ELSEVIER

A Bayesian model for joint segmentation and registration

Kilian M. Pohl,[a,b,*] John Fisher,[a] W. Eric L. Grimson,[a] Ron Kikinis,[b] and William M. Wells[b]

[a]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
[b]Surgical Planning Laboratory, Harvard Medical School and Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, USA

A statistical model is presented that combines the registration of an atlas with the segmentation of magnetic resonance images. We use an Expectation Maximization-based algorithm to find a solution within the model, which simultaneously estimates image artifacts, anatomical labelmaps, and a structure-dependent hierarchical mapping from the atlas to the image space. The algorithm produces segmentations for brain tissues as well as their substructures. We demonstrate the approach on a set of 22 magnetic resonance images. On this set of images, the new approach performs significantly better than similar methods which sequentially apply registration and segmentation.
© 2005 Elsevier Inc. All rights reserved.

Introduction

To better understand brain diseases, many neuroscience studies focus on the anatomical differences between control and diseased subjects. In order to find these differences, scientists often analyze medical images for brain structures which seem to be influenced by the disease. The analysis is frequently based on segmentations of the structures of interest that are mostly performed by human experts. However, this manual process is not only very expensive, but in addition, it increases risks related to inter- and intra-observer reliability (Kikinis et al., 1992). Neuroscientists are keenly interested in automatic methods, which often rely on prior information, to perform this task (Collins et al., 1999; Leventon et al., 2000; Marroquin et al., 2003; Fischl et al., 2004; Pohl et al., 2004a; Ashburner and Friston, 2005). With notable exceptions, these methods first register the prior information, i.e., an atlas, to the medical image and then segment the medical image into anatomical structures based on that aligned information. The goal of this work is to unify this process into a single Bayesian framework in order to overcome biases caused by commitment to the initial registration.

When automatic segmentation methods are guided by prior information, they frequently are used to segment anatomical structures defined by weakly visible boundaries in medical images. For example, the intensity properties of the thalamus in T1-weighted magnetic resonance (MR) images are very similar to those of the neighboring white matter (Fig. 1). Algorithms cannot rely on the MR images alone in order to distinguish these two structures. However, the ventricles, the dark structures above the thalamus, are more easily identified. In order for the ventricles to guide the detection of the boundary between the thalamus and the white matter, automatic segmentation algorithms use spatial priors (Mazziotta et al., 1995; Thompson et al., 1996). These spatial priors capture the relationship between structures such as the fact that the ventricles are above the thalamus.

As mentioned previously, most atlas-based algorithms perform registration and segmentation sequentially (Cocosco et al., 2003; Van Leemput et al., 1999; Fischl et al., 2002). In this paper, we present an approach that exploits complementary aspects of both problems. In the example of the thalamus, segmentation simplifies the registration of the region defined by ambiguous intensity patterns between thalamus and white matter (Pohl et al., 2004a). On the other hand, aligning an atlas aids the detection of thalamus as discussed previously.

The idea of joining registration and segmentation has been utilized by boundary localization techniques using level set representation, such as Leventon et al. (2000) and Tsai et al. (2003), which align an atlas to the subject and simultaneously estimate the shape of a structure. These methods relate both problems to each other by extending the definition of the shape to include its pose. This paper describes an integrated segmentation and registration approach related to voxel-based classification methods, which considers the anatomical structure associated with each voxel within a Bayesian framework (Wells et al., 1996; Van Leemput et al., 1999; Marroquin et al., 2003; Pohl et al., 2004a). These voxel-based classification methods explicitly model image artifacts in order to segment large data sets without manual intervention.

Voxel-based classification methods that have coupled registration and segmentation of misaligned spectral images include

---
* Corresponding author. Surgical Planning Laboratory, Harvard Medical School and Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, USA.
E-mail address: pohl@csail.mit.edu (K.M. Pohl).
URL: http://www.csail.mit.edu/~pohl (K.M. Pohl).
Available online on ScienceDirect (www.sciencedirect.com).
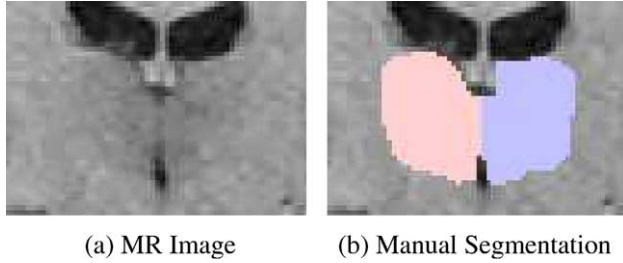
(a) MR Image   (b) Manual Segmentation

Fig. 1. Panel (a) shows an MR image (1.5-T MRI system (GE Medical Systems, Milwaukee), three-dimensional Fourier transformation spoiled gradient-recalled acquisition sequence, echo time (TE) = 5 ms, repetition time (TR) = 35 ms, repetition = 1, nutation angle = 45, field of view = 24 cm, acquisition matrix = 256 × 256 × 124, voxel dimension = 0.9375 × 0.9375 × 1.5 mm) of the area around the thalamus, which in this image is the structure below the dark ventricles. The corresponding manual segmentation of the thalamus is shown in panel (b). From the intensity pattern in panel (a), the boundary of the thalamus and the neighboring white matter is not clearly defined.

(Wyatt and Noble, 2002; Xiaohua et al., 2005). Here, however, our purpose is to align an atlas to MR images and separate the images into anatomical structures, as also suggested by (Ashburner and Friston, 2005). The proposed non-rigid registration framework of Ashburner and Friston (2005) performs a local deformation of the atlas space under the assumption that the atlas is globally aligned to the image space. This is an assumption which does not hold in practice and which we address in our approach.

Section 2 describes the Bayesian framework that models the relationship between the unknown segmentation, the observed MR image data, and the registration parameters. The registration parameters capture the deformation between atlas and image space. The solution within our joint registration and segmentation framework is defined by a Maximum A posteriori Probability (MAP) estimation problem and is found by adopting an instance of the expectation maximization (EM) algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997).

In Section 3, we specify a hierarchical description of the registration parameters. Global registration parameters capture the correspondence between the entire atlas and image space. Structure-dependent registration parameters represent the local deformations within a region of an anatomical structure. This hierarchical model simplifies the correspondence problem between atlas and image space.

We describe in Section 4 a robust implementation of our algorithm, where the registration parameters capture affine mappings. In Section 5, we perform an experiment comparing the robustness of this implementation to other EM segmenters. In this experiment, the automatic methods outline a set of 22 MR images into the major brain tissue classes as well as the thalamus, which is a structure with indistinct boundaries. When comparing the automatic results to manual segmentations, our new method significantly outperforms the other implementations.

A Bayesian model combining registration and segmentation

The accuracy of segmenting structures that have indistinct boundaries in MR images, using tissue classification methods, significantly depends on properly modeling image artifacts and accurately registering prior information to the subject. In this section, we develop a unified framework which performs registration, segmentation, and image artifact estimation simultaneously.

In this paper, we are primarily interested in studying the dependencies between segmentation and registration. We therefore view image artifacts as nuisance parameters[1], which, if known, would greatly simplify the segmentation of MR images. For example, nuisance parameters can represent the image inhomogeneity caused by the acquisition of the MR images or the subject specific intensity histogram of an anatomical structure. The proper modeling of these nuisance parameters is an ongoing discussion in neuroimaging (Wells et al., 1996; Leventon, 2000; Kapur, 1999; Marroquin et al., 2003; Ashburner and Friston, 2005).

Due to their complex dependencies, it is difficult to extract the nuisance parameters θ and the registration parameters $\mathcal{R}$ from the MR images $\mathcal{I}$ without explicit knowledge of the unknown true segmentation. However, this problem is greatly simplified when the solution is determined within an EM framework. In this framework, the solution is formulated as an incomplete data problem marginalizing over all possible segmentations $\mathcal{T}$. The remainder of this section derives an instance of the EM algorithm by modeling the dependencies between the nuisance parameters θ, the registration parameters $\mathcal{R}$, the hidden labelmap $\mathcal{T}$, and the observed image data $\mathcal{I}$.

To determine θ and $\mathcal{R}$ within this framework, we define the following MAP estimation problem:

$$(\hat{\theta}, \hat{\mathcal{R}}) = \arg\max_{\theta, \mathcal{R}} \log P(\theta, \mathcal{R} | \mathcal{I}). \tag{1}$$

In general, this results in a system of equations for which there is no analytical solution.

We simplify the problem by adding the labelmap $\mathcal{T}$ to the model. The unknown labelmap $\mathcal{T}$ assigns each voxel in the image to an anatomical structure. Combining $\mathcal{T}$ with the known intensity histograms of anatomical structures produces images, which then can be compared to the observed MR images $\mathcal{I}$ in order to determine the solution to Eq. (1). We can restate Eq. (1) by marginalizing with respect to all possible labelmaps $\mathcal{T}$

$$(\hat{\theta}, \hat{\mathcal{R}}) = \arg\max_{\theta, \mathcal{R}} \log \left(\sum_{\mathcal{T}} P(\theta, \mathcal{R}, \mathcal{T} | \mathcal{I}) \right). \tag{2}$$

Next, we incorporate the conditional labelmap probability $P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')$, where $(\theta', \mathcal{R}')$ are estimates of $(\hat{\theta}, \hat{\mathcal{R}})$, into Eq. (2)

$$(\hat{\theta}, \hat{\mathcal{R}}) = \arg\max_{\theta, \mathcal{R}} \log \left(\sum_{\mathcal{T}} \frac{P(\theta, \mathcal{R}, \mathcal{T} | \mathcal{I}) P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')}{P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')} \right)$$

$$= \arg\max_{\theta, \mathcal{R}} \log E_{\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}'} \left(\frac{P(\theta, \mathcal{R}, \mathcal{T} | \mathcal{I})}{P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')} \right). \tag{3}$$

The purpose of these operations is to put Eq. (3) into a form such that we can exploit the following bound derived via Jensen's Inequality (Neal and Hinton, 1998)

$$\log E_{\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}'} \left(\frac{P(\theta, \mathcal{R}, \mathcal{T} | \mathcal{I})}{P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')} \right) \geq E_{\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}'} \left(\log \frac{P(\theta, \mathcal{R}, \mathcal{T} | \mathcal{I})}{P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')} \right). \tag{4}$$

---

[1] It is common in discussions of statistical estimation to refer to parameters, which are not of primary importance, as nuisance parameters in formulating an estimation problem. For the purposes of this work, the relevant nuisance parameters are the values of the intensity correction at every voxel.

The right side of Eq. (4) defines a lower bound on the objective function (the left side). The EM approach uses this lower bound to find a local maximum to the objective function. The Expectation Step (E-Step) establishes a new objective function or the lower bound (McLachlan and Krishnan, 1997)

$$Q[(\theta,\mathcal{R});(\theta',\mathcal{R}')] \triangleq E_{\mathcal{T}|\mathcal{I},\theta',\mathcal{R}'}\left(\log \frac{P(\theta,\mathcal{R},\mathcal{T}|\mathcal{I})}{P(\mathcal{T}|\mathcal{I},\theta',\mathcal{R}')}\right). \qquad (5)$$

The Maximization Step (M-Step) updates the approximations θ' and $\mathcal{R}'$ with the maximum of the lower bound

$$(\theta',\mathcal{R}') \leftarrow \arg\max_{\theta,\mathcal{R}} Q[(\theta,\mathcal{R});(\theta',\mathcal{R}')]. \qquad (6)$$

For many problems, Eq. (6) defines a simpler maximization problem than Eq. (3).

Finding the solution to Eq. (6) is still complicated, as we simultaneously estimate the nuisance parameters θ and registration parameters $\mathcal{R}$. We will make certain simplifying assumptions. Before we do so, we rewrite Eq. (6) by applying Bayes' rule

$$(\theta',\mathcal{R}') \leftarrow \arg\max_{\theta,\mathcal{R}} E_{\mathcal{T}|\mathcal{I},\theta',\mathcal{R}'}(\log P(\mathcal{I}|\mathcal{T},\theta,\mathcal{R}) + \log P(\mathcal{R}|\mathcal{T},\theta)$$
$$+ \log P(\theta|\mathcal{T}) - \log P(\mathcal{T}|\mathcal{I},\theta',\mathcal{R}') - \log P(\mathcal{I}) + \log P(\mathcal{T}))$$

and dropping terms that do not depend on $(\theta,\mathcal{R})$

$$(\theta',\mathcal{R}') \leftarrow \arg\max_{\theta,\mathcal{R}} E_{\mathcal{T}|\mathcal{I},\theta',\mathcal{R}'}(\log P(\mathcal{I}|\mathcal{T},\theta,\mathcal{R}) + \log P(\mathcal{R}|\mathcal{T},\theta)$$
$$+ \log P(\theta|\mathcal{T})). \qquad (7)$$

The optimization procedure decomposes nicely as a consequence of the following independence assumptions (see Fig. 2): First, we assume that $\mathcal{I}$ is independent of $\mathcal{R}$ conditioned on $\mathcal{T}$ and θ. We therefore characterize each anatomical structure with a stationary intensity distribution which is not influenced by the mapping between atlas and image space such as described by the models (Wells et al., 1996; Van Leemput et al., 1999; Kapur, 1999; Ashburner and Friston, 2005). Based on the same argument, we also assume $\mathcal{R}$ is independent of θ conditioned $\mathcal{T}$. Finally, we assume independence of θ and $\mathcal{T}$.[2] Note that in the graphical model shown in Fig. 2, the link joining the registration parameters $\mathcal{R}$ with the labelmap $\mathcal{T}$ could be reversed by the use of Bayes' rule, which would yield an equivalent model (this conditioning of $\mathcal{T}$ on $\mathcal{R}$ is used below in the development of the EM iteration).

Based on the three independence assumptions, Eq. (7) simplifies to

$$(\theta',\mathcal{R}') \leftarrow \arg\max_{\theta,\mathcal{R}} E_{\mathcal{T}|\mathcal{I},\theta',\mathcal{R}'}(\log P(\mathcal{I}|\mathcal{T},\theta) + \log P(\mathcal{R}|\mathcal{T})$$
$$+ \log P(\theta)). \qquad (8)$$

We demonstrate the utility of the independence assumptions in the experiment of Section 5 when we segment the MR images of 22 subjects into white matter, gray matter, cerebrospinal fluid, and cerebrospinal structures.

In its current form, Eq. (8) contains the expected value, which sums over all possible settings of $\mathcal{T}$. To further simplify the problem, we assume spatial independence of $\mathcal{T}$ and independence

---

[2] Note, if we incorporate a nuisance parameter model into our algorithm that does not fulfill these independence requirements, e.g., we use spatial varying intensity distributions as suggest by Fischl et al. (2004), then we perform the maximization for the nuisance and registration parameters sequentially. The resulting algorithm belongs to the class of generalized-EM algorithms, which have the same convergence properties as EM.
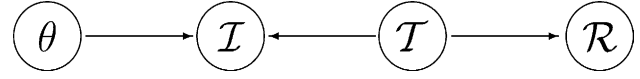

Fig. 2. A graphical model defining our Bayesian framework as presented in Eq. (8). In this model, the observed image data $\mathcal{I}$ is directly linked to the labelmap $\mathcal{T}$ and the nuisance parameters θ. In addition, the labelmap $\mathcal{T}$ is linked to the registration parameter $\mathcal{R}$.

between $\mathcal{R}$ and θ on the basis of the previous discussion. Eq. (8), which characterizes the EM approach, is restated as the sum over all voxel locations x of the expected value with respect to indicator random vector $\mathcal{T}_x \in \{e_1, \ldots, e_N\}$. The vector e_a is zero at every position but a, where its value is one. For example, if $\mathcal{T}_x = e_a$ then voxel x is assigned to the anatomical structure a. If we now assume that the two conditional probabilities $P(\mathcal{I}|\mathcal{T},\theta)$ and $P(\mathcal{R}|\mathcal{T})$ are defined by the product of the corresponding conditional probabilities over all the voxels in the image space, then we can rewrite Eq. (8) as

$$(\theta',\mathcal{R}') \leftarrow \arg\max_{\theta,\mathcal{R}} \sum_x E_{\mathcal{T}_x,\mathcal{I},\theta',\mathcal{R}'}(\log P(\mathcal{I}|\mathcal{T},\theta)$$
$$+ \log P(\mathcal{R}|\mathcal{T}) + \log P(\theta))$$
$$= \arg\max_{\theta,\mathcal{R}} \sum_x \sum_a P(\mathcal{T}_x = e_a|\mathcal{I},\theta',\mathcal{R}')$$
$$\times [\log P(\mathcal{I}_x|\mathcal{T}_x = e_a,\theta_x) + \log P(\mathcal{R}|\mathcal{T}_x = e_a)$$
$$+ \log P(\theta)]$$
$$= \arg\max_{\theta,\mathcal{R}} \sum_x \sum_a \mathcal{W}_x(a)[\log P(\mathcal{I}_x|\mathcal{T}_x = e_a,\theta_x)$$
$$+ \log P(\mathcal{R}|\mathcal{T}_x = e_a) + \log P(\theta)]. \qquad (9)$$

The weights, which will be calculated in the E-Step instead of Eq. (5), capture the posterior probability of the structure a being present at voxel x and are now defined as

$$\mathcal{W}_x(a) \triangleq P(\mathcal{T}_x = e_a|\mathcal{I},\theta',\mathcal{R}')$$
$$= \frac{P(\mathcal{I}_x|\mathcal{T}_x = e_a,\theta'_x) \cdot P(\mathcal{T}_x = e_a|\mathcal{R}')}{P(\mathcal{I}_x|\theta'_x,\mathcal{R}')}. \qquad (10)$$

As each factor in Eq. (9) only depends on either $\mathcal{R}$ or θ, the M-Step updates the approximations of the inhomogeneities and the registration parameters separately

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_x \sum_\alpha \mathcal{W}_x(a) \cdot \log P(\mathcal{T}_x = e_a|\mathcal{R})$$
$$+ \log P(\mathcal{R}) \qquad (11)$$

$$\theta' \leftarrow \arg\max_{\theta} \sum_x \sum_\alpha \mathcal{W}_x(a) \cdot \log P(\mathcal{I}|\mathcal{T}_x = e_a,\theta)$$
$$+ \log P(\theta). \qquad (12)$$

The EM algorithm iterates between the E-Step (Eq. (10)) and the M-Steps (Eqs. (11) and (12)) until the lower bound function converges to a local extrema, which is guaranteed by the EM framework if the iteration sequence has an upper bound (McLachlan and Krishnan, 1997).

Based on Eq. (11), the algorithm adjusts the registration parameters so that the atlas matches the weights defined by the E-Step. However, the weights might mislead the algorithm in cases where the alignment between atlas and image coordinate system is difficult. For example, the MR image of Fig. 3 shows a subject with an unusual head position. In this case, the weights wrongly assert that the neck is part of the brain. This causes the algorithm
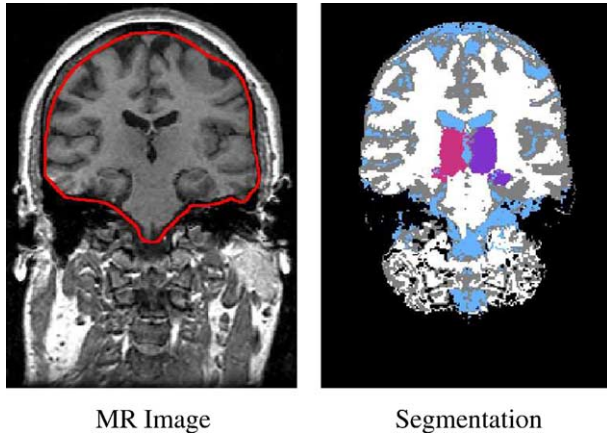
MR Image   Segmentation

Fig. 3. For this MR image, our approach achieved a suboptimal result as it failed to differentiate the neck from the brain (outlined by the line). We refer to the text for a detailed discussion.

to adjust the nuisance parameters so that intensity histogram of the neck is equivalent to the ones inside the brain. The registration parameters scale the atlas of the brain to cover both brain and neck. The algorithm, therefore, converges to a suboptimal solution, which does not separate the neck from the brain.

In summary, this section defined the problem of mapping the atlas to the image space as a MAP estimation problem. Finding the solution to this problem is generally too difficult, so we iteratively solve the simpler lower bound optimization problem of Eq. (4) in an instance of an EM algorithm. Even computing the solution to the optimization problem is generally too complicated as we have to simultaneously maximize the underlying objective with respect to the nuisance parameters θ and the registration parameters $\mathcal{R}$. We therefore split Eq. (4) into two separate estimation problems for θ and $\mathcal{R}$ by making certain simplifying independence assumptions. The resulting EM approach is composed of an E-Step determining the $\mathcal{W}$ of Eq. (10) and the M-Step, which solves the estimation problems of Eqs. (11) and (12). While the solution to Eq. (12) has been discussed in detail in the literature (Wells et al., 1996; Leventon, 2000; Kapur, 1999; Marroquin et al., 2003; Ashburner and Friston, 2005), the next section focuses on determining the registration parameters of Eq. (11).

A hierarchical registration model

To solve the estimation problem of Eq. (11), we define the registration parameters $\mathcal{R}$, the conditional log probability log $P(\mathcal{T}|\mathcal{R})$, and the prior probability $P(\mathcal{R})$. We model the parameters $\mathcal{R}$ as a hierarchical registration framework, which distinguishes between global- and structure-dependent deformations. We then apply the registration framework to the estimation problem and determine its solution with another optimization algorithm.

The hierarchical registration parameters $\mathcal{R}$ capture the correspondence between atlas, brain, and structures within the brain. The mapping of the atlas to the image space is performed by an interpolation function $r(\mathcal{R}, \cdot x)$, which maps voxel x into the coordinate system defined by $\mathcal{R}$. $r(\mathcal{R}, \cdot)$ can be rigid, affine, or more general non-rigid. As a rigid registration can be interpreted as a special case of an affine transformation, we focus our discussion on affine and non-rigid registration methods.

The parameters $\mathcal{R}$ can be structure-dependent or -independent. The structure-independent parameters capture the correspondence between atlas and image space without knowledge of the underlying brain anatomy. If constrained to affine interpolation, the degrees of freedom of $\mathcal{R}$ are too low to capture the characteristics of individual brain structures (Pohl et al., 2002; Srivastava et al., 2004). The alternative is a more general non-rigid framework, which often has problems aligning structures with weakly visible boundaries (Pohl et al., 2004b). Instead, we favor an approach that registers each anatomical structure based on the specific requirements of that structure.

Our structure-dependent registration parameters treat the relationship between the atlas and image space for each structure independently. However, this independence assumption does not hold in practice. Our registration problem is mostly characterized by the structure-independent misalignment between the coordinate system of the image space (represented by the position of the head) and the atlas coordinate system. Thus, the difference between different structure-dependent parameters should be relatively small. We therefore propose a registration framework that consists of structure-dependent and -independent parameters.

We model dependency across structures with a hierarchical registration framework $\mathcal{R} = (\mathcal{R}_G, \mathcal{R}_C)$ as shown in Fig. 4. $\mathcal{R}_G$ are the global registration parameters, which describe the non-structure-dependent deformations between atlas and image. The

Atlas Space   Image Space   Local Deformation

$$\mathbf{R_G} \qquad \mathbf{R_a}$$
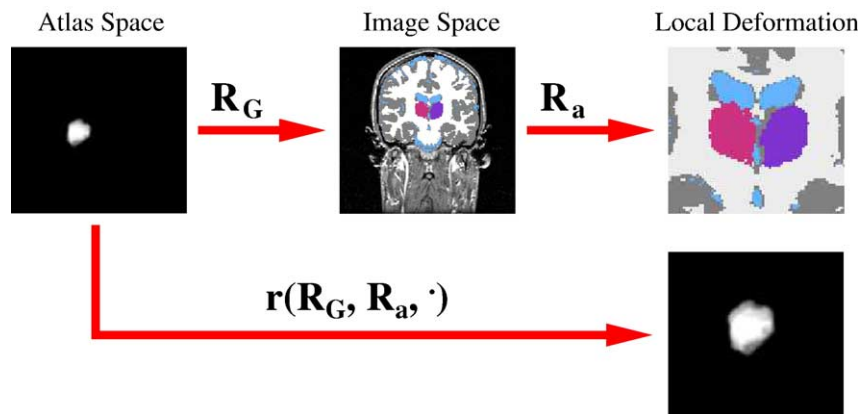
$$\mathbf{r(R_G, R_a, \cdot)}$$

Fig. 4. The graph represents the hierarchical affine registration framework developed throughout this section. In this approach, the registration parameter $\mathcal{R}_G$ captures the global deformations of the atlas coordinate system to the image space. $\mathcal{R}_a$ represents the structure-specific alignment of structure a. The atlas is aligned to the structure-specific coordinate system in the image space by applying the interpolation function $r(\mathcal{R}_G, \mathcal{R}_a, \cdot)$.
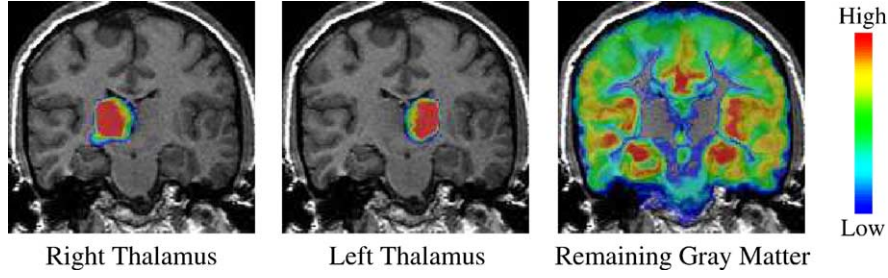
Fig. 5. Examples of space conditioned probabilities of anatomical structures. In our framework, these probabilities are represented by $f_a(\cdot)$. Blue indicates low and red high probability with respect to presence of a structure at that voxel location.

structure-dependent parameters $\mathcal{R}_C \triangleq (\mathcal{R}_1, \ldots, \mathcal{R}_N)$ are defined in relation to $\mathcal{R}_G$ and capture the residual structure-specific deformations that are not adequately explained by $\mathcal{R}_G$. In the remainder of this work, we refer to $\mathcal{R}_a$, the ath entry of $\mathcal{R}_C$, as the registration parameters specific to structure a with $a \in \{1, \ldots, N\}$.

The similarity between structure-specific parameters is encoded in Eq. (11) through the prior probability $P(\mathcal{R})$. Section 4 defines $\mathcal{R}$ explicitly and models $P(\mathcal{R})$ with a simple Gaussian distribution. This model enforces tight bounds on $\mathcal{R}_C$ and weak constraints on $\mathcal{R}_G$.

To further specify the model, we now define the conditional probability $P(\mathcal{T}_x = e_a | \mathcal{R})$, which is the spatial prior probability of structure a in the image space. In our model, this prior is composed by an atlas of the spatial distributions across all structures. The spatial distribution for a single structure a is represented in our model by $f_a(\cdot)$.

Fig. 5 shows examples of $f_a(\cdot)$ for the right and left thalamus as well as the remaining gray matter in the images space. These examples were generated by computing the overlap of between the segmentations of structures of interest according to (Warfield et al., 2001). Blue indicates low and red high probability of the structure at that voxel location.

$f_a(\cdot)$ is defined in the coordinate system of the atlas space, which is in general different from the image space. We align the atlas to the image space by making use of the interpolation function $r(\mathcal{R}_G, \mathcal{R}_a,)$. The conditional structure probability is then defined as

$$P(\mathcal{T}_x = e_a | \mathcal{R}) \triangleq \frac{f_a(r(\mathcal{R}_G, \mathcal{R}_a, x)}{\sum\limits_{a'} f_{a'}(r(\mathcal{R}_G, \mathcal{R}_{a'}, x))}, \tag{13}$$

which is the conditional probability of the presence of an individual structure normalized by the spatial prior probability

of all structures. The normalization across all structures is necessary as the coordinate system of each structure is characterized by the structure-dependent registration parameters $\mathcal{R}_a$. Unlike global affine registration methods, this results in structure-dependent coordinate systems that are not aligned with each other. In other words, multiple voxels in the atlas space can be mapped to one location in the image space.

Substituting Eq. (13) into Eq. (11) changes the MAP estimation problem to

$$\begin{aligned}
\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_x \Bigg(\sum_a \mathcal{W}_x(a) \cdot \Big(\log f_a[r(\mathcal{R}_G, \mathcal{R}_a, x)] \\
- \log \sum_{a'} f_{a'}[r(\mathcal{R}_G, \mathcal{R}_{a'}, x)] \Big) \Bigg) + \log P(\mathcal{R}) \\
= \arg\max_{\mathcal{R}} \sum_x \Bigg(\sum_a \mathcal{W}_x(a) \cdot \log f_a[r(\mathcal{R}_G, \mathcal{R}_a, x)] \Bigg) \\
\underbrace{- \log \Big(\sum_a f_a[r(\mathcal{R}_G, \mathcal{R}_a, x)] \Big) + \log P(\mathcal{R})}_{\triangleq Q(\mathcal{R})}
\end{aligned} \tag{14}$$

The spatial conditioned probabilities $f_a(r(\mathcal{R}_G, \mathcal{R}_a, x))$ are not sufficient to account for all occurrences within the model. For example, the product $\mathcal{W}_x(a) \cdot \log f_a(r(\mathcal{R}_G, \mathcal{R}_a, x))$ is not defined in the case where $\mathcal{W}_x(a) \neq 0$ and $f_a(r(\mathcal{R}_G, \mathcal{R}_a, x)) = 0$. To overcome this problem, one can simply add a uniform distribution to the spatial conditioned probabilities by replacing $\log f_a(r(\mathcal{R}_G, \mathcal{R}_a, x))$ with $\log f_a(r(\mathcal{R}_G, \mathcal{R}_a, x)) = \epsilon$. Note, that our approach is relatively insensitive to the exact value of the ϵ.

The objective function $Q(\cdot)$ measures the disagreement between $f_a(\cdot)$ and $\mathcal{W}_a()$ (see Fig. 6). We show in Appendix A that the MAP estimation is composed by the Kullback–Leibler (KL) divergence

(a) MR Image   (b) $f_a(\cdot)$ (c) $W_a(\cdot)$ (d) $Q(\cdot)$
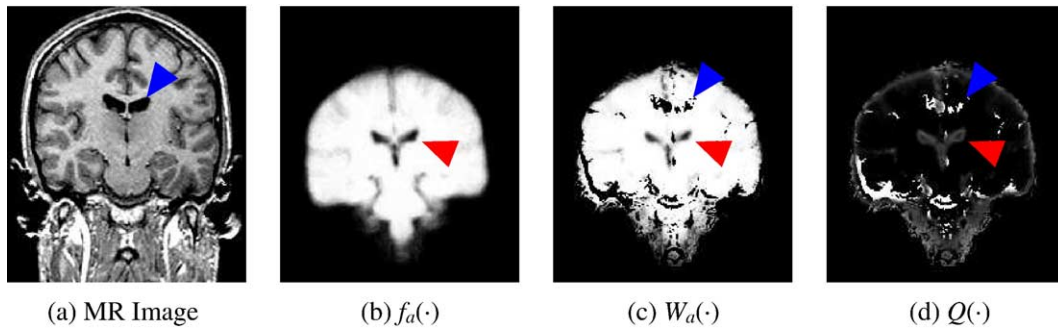
Fig. 6. We applied the EM approach to the MR image (a) and the misaligned spatial atlas of white matter and gray matter in panel (b). The result of the analysis is shown in panel (c), which is the weights $\mathcal{W}$ of the E-Step, and (d), which is the initial objective function $Q(\cdot)$ of the M-Step. The impact of the misalignment between atlas and image space is especially apparent in the region of the ventricles, whose location are pointed out in the image space and in the atlas space by arrowheads.

$D(\cdot\|\cdot)$ with respect to the labelmap $\mathcal{T}$ and the prior $P(\mathcal{R})$ of the registration parameters $\mathcal{R}$:

$$\mathcal{R}' \leftarrow \arg\min_{\mathcal{R}} D(\mathcal{W}_{\mathcal{T}} \| P(\mathcal{T}|\mathcal{R})) - \log P(\mathcal{R}).$$

Ignoring the log prior $\log P(\mathcal{R})$, the solution of the estimation problem is the registration parameters $\mathcal{R}$, which minimize the difference between the weights $\mathcal{W}_{\mathcal{T}}$ (see Appendix A) and the spatial prior $P(\mathcal{T}|\mathcal{R})$ with respect to all labelmaps $\mathcal{T}$, in the KL sense.

The example of Fig. 6 shows how the misalignment between atlas and image space impacts the analysis of the EM approach. The intensity of the images corresponds to the value of the underlying function with white indicating high and black low values. In this example, we especially focus on the region around the ventricles. Arrowheads point out their location in the MR image in (a) and the atlas space in (b), which shows the combined spatial atlas of white matter and gray matter. The misalignment between (a) and (b) impacts the weights $\mathcal{W}$ of the brain in (c) as it detects the ventricles twice—once according to the location of the image and once according to the spatial prior. The initial objective function $Q(\cdot)$ defined by the M-Step in (d) detects this ambiguity as it measures the disagreement between spatial prior of the brain $f_a(\cdot)$ in (b) and the weights $\mathcal{W}_a(\cdot)$ in (c). The objective function shows large disagreement in the area around the location of the ventricles in the image space as (b) assigns this region to the brain and (c) does not. In order to resolve this disagreement, the EM approach aligns the atlas to the image space as we will observe later in the example of Fig. 7.

In summary, we developed an EM approach based on the newly defined registration parameters $\mathcal{R}$. The E-Step calculates the weights $\mathcal{W}$ at voxel x of Eq. (10) based on the aligned spatial
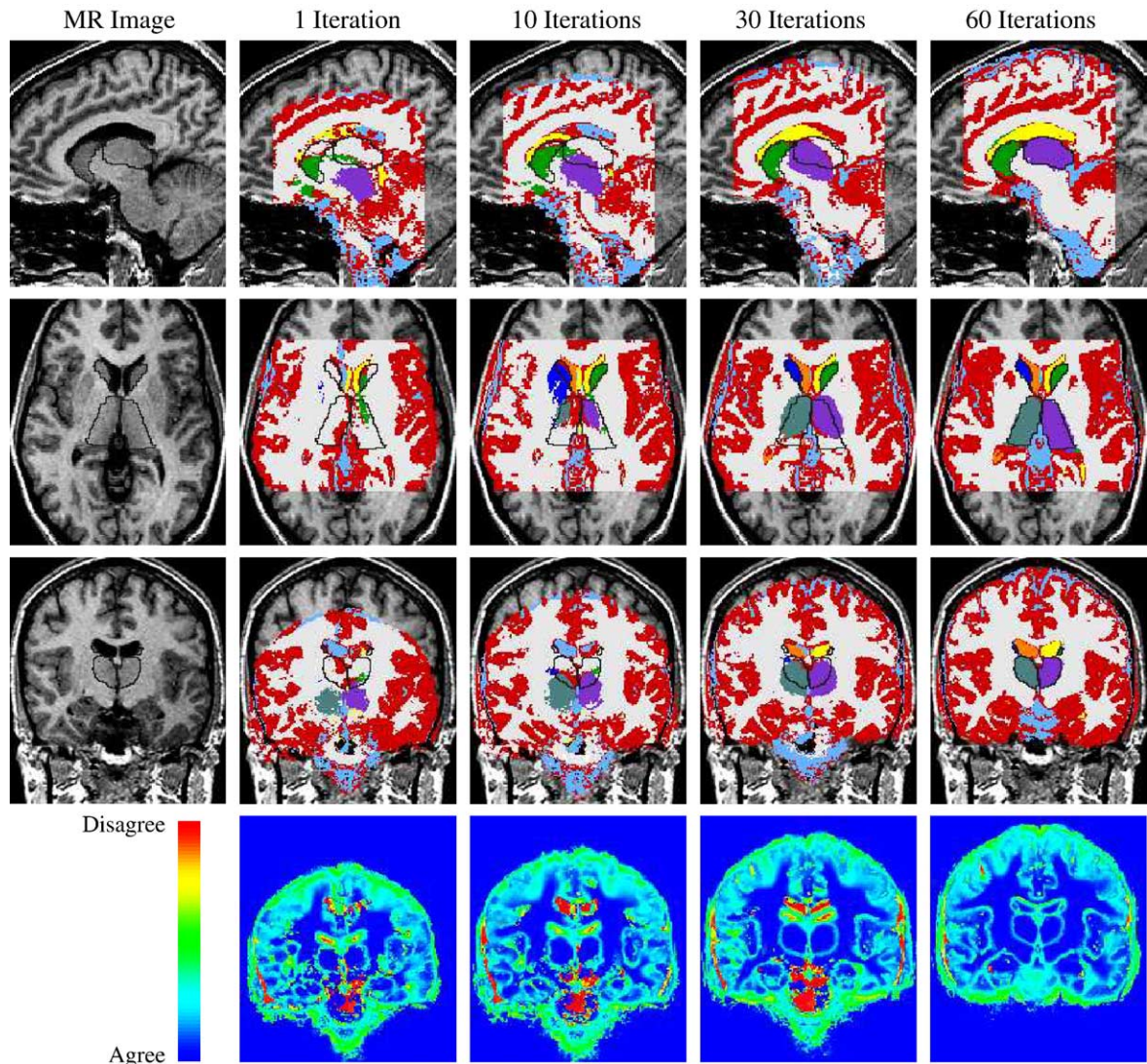

Fig. 7. The first three rows show a sagittal, axial, and coronal view of an MR volume. The thalamus and the caudate are outlined in black in these MR images. The segmentations to the right are the intermediate results of our approach. The images in the bottom row are the corresponding objective function $Q(\cdot)$ in the coronal view. Red indicates large and blue small disagreement measured by the objective function values. Initially, only the ventricles and cerebrospinal fluid close to the neck are correctly outlined, which is also indicated by red and yellow in the objective function. In the final iteration, the atlas is aligned to the image space so that disagreement is mostly due to the smooth boundaries of the spatial priors.

priors $f_a(r(\mathcal{R}_G',\mathcal{R}_a', \cdot))$, intensity $\mathcal{I}$, and the approximation of the nuisance parameters θ':

$$\mathcal{W}_x(a) = \frac{P(\mathcal{I}_x|\mathcal{T}_x = e_a,\theta_x') \cdot P(\mathcal{T}_x = e_a|\mathcal{R}')}{P(\mathcal{I}_x|\theta_x',\mathcal{R}')}$$

$$= \frac{P(\mathcal{I}_x|\mathcal{T}_x = e_a,\theta_x') \cdot f_{a'}\left(r(\mathcal{R}_G',\mathcal{R}_a',x)\right)}{\sum_{a'} P(\mathcal{I}_x|\mathcal{T}_x = e_a',\theta_x') \cdot f_{a'}\left(r(\mathcal{R}_G',\mathcal{R}_{a'}',x)\right)} \quad (15)$$

The M-Step updates the approximation of the nuisance parameters θ' and registration parameters $\mathcal{R}'$ based on the current weights $\mathcal{W}$. These parameters will be described in more detail in the next section. The update rule for $\mathcal{R}$ is defined by Eq. (14) as the maximum of objective functions $Q(\cdot)$

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} Q(\mathcal{R})$$

$$= \arg\max_{\mathcal{R}} \sum_x \left[\sum_a \mathcal{W}_x(a) \cdot \log\left(f_a[r(\mathcal{R}_G,\mathcal{R}_a,x)] + \epsilon \right) \right.$$

$$\left. - \log\left(\sum_a f_a[r(\mathcal{R}_G,\mathcal{R}_a,x)] + \epsilon \right) \right] + \log P(\mathcal{R}).$$

The solution to this update rule depends on the underlying model with respect to $\mathcal{R}_G$ and $\mathcal{R}_a$, which is discussed in detail in the next section.

An implementation of the joint registration and segmentation model

This section describes an implementation of the previously derived EM approach. We give an example for the interpolation function $r(\cdot,\cdot,\cdot)$, the corresponding registration parameters $\mathcal{R}$, the prior of the registration parameter $P(\mathcal{R})$, and the maximization algorithm to solve the MAP estimation problem defined in Eq. (14).

The interpolation function $r(\cdot,\cdot,\cdot)$ of Section 3 can represent a variety of mapping approaches. For simplicity, we choose a restricted affine interpolation function. The global and structure-specific parameters are defined as $\mathcal{R}_z = \left(\vec{t_z^t}, \vec{r_z^t}, \vec{s_z^t}\right)^t \in \mathbb{R}^9$ with $z \in \{G, 1,\ldots,N\}$, where G is the index of the global parameters and $\{1,\ldots,N\}$ represents the list of indices for the structure-specific parameters. The parameters capture displacement $\vec{t_z} \in \mathbb{R}^3$, rotation $\vec{r_z} \in \mathbb{R}^3$, and scaling $\vec{s_z} \in \mathbb{R}^3$. The mapping is defined by the function

$$r(\cdot,\cdot,\cdot) : \mathbb{R}^{9 \times 9 \times 3} \rightarrow \mathbb{R}^3, (\mathcal{R}_G,\mathcal{R}_a,x) \rightarrow A_{\mathcal{R}_G} \cdot A_{\mathcal{R}_a} \cdot \left(x^t,1\right)^t$$

where $A_{\mathcal{R}_z}$ is an affine transformation matrix based on the parameter setting $\mathcal{R}_z$.

This framework makes no assumptions about the correspondence between the atlas and the image space by uniformly distributing the global registration parameter $\mathcal{R}_G$. As opposed to $\mathcal{R}_G$, the structure or class specific parameters $\mathcal{R}_C \triangleq (\mathcal{R}_1,\ldots,\mathcal{R}_N)$ describe the residual of structure-specific deformations that are not well explained by $\mathcal{R}_G$. In general, our model should penalize large deviations of $\mathcal{R}_C$ from the expected mean, which is approximated by the average structure-specific registration parameters of the training data. We enforce this penalty by modeling the probability density function of $\mathcal{R}_C$ as a Gaussian distribution $\mathcal{N}(\mu_{\mathcal{R}_C},\Upsilon_{\mathcal{R}_C})$ with structure-dependent mean $\mu_{\mathcal{R}_C}$ and variance $\Upsilon_{\mathcal{R}_C}$ based on the mapping parameters of the training data. We choose a Gaussian distribution, as small variance $\Upsilon_{\mathcal{R}_C}$ discourages large deformations from the mean $\mu_{\mathcal{R}_C}$. In addition, Gaussian distributions simplify the calculations in the M-Step (McLachlan and Krishnan, 1997).

Based on the previous modeling assumptions, the objective function $Q(\cdot)$ of Eq. (14) is defined as

$$Q(\mathcal{R}) \triangleq \sum_x \left(\sum_a \mathcal{W}_x(a) \cdot \log\left[f_a\left(A_{\mathcal{R}_G} \cdot A_{\mathcal{R}_a} \cdot (x^t,1)^t\right) + \epsilon \right] \right.$$

$$- \log\left[\sum_a f_a\left(A_{\mathcal{R}_G} \cdot A_{\mathcal{R}_a} \cdot (x^t,1)^t\right) + \epsilon \right] \right)$$

$$- \frac{1}{2}\left(\mathcal{R}_C - \mu_{\mathcal{R}_C}\right)' \Upsilon_{\mathcal{R}_C}^{-1}\left(\mathcal{R}_C - \mu_{\mathcal{R}_C}\right). \quad (16)$$

To find a solution to the MAP problem defined by $Q(\cdot)$, we decouple the search for $\mathcal{R}_G$ and $\mathcal{R}_S$ as their dependencies can cause instability. It is generally difficult to determine a closed form solution for this problem so that we estimate the solutions through the Powell's method (Press et al., 1992).

The second MAP estimation problem (Eq. (12)) of the M-Step updates the approximation of the nuisance parameters

$$\theta' \leftarrow \arg\max_{\theta} \sum_x \sum_a \mathcal{W}_x(a) \cdot \log P(\mathcal{I}|\mathcal{T}_x = e_a,\theta) + \log P(\theta).$$

In our implementation, the nuisance parameters θ are defined by the image inhomogeneity model of Wells et al. (1996). We choose this model as it is quite simple but also flexible enough for the given domain. Note that other models, such as Leventon (2000); Kapur (1999); Marroquin et al. (2003); and Ashburner and Friston (2005), are appropriate choices for the nuisance parameters. The choice of model influences the behavior of our algorithm. In this paper, however, we will not further investigate the impact of these different models on our algorithm as we are primarily interested in studying the dependencies between segmentation and registration.

One of Wells' models defines the likelihood $P(\mathcal{I}|\mathcal{T}_x = e_a,\theta)$ of Eq. (12) by the following Gaussian distribution:

$$P(\mathcal{I}_x|\mathcal{T}_x = e_a,\theta_x) \triangleq \mathcal{N}(\mathcal{I}_x - \theta_x - \mu_a,\Upsilon_a).$$

(μ_a, Υ_a) represents the mean and variance of the intensity distribution of the anatomical structure a. As a consequence of this definition, Eq. (12) reduces to a system of linear equations. In Appendix B, we explain in further detail the module of the image inhomogeneities and derive the solution of this system of linear equations.

The solution is the product of the weighted residual $\bar{R}_x = \sum_a \mathcal{W}_a(a)\Upsilon_a^{-1}(\mathcal{I}_x - \mu_a)$ and a low pass filter represented by the matrix H. The E-Step, characterized by Eq. (15), calculates the weights

$$\mathcal{W}_x(a) = \frac{\mathcal{N}(\mathcal{I}_x - \theta_x - \mu_a,\Upsilon_a) \cdot f_a(r(\mathcal{R}_G',\mathcal{R}_a',x))}{\sum_{a'} \mathcal{N}(\mathcal{I}_x - \theta_x - \mu_a,\Upsilon_a) \cdot f_{a'}(r(\mathcal{R}_G',\mathcal{R}_{a'}',x))} \quad (17)$$

The entire implementation is summarized by the pseudo code below:

Algorithm 1. Segmentation and Registration
  repeat
    E-Step: Update soft assignment according to Eq. (17)
      $\mathcal{W}_x(a) \leftarrow \frac{1}{Z}\mathcal{N}(\mathcal{I}_x - \theta_x - \mu_a,\Upsilon_a) \cdot f_a(r(\mathcal{R}_G',\mathcal{R}_a',x))$
    M-Step: Update parameter space
      $\theta' \leftarrow$ Estimation of the inhomogeneities by calculating $H \cdot \bar{R}$
      $\mathcal{R}_G' \leftarrow$ Result of Powell's method with $Q((\cdot,\mathcal{R}_S'))$
      $\mathcal{R}_S' \leftarrow$ Result of Powell's method with $Q((\mathcal{R}_G',\cdot))$
  until θ' and $\mathcal{R}'$ converge
  define labelmap: $\hat{\mathcal{T}}_x \leftarrow \arg\max_a \mathcal{W}_x(a)$

We end this section with an example illustrating the dependencies between registration and segmentation. The first three rows of

Fig. 7 show a sagittal, axial, and coronal view of a MR volume. Black lines outline the thalamus and the caudate. The segmentations to the right are the results of our algorithm after 1, 10, 30, and 60 iterations. To show the strength of our model, we initialized the approach with an unusual position of the atlas space. This also explains why the method needs 60 iterations instead of the normal 20 iterations until convergence.

The last row shows the objective function $Q(\cdot)$ corresponding to the segmentations above. Red indicates high and blue low disagreement between the atlas space and the weights of the M-Step. Initially, the area around the ventricles and cerebrospinal fluid close to the neck are highlighted in red, as they are correctly segmented regions that disagree with their location within the atlas space. As the method progresses, the overall accuracy of the registration as well as segmentation increases. The disagreement captured by the objective function in the final iteration is mostly caused by the smooth boundaries of the spatial priors.

In summary, this section described an implementation of the joint registration and segmentation algorithm. Affine registration parameters represent the global and structure-specific deformations between atlas and image space. The nuisance parameters capture the inhomogeneity of the MR images caused by the acquisition of these images. The next section discusses the accuracy of our implementation in outlining the thalamus.

Comparative study on 22 test cases

The experiment described in this section compares the accuracy of four EM implementations which differ in how they map the atlas to the patient. The first approach (EM-Nonrigid) maps the atlas to the patient using an intensity based non-rigid registration approach and then runs our EM implementation without registration parameters (Pohl et al., 2004a). The second approach (EM-Affine) is similar to EM-Nonrigid but uses the affine mapping method by (Warfield et al., 2001). The remaining two approaches are based on our new algorithm, which solves the registration and segmentation problem simultaneously. EM-Sim-Affine uses the same global affine registration model as EM-Affine and EM-Sim-Hierarchical incorporates the hierarchical registration approach of Section 3. All four methods utilize the same atlas, which captures the spatial distributions of the anatomical structures. The training data of the atlas consist of semi-automatic as well as manual segmentations of up to 80 different subjects[3]. The atlas is generated through the approach suggested by Warfield et al. (2001), which aligns the training subjects to a preselected training case and then measures the overlap between the corresponding segmentations. Examples of the atlas are shown in Fig. 5.

Each method segments 22 test cases (1.5-T MRI system (GE Medical Systems, Milwaukee), three-dimensional Fourier transformation spoiled gradient-recalled acquisition sequence, echo time (TE) = 5 ms, repetition time (TR) = 35 ms, repetition = 1, nutation angle = 45, field of view = 24 cm, acquisition matrix = $256 \times 256 \times 124$, voxel dimension = $0.9375 \times 0.9375 \times 1.5$ mm) into the three brain tissue classes – white matter, gray matter, and cerebrospinal fluid – and further parcellates gray matter into the right and left thalamus. We evaluate the accuracy of the

approaches by measuring the agreement of the automatic segmentations of the thalamus to manual ones, which we view as ground-truth. Note that our EM approach segments anatomical structure by considering all neighboring structures. Thus, the accuracy of segmenting the thalamus greatly depends on the precise segmentation of white matter, the remaining gray matter, and the cerebrospinal fluid.

We focus on the thalamus because it is a challenging structure for registration and segmentation. Many fiber tracks pass through the thalamus so that its intensity properties in MR images are very similar to the neighboring white matter (Fig. 8). Intensity based registration methods have difficulties aligning this structure because of its weakly visible boundary. In our experience, EM without spatial priors will fail and EM with spatial priors heavily relies on these priors. Thus, registration errors greatly influence the segmentation quality of our three methods.

To measure the quality of the automatic generated results, we compare them to the manual segmentations using the volume overlap measure DICE (Dice, 1945). The results of the experiment are summarized in the Fig. 9. The graph shows the mean and standard error of the DICE measure for the three algorithms in the 22 cases. For the thalamus, EM-Nonrigid performed worst ($82.35 \pm 1.2\%$ – average DICE score ± standard error) because the intensity based registration method is too unreliable for structures with smooth boundaries. The method generally overestimates white matter and underestimated the thalamus in this region, which can be also observed by the segmentation in Fig. 8 (EM-Nonrigid). EM-Affine ($87.28 \pm 1.2\%$) performs much better than EM-Nonrigid, but the method is sensitive towards initial misalignments. In Fig. 8 (EM-Affine), this sensitivity causes a vertical offset in the segmentation generated by EM-Affine. Note that the average DICE score of EM-Affine for the thalamus is already excellent as we experienced large disagreement even between different expert segmentations. However, EM-Sim-Affine and EM-Sim-Hierarchical, the approaches based on our new framework, achieve a higher average score than the other two methods.

EM-Sim-Affine is more robust ($88.65 \pm 0.4\%$) than EM-Affine even though both methods use the same affine registration framework. This empirically demonstrates the utility of performing registration in an EM segmenter. EM-Sim-Affine is constrained by the dependency between segmentation and registration. In comparison to EM-Affine, these constraints further reduce the space of possible solutions, which simplifies the search for the optimal solution. In addition, the spatial priors of the structures are directly mapped to the segmentation model. In contrast, EM-Affine and EM-non rigid align an MR image in the atlas space to the image of the patient. They use the resulting deformation map to align the spatial priors, which inherently increases the risk of systematic biases in the model.

EM-Sim-Hierarchical ($89.40 \pm 0.3\%$), however, significantly outperforms the other three approaches. The hierarchical registration framework has greater flexibility than the single global affine registration models of EM-Affine and EM-Sim-Affine. We believe that the structure-specific registration framework of EM-Sim-Hierarchical better incorporates the anatomical knowledge gained from the segmentation domain than the continuous model of EM-Sim-Affine.

We end the discussion by further investigating the previously made observation with respect to the ventricles. As pointed out in

---

[3] The data were provided by the Schizophrenia Research Project, Surgical Planning Laboratory, Brigham and Women's Hospital, Harvard.
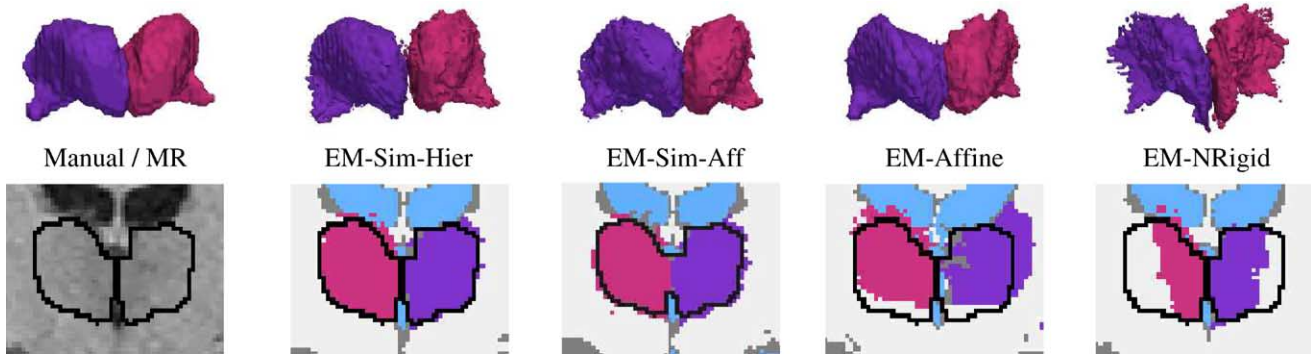
Fig. 8. The images zoom in on the automatic segmentations of the thalamus region. In the MR image as well as 2D segmentations, the thalamus is outlined in black. The registration results of EM-Affine shows an offset in the vertical direction and EM-Nonrigid (EM-NRigid) greatly underestimates the thalamus. Our two new approach EM-Sim-Hierarchical (EM-Sim-Hier) and EM-Affine (EM-Sim-Aff) achieve the most accurate segmentations, which was generally true in the 22 test cases (see Fig. 9).

Section 3, the ventricles are very important in guiding the joint registration and segmentation method. In order to validate this statement, we repeat the previous experiment. This time gray matter, white matter, and the ventricles are combined into one structure. This change in our model causes the registration cost function to ignore the misalignment of the location of the ventricles between atlas and image space (see also Fig. 6). As a result, the average DICE score for the thalamus reduces by 5.2% and the standard error increases by a factor of 2.5. We therefore conclude that the accuracy of our new approach greatly depends on anatomical structures that are easily identifiable in the segmentation problem.

We have demonstrated that our new approach EM-Sim-Hierarchical and EM-Sim-Affine achieve higher accuracy than the methods we denoted as EM-Affine and EM-Nonrigid as a consequence of the fact that our approach directly maps the spatial priors of the structures to the segmentation model. In contrast, EM-Affine and EM-Nonrigid align an MR image in the atlas space to the image of the patient (Pohl et al., 2004a), using the resulting deformation map to align the spatial priors. This inherently increases the risk of systematic biases in the model. Furthermore, EM-Sim-Hierarchical performs significantly better then the other three methods as the hierarchical registration model utilizes well the anatomical knowledge gained from the segmentation domain.

Conclusion

We presented a statistical framework combining atlas registration and segmentation of MR images. Unlike other voxel-based classification methods, our framework models these problems as a single MAP estimation problem. We implemented the framework as an instance of an EM algorithm using a hierarchical affine mapping approach for anatomical structures. Our approach was validated by automatically segmenting 22 sets of MR images into the major brain tissue classes and the thalamus, a structure with indistinct boundaries. Using manual segmentations, we then compared our results to other EM implementations which sequentially register and segment. The study empirically demonstrated the utility of combining registration and segmentation in an EM framework. In addition, our hierarchical registration approach achieved a significantly higher score than the single continuous registration models.
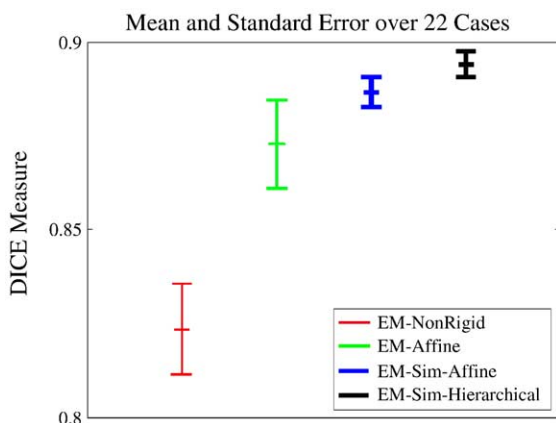
Fig. 9. The graph shows the average DICE score over 22 cases for the pipeline based approaches EM-Affine and EM-Non rigid, and approaches based on our integrated framework EM-Sim-Affine and EM-Sim-Hierarchical. EM-Sim-Affine achieves a higher score than the pipeline based approaches and EM-Sim-Hierarchical significantly outperforms the other three approaches for the thalamus.

Appendix A. Relationship between MAP estimate of Registration Parameters and KL Divergence

We now derive the equality between the MAP estimation problem of Eq. (14) and minimization of the KL divergence $D(\cdot\|\cdot)$ with respect to the labelmap $\mathcal{T}$:

$$\arg\max_{\mathcal{R}} \sum_{x \in \mathcal{M}} \left[\sum_a P(\mathcal{W}_x(a)) \cdot \log P(\mathcal{T}_x = e_a | \mathcal{R}) \right] + \log P(\mathcal{R})$$

$$= \arg\min_{\mathcal{R}} D(\mathcal{W}_\mathcal{T} \| P(\mathcal{T}|\mathcal{R})) - \log P(\mathcal{R})$$

where $\mathcal{W}_T$ is the product of the weights $\mathcal{W}_x$ over the entire image space conditioned on $\mathcal{T}$.

For simplification, we use the following notation for the sum over all labels at the voxel x

$$\sum_{\mathcal{T}_x} P(\mathcal{T}_x | \mathcal{I}, \theta', \mathcal{R}') \cdot \log P(\mathcal{T}_x | \mathcal{R})$$
$$\triangleq \sum_a P(\mathcal{T}_x = e_a | \mathcal{I}, \theta', \mathcal{R}') \cdot \log P(\mathcal{T}_x = a | \mathcal{R}).$$

In addition, we represent the set of voxels of the image space with $\mathcal{M} = \{1, \ldots, n\}$. The MAP estimation of Eq. (14) is therefore defined as

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_{x \in \mathcal{M}} \left[\sum_{\mathcal{T}_x} P(\mathcal{T}_x | \mathcal{I}, \theta', \mathcal{R}') \cdot \log P(\mathcal{T}_x | \mathcal{R}) \right] + \log P(\mathcal{R}).$$

The total probability theorem states that

$$\sum_{\mathcal{T}_1, \ldots, \mathcal{T}_{x-1}, \mathcal{T}_{x+1}, \ldots, \mathcal{T}_n} \prod_{y \in \mathcal{M} \setminus \{x\}} P(\mathcal{T}_y | \mathcal{I}, \theta', \mathcal{R}') = 1,$$

where we use the notation $\sum_{\mathcal{T}_1, \ldots, \mathcal{T}_z} \triangleq \sum_{\mathcal{T}_1, \ldots \mathcal{T}_z}$. We now extend the product of the estimation problem with the previous sum and apply the following fact $a \cdot \sum_l b_l = \sum_l a \cdot b_l$:

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_{x \in \mathcal{M}} \left[\sum_{\mathcal{T}_x} P(\mathcal{T}_x | \mathcal{I}, \theta', \mathcal{R}') \cdot \log P(\mathcal{T}_x | \mathcal{R}) \right.$$
$$\left. \cdot \left(\sum_{\mathcal{T}_1, \ldots, \mathcal{T}_{x-1}, \mathcal{T}_{x+1}, \ldots, \mathcal{T}_n} \prod_{y \in \mathcal{M} \setminus \{x\}} P(\mathcal{T}_y | \theta', \mathcal{R}') \right) \right] + \log P(\mathcal{R})$$
$$= \arg\max_{\mathcal{R}} \sum_{x \in \mathcal{M}} \left[\sum_{\mathcal{T}_x} \sum_{\mathcal{T}_1, \ldots, \mathcal{T}_{x-1}, \mathcal{T}_{x+1}, \ldots, \mathcal{T}_n} \log P(\mathcal{T}_x | \mathcal{R}) \right.$$
$$\left. \cdot \prod_{y \in \mathcal{M}} P(\mathcal{T}_y | \mathcal{I}, \theta', \mathcal{R}') \right] + \log P(\mathcal{R})$$

We rephrase the estimation problem by combining the sums over the labeling of the image

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_{x \in \mathcal{M}} \left[\sum_{\mathcal{T}_1, \ldots, \mathcal{T}_n} \prod_{y \in \mathcal{M}} P(\mathcal{T}_y | \mathcal{I}, \theta', \mathcal{R}') \cdot \log P(\mathcal{T}_x | \mathcal{R}) \right]$$
$$+ \log P(\mathcal{R}).$$

We replace $\sum_{\mathcal{T}} \triangleq \sum_{\mathcal{T}_1, \ldots, \mathcal{T}_n}$ and again make use of the fact $\sum_l a \cdot b_l = a \cdot \sum_l b_l$:

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_{\mathcal{T}} \left(\prod_{y \in \mathcal{M}} P(\mathcal{T}_y | \mathcal{I}, \theta', \mathcal{R}') \cdot \left[\sum_{x \in \mathcal{M}} \log P(\mathcal{T}_x | \mathcal{R}) \right] \right)$$
$$+ \log P(\mathcal{R}).$$

In Section 2, we assumed spatial independence of the labelmap $\mathcal{T}$ so that

$$\prod_{y \in \mathcal{M}} P(\mathcal{T}_y | \mathcal{I}, \theta', \mathcal{R}') = P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')$$

and

$$\sum_{x \in \mathcal{M}} \log P(\mathcal{T}_x | \mathcal{R}) = \log \prod_{x \in \mathcal{M}} P(\mathcal{T}_x | \mathcal{R}) = \log P(\mathcal{T} | \mathcal{R}).$$

Thus, the estimation problem changes to

$$\mathcal{R}' \leftarrow \arg\max_{\mathcal{R}} \sum_{\mathcal{T}} P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}') \cdot \log P(\mathcal{T} | \mathcal{R})$$
$$+ \log P(\mathcal{R}) = \arg\max_{\mathcal{R}}$$
$$- \sum_{\mathcal{T}} \left[P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}') \left(\log \frac{P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')}{P(\mathcal{T} | \mathcal{R})} - \log P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}') \right) \right]$$
$$+ \log P(\mathcal{R})$$

We drop from the estimation problem the constant $\log P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')$ with respect to $\mathcal{R}$ and replace the summation over all possible labelmaps with the KL divergence:

$$\mathcal{R}' \leftarrow \arg\min_{\mathcal{R}} \sum_{\mathcal{T}} \left[P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}') \log \frac{P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')}{P(\mathcal{T} | \mathcal{R})} \right] - \log P(\mathcal{R})$$
$$= \arg\min_{\mathcal{R}} D(P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}') \| P(\mathcal{T} | \mathcal{R})) - \log P(\mathcal{R})$$

If we now define the weights with respect to a specific labelmap $\mathcal{T}$ as

$$\mathcal{W}_T \triangleq \prod_x \mathcal{W}_x^T \mathcal{T}_x = \prod_x P(\mathcal{T}_x | \mathcal{I}, \theta', \mathcal{R}') = P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}')$$

then the KL divergence

$$D(P(\mathcal{T} | \mathcal{I}, \theta', \mathcal{R}') \| P(\mathcal{T} | \mathcal{R})) = D(\mathcal{W}_T \| P(\mathcal{T} | \mathcal{R}))$$

measures the disagreement between the weights and spatial priors, in the KL sense.

Appendix B. Modeling Image Inhomogeneities

One set of parameters determined in the M-Step of our algorithm is the estimate of the image inhomogeneity θ'. The estimate θ' is updated at each iteration according to Eq. (12)

$$\theta' \leftarrow \arg\max_{\theta} \sum_x \sum_a \mathcal{W}_x(a) \cdot \log P(\mathcal{I} | \mathcal{T}_x = e_a, \theta) + \log P(\theta).$$

In this appendix, we describe a closed form of the estimator and a fast approximation of it.

We use a model of the intensity inhomogeneity that is based on Wells et al. (1996). In addition, the probability of the intensity of image $\mathcal{I}$ at voxel x conditioned on the anatomical structure and the image inhomogeneity is modeled by the Gaussian distribution $P(\mathcal{I} | \mathcal{T}_x = e_a, \theta) \triangleq \mathcal{N}(\mathcal{I}_x - \theta_x - \mu_a, \Upsilon_a)$. The prior of the image inhomogeneity $P(\theta) \triangleq \mathcal{N}(\theta, \Upsilon_B)$ is defined as a very large zero-mean multi-variate Gaussian that has a value at every voxel. The structure-specific intensity parameters (μ_a, Υ_a), and the covariance of the image inhomogeneity model Υ_B, are not adjusted within the EM iteration.

We now specify Eq. (12) with respect to this model. The log conditional probabilities of Eq. (12) are now

$$\log P(\mathcal{I} | \mathcal{T}_x = e_a, \theta) = \log \left(\frac{1}{Z} e^{-\frac{1}{2}(\mathcal{I}_x - \theta_x - \mu_a)' \Upsilon_a^{-1} (\mathcal{I}_x - \theta_x - \mu_a)} \right)$$
$$= -\log Z - \frac{1}{2}(\mathcal{I}_x - \theta_x - \mu_a)^t \Upsilon_a^{-1} (\mathcal{I}_x - \theta_x - \mu_a)$$

and

$$\log P(\theta) = -\log Z - \frac{1}{2} \theta^t \Upsilon_B^{-1} \theta.$$

Defining

$$Q_B(\theta) \triangleq \sum_x \sum_a \mathcal{W}_x(a) \cdot (\mathcal{I}_x - \theta_x - \mu_a)^t \Upsilon_a^{-1} (\mathcal{I}_x - \theta_x - \mu_a)$$
$$+ \theta^t \Upsilon_B^{-1} \theta,$$

Eq. (12) becomes

$$\theta' \leftarrow \arg\max_\theta Q_B(\theta). \tag{B.1}$$

An attraction of these Gaussian modeling assumptions is that Eq. (B.1) can be solved in closed form. The objective function of Eq. (B.1) is defined as so that the necessary zero gradient condition of the maximum θ' at each voxel x is expressed by

$$0 = \frac{\partial}{\partial \theta_x} Q_B(\theta)_{\theta = \theta'}. \tag{B.2}$$

For each voxel x Eq. (B.2) resolves to a system of linear equations:

$$0 = \frac{\partial}{\partial \theta_x} Q_B(\theta) = \sum_a \mathcal{W}_x(a) \cdot \Upsilon_a^{-1} \cdot (\mathcal{I}_x - \theta_x - \mu_a) + [\Upsilon_B^{-1} \cdot \theta]_x$$

$$= \sum_\alpha \mathcal{W}_x(a) \cdot \Upsilon_a^{-1} (\mathcal{I}_x - \mu_a)$$

$$- \left(\sum_\alpha \mathcal{W}_x(a) \cdot \Upsilon_a^{-1} \right) \cdot \theta_x - [\Upsilon_B^{-1} \cdot \theta]_x. \tag{B.3}$$

If we now define the weighted residual at voxel x as

$$\bar{R}(x) \triangleq \sum_a \mathcal{W}_x(a) \cdot \Upsilon_a^{-1} \cdot (\mathcal{I}_x - \mu_a) \tag{B.4}$$

and the non-zero entries of the matrix $\bar{\Upsilon}^{-1}$ along the diagonal by the submatrix

$$\bar{\Upsilon}_{xx}^{-1} \triangleq \sum_a \mathcal{W}_x(a) \cdot \Upsilon_a^{-1}$$

then Eq. (B.3) simplifies to

$$\bar{\Upsilon}^{-1} \cdot \theta + \Upsilon_B^{-1} \cdot \theta = \bar{R}$$

The solution to Eq. (B.1) may be written as

$$\theta' = \left(\bar{\Upsilon}^{-1} + \bar{\Upsilon}_B^{-1} \right)^{-1} \cdot \bar{R} \tag{B.5}$$

Even though Eq. (B.5) defines a closed form solution to our initial estimation problem of Eq. (12) it is, in general, computationally very challenging to solve a linear system of this size (the voxel count). However, a major savings in computation can be achieved when modeling assumptions are made so that $(\bar{\Upsilon}^{-1} + \bar{\Upsilon}_B^{-1})^{-1}$, which is a linear operator on the residual image, is a spatially stationary unity-gain low-pass filter.

This approach depends on the assumption that each structure a has the same variance $\Upsilon_a = \tilde{\Upsilon}$. Eq. (B.3) now simplifies to

$$0 = \sum_\alpha \mathcal{W}_x(a) \cdot \tilde{\Upsilon}^{-1} (\mathcal{I}_x - \mu_a) - \left(\sum_\alpha \mathcal{W}_x(a) \tilde{\Upsilon}^{-1} \right) \cdot \theta_x - [\Upsilon_B^{-1} \cdot \theta]_x$$

$$= \tilde{\Upsilon}^{-1} \cdot \left(\sum_\alpha \mathcal{W}_x(a) \cdot (\mathcal{I}_x - \mu_a) - \left(\sum_\alpha \mathcal{W}_x(a) \right) \cdot \theta_x - \tilde{\Upsilon} \cdot [\Upsilon_B^{-1} \cdot \theta]_x \right)$$

$$= \tilde{\Upsilon}^{-1} \cdot \left(\sum_a \mathcal{W}_x(a) \cdot (\mathcal{I}_x - \mu_x) - \theta_x - \tilde{\Upsilon} \cdot [\Upsilon_B^{-1} \cdot \theta]_x \right). \tag{B.6}$$

In the case of unequal class variances it is not feasible to implement, in this style, an exact estimator that has spatial stationarity, however, it has proven to be a useful approximation in such applications.

We now define the weighted residual for the simplified case as

$$\bar{R}(x) \triangleq \sum_a \mathcal{W}_x(a) \cdot (\mathcal{I}_x - \mu_a) \tag{B.7}$$

and denote with I the identity matrix. Thus, Eq. (B.6) becomes

$$\theta + \tilde{\Upsilon} \otimes \Upsilon_B^{-1} \cdot \theta = \bar{R}$$

($\otimes$ is the Kronecker product) and the solution for θ is defined as

$$\theta = \left(I + \tilde{\Upsilon} \otimes \Upsilon_B^{-1} \right)^{-1} \cdot \bar{R}$$

Within the EM iterations, the residual $\bar{R}$ represents an estimate of the image inhomogeneities that is based on the current results of the E-Step as captured by the weights $\mathcal{W}$. One important characteristic of image inhomogeneities is that they are slowly varying functions of space. $\bar{R}$, however, is corrupted by the white noise in the image and therefore does not share this property. Consequently, one may view $(I + \tilde{\Upsilon} \otimes \Upsilon_B^{-1})^{-1}$ as acting like a low pass filter in order for θ' to have the properties of an image inhomogeneity field. We characterize such a filter by the following properties of its frequency response (the Fourier transform of its impulse response): it has value one at spatial frequency zero and the magnitude decreases as the frequency moves away from zero (a *unity DC gain* low pass filter).

We now analyze the impact of this characterization on our statistical model. We focus on the prior probability $P(\theta)$. First, we assume that $\tilde{\Upsilon} = c \cdot I$ to simplify the analysis. Let $H \approx (I + c \cdot \Upsilon_B^{-1})^{-1}$, then $\Upsilon_B^{-1} \approx \frac{1}{c}(H^{-1} - I)$. Thus, the prior probability is approximated by

$$P(\theta) = \frac{1}{Z} \cdot e^{-\frac{1}{2}\theta^t \Upsilon_B^{-1} \theta} \approx \frac{1}{Z} \cdot e^{-\frac{1}{2}\theta^t \frac{1}{c}(H^{-1} - I)\theta}.$$

If, as suggested above, we choose a unity DC gain spatial low pass filter for the linear operator H, then being the inverse of the low pass filter, H^{-1} is a filter with response one at spatial frequency zero and increasing magnitudes as the frequency moves away from zero. Thus, $(H^{-1} - I)$ is a high pass filter with response zero at frequency zero and increasing values for frequencies away from zero. Bearing these properties of the implied inverse covariance in mind, we can see that the implied prior on bias fields, $P(\theta)$, assigns the highest probability to constant bias fields (the exponent will be zero), relatively large values to θ with low spatial frequency components and decreasing probabilities for θ with higher spatial frequencies. Thus, our modeling choices imply a prior model that is consistent with observed bias fields in that it discourages high frequency components in the image inhomogeneities. In our experiments we have used rectangular smoothing filters for H. In some applications, the filter is applied several times, which approximates Gaussian smoothing.

Appendix C. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2005.11.044.

References

Ashburner, J., Friston, K., 2005. Unified segmentation. NeuroImaging 15 (13), 839–851.

Cocosco, C., Zijdenbos, A., Evans, A., 2003. A fully automatic and robust brain MRI tissue classification method. Med. Image Anal. 7, 513–527.

Collins, D., Zijdenbos, A., Barre, W., Evans, A., 1999. ANIMAL + INSECT: improved cortical structure segmentation. Inf. Process. Med., 1613.

Dempster, A., Laird, N., Rubin, D., 1977. Maximal likelihood form incomplete data via the EM algorithm. Proceedings of the Royal Statistical Society vol. 39. , pp. 1–38.

Dice, L.R., 1945. Measure of the amount of ecological association between species. Ecology 26, 297–302.

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

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Sgonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.

Kapur, T., 1999. Model based three dimensional medical imaging segmentation. PhD thesis, Massachusetts Institute of Technology.

Kikinis, R., Shenton, M.E., Gering, G., Martin, J., Anderson, M., Metcalf, D., Guttmann, C., McCarley, R.W., Lorensen, W., Line, H., Jolesz, F.A., 1992. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. Magn. Reson. Imaging 2 (6), 619–629.

Leventon, M.E., 2000. Statistical models in medical image analysis. PhD thesis, Massachusetts Institute of Technology.

Leventon, M., Grimson, W., Faugeras, O., 2000. Statistical shape influence in geodesic active contours. IEEE Conference on Computer Vision and Pattern Recognition. , pp. 1316–1323.

Marroquin, J., Santana, E., Botello, S., 2003. Hidden Markov measure field models for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1380–1387.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The international consortium for brain mapping (ICBM). NeuroImaging 2 (2), 89–101.

McLachlan, G.J., Krishnan, T., 1997. The EM Algorithm and Extensions. John Wiley and Sons, Inc..

Neal, R., Hinton, G., 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan, M. (Ed.), Learning in Graphical Models. Kluwer Academic Press, pp. 355–368.

Pohl, K., Wells, W., Guimond, A., Kasai, K., Shenton, M., Kikinis, R., Grimson, E., Warfield, S., 2002. Incorporating non-rigid registration into expectation maximization algorithm to segment MR images. Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science vol. 2488. Springer-Verlag, Tokyo, Japan, pp. 564–572.

Pohl, K., Bouix, S., Kikinis, R., Grimson, W., 2004a. Anatomical guided segmentation with non-stationary tissue class distributions in an expectation-maximization framework. IEEE International Symposium on Biomedical Imaging. IEEE, Arlington, VA, USA, pp. 81–84.

Pohl, K., Warfield, S., Kikinis, R., Grimson, W., Wells, W., 2004b. Coupling statistical segmentation and PCA shape modeling. Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science vol. 3216/2004. Springer-Verlag, Rennes/St-Malo, France, pp. 151–159.

Press, W., Flannery, B., Teukolsky, S., Vetterling, W., 1992. Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. Cambridge Univ. Press.

Srivastava, S., Maes, F., Vandermeulen, D., Paesschen, W.V., Dupont, P., Suetens, P., 2004. Effects of anatomical asymmetry in spatial priors on model-based segmentation of the brain MRI: a validation study. Medical Image Computing and Computer-Assisted Intervention. No. 3216 in Lecture Notes in Computer Science. Springer-Verlag, pp. 327–334.

Thompson, P., Schwartz, C., Toga, A., 1996. High-resolution random mesh algorithms for creating a probabilistic 3 D surface atlas of the human brain. NeuroImaging 3 (1), 19–34.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imaging 22 (2), 137–154.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based bias field correction of MR images of the brain. IEEE Trans. Med. Imaging 18 (10), 885–895.

Warfield, S., Rexilius, J., Huppi, P., Inder, T., Miller, E., Wells, W., Zientara, G., Jolesz, F., Kikinis, R., 2001. A binary entropy measure to assess nonrigid registration algorithm. Medical Image Computing and Computer-Assisted Intervention, pp. 266–274 (Oct.).

Wells, W., Grimson, W., Kikinis, R., Jolesz, F., 1996. Adaptive segmentation of MRI data. IEEE Trans. Med. Imaging 15, 429–442.

Wyatt, P.P., Noble, J.A., 2002. MAP MRF joint segmentation and registration. Medical Image Computing and Computer-Assisted Intervention, pp. 580–587.

Xiaohua, C., Brady, M., Lo, J.L.-C., More, N., 2005. Simultaneous segmentation and registration of contrast-enhanced breast MRI. Information Processing in Medical Imaging, pp. 126–137.