposed method, some steps could be easily replaced by other ones. For instance the final segmentation, which follows the approach in Ref. [2], could be replaced by another method such as minimal surface or level sets for instance. Similarly, the computation of saliency could include other features.

The proposed approach shows that image analysis and interpretation can benefit from visual attention models. The proposed optimization relies on a structural model involving spatial relations, which implies that some expert prior knowledge is available to build this model. This is the case for the considered example in brain imaging. Further work could investigate this type of approaches in the case of imprecise and/or incomplete knowledge description of the scene.

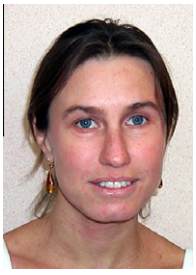
References

[1] I. Bloch, T. Géraud, H. Maître, Representation and fusion of heterogeneous fuzzy information in the 3D Space for model-based structural recognition – application to 3D brain imaging, Artif. Intell. 148 (2003) 141–175.
[2] O. Colliot, O. Camara, I. Bloch, Integration of fuzzy spatial relations in deformable models – application to brain MRI segmentation, Pattern Recogn. 39 (2006) 1401–1414.

[3] O. Colliot, Representation, évaluation et utilisation de relations spatiales pour l'interprétation d'images. Applications à la reconnaissance de structures anatomiques en imagerie médicale, Ph.D. thesis, ENST, 2003.
[4] U. Neisser, Cognitive psychology, Appleton-Century-Crofts, 1967.
[5] A. Treisman, G. Gelade, A feature-integration theory of attention, Cognitive Psychol. 12 (1980) 97–136.
[6] A. Treisman, Preattentive processing in vision, Comput. Vision Graph. Image Process. 31 (2) (1985) 156–177. doi:http://www.dx.doi.org/10.1016/S0734-189X(85)80004-9.
[7] D. Crevier, R. Lepage, Knowledge-based image understanding systems: a survey, Comput. Vis. Image Understand. 67 (2) (1997) 161–185, doi:10.1006/cviu.1996.0520.
[8] F. Le Ber, J. Lieber, A. Napoli, Les systèmes à base de connaissances, in: J. Akoka, I. Comyn Wattiau (Eds.), Encyclopédie de l'informatique et des systèmes d'information, Vuibert, 2006, pp. 1197–1208.
[9] J. Talairach, P. Tournoux, Co-Planar Stereotaxic Atlas of the Human Brain 3-Dimensional Proportional System: An Approach to Cerebral Imaging, Thieme, 1988.
[10] D. Bowden, M. Dubach, Neuroanatomical Nomenclature and Ontology, John Wiley and Sons, Inc., 2005. Ch. Databasing the Brain.
[11] C. Rosse, J.L. Mejino, Anatomy Ontologies for Bioinformatics: Principles and Practice, Springer, 2007, Ch. The Foundational Model of Anatomy Ontology, pp. 59–117.
[12] J. Atif, C. Hudelot, O. Nempont, N. Richard, B. Batrancourt, E. Angelini, I. Bloch, GRAFIP: A Framework for the Representation of Healthy and Pathological Cerebral Information, in: IEEE International Symposium on Biomedical Imaging (ISBI), Washington DC, USA, 2007, pp. 205–208.
[13] D. Conte, P. Foggia, C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition, Int. J. Pattern Rec. Art. Intell. 18 (3) (2004) 265–298.
[14] A. Perchant, Morphisme de graphes d'attributs flous pour la reconnaissance structurelle de scènes, Ph.D. thesis, Ecole nationale supérieure des télécommunications, Paris, France, 2000.
[15] A. Perchant, I. Bloch, Fuzzy morphisms between graphs, Fuzzy Sets Syst. 128 (2) (2002) 149–168.
[16] E. Bengoetxea, P. Larranaga, I. Bloch, A. Perchant, C. Boeres, Inexact graph matching by means of estimation of distribution algorithms, Pattern Recogn. 35 (2002) 2867–2880.
[17] R. Cesar, E. Bengoetxea, I. Bloch, P. Larranaga, Inexact graph matching for model-based recognition: evaluation and comparison of optimization algorithms, Pattern Recogn. 38 (2005) 2099–2113.
[18] A. Deruyver, Y. Hodé, E. Leammer, J.-M. Jolion, Adaptive pyramid and semantic graph: knowledge driven segmentation, in: Graph-Based Representations in Pattern Recognition: 5th IAPR International Workshop, vol. 3434, Springer-Verlag GmbH, Poitiers, France, 2005, pp. 213–222.
[19] A. Deruyver, Y. Hodé, L. Brun, Image interpretation with a conceptual graph: labeling over-segmented images and detection of unexpected objects, Artif. Intell. 173 (14) (2009) 1245–1265. doi:http://www.dx.doi.org/10.1016/j.artint.2009.05.003..
[20] O. Nempont, Modèles structurels flous et propagation de contraintes pour la segmentation et la reconnaissance d'objets dans les images. application aux structures normales et pathologiques du cerveau en IRM, Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications (Mars 2009).
[21] O. Nempont, J. Atif, E. Angelini, I. Bloch, Structure Segmentation and Recognition in Images Guided by Structural Constraint Propagation, in: European Conference on Artificial Intelligence ECAI, Patras, Greece, 2008, pp. 621–625.
[22] I. Bloch, Fuzzy spatial relationships for image processing and interpretation: a review, Image Vision Comput 23 (2) (2005) 89–110.
[23] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New-York, 1980.
[24] I. Bloch, Fuzzy relative position between objects in image processing: a morphological approach, IEEE Trans. Pattern Anal. Mach. Intell. 21 (7) (1999) 657–664.
[25] J. Atif, C. Hudelot, G. Fouquier, I. Bloch, E. Angelini, From generic knowledge to specific reasoning for medical image interpretation using graph-based representations, in: International Joint Conference on Artificial Intelligence IJCAI'07, Hyderabad, India, 2007, pp. 224–229.
[26] A. Treisman, Search, similarity, and integration of features between and within dimensions, J. Exp. Psychol.: Hum. Percept. Perform. 17 (3) (1991) 652–676.
[27] C. Healey, Perception in visualization, available at: http://www.csc.ncsu.edu/faculty/healey/PP/index.html (2007).
[28] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[29] C. Koch, S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry, Hum. Neurobiol. 4 (4) (1985) 219–227.
[30] J.-F. Mangin, O. Coulon, V. Frouin, Robust brain segmentation using histogram scale-space analysis and mathematical morphology, in: Medical Image Computing and Computer-Assisted Interventation, 1998, p. 1230.
[31] P. Thevenaz, T. Blu, M. Unser, Interpolation revisited, IEEE Trans. Med. Imag. 19 (7) (2000) 739–758.
[32] T.R. Reed, Motion analysis using the 3D Gabor transform, IEEE (1997) 506–509.
[33] Y. Wang, C. Chua, Face recognition from 2D and 3D images using 3d gabor filters, Image Vision Comput 23 (2005) 1018–1028.
[34] L. Itti, C. Koch, Feature combination strategies for saliency-based visual attention systems, J. Electr. Imag. 10 (1) (2001) 161–169.
[35] Y. Rubner, C. Tomasi, L. Guibas, A metric for distributions with applications to image databases, in: Sixth International Conference on Computer Vision, Bombay, India, 1998, pp. 59–66.
[36] C. Villani, Topics in optimal transportation, American Math. Soc., 2003.
[37] F. Poupon, J.-F. Mangin, D. Hasboun, C. Poupon, I. Magnin, V. Frouin, Multi-object deformable templates dedicated to the segmentation of brain deep structures, in: Medical Image Computing and Computer-Assisted Interventation, vol. 1496, 2008, p. 1134.
[38] B. Bouchon-Meunier, M. Rifqi, S. Bothorel, Towards general measures of comparison of objects, Fuzzy Sets Syst. 84 (2) (1996) 143–153.

Geoffroy Fouquier received a PhD in image processing from Telecom ParisTech in 2010 and the M.Sc. degree in artificial intelligence from the Paris VI university in 2006. His PhD was followed by a post-doctoral position about semi-automatic whole-body MRI segmentation dedicated to dosimetry studies. He is now research engineer at eXenSa working on collaborative filtering and latent semantic analysis. His main research interests include spatial reasoning, graph representations, model and knowledge-based recognition of image structures, pre-attentive mechanisms, medical images and anatomical models.

Jamal Atif is an associate professor of computer sciences at Paris-Sud XI University performing his research within the Machine Learning and Optimization team (project team TAO, INRIA) at the Computer Sciences Lab (LRI, CNRS/Paris-Sud XI University). From 2006 to 2010, he was a research scientist at IRD (Institut de Recherche pour le Développement), Unité ESPACE S140 and associate professor of computer sciences at the University of French West Indies. He received a master degree and PhD in computer sciences and medical imaging from the University of Paris-XI in 2000 and 2004. His research interests focus on computer vision and knowledge based image understanding (semantic image interpretation) for medical and earth observation applications. He works on fields arising from information theory, graph theory, uncertainty management (fuzzy sets), ontological engineering (description logics), mathematical morphology, spatial reasoning and machine learning.



Isabelle Bloch is professor at the Signal and Image Processing Department of Telecom ParisTech – CNRS LTCI, in charge of the Image Processing and Understanding group. Her research interests include 3-D image and object processing, computer vision, 3-D, fuzzy and logics mathematical morphology, information fusion, fuzzy set theory, structural, graph-based and knowledge-based object recognition, spatial reasoning, and medical imaging.