Unsupervised Segmentation, Clustering, and Groupwise Registration of Heterogeneous Populations of Brain MR Images

Annemie Ribbens*, Jeroen Hermans, Frederik Maes, *Senior Member, IEEE*, Dirk Vandermeulen, Paul Suetens, *Member, IEEE*, and Alzheimer's Disease Neuroimaging Initiative

***Abstract*—Population analysis of brain morphology from magnetic resonance images contributes to the study and understanding of neurological diseases. Such analysis typically involves segmentation of a large set of images and comparisons of these segmentations between relevant subgroups of images (e.g., "normal" versus "diseased"). The images of each subgroup are usually selected in advance in a supervised way based on clinical knowledge. Their segmentations are typically guided by one or more available atlases, assumed to be suitable for the images at hand. We present a data-driven probabilistic framework that simultaneously performs atlas-guided segmentation of a heterogeneous set of brain MR images and clusters the images in homogeneous subgroups, while constructing separate probabilistic atlases for each cluster to guide the segmentation. The main benefits of integrating segmentation, clustering and atlas construction in a single framework are that: 1) our method can handle images of a heterogeneous group of subjects and automatically identifies homogeneous subgroups in an unsupervised way with minimal prior knowledge, 2) the subgroups are formed by automatical detection of the relevant morphological features based on the segmentation, 3) the atlases used by our method are constructed from the images themselves and optimally adapted for guiding the segmentation of each subgroup, and 4) the probabilistic atlases represent the morphological pattern that is specific for each subgroup and expose the groupwise differences between different subgroups. We demonstrate the feasibility of the proposed framework and evaluate its performance with respect to image segmentation, clustering and atlas construction on simulated and real data sets including the publicly available BrainWeb and ADNI data. It is shown that combined segmentation and atlas construction leads to improved segmentation accuracy. Furthermore, it is demonstrated that the clusters generated by our unsupervised framework largely coincide with the clinically determined subgroups in case of disease-specific differences in brain morphology and that the differences between the cluster-specific atlases are in agreement with the expected disease-specific patterns, indicating that our method is capable of detecting the different modes in a population. Our method can thus be seen as a comprehensive image-driven population analysis framework that can contribute to the detection of novel subgroups and distinctive image features, potentially leading to new insights in the brain development and disease.**

***Index Terms*—Atlas stratification, image clustering, pattern recognition, population analysis, registration, segmentation.**

Manuscript received April 04, 2013; revised June 08, 2013; accepted June 10, 2013. Date of publication June 19, 2013; date of current version January 30, 2014. This work is financially supported by KU Leuven's Concerted Research Action GOA/11/006. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. Part of this work has been presented in: A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen, and P. Suetens, "SPARC: Unified Framework for Automatic Segmentation, Probabilistic Atlas construction, Registration and Clustering of Brain MR Images," IEEE Int. Symp. Biomed. Imag,: From Nano to Macro, pp. 856–859, 2010. *Asterisk indicates corresponding author.*

*A. Ribbens is with the KU Leuven, Faculty of Engineering, Department of Electrical Engineering (ESAT/PSI), Medical Image Computing group; Medical Imaging Research Center, KU Leuven and UZ Leuven; and iMinds-KU Leuven, Future Health Department, Medical Imaging Research Center, UZ Gasthuisberg, B-3000 Leuven, Belgium (e-mail: annemie.ribbens@uz.kuleuven.ac.be).

J. Hermans, F. Maes, D. Vandermeulen, and P. Suetens are with the KU Leuven, Faculty of Engineering, Department of Electrical Engineering (ESAT/PSI), Medical Image Computing group; Medical Imaging Research Center, KU Leuven and UZ Leuven; and iMinds-KU Leuven, Future Health Department, Medical Imaging Research Center, UZ Gasthuisberg, B-3000 Leuven, Belgium.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2013.2270114

I. Introduction

DIAGNOSIS of neurological diseases based on standardized clinical tests is often difficult due to the complex spectrum of symptoms and the overlap of symptoms across different diseases [1]. However, many neurological diseases can be characterized by their gradual modification of the cellular environment resulting in macroscopic effects visible in brain magnetic resonance (MR) images such as changes in shape, size or image intensity of anatomical structures. For instance, changes in the basal ganglia have been linked to the neurodegenerative hypothesis in Huntington's disease (HD) [2], while evaluating the atrophy of temporal lobe structures can improve the diagnostic accuracy of Alzheimer's disease (AD) [3]. Analysis of brain MR images can thus provide disease-related features

which contribute to the classification of the images according to disease [4].

Accurate segmentation of brain MR images according to tissue types [white matter (WM), gray matter (GM), cerebrospinal fluid (CSF)] or to relevant anatomical structures (e.g., deep gray matter structures, ventricles, hippocampus, cortical lobes, etc.) is often a first step to extract features useful for the identification of disease-specific morphological differences [3], [5]–[7]. However, manual segmentation is time consuming and susceptible to inter- and intra-observer variability. Fully automated techniques [8]–[16] may provide a solution. To cope with the complex structure of the human brain, they typically rely on prior knowledge in the form of (possibly multiple) brain atlases, i.e., templates representing the mean brain anatomy of the population of interest and possibly its variability. These atlases are spatially warped to each individual study image using nonrigid atlas-to-image registration techniques. Segmentations are then obtained by label propagation [13], [17] and, in case of homogeneous regions, often further refined by combining atlas-to-image registration with an intensity model [18], [19]. In [12] and [14], algorithms have been presented for tissue segmentation which combine atlas-based registration with an intensity-based segmentation model in one unified probabilistic framework whereby both techniques cooperate resulting in an improved segmentation (and registration) performance.

Several methods have been proposed for constructing a probabilistic atlas from a group of images [20]–[28]. These methods differ in the similarity measure and the deformation model used for warping all images to a common space, as well as in the possible use of a target image to guide the groupwise registration, whereby different strategies have been proposed to minimize the impact of the choice of the target image on the final atlas in order to obtain unbiased atlases. Many studies have emphasized the influence of the atlas construction procedure on the performance of atlas-guided segmentation [28]–[30], in particular the relation between atlas sharpness (or blurriness) and the flexibility of the atlas-to-image registration. Moreover, also the choice of an appropriate atlas template is important for an optimal segmentation performance, given that the similarity between a particular atlas and an individual study image can be rather poor as a single mean shape atlas is not sufficient to adequately summarize the variability within the heterogeneous (e.g., healthy and diseased) human population [31], [32]. To overcome problems with heterogeneity in a population, some atlas construction approaches perform a groupwise registration without computing a mean shape template [33], [34]. In [35], images are assigned a weight such that images closer to the "real" population mean get a larger contribution in the construction of the atlas. Another possibility is to create atlases for the different modes in a heterogeneous population, i.e., disease-specific atlases [31], [32].

In case multiple atlases are available, each atlas in turn can be used as prior for the segmentation [13], [36]. To cope with the fact that some of the atlases may deviate substantially from the image to be segmented, several methods for (locally) selecting suitable atlases have been proposed [18], [19], [37]–[41].

However, the assumption that multiple atlases are already available implies that the procedures used for atlas construction and atlas-guided segmentation are not necessarily adapted to each other. Moreover, the criteria for atlas selection are typically chosen heuristically, although intuitively the atlas selection criterion should be identical to the criterion used to select the images for atlas construction. This is not guaranteed when atlas construction and atlas selection are treated separately. A possible solution is to avoid the use of an atlas by segmenting a homogeneous group of images simultaneously. Here, models (priors) are built implicitly or explicitly by the segmentation algorithm itself [42]–[44]. However, similar criteria as used for atlas construction and selection, are still needed for defining such homogeneous groups of images.

A possible solution for creating such homogeneous groups of images is the use of classification methods based on image features. As stated before, these image features are often based on the segmentations of the images and the atlas-to-image registrations [45]–[49]. Most classification approaches identify the image features that are characteristic for the disease-specific morphological differences in advance, based on a training set of already classified images. Afterwards these features are used for classifying a new set of images [45]–[49]. The mutual dependency of image feature-based classification and identification of discriminative image features induces the combination of both aspects in one unified framework. Such methods, as proposed in [50]–[52], try to model the heterogeneity of a population by simultaneously searching for the major modes in the population and their discriminating mode-specific image features, by simultaneously creating an atlas template for each of the modes. This process is called atlas stratification and avoids the need for prior knowledge about the relevant features, and of the classifications of any of the images. Consequently, such data-driven algorithms cluster the images in subgroups based on differences in morphology as detected in the image data itself. This can result in the discovery of new subgroups differing in unexpected ways.

The mutual dependency between atlas-guided segmentation, atlas construction and clustering induces the combination of all these aspects in a unified framework, as proposed in this paper. As such the individual methods can benefit from each other and become more data-driven.

In this paper, we propose such a unified framework for population analysis of large heterogeneous image data sets. The framework searches different modes in brain morphology within a large population of brain MR images and tries to model them by constructing a probabilistic atlas for each mode (atlas stratification). The duality between the need of a segmented set of brain MR images to construct probabilistic atlases and atlas-guided brain segmentation techniques requiring well adapted atlases, leads to including a segmentation procedure within the framework, which is the main advantage of our approach compared to related methods [50], [52]. Hence, our proposed model combines and integrates segmentation, clustering, groupwise registration and atlas construction within one framework such that all aspects can benefit from one another. In particular, such combinational approach improves the segmentation performance as atlases are optimally adapted to the images to be segmented, specifically as the same selection criteria as well as the same flexibility in registration model are used in both atlas construction and segmentation. The

atlas stratification is also helped by the integrated segmentation process as problems with image artifacts are avoided and as the segmentations are consistent with the constructed atlases. The combinational approach also results in a minimal requirement of prior knowledge what makes the framework more data-driven instead of hypothesis-driven. Therefore, we hypothesize that such comprehensive data-driven framework could contribute to getting new insights in neurological diseases and their development and can therefore become an important tool in the understanding and analysis of structural changes due to pathology or aging.

The paper is organized as follows. In Section II, our unified probabilistic framework is introduced as a maximum a posteriori (MAP) problem and the optimization based on an expectation maximization algorithm is described. Furthermore, the framework is situated in relation to the current state-of-the-art, demonstrating the generality of our framework to other published work. Section III describes the experiments and data sets used for validation of the segmentation and clustering performance of our framework, and of the quality of the obtained atlases in terms of sharpness and stratification, i.e., the clear visualization by the atlases of apparent disease-specific morphology. The results are reported in Section IV. The paper concludes with a discussion concerning the advantages and drawbacks of the proposed method as well as an outlook for future work (Section V).

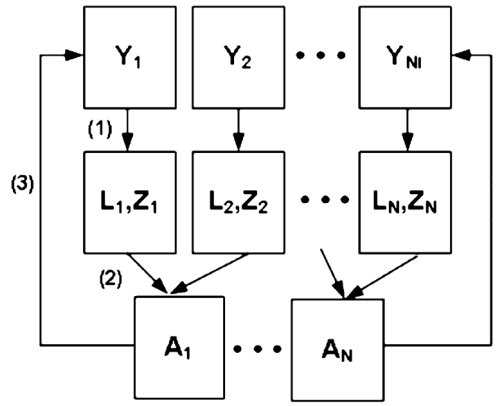


Fig. 1. Schematic overview of the proposed framework for joint segmentation and atlas stratification of a heterogeneous group of images. (1) The images are segmented based on a Gaussian mixture model on the intensities Y guided by cluster-specific atlases A (E-step). The images are clustered based on the similarity between the segmentations L and the cluster-specific atlases A (E-step). (2) Cluster-specific atlases are constructed from the segmentations L based on the cluster memberships Z (M-step). (3) The atlases are deformed towards the images using segmentation-based registration (M-step).

II. Methods

In this section, we describe our probabilistic model for simultaneous segmentation of all images of a heterogeneous data set and clustering of the images into different homogeneous subgroups. The model incorporates the iterative construction of a probabilistic brain atlas for each cluster and the alignment of each atlas to all of the images, which guide their segmentation and clustering. A schematic overview of our approach is given in Fig. 1. The proposed framework is called **SPARC** as it iteratively estimates the **S**egmentation of the images, the **P**robabilistic **A**tlases per cluster, the atlas-to-image **R**egistrations and the **C**luster memberships of each image. We first describe an unsupervised version of our framework in Sections II-A to II-D. Different strategies for (semi-)supervised image clustering based on prior information that might be available, are described in Section II-E. Section II-F explains the relation of the proposed method to the current state-of-the-art.

A. Model Assumptions

Given a heterogeneous set of brain MR images, we try to find the different modes (clusters) in a population. Different clusters in the data set can correspond to the presence or absence of disease, different stages or types of disease, different age groups, etc. We wish to cluster the images based on morphological features derived by segmentation. In this paper, the segmentation focuses on the three major brain tissue classes, i.e., white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). However, the method will be generally formulated such that more tissue classes can be estimated, e.g., to separate internal and external CSF. Thereto, we define a model that tries to explain the intensities of all images (i.e., observed variables) based on knowledge of the segmentations and cluster memberships. Optimization of this model will then result in an estimation of the tissue segmentations and the cluster memberships.

Let $i = \{1, \ldots, N_I\}$ be the indices for the images and $j = \{1, \ldots, N_J\}$ for the voxels in an individual image (assuming all images to have the same number of voxels for notational simplicity). Furthermore, let $k = \{1, \ldots, N_K\}$ be the indices for the tissue classes of an image and let $t = \{1, \ldots, N_T\}$ be the indices for the clusters (modes) in the heterogeneous data set of brain MR images. The number of tissue classes N_K and the number of clusters N_T are fixed by the user. The tissue class labels of all voxels are denoted as $L = \{l_{ijk} | \forall i, j, k\}$ with $l_{ijk} = 1$ if voxel j in image i has tissue class k and $l_{ijk} = 0$ otherwise. The cluster memberships $Z = \{z_{ijt} | \forall i, j, t\}$ are specified as $z_{ijt} = 1$ if voxel j in image i belongs to cluster t and $z_{ijt} = 0$ otherwise. The cluster memberships z_{ijt} are specified voxel-wisely instead of assigning each image globally to a single cluster. This choice is made as it takes into account that different regions of the brain can evolve differently, e.g., not all regions of the brain will be affected in the same way by the neurological disease. Hence, the cluster memberships Z highlight the spatial locations of the features that discriminate between the clusters, e.g., regions specifically affected by disease or aging.

We specify now our model and make some independency assumptions. We assume a Gaussian mixture model on the intensities of each image including bias field correction analogously to several models in literature [8], [9], [16]. For each image i, let the image intensities of each tissue type k be described by a normal distribution with parameters $\theta_{ik} = \{\mu_{ik}, \sigma^2_{ik}\}$, i.e., the mean and variance respectively. Furthermore, we assume that the intensities are corrupted by a smoothly varying multiplicative effect to account for the bias field induced by MR field inhomogeneities. We model the bias field as a linear combination of smooth basis functions, with parameters C_i representing the set of parameters of the bias field for image i. For mathematical stability, we perform a log-transformation of the

data to make the bias additive and assume that the log-transformed image intensities are normally distributed. Denoting the log-transformed image intensities as $Y = \{y_{ij} | \forall i, j\}$, then the probability density function on the intensities is given by the following Gaussian mixture model:

$$P(Y|L,\theta,C) = \prod_{i=1}^{N_I}\prod_{j=1}^{N_J}\prod_{k=1}^{N_K} P(y_{ij}|\theta_{ik}, C_i)^{l_{ijk}}$$
$$= \prod_{i,j,k}\left[\frac{1}{\sqrt{2\pi\sigma_{ik}^2}}\exp\left(-\frac{1}{2}\frac{(y_{ij} - B(x_{ij}|C_i) - \mu_{ik})^2}{\sigma_{ik}^2}\right)\right]^{l_{ijk}} \quad (1)$$

with $B(x_{ij}|C_i)$ the bias field evaluated in voxel j of image i with spatial coordinates x_{ij}.

The segmentation is guided by a prior probability for the class label l_{ijk} of voxel j in image i. This prior distribution is obtained from the probabilistic atlas of cluster t to which this voxel belongs:

$$\forall i : P(l_{ijk} = 1 | z_{ijt} = 1) = A_{kt}\left(R_{it}(x_{ij})\right) \quad (2)$$

with A_{kt} the probabilistic atlas for tissue class k and cluster t. The parameter R_{it} denotes the deformation field representing the atlas t-to-image i registration. We denote $R_{it}(x_{ij}) = R_{ijt} = x_{ij} + dx_{ijt}$ for each image i and each cluster t with dx_{ijt} the displacement of voxel j in image i to the space of the atlas of cluster t. Assuming all voxels and all images to be independent, the prior on the segmentations L is then given by

$$P(L|Z,A,R) = \prod_{i=1}^{N_I}\prod_{j=1}^{N_J}\prod_{k=1}^{N_K}\prod_{t=1}^{N_T} P(l_{ijk} = 1|A_{kt}, R_{ijt})^{z_{ijt}}$$
$$= \prod_{i,j,k,t}\left(A_{kt}\left(R_{it}(x_{ij})\right)\right)^{l_{ijk}z_{ijt}}. \quad (3)$$

In addition to the diffeomorphic regularizer embedded in the registration procedure (see further), we define a prior distribution on the atlas-to-image registrations R_{ijt}:

$$P(R|Z,G,\epsilon^2)$$
$$= \prod_{i=1}^{N_I}\prod_{j=1}^{N_J}\prod_{t=1}^{N_T} P\left(R_{ijt}|G_{jt},\epsilon_{jt}^2\right)^{z_{ijt}}$$
$$= \prod_{i,j,t}\left[\frac{1}{\sqrt{2\pi\epsilon_{jt}^2}}\exp\left(-\frac{1}{2}\frac{\|R_{ijt} - G_{jt}\|^2}{\epsilon_{jt}^2}\right)\right]^{z_{ijt}} \quad (4)$$

i.e., a Gaussian distribution with mean G_{jt} and variance ϵ_{jt}^2. The underlying assumption is that we expect the individual atlas-to-image registrations R_{ijt} within one cluster t to be close to a cluster-specific mean registration $G_t = \{G_{jt}\}$. This mean registration is the groupwise registration between the atlas of cluster t and all images belonging to that cluster. In case the atlas of cluster t is in minimum deformation space, i.e., the space where a minimum amount of displacement is necessary to transform the atlas to all of the images individually [21], [26], then G_t equals the identity transform. The Gaussian distribution describes the L^2-norm between an individual and the mean registration of a cluster t. The parameter ϵ_{jt}^2 is a weighting term representing the (voxelwise and cluster-specific) variability of the individual atlas-to-image deformations R_{ijt} around the cluster-specific mean deformation G_{jt}.

The prior probability on the tissue class labels L and the registrations R requires knowledge of the cluster memberships Z. In this unsupervised version of the framework, we aim at discovering relevant clusters within a heterogeneous population of images based on the image data itself. Without specific prior knowledge of the regions that discriminate between the different clusters (i.e., the disease-specific patterns), we assume *a priori* that the overall morphology changes (i.e., the cluster-specific differences are present in all voxels) and that all voxels of the same image are likely to belong to the same cluster. Hence, the prior distribution on the cluster memberships Z is formulated as a uniform distribution over the voxels within each image

$$P(Z|\pi) = \prod_{i=1}^{N_I}\prod_{j=1}^{N_J}\prod_{t=1}^{N_T} P(z_{ijt} = 1|\pi_{it}) = \prod_{i,j,t}\pi_{it}^{z_{ijt}}. \quad (5)$$

Depending on the research question, more specific assumptions concerning the prior probability of the cluster memberships can be made. Some examples are given in Section II-E.

B. Maximum a Posteriori Problem

Our model with parameters $\Upsilon = \{\theta, C, A, R, G, \epsilon^2, \pi\}$, needs to be fitted to the observed image intensities Y, through the following maximum a posteriori (MAP) problem:

$$\hat{\Upsilon} = \arg\max_{\Upsilon} \log P(\Upsilon|Y) \propto \arg\max_{\Upsilon} \log P(Y, \Upsilon). \quad (6)$$

Optimal estimation of the model parameters Υ is facilitated by knowledge of the class labels L and cluster memberships Z. In turn, estimation of the class labels L and cluster memberships Z is straightforward once the model parameters Υ are known. Therefore, the MAP problem is rewritten as

$$\hat{\Upsilon} = \arg\max_{\Upsilon} \log\left[\sum_{L,Z} P(Y, L, Z, \Upsilon)\right] \quad (7)$$

by introducing L and Z as latent (hidden) variables and with

$$P(Y, L, Z, \Upsilon)$$
$$= P(Y|L, Z, \theta, C)P(L|Z, A, R)P(R|Z, G, \epsilon^2)P(Z|\pi) \quad (8)$$

making use of the model assumptions defined in Section II-A. To solve (7), we derive a lower bound of the log-likelihood function using Jensen's inequality

$$\log\sum_{L,Z} P(Y, L, Z, \Upsilon)$$
$$= \log\sum_{L,Z} P(L, Z|Y, \Upsilon')\frac{P(Y, L, Z, \Upsilon)}{P(L, Z|Y, \Upsilon')}$$
$$\geq \sum_{L,Z} P(L, Z|Y, \Upsilon')\log P(Y, L, Z, \Upsilon)$$
$$- P(L, Z|Y, \Upsilon')\log P(L, Z|Y, \Upsilon') \quad (9)$$

with Υ' denoting an estimate for the parameter set Υ. This lower bound can be optimized using the expectation maximization (EM) algorithm which alternates between an expectation step, inferring the posterior distribution over the hidden variables given the data and the model parameters, and a maximization step, in which the expectation of the log-likelihood function is maximized with respect to the parameters Υ.

C. Expectation Step

Based on (9) and the assumptions stated in Section II-A, the expectation Q of the log-likelihood function is derived as

$$\begin{aligned} Q(\Upsilon|\Upsilon^\eta) &= E_{L,Z|Y,\Upsilon^\eta}\left[\log P(Y,L,Z,\Upsilon)\right] \\ &= \sum_{L,Z} P(L,Z|Y,\Upsilon^\eta)\log P(Y,L,Z,\Upsilon) \\ &= \sum_{i=1}^{N_I}\sum_{j=1}^{N_J}\sum_{k=1}^{N_K}\sum_{t=1}^{N_T} P\left(l_{ijk}=1, z_{ijt}=1|y_{ij},\Upsilon^\eta_{ijkt}\right) \\ &\quad \cdot \left[\log P(y_{ij}|\theta_{ik},C_i) + \log P(l_{ijk}=1|A_{kt},R_{ijt}) \right. \\ &\quad + \log P(z_{ijt}=1|\pi_{it}) \\ &\quad \left. + \log P\left(R_{ijt}|G_{jt},\epsilon^2_{jt}\right)\right] \end{aligned} \tag{10}$$

making use of (8) and with $P(l_{ijk}, z_{ijt}|y_{ij},\Upsilon_{ijkt})$ the joint posterior probability in voxel j of image i of the class label k and the cluster membership t given the (log-transformed) intensities y_{ij} and the current estimate of the model parameters Υ^η (with η indexing the EM iterations). An expression for the posterior can be obtained using Bayes' rule

$$\begin{aligned} &P(l_{ijk}=1, z_{ijt}=1|y_{ij},\Upsilon_{ijkt}) \\ &\qquad = \frac{P(y_{ij}, l_{ijk}=1, z_{ijkt}=1,\Upsilon_{ijkt})}{\sum_{k,t} P(y_{ij}, l_{ijk}, z_{ijkt},\Upsilon_{ijkt})} \end{aligned} \tag{11}$$

with

$$\begin{aligned} &P(y_{ij}, l_{ijk}=1, z_{ijkt}=1,\Upsilon_{ijkt}) \\ &\quad = P(y_{ij}|\theta_{ik},C_i)\cdot P(l_{ijk}=1|A_{kt},R_{ijt}) \\ &\qquad \cdot P(z_{ijt}=1|\pi_{it})\cdot P\left(R_{ijt}|G_{jt},\epsilon^2_{jt}\right) \end{aligned} \tag{12}$$

following (8). The posterior can now be rewritten as

$$\begin{aligned} &P(l_{ijk}=1, z_{ijt}=1|y_{ij},\Upsilon_{ijkt}) \\ &\quad = \frac{\sum_k P(y_{ij}, l_{ijk}=1, z_{ijt}=1,\Upsilon_{ijkt})}{\sum_{k,t} P(y_{ij}, l_{ijk}=1, z_{ijt},\Upsilon_{ijkt})} \\ &\qquad \cdot \frac{P(y_{ij}, l_{ijk}=1, z_{ijt}=1,\Upsilon_{ijkt})}{\sum_k P(y_{ij}, l_{ijk}=1, z_{ijt}=1,\Upsilon_{ijkt})} \\ &\quad = \rho_{ijt}\cdot \frac{P(y_{ij}|\theta_{ik},C_i)P(l_{ijk}=1|A_{kt},R_{ijt})}{\sum_k P(y_{ij}|\theta_{ik},C_i)P(l_{ijk}=1|A_{kt},R_{ijt})} \\ &\quad = \rho_{ijt}\cdot p_{ijkt}. \end{aligned} \tag{13}$$

The marginal posteriors of the class labels and the cluster memberships respectively can be obtained by integration of the joint posterior, i.e.,

$$\begin{aligned} P(l_{ijk}=1|y_{ij},\Upsilon_{ijkt}) &= \sum_t P(l_{ijk}=1, z_{ijt}=1|y_{ij},\Upsilon_{ijkt}) \\ &= \sum_t \rho_{ijt}\cdot p_{ijkt} = p_{ijk} \end{aligned} \tag{14}$$

and

$$\begin{aligned} P(z_{ijt}=1|y_{ij},\Upsilon_{ijkt}) &= \sum_k P(l_{ijk}=1, z_{ijt}=1|y_{ij},\Upsilon_{ijkt}) \\ &= \sum_k \rho_{ijt}\cdot p_{ijkt} = \rho_{ijt}. \end{aligned} \tag{15}$$

From (13), it can be seen that p_{ijkt} are the probabilistic tissue segmentations of image i obtained by the conventional approach [12], [15] of combining the Gaussian mixture model on the (bias-field corrected) intensities with prior information derived from an atlas, namely the cluster-specific atlas A_{kt}. Multiple such segmentations for image i are computed using the different atlases of all clusters t and these are subsequently combined by a weighted sum to obtain the final segmentation for image i, i.e., the probabilities p_{ijk} of voxel j in image i to belong to class k [see (14)].

The weights ρ_{ijt} are the probabilistic voxelwise cluster memberships, i.e., the probability of voxel j in image i to belong to cluster t. From (15) and (12), it follows that the cluster memberships are determined by three aspects. The first aspect is the amount of deformation needed to transform the atlas to the image. The second aspect is the likelihood of the image data based on the intensity model given the deformed atlas templates. This can be interpreted as a similarity measure between atlas and image as it expresses the overlap between a segmentation purely based on the intensity model and the segmentation based on the deformed atlas. The third aspect that influences the voxelwise cluster memberships is the prior model that is assumed. The prior distribution on the cluster memberships Z allows to implement various types of spatial regularization on the final estimation of the voxelwise cluster memberships. It can guide the cluster memberships to be locally or globally similar by making the clustering more region-based or image-based instead of voxel-based. Some examples of such clustering priors are given in Section II-E.

Finally, a probabilistic cluster membership per image ρ_{it} can be obtained by averaging the voxelwise cluster memberships per image

$$\rho_{it} = \frac{1}{N_J}\sum_j \rho_{ijt}. \tag{16}$$

Each image can thus be assigned to a single cluster $t_i = \max_t \rho_{it}$, that is hypothesized to indicate the disease state or stage of the corresponding subject.

D. Maximization Step

In the maximization step of the EM algorithm, the Q-function, given the posterior distribution for the hidden variables L and Z, is optimized with respect to the model parameters Υ, resulting in new parameter estimates $\Upsilon^{(\eta+1)}$. As the probability density function on the image intensities $P(y_{ij}|l_{ijk},\theta_{ik},C_i)$ is independent of the cluster memberships and as $\sum_t p(l_{ijk}, z_{ijt}|y_{ij},\Upsilon_{ijkt}) = p_{ijk}$ [see (14)], the Gaussian mixture parameters θ_{ik} and the bias field parameters C_i are determined in a completely similar way as in [8].

A coordinate ascent strategy is used to update the atlases A_{kt}, the atlas-to-image registrations R_{ijt} and the groupwise registrations G_{jt}, as all these parameters depend on each other. We start with updating the atlases and the groupwise registrations, making use of the atlas-to-image registrations of the previous iteration. The atlases can then be determined as follows:

$$\frac{\partial Q}{\partial A_{ukt}} = 0, \text{ subject to } \sum_k A_{ukt} = 1$$

with u indexing the voxels in the atlas space. For each voxel j in image i at location x_{ij}, $u = R_{it}(x_{ij})$ is the corresponding voxel in the atlas space. The following formulation for the atlas is then found:

$$A_{kt}(u) = \frac{\sum_i \rho_{it}\left(R_{it}^{-1}(u)\right) p_{ikt}\left(R_{it}^{-1}(u)\right) \left|\text{Jac}\left(R_{it}^{-1}(u)\right)\right|}{\sum_i \rho_{it}\left(R_{it}^{-1}(u)\right) \left|\text{Jac}\left(R_{it}^{-1}(u)\right)\right|} \tag{17}$$

with $R_{it}^{-1}(u)$ the mapping of the atlas voxels u onto image i, with Jac the Jacobian and $|\cdot|$ the determinant. Evaluation of (17) assumes a suitable scheme for interpolation of ρ_{it} and p_{ikt} at the transformed locations $R_{it}^{-1}(u)$. A detailed derivation can be found in Appendix A.

The atlases are thus constructed as the weighted sum of the segmentations of the images deformed towards the atlas space. The weights are determined by the (voxelwise) cluster membership probabilities. Moreover, the atlas construction implies a modulation step, i.e., a scaling of the deformed segmentation maps with the Jacobian determinant to preserve their local probabilistic volumes. Due to the fact that the cluster memberships are determined voxelwisely, the set of images that contributes to the construction of an atlas is locally varying. In regions with large inter-cluster differences, a small set of locally similar images will be selected for each cluster, while in case of small inter-cluster differences, many images will contribute to the atlases as the cluster memberships will be rather equal between images and between clusters. This results in a better capturing of the inter-subject variability as a larger data set is available.

For the groupwise deformation field G_{jt}, we find

$$\frac{\partial Q}{\partial G_{jt}} = 0 \Rightarrow G_{jt} = \frac{\sum_i \rho_{ijt} R_{ijt}}{\sum_i \rho_{ijt}} \tag{18}$$

i.e., G_{jt} is constructed as the weighted sum of the deformations corresponding to the individual atlas-to-image registrations. The weights are given by the probabilistic cluster memberships.

Next, the atlas-to-image registrations R_{ijt} can be updated. We find the following derivative:

$$\frac{\partial Q}{\partial R_{ijt}} = \rho_{ijt}\left[\left(\sum_k \frac{p_{ijkt}}{A_{kt}\left(R_{it}(x_{ij})\right)} \nabla_{R_{ijt}} A_{kt}\left(R_{it}(x_{ij})\right)\right) - \frac{(R_{ijt} - G_{jt})}{\epsilon_{jt}^2}\right]. \tag{19}$$

It is clear that there is no closed form solution for R_{ijt} such that $\partial Q/\partial R_{ijt} = 0$, but the derivative provides a direction for iteratively updating R_{ijt} in order to maximize the Q-function. Some form of regularization is required to impose spatial smoothness for obtaining a physically acceptable deformation field. To this end, we interpret the derivative $\partial Q/\partial R_{ijt}$ as a force field that we use to drive the diffeomorphic demons registration approach of [53] (although several other regularizers can be chosen instead). The force field in (19) is modulated by ρ_{ijt}, such that image voxels with low probability to belong to cluster t contribute only little to driving the registration of this image to that particular cluster.

The parameter π_{it} of the clustering prior is updated as follows:

$$\frac{\partial Q}{\partial \pi_{it}} = 0, \text{ s.t. } \sum_t \pi_{it} = 1 \Rightarrow \pi_{it} = \frac{1}{N_J}\sum_j \rho_{ijt} \tag{20}$$

which equals the mean of the voxelwise cluster membership probabilities per image, and corresponds to the probabilistic assignment of an image to a cluster [see (16)]. Finally, the parameter ϵ_{jt}^2 of the registration prior is not updated, to avoid that the MAP problem becomes underdetermined as the parameter is only dependent on two other parameters, i.e., R_{ijt} and G_{jt}. Specific choices for ϵ_{jt}^2 are suggested in Section III.

E. Clustering Priors

From (15), it follows that the cluster memberships are determined by the similarity between atlas and intensity model, the amount of deformation needed for atlas-to-image registration and a prior model on the cluster memberships (Section II-C). This prior model allows to make the clustering more or less supervised. In Section II-A, we specified the prior on the cluster memberships to be uniform for each image [see (5)]. The parameters of the prior model (i.e., π_{it}) as well as the cluster memberships themselves are iteratively estimated from the data itself as described above, such that the clustering is completely unsupervised. However, in case a fixed binary prior is used, i.e., $\forall i\, \pi_{it} = 1$ for some $t = t_i$ and zero otherwise, the framework becomes completely supervised by assigning each image *a priori* to a single fixed subgroup t'. Using such a prior is comparable to executing the framework multiple times, i.e., on each of the predefined subgroups separately, using each time one cluster. In case fixed nonbinary prior probabilities π_{it} are specified for some of the images, the algorithm becomes semisupervised. The final cluster memberships for each image may differ from the initial clustering as specified by the prior, depending on the agreement between atlas and intensity model and the amount of deformation needed to map the atlas onto the image.

Both the unsupervised and the supervised approach have their advantages. Unsupervised clustering allows to detect new subgroups differing in unexpected ways which could contribute to gaining new insights in neurological disease(s) as no prior hypothesis is assumed. Moreover, as only image features are considered during image clustering and atlas construction, the atlases are driven towards optimally representing the images under study and their variability, which in turn improves the segmentation performance. On the other hand, supervised methods are often better suited for specific clinical research questions as in unsupervised methods the control over the representation of the clinical subgroups is restricted to the initialization of the framework (see further).

We now suggest a semisupervised approach in which the clustering still relies to a large extent on the image data itself, but

takes prior information into account as well (in a more intelligent way compared to fixing the uniform distribution to prespecified values). Two types of prior information are considered here. A first type concerns prior knowledge about the cluster memberships of some of the images. This prior knowledge can be based on clinical information and demographics or on explicit knowledge about the brain morphology. In this case, we expect images with similar properties to belong to the same cluster. A second type of prior information, denoted as spatial prior information, is the availability of a hypothesis about the regions where morphological features can be found, indicative for a certain disease state or stage. For instance, hippocampal atrophy has been proven to be indicative for patients suffering from AD (e.g., [3]). Focussing on the hippocampal region when assigning the cluster memberships would therefore be beneficial when the aim is to separate AD patients from patients suffering from other dementias. Such a regional focus in the voxelwise clustering performed by our framework can be achieved by assigning a larger weight to this region and less to other regions when updating the clustering prior.

Both types of prior information can be specified in the form of must-link and cannot-link constraints, which can be modeled by a Markov Random field (MRF) on the cluster memberships [54]. Such MRF prior is defined by the Gibbs distribution

$$P(z_{ijt}=1|\Phi_{z_{ij}}) = Z(\Phi_{z_{ij}})^{-1} \exp\left[-U\left(z_{ijt}=1|z_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}}\right)\right] \tag{21}$$

with $\Phi_{z_{ij}}$ the MRF parameters, $z_{\mathcal{N}_{ij}}$ the cluster memberships of voxels in a neighborhood $\mathcal{N}_{ij}$ of voxel j of image i and $Z(\Phi_{z_{ij}})$ the normalization constant. The function U is an energy function with parameters Φ that is defined as the sum of clique potentials over the neighborhood $\mathcal{N}_{ij}$, and is generally defined by the following Potts model:

$$U\left(z_{ijt}|z_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}}\right) = \sum_{i',j' \in \mathcal{N}_{ij}} \sum_{t'} z_{ijt} \alpha_{ii'jj'tt'} z_{i'j't'} \tag{22}$$

with $\mathcal{N}_{ij}$ the neighborhood of voxel j in image i and $\Phi_{z_{ijt}} = \{\alpha_{ii'jj'tt'} | i', j' \in \mathcal{N}_{ij}, \forall t'\}$ the MRF parameters describing the transition energy between cluster memberships t and t' specified for voxel j in image i in relation to voxel j' in image i'. The way the neighborhoods and MRF parameters are defined, is dependent on the available prior knowledge, as they allow stimulating or penalizing voxels/images to belong to the same cluster and/or emphasizing the dependency of the cluster memberships of certain voxels on those of other voxels/regions. A specific example is described in the experiments (Section III). A particular case of such a MRF on the cluster memberships, has also been proposed in [41], to stimulate local similarity (spatially in the image) in the selection of a training image for label fusion.

The calculation of the Q-function by the EM algorithm requires all possible realizations of the MRF, both in the expectation step (posterior) and maximization step (update of the prior distribution). However, this is computationally not feasible. Therefore, we use the mean field approximation [55], [56] to estimate the prior distribution on the cluster memberships, resulting in the following expression for iteration $\eta + 1$ of the EM algorithm:

$$P\left(z_{ijt}=1|\Phi_{z_{ijt}}\right) = Z\left(\Phi_{z_{ij}}\right)^{-1} \exp\left[-U\left(z_{ijt}=1|z^{\eta}_{\mathcal{N}_{ij}}, \Phi_{z_{ijt}}\right)\right]$$

with $z^{\eta}_{\mathcal{N}_{ij}}$ the probabilistic estimation of the cluster memberships z_{ijt} in the neighborhood of voxel j in image i, i.e., $i, j \in \mathcal{N}_{ij}$, in the previous iteration η of the EM algorithm.

F. Model Variations Versus Related Work

In this section, we illustrate the generality of our framework by showing how the framework described above reduces to some well known methods in the literature under certain simplifying assumptions.

- The method of [15] combines atlas-based image segmentation with atlas-to-image registration. Similar approaches are presented in [12], [14]. In our framework, this comes down to considering only one image, using a given atlas without updating it during iterations and specifying only one cluster. The expected log-likelihood of our framework reduces to

$$Q = \sum_{j=1}^{N_J} \sum_{k=1}^{N_K} P\left(l_{jk}|y_j, \theta^{\eta}_{jk}, C^{\eta}, A_k, R^{\eta}_j\right) \cdot \left[\log P(y_j|\theta_k, C) + \log P(l_{jk}|A_k, R_j)\right] \tag{23}$$

  which corresponds indeed to the one given in [15]. If one assumes that the atlas and the image are well enough aligned *a priori*, the atlas-to-image registration can also be left out from the framework, resulting in the model proposed in [8].
- In [43], a method is proposed to simultaneously segment a homogeneous data set of brain MR images (in atlas space) while performing probabilistic atlas construction. The segmentation is obtained based on a Gaussian mixture model and the estimated atlas. In terms of our framework, the expected loglikelihood function can be written as

$$Q = \sum_{i=1}^{N_I} \sum_{j=1}^{N_J} \sum_{k=1}^{N_K} P\left(l_{ijk}|\theta^{\eta}_{ik}, A^{\eta}_{kt}, R^{\eta}_{ijt}\right) \cdot \left[\log P(y_{ij}|\theta_{ik}) + \log P(l_{ijk}|A_{kt}, R_{ijt})\right]. \tag{24}$$

  Hence, the method comes down to a simplified version of our framework, whereby the images are assumed to be homogeneous (i.e., only a single cluster is considered) and no bias field correction is performed. Another fundamental difference, is that in our case, the segmentations are performed in the space of the images. In the method of [43], the parameter R equals an image-to-atlas registration, compared to R being an atlas-to-image registration in the method proposed in our framework.
- The method of [52] considers a heterogeneous set of images which one wants to split in homogeneous clusters whereby the segmentations are of no interest. In our framework, this means that the only hidden variables are the cluster memberships. In [52], one therefore assumes that the intensities of all images belonging to a certain cluster are generated, per voxel, from a Gaussian distribution, with the mean equal to a mean intensity template of that cluster. The method also includes a warping between the atlases

and each of the images to correctly indicate the corresponding voxels. The expected log-likelihood equals

$$Q = \sum_{i=1}^{N_I} \sum_{t=1}^{N_T} P\left(z_{it}|y_{ij}, A_t^\eta, \epsilon_t^{2(\eta)}, R_{ijt}^\eta, \pi_{it}^\eta\right) \cdot \left[\log P\left(y_{ij}|A_t, \epsilon_t^2, R_{ijt}\right) + \log P(z_{it}|\pi_{it})\right] \quad (25)$$

with $P(y_{ij}|A_t, \epsilon_t^2, R_{ijt})$ the Gaussian distribution. In [52], the cluster memberships are determined per image instead of per voxel as in our approach.

- Our framework is also closely related to multi-atlas segmentation methods. As described in Section II-C, the segmentation algorithm within our framework comes down to a local-selection multi-atlas technique combined with an intensity model. Therefore, our segmentation method follows state-of-the-art segmentation techniques such as [18], [19], [37]–[41], [57] that apply (local) atlas selection to cope with possibly substantial differences between the atlases and the images to be segmented. While most of these proposed segmentation techniques are formulated rather heuristically, the method of [41] uses a probabilistic framework to generally describe such segmentation methods based on multiple (local) atlas selection, comparable as in our approach. Some of the proposed (local) atlas selection techniques used for the segmentation of homogeneous regions (e.g., [18], [19]) are further improved by the use of an explicit intensity model, which is also included in our framework. However, we follow the approach of [12], [14], [15], developed for atlas-guided segmentation using a single atlas, which simultaneously optimize the intensity model and the atlas-to-image registration as this has proven to result in improved segmentation (and registration) results. In contrast to most state-of-the-art methods, the weights used by our method to fuse the segmentations are not determined heuristically, but result directly from our probabilistic model. These weights ρ_{ijt} are the cluster membership probabilities, i.e., the probability of voxel j in image i to belong to cluster t. Moreover, the same weights are used by our method when constructing the atlases as a weighted average of the segmentations.
- Finally, from Section II-D it follows that the atlases of our framework are constructed in a similar way as in [25], [27]. Applying the groupwise registrations to the atlases results in a similar method as in [21] where a minimal deformation atlas is constructed. However, these methods are developed for single atlas construction. Our algorithm considers multiple clusters which has the advantage that images deviating from the group mean are given a smaller weight, such that our atlases are better representations of the true group means (of each cluster).

G. Initialization

All images are first globally normalized in a common coordinate frame by affine registration of all images to a certain reference image using maximization of mutual information [58]. In this paper, we use the ICBM452 atlas in the MNI standard coordinate space as [59]. Furthermore, brain masking of all images is preferred as it avoids that non-brain voxels are erroneously classified as brain tissue due to similarities in intensity (e.g., dura and GM) what can also influence our intensity model. Brain masking can be achieved for instance using a brain extraction tool such as [60], but for the images in our experiments a brain mask was already available.

The EM algorithm must be initialized by providing an initial estimation of the model parameters or of the hidden variables, i.e., the segmentation and cluster memberships. In this work, initial values for the model parameters are provided. Initial values for the Gaussian mixture parameters of the intensity model θ_{ik} are estimated in the same way as within the framework, with the required image segmentation p_{ijk} replaced by the probabilistic tissue class maps of the ICBM452 atlas [59] available in SPM (http://www.fil.ion.ucl.ac.uk/spm/). The bias field coefficients C_i are initially all set to zero and all registrations are initially set to the identity transformation. All parameters of the uniform prior distribution on the cluster memberships are initialized as the inverse of the total number of clusters, i.e., $\pi_{it} = 1/N_t$ for all i, t. Finally, the parameters A_{kt}, i.e., the cluster-specific atlases, are initialized from a crude segmentation of a subset of the images. These segmentations are obtained by applying the Gaussian mixture model to the image intensities (no atlas is used) using the initial estimations of the Gaussian mixture parameters (see above). Subsequently initial estimates of the atlases A_{kt} are obtained as weighted means of these roughly segmented images [analogously to (17)]. The subset of images and the weights used when constructing the initial atlases for each cluster are randomly selected or determined based on (clinical) prior knowledge, depending on the specific research question. We refer to the experiments for specific choices.

H. Implementation Details

We implemented the method in MATLAB (The Mathworks Inc., Natick, MA, USA). To reduce the calculation time, time-consuming parts are implemented in C++ and MATLAB's parallel toolbox is used. The implementation consists of four main parts: the intensity model, the atlas construction, the atlas-to-image registration and the update of the prior on the cluster memberships.

For the intensity model, analogously to [8], the number of tissue classes has to be fixed in advance. In this work, the number of tissue classes is fixed to three (WM, GM, CSF) assuming that the available brain masks are sufficiently accurate to rule out background and non-brain voxels. Furthermore, two Gaussians per tissue class and a 3-D fourth order polynomial bias field model are used, as in [8].

For updating the atlases and groupwise registrations, we follow the formulations of (17) and (20). To update the atlases, the inverse of the registrations R_{it} is required. However, in this paper, we perform a forward image-to-atlas registration H_{it} instead as discussed in Appendix A. Furthermore, the atlases can be brought to the minimal deformation space [21], i.e., the space where the least amount of deformation is needed to transform the atlas to all images in the set, by applying the cluster-specific

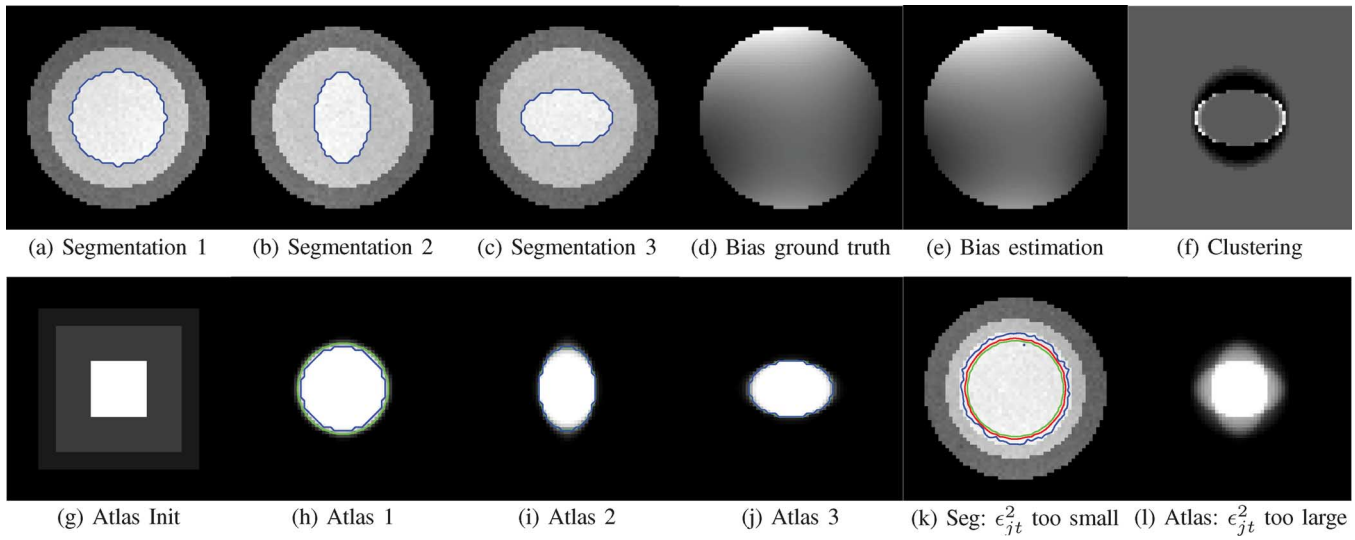


Fig. 2. Exp. 1: (a)–(c): Synthetic images each generated from a different template with segmentation of the inner structure by SPARC indicated in blue. (d)–(e) Bias field of an image: ground truth and estimation by SPARC. (f) Voxelwise cluster memberships of an image generated from cluster 3 [i.e., image of (c)] to belong to cluster 1. (g)–(j) Probabilistic atlas for the inner structure: initialization and final atlases for cluster 1, 2, and 3 [blue contour = ground truth, green = SPARC (after binairizing the probabilistic map)]. (k)–(l): Poor performance in case ϵ_{jt}^2 is chosen (k) too small (insufficient atlas deformation resulting in poor segmentation: blue = segmentation, red = deformed atlas, green = original atlas) or (l) too large (poor clustering resulting in a fuzzy atlas).

groupwise deformations G_t onto the corresponding probabilistic atlas A_{kt}. In practice, the groupwise registration is enforced to be the identity transformation per cluster t when updating the individual atlas-to-image registrations, aiming to construct the atlas directly in the minimal deformation space. However, to correct for deviations from this constraint to keep G_t fixed to the identity transform, the groupwise registrations G_t are computed anyway in each iteration of the EM algorithm according to (18) and directly applied on the constructed atlases A_{kt}. The individual atlas-to-image registrations are adapted accordingly as $R_{it}^{\text{new}}(j) = (R_{it}^{\text{old}} \circ G_t^{-1})(j)$.

To update Q with respect to the registrations R_{ijt}, we run five iterations of [53] in each EM iteration using the force field of (19). To equalize the magnitude of the force fields between different images with respect to a particular cluster, each force field is normalized by a factor ρ_{it} [see (16)]. An initial value of four voxels was empirically chosen for the variance of the Gaussian kernel used by this method to smooth subsequent updates of the deformation field, but we reduce this value in a systematic way each time the relative increase in the loglikelihood [see (10)] drops below 0.5%. In this way, the registration is first performed at a more global scale and gradually more locally as the method converges, what corresponds to a kind of multi-resolution approach.

Finally, the EM framework is stopped when the expected loglikelihood function [see (10)] no longer increases significantly (relative change smaller than 0.1%) or when the total number of EM iterations exceeds 350.

III. Data Sets and Experiments

Three different aspects of the proposed method need to be validated: the clustering performance, the segmentation performance, and the quality of the constructed probabilistic atlases. Experiment 1 provides a proof-of-concept of the method based on synthetic images. Experiment 2 focuses on combined segmentation and atlas construction making use of the publicly available BrainWeb data set. Experiment 3 validates the clustering performance of our algorithm on a clinical data set of Huntington's disease and normal control subjects. Finally, experiment 4 involves a global evaluation of the framework on the publicly available ADNI data set, with a focus on the voxelwise cluster memberships and providing a practical example of the benefits of combined segmentation and atlas stratification.

A. Experiment 1

To provide a proof-of-concept of our method, we first apply our method to a heterogeneous group of simulated 2-D images with known ground truth for the segmentation of the images, their cluster memberships and the cluster-specific atlases. The test images are generated from three templates containing three structures (and background) mimicking three different tissue classes as illustrated in Fig. 2(a), (b), and (c). The first template contains a ring with an inscribed ring and in this second ring an inscribed circle (radius $\mathfrak{r}$). In the second template, the inner circle is elongated along the vertical axis (long axis $\mathfrak{l}$) forming an vertical ellipse and in the third template, the inner circle is elongated along the horizontal axis (long axis $\mathfrak{l}$) forming a horizontal ellipse. The test images are generated by varying the radius of the inner circle (in case of template 1) or the length of the long axis of the inner ellipse (in case of templates 2 and 3) by sampling these from a Gaussian distribution with mean equal to $\mathfrak{r}$ and $\mathfrak{l}$, respectively. The intensities of the different structures in each test image are chosen to be similar to WM (inner circle\ellipse), GM (inner ring) and CSF (outermost ring) in a brain MR image, including 3% noise following a Rayleigh distribution similarly as in [61]. Also a smoothly varying multiplicative bias field modeled as a fourth order polynomial is included, with values ranging between approximately 0.9 and 1.1

[e.g., Fig. 2(d)]. A total of 150 test images are generated this way (i.e., 50 per template).

Our unsupervised framework is applied to this data set with $N_T = 3$ and $N_K = 3$. The modes (clusters) in the population are expected to follow the morphology induced by the generating templates, while the atlases are expected to coincide with the templates. The segmentations are compared with their ground truth by the Dice overlap measure of their inner circle/ellipses. The atlases obtained for each cluster are compared with their ground truth, i.e., the template corresponding to that cluster, after binarizing the atlases by assigning each pixel in the atlas to its most likely tissue class. The clustering is evaluated by the percentage of images that is assigned to the correct cluster, i.e., the cluster corresponding to the template from which the image was generated. However, an incorrectly clustered image can have a cluster membership close to 0.5. Also, even if all images have been assigned to the correct cluster, the separation between both clusters can be limited as all cluster membership probabilities can be rather close to 0.5. Therefore, we also measure the clustering accuracy by the following inner product (as also used in [52]):

$$\frac{1}{N_I}\frac{1}{N_T}\sum_i \sum_t P(z_{it}) z^g_{it} \tag{26}$$

with z^g the ground truth cluster memberships. To get insight in the influence of the parameter ϵ^2_{jt} of the Gaussian prior distribution imposed on the atlas-to-image registrations, the experiment is performed with different values for this parameter, namely $\epsilon^2_{jt} = 1$, 3, 5, 8, and ∞, i.e., no prior on the registrations. All experiments are repeated five times, each time using a different set of 150 test images, and the average performance over these five runs is reported.

B. Experiment 2

The purpose of this experiment is to investigate the benefits of simultaneous segmentation and atlas construction over doing both processes separately and sequentially. We apply our framework to the publicly available BrainWeb data set [61] which consists of 20 simulated MR images representing normal brains with ground truths for the major brain tissue classes (WM and GM). We run our algorithm using only one cluster ($N_T = 1$) such that the prior distribution on the cluster memberships equals one ($P(z_{ijt} = 1) = 1$ for $t = 1, \forall i, j$). With only one cluster, the prior on the atlas-to-image registrations [see (4)] reduces to an additional regularization of the deformation fields. As all images are from normal brains, this prior is not useful here. Hence, for fair comparison with other atlas-guided methods (see below), we remove this prior by setting ϵ^2_{jt} to be ∞ in this experiment.

To investigate the benefits of combined segmentation and atlas construction, we compare SPARC with a simplified version of our algorithm (denoted as SPARC-a) whereby the atlas itself is not updated, but a previously constructed atlas is used instead, which is obtained by first running the full version of SPARC on the same images. This comes down to sequential atlas construction and segmentation, whereby segmentation is performed by simultaneously estimating the Gaussian mixture model parameters and the atlas-to-image registrations similar to [12] (see Section II-F). To avoid the introduction of bias towards the atlas in the sequential algorithm, as well as to make a fair comparison with state-of-the-art techniques using predefined prior information, we also run SPARC-a with a different atlas, i.e., not generated by our method itself, namely the atlas available from SPM8b (http://www.fil.ion.ucl.ac.uk/spm/software/spm8b/). We denote these results as SPARC-a-spm.

To further assess the quality of the segmentations obtained from SPARC, we compare our results with two widely used single atlas segmentation tools, i.e., SPM8b (http://www.fil.ion.ucl.ac.uk/spm/software/spm8b/, [12]) and FSL (http://www.fmrib.ox.ac.uk/fsl/fsl/, [9]). The segmentations obtained by SPARC are compared to the other methods in this experiment by paired t-tests of their Dice overlap measures with the ground truth segmentations.

C. Experiment 3

The goal of this experiment is to evaluate the clustering performance of our framework on a heterogeneous set of clinical brain MR images. To this end, we try to separate between normal subjects and patients with neurodegenerative disease by applying our method on 3-D T1-weighted MR images of eight clinically confirmed Huntington's disease (HD) patients (two males, six females, age between 33–57 years with the average age 46 year) and eight normal controls (two males, six females, age between 28–48 years with the average age 38 years) [62]. All images were acquired on the same 3T MRI scanner (Philips Achieva) at the radiology department of the university hospital of our institute (UZ Gasthuisberg, Leuven, Belgium), with dimensions of $256 \times 256 \times 182$ and voxel sizes around $1\ \mathrm{mm}^3$. Manual segmentations are available for nucleus caudatus for all images. We run our algorithm using two clusters ($N_T = 2$) and set ϵ^2_{jt} empirically to 4 mm (based on an assessment of the inter-subject variance of the atlas-to-image deformations in experiment 2). We initialize the atlases A_{kt} as described in Section II-G, i.e., as the weighted sum of crude segmentations of all the images, whereby for each cluster one image of the set of 16 images is randomly selected which gets a larger weight than all others, i.e., a weight of 2/17 for the selected image and 1/17 for all other images respectively. We evaluate the clustering performance of our method by comparing with the clinical diagnosis and by assessing the ability of our method to pick up morphological features that are relevant for HD. Thereto, we relate the clustering obtained from SPARC to the (dis)similarity in morphology of the nucleus caudatus between the images, as nucleus caudatus has been indicated as a region largely affected by HD [2], [62]. We quantify the (dis)similarity by the Dice overlap measure between the available manual segmentations for nucleus caudatus of all images. Furthermore, the experiment contributes to assessing the segmentation performance of our framework on clinical images. However, as no ground truth is available for the segmentation of WM, GM and CSF, segmentation accuracy of our method can only be verified visually in this experiment. The quality of the atlases obtained for both clusters is also assessed visually for this experiment in terms of sharpness and stratification, i.e., is the disease-specific morphology clearly visualized by the atlases.

D. Experiment 4

In this experiment, we show that the definition of voxelwise cluster memberships can be useful for the study of specific research questions. We also show how clustering based on image features of brain morphology can provide new insights in the available clinical and demographical information. In addition, we use this experiment to provide a practical example of the benefits of combined segmentation and atlas stratification.

Data used for this experiment were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a $60 million, five-year public private partnership. The primary goal of ADNI is to test whether imaging measures, biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California—San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations. For up-to-date information, see www.adni-info.org.

For this experiment, structural brain MR images of 45 AD patients (aged 72.5 ± 3.1 years on average, range 65–77 years) are selected as well as 45 images of age-matched normal controls (aged 72.1 ± 2.1 years, range 65–75 years) from the ADNI database. All images are 1.5T 3-D T1-weighted baseline scans with voxelsizes around $1\ \mathrm{mm}^3$ that were corrected for MR acquisition artifacts as described in [63]. For all images, a segmentation of the hippocampus is available as provided by ADNI. As both groups in this experiment have the same age range, patterns of normal brain aging can be assumed to be similar for both, such that clusters detected by our method can be expected to expose patterns that are characteristic for AD. Finally, the images of three normal controls and of 16 AD patients showed clear signs of leukoaraiosis, i.e., a WM deficit that shows hypo-intense T1-contrast compared to the normal WM, as illustrated in Fig. 15(a). This might influence the segmentation and clustering process and is therefore studied more into detail.

We run SPARC using two clusters ($N_T = 2$) and set ϵ_{jt}^2 equal to 3 mm if $j \in \Omega$ and to 5 mm elsewhere, with Ω the hippocampal region (region of interest), in which a smaller intersubject variability can be assumed than in the rest of the brain (e.g., the cortex). We initialize the atlases (Section II-G) making use of some clinical prior knowledge, i.e., by averaging the initial segmentation of three images with respectively the lowest (AD cluster) and highest (normal cluster) Mini-Mental State Examination (MMSE) score. As hippocampal changes have been linked to AD [3], we wish to exploit and steer the voxelwise clustering mechanism of our algorithm to focus in particular on the hippocampal region. We can impose that the cluster memberships of all voxels are determined to a large extent by the cluster memberships of voxels in the hippocampal region, by using an MRF prior on the cluster memberships, as described in Section II-E [see (21)]

$$U\left(z_{ijt}=1|z_{\mathcal{N}_{ij}},\Phi_{z_{ijt}}\right)=\sum_{j'}\sum_{t'}z_{ijt}\alpha_{j'tt'}z_{ij't'} \tag{27}$$

with $\alpha_{j'tt'} = \beta_{j'} \cdot \gamma_{tt'}$. We set $\gamma_{tt'} = 0$ if $t = t'$, and $\gamma_{tt'} = 2$ otherwise. Hence, $\gamma_{tt'}$ is the MRF field strength, i.e., the amount of penalizing the clustering configurations in which the cluster membership of voxel j is different from those of voxels j' in neighborhood $\mathcal{N}_{ij}$. We define the neighborhood $\mathcal{N}_{ij}$ to contain all voxels of image i and choose $\beta_{j'}$ to be different for voxels j' in the hippocampal region Ω compared to voxels j' in the remainder of the image. If we wish that the clustering is primarily determined by the hippocampal region, a suitable choice is for instance

$$\beta_{j'}=\frac{0.1}{N_J-\sharp\Omega},\ \text{if } j'\notin\Omega \tag{28}$$

$$\beta_{j'}=\frac{0.9}{\sharp\Omega},\ \text{if } j'\in\Omega \tag{29}$$

with $\sharp\Omega$ the number of voxels in the hippocampal region. Thus, the prior on the cluster membership of each voxel is determined by the cluster memberships of all other voxels of the same image, however, the voxels in the hippocampal region are dominant over all other regions (90%–10%). To avoid the need of predetermined manual segmentations of the hippocampus for all images, we define the hippocampal region Ω generally within the MNI space and hence the same for all images. Its location is determined based on the union of an independent set of images (after affine alignment to MNI space) with available hippocampus segmentations (also extracted from ADNI). This union is then smoothed with a Gaussian kernel (kernel width 2) to account for the variability that is not captured within the given set, and subsequently a threshold is taken at 0.25.

To get more insight in our atlas stratification (and in its impact on the segmentation), we compare our stratification approach (SPARC) using image-based clustering with a simplified version of our method in which the clustering is kept fixed and identical to the clinical classification (denoted as SPARC-c). This means that the clustering process becomes supervised and the method boils down to running SPARC using a single cluster ($N_T = 1$) on each of the two clinical subgroups of 45 images separately. To evaluate the impact of the regional focus on the clustering, we also run the unsupervised version of SPARC (i.e., without MRF on the cluster memberships), which will be denoted as SPARC-u in this experiment. Finally, we show how our algorithm for combined segmentation and atlas stratification copes with the images showing leukoaraiosis, compared to techniques that perform the atlas construction and segmentation separately and sequentially, including SPM8b [12], SPARC-a-spm (SPARC-a using the SPM99 atlas (MNI305), i.e., a completely affinely constructed atlas) and SPARC-a-Wang (SPARC-a using the nonrigidly constructed atlas from [26]). In particular, the presence of leukoaraiosis in some images is likely to affect the atlas-guided segmentation, which may lead to inconsistencies

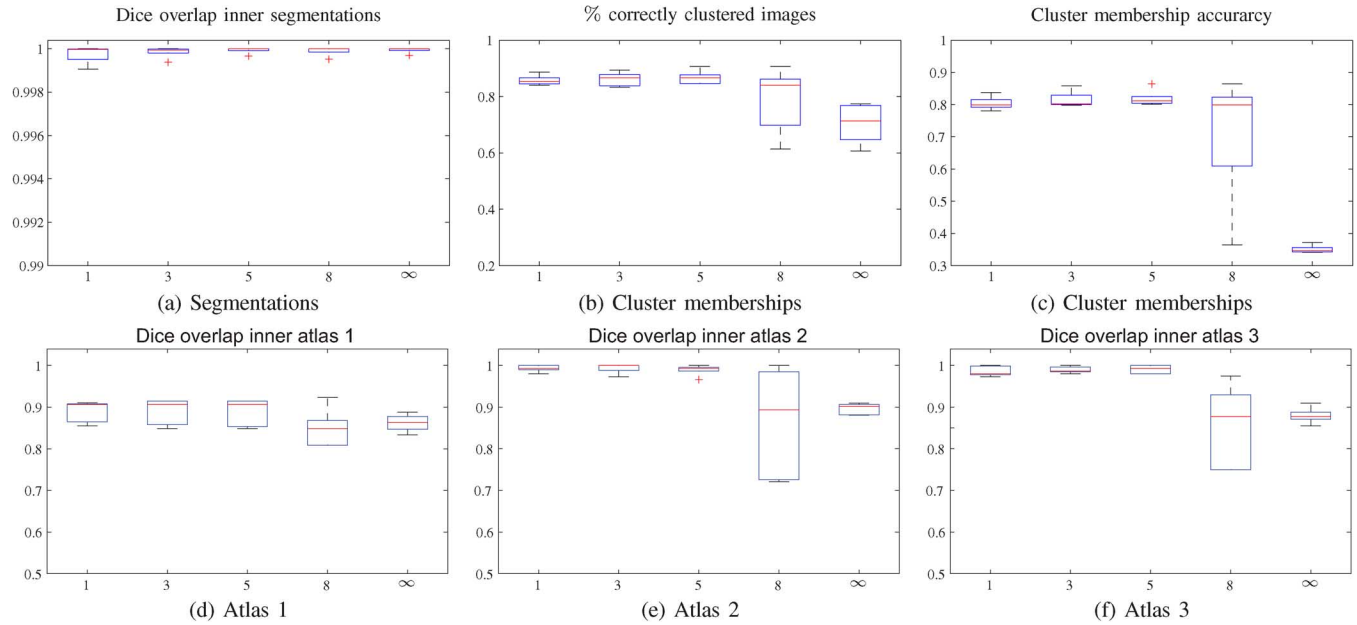


Fig. 3. Exp. 1: Boxplots indicating the performance of SPARC on the sets of synthetic images. The performance of SPARC is given for different values of the parameter ϵ_{jt}^2 (x-axis): (a) the segmentation accuracy (Dice overlap for inner structure), (b)–(c) the cluster membership accuracy [the percentage of correctly clustered images and by (26)], (d)–(f) the quality of the atlas for cluster 1, 2, and 3 (Dice overlap of ground truth and binarized segmentation for inner structure).

between different images that could complicate groupwise assessment of morphological differences. We use this experiment as a practical example of the benefits of combined segmentation and atlas construction for stratification as implemented in SPARC, showing that SPARC creates consistent segmentations over all subjects despite the intensity abnormalities present in some subjects contributing to the clustering process.

IV. Results

A. Experiment 1

In this experiment, our unsupervised framework using three clusters is performed on a data set of artificial images, for different values of the parameter ϵ_{jt}^2. The results are illustrated in Fig. 2 and summarized in the boxplots in Fig. 3.

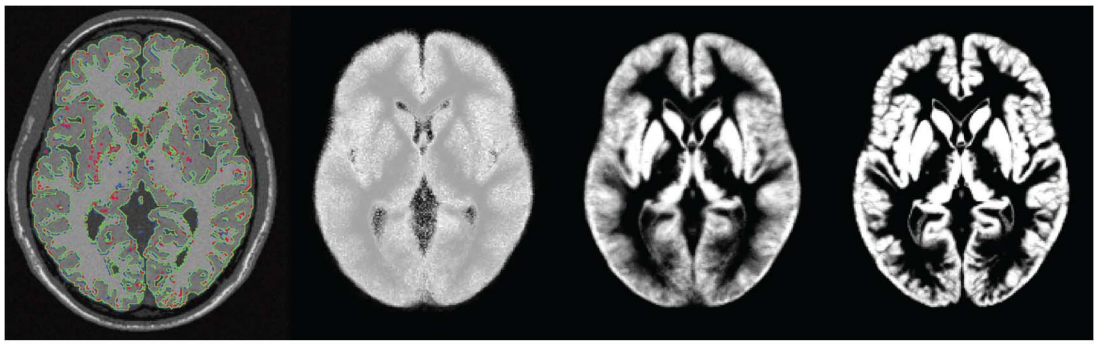
Overall SPARC segments the images well [e.g., Fig. 2(a), (b), and (c)]. The Dice scores for the inner circle/ellipse of all images are close to one [Fig. 3(a)]. Also the bias fields are well estimated [Fig. 2(e) and (d)]. The obtained cluster memberships are sharp and about 85% of the images are correctly classified [Fig. 3(c) and (b)]. We found that the misclassified images are all generated from the first template, as these are easily confused with images of the other clusters in case the radius of the inner circle is small.

The overlap of the atlases with their corresponding ground truth is large, resulting in high Dice coefficients when binairizing the probabilistic atlas maps [Fig. 3(d), (e), and (f)]. The obtained atlases are sharp and represent well the cluster-specific morphology (circle/vertical ellipse/horizontal ellipse), mimicking the original templates from which the images were created [Fig. 2(h), (i), and (j)]. Furthermore, the voxelwise cluster memberships clearly indicate the cluster-specific morphological features per image. For instance, in Fig. 2(f) the voxelwise cluster membership probabilities to belong to cluster 1 (circle) are given for an image that is generated from template 3 (horizontal ellipse). Most regions in the image have a similar morphology for all three clusters (grey). A larger difference between the image and template 3 compared to the other templates is visible in the vertical direction (black). In the horizontal direction, the image and template 3 have a similar morphological pattern, while it differs from template 2 (vertical ellipse). Therefore, the voxels are slightly highlighted in this region (white). We can thus conclude that our framework is capable of segmenting a set of images in tissue classes and detecting the major structural modes as well as the discriminative imaging features between the modes in the heterogeneous data set.

From the boxplots in Fig. 3, it is clear that the influence of the registration prior is quite important. A value of the variance ϵ_{jt}^2 equal to 3–5 mm seems to be preferable, what is conform with the way the images are generated. A too small value of ϵ_{jt}^2 can result in quite fuzzy atlases and/or in inferior segmentation results [e.g., Fig. 2(k)]. When the value is too large, large deformations are allowed and are not steering the clustering. The clustering performance decreases significantly and the atlases stay fuzzy as they do not show the template-specific morphology [Fig. 2(l)]. However, it is clear that a fairly similar performance is obtained for a relatively large range of values for ϵ_{jt}^2. This is an important result as we want the prior information needed in our unsupervised framework to be limited, to rely as much as possible on the data itself.

B. Experiment 2

The segmentation results for the BrainWeb images obtained by the different segmentation methods are summarized in Table I by the average Dice overlap measures per tissue class (and standard deviations), showing that our method performs superiorly to current state-of-the-art methods. An example of a



(a) Segmentations (b) Atlas: iteration 1 (c) Atlas: iteration 15 (d) Atlas: iteration 100

Fig. 4. Exp. 2: BrainWeb data set: (a) GM segmentation for one example image as obtained with SPARC (green) compared to SPM8b (red) and ground truth (blue). (b)–(d) Probabilistic GM atlas constructed during iterations 1, 15, and 100 of SPARC.

TABLE I
EXP. 2: BRAINWEB DATA—SEGMENTATION ACCURACY FOR THE 20 SIMULATED IMAGES IN TERMS OF % DICE OVERLAP (MEAN ± STANDARD DEVIATION). BOLD = HIGHEST VALUE, * = VALUES SIGNIFICANTLY DIFFERENT FROM SPARC (PAIRED T-TEST WITH 5% SIGNIFICANCE LEVEL)

	WM	GM
SPARC	**93.48±0.35**	**92.84±0.52**
SPARC-a	93.47±0.36*	92.30±0.63*
SPARC-a-spm	93.22±0.35*	92.30±0.53*
SPM8b	93.19±0.48*	92.20±0.85*
FSL5.1.0	93.44±0.36*	90.91±0.72*

segmented image is given in Fig. 4(a). Furthermore, the results indicate that combined segmentation and atlas construction is beneficial over sequential atlas construction and atlas-guided segmentation (comparing SPARC vs. the sequential methods SPARC-a, SPARC-a-spm and SPM). The atlases constructed over different iterations of SPARC are shown in Fig. 4(b)–(d). The atlases gradually converge over EM iterations to a sharp representation of the mean of the normal population.

C. Experiment 3

In this experiment, we investigate whether SPARC is capable of discriminating brain MR images of HD patients from normal controls by unsupervised image clustering. In Fig. 5, the image-wise cluster membership probabilities [see (16)] are plotted for all 16 images (eight normal controls, eight HD patients) in this study. A sharp clustering is obtained for most images and the obtained clustering corresponds to the clinical classification for 15 of the 16 images. Fig. 6 depicts the similarity between any pair of images as evaluated by the Dice overlap of their nucleus caudatus segmentations, one of the regions most affected by HD [2]. Comparing this to the cluster memberships in Fig. 5, similar trends are visible, e.g., image 10 and 11 (clinical HD) show larger similarity to the normal images than to the HD images, but smaller than the similarities among the normal images. This indicates that SPARC indeed detects the major clusters in a heterogeneous set of images based on relevant morphological features. The cluster-specific features are highlighted by the voxelwise cluster membership probabilities as can be seen in Fig. 7(a) and (b). Mainly voxels in the ventricles, nucleus caudatus and putamen are highlighted, indicating that these voxels drive the clustering process. The corresponding atlases, shown

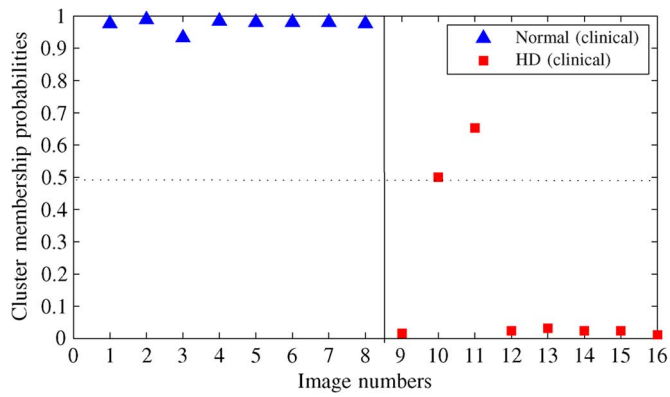


Fig. 5. Exp. 3: Cluster membership probabilities of all images to belong to the normal cluster. Image 1–8 are normal controls (blue triangles) and image 9–16 are clinically diagnosed with HD (red squares).

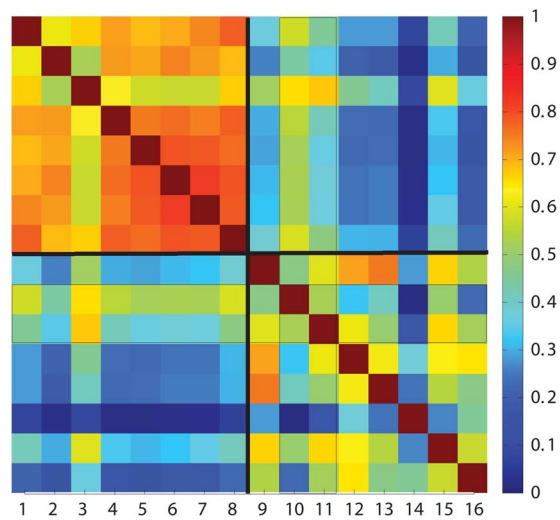
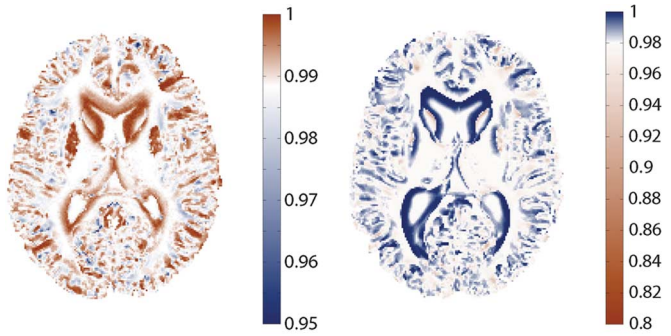


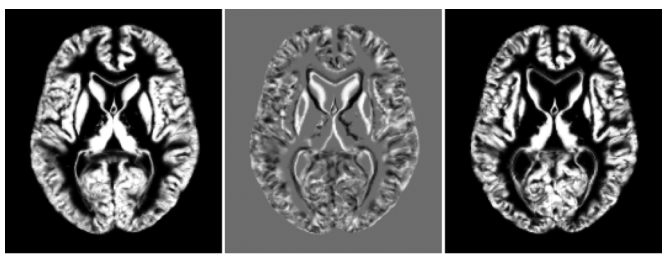
Fig. 6. Exp. 3: Pairwise Dice overlap of nucleus caudatus between all 16 images. Images are ordered as in Fig. 5, i.e., image 1–16 can be found from left to right and from top to bottom.

in Fig. 8(b), are sharp and clearly reveal the cluster-specific morphological features. The atlas of the HD patients has enlarged ventricles and clearly shows brain atrophy in putamen and nucleus caudatus and to a lesser extent in the cortex, when compared to the atlas of the normal controls.



(a) Normal (b) HD

Fig. 7. Exp. 3: Voxelwise cluster memberships for (a) a normal image to belong to the normal cluster and (b) a HD image to belong to the HD cluster. Identical regions (e.g., nucleus caudatus, putamen, ventricles) are highlighted in these images, indicating that these regions are affected by HD. Note that the images are illustrated at a different scale, related to the prior probability on the cluster memberships of both shown images which results in cluster membership probabilities that are on average lower for the shown HD image compared to the shown normal image.



(a) Normal Atlas (b) Difference (c) HD Atlas

Fig. 8. Exp. 3: Probabilistic GM atlas for (a) cluster 1 ("normal") and (c) cluster 2 ("HD") and (b) their difference image.

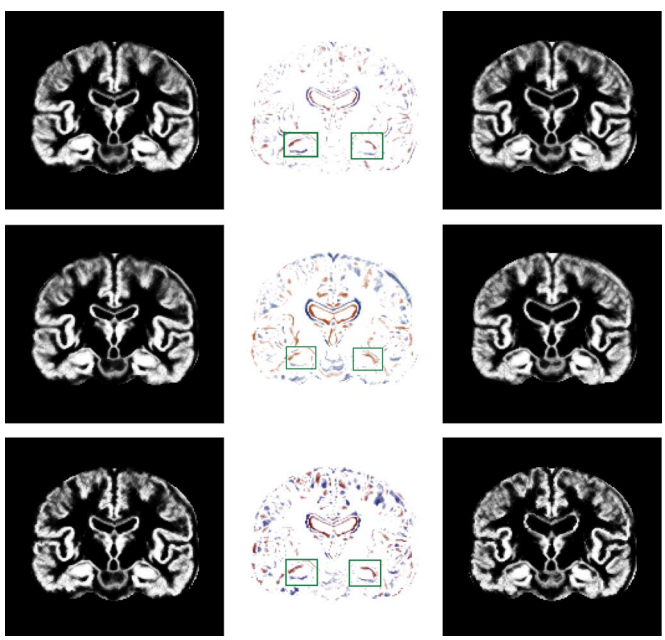


Fig. 9. Exp. 4: ADNI data set: probabilistic GM atlases constructed based on the clinical diagnosis (SPARC-c (supervised), top), based on global morphology-based stratification (SPARC-u (unsupervised), middle) and based on the local morphology-based stratification (SPARC (semisupervised), bottom). Left: group corresponding to healthy controls, right: group corresponding to AD, middle: difference map between both atlases ("AD"—"normal"), scaled between -1 (red) and $+1$ (blue) with the interval $[-0.2\ 0.2]$ set to white (i.e., no difference). The green boxes provide a rough outline of the hippocampal region. In general, differences are more pronounced by SPARC than by SPARC-c.

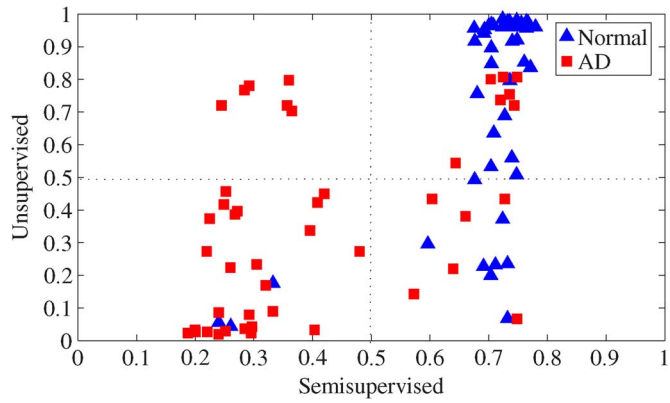


Fig. 10. Exp. 4: Probabilistic cluster membership to belong to the "normal" cluster obtained from semisupervised SPARC compared to those obtained from unsupervised SPARC. Blue triangles indicate the subjects clinically diagnosed as healthy, while the red squares indicate the subjects clinically diagnosed as AD.

D. Experiment 4

We applied SPARC to a set of 90 images from the ADNI database (45 AD patients, 45 normal controls), focusing largely on the hippocampal region to guide the clustering. Fig. 11 plots the clustering by SPARC, in terms of the image-wise probability to belong to the normal cluster, versus the subjects' age [Fig. 11(a)] and hippocampal volume [Fig. 11(b)]. Fig. 11(a) shows a similar distribution of ages for both clinical groups as well as for the two clusters obtained by SPARC, indicating that the influence of normal ageing patterns, which can exhibit those of AD, on the clustering can be neglected. The clustering obtained by SPARC agrees with the clinical diagnosis for 74 of the 90 images (82%) if a threshold of 50% on the image-wise "normal" cluster membership probabilities is applied. Even better agreement is obtained by adopting a larger threshold of 66% (78 of 90 images correctly classified, or 87%), which can be explained by the fact that the variability within the diseased cluster is expected to be larger than the one in the normal cluster. Fig. 11(b) relates the clustering by SPARC, which was largely influenced by the appearance of the hippocampal region, to the volume of the hippocampus as derived from the hippocampus segmentations provided by ADNI. Three clinically normal subjects that SPARC classified as AD have a smaller hippocampus volume than most other clinically normal images and six AD cases, classified as normal by SPARC, have a larger hippocampus volume than most of the other AD subjects. On the other hand, three images clinically diagnosed as normal are correctly classified as belonging to the "normal" cluster by SPARC although their hippocampus volume is also small. This can be explained by the fact that our clustering method considers not just the volume, but the entire appearance of the hippocampus, including position and shape, and (to a smaller extent in this experiment) the appearance of all other brain regions as well.

To study, the impact of the focus on the hippocampal region, we not only compare with the clinical diagnosis (i.e., supervised), but also perform unsupervised SPARC (without the MRF) clustering the images based on the total brain morphology. In Fig. 10, the probabilistic cluster memberships

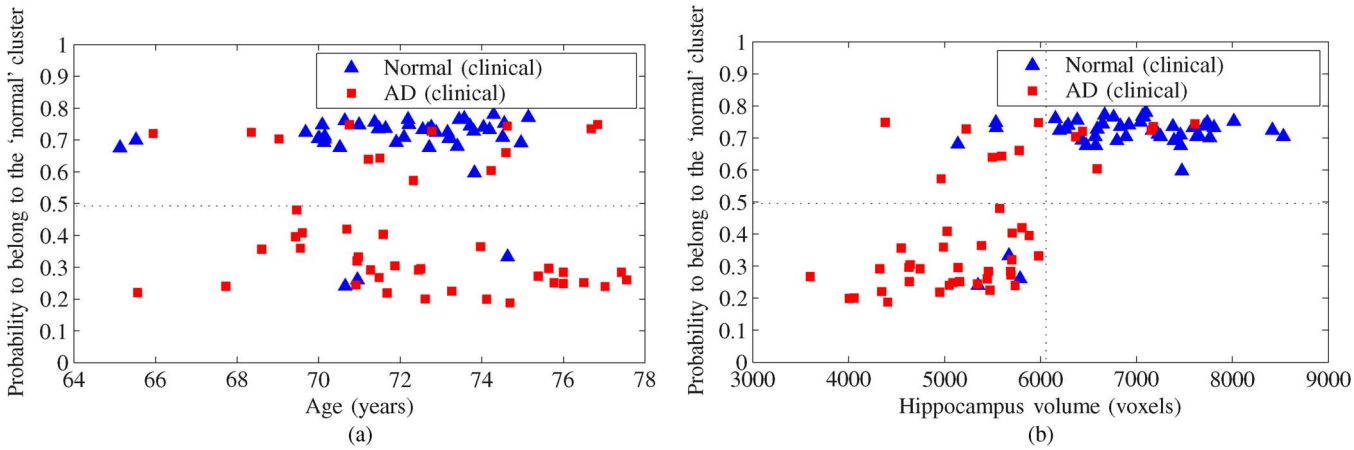


Fig. 11. Exp. 4: (a) Age and (b) hippocampus volume in terms of the probabilistic cluster memberships obtained from semisupervised SPARC. Blue triangles indicate the subjects clinically diagnosed as healthy, while the red squares indicate the subjects clinically diagnosed as AD.

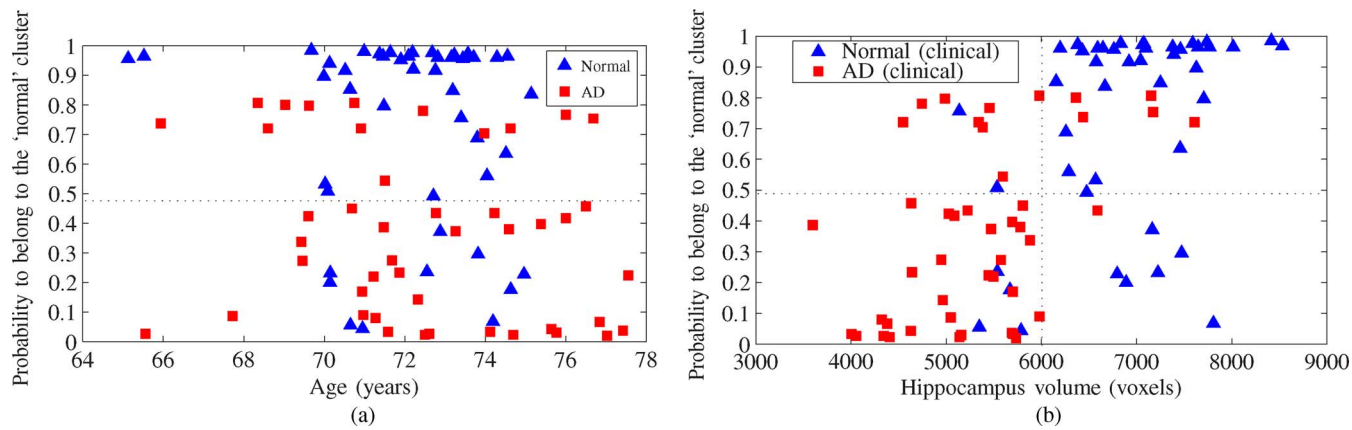
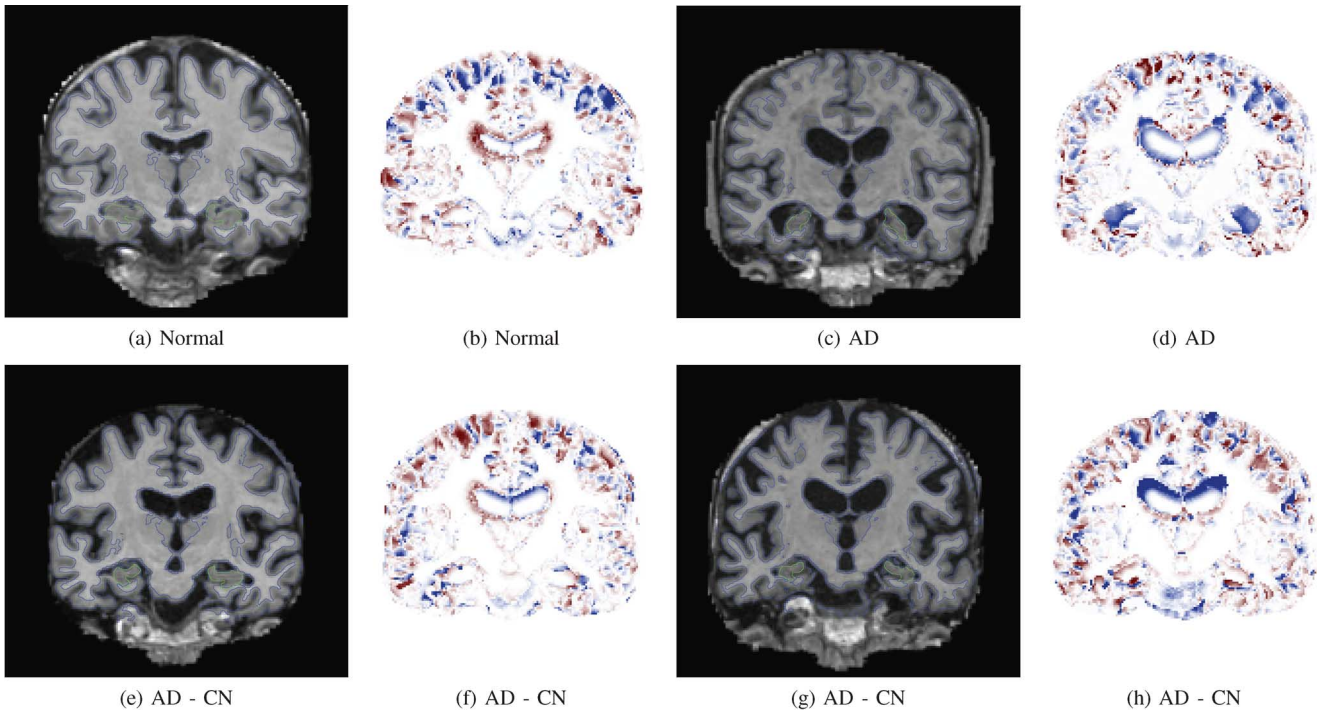


Fig. 12. Exp. 4: (a) Age and (b) hippocampus volume in terms of the probabilistic cluster memberships obtained from unsupervised SPARC. Blue triangles indicate the subjects clinically diagnosed as healthy, while the red squares indicate the subjects clinically diagnosed as AD.

obtained from SPARC with and without focus on the hippocampal region (i.e., semisupervised SPARC compared to unsupervised SPARC) are plotted. We notice that there is no strong correlation between the results of the unsupervised and semisupervised model. Furthermore, the overlap between the unsupervised model and the clinical diagnosis is smaller compared to that of the semisupervised model with the clinical diagnosis. In the unsupervised model, 66 of the 90 images (or 73%) are clustered according to the clinical diagnosis of the images, when applying a threshold of 50% on the image-wise "normal" cluster membership probabilities, compared to the 82% in the semisupervised model. In Fig. 12, we plot the image-wise probability to belong to the normal cluster obtained from SPARC-u versus the subject's age [Fig. 11(a)] and versus the hippocampus volumes [Fig. 11(b)]. The similar distribution of the ages for both obtained subgroups indicates that the influence of normal aging patterns can again be neglected. Furthermore, from Fig. 11(b), it follows that the correspondence between the clustering obtained from SPARC-u and a classification based on hippocampus volume, is a lot smaller than in case of our semisupervised approach [Fig. 11(b)]. This indicates that our semisupervised model is indeed largely driven by changes in the hippocampal region, while larger morphological variations might be present in different regions of the brain MR images of the heterogeneous data set, driving the unsupervised model.

To study major morphological changes between the constructed subgroups of both methods, which drive the individual cluster memberships, we analyze the constructed subgroup-specific atlases in more detail. These cluster-specific atlases represent the global morphological patterns for the different subgroups in the population. Fig. 9 shows corresponding coronal slices of the GM atlases as constructed by SPARC [morphological-based stratification focusing on the hippocampal region (semisupervised)], SPARC-u [morphological-based stratification without prior knowledge (unsupervised)] and SPARC-c [predefined subgroups based on clinical classification (supervised)]. We observe that the constructed atlases are largely similar between the different levels of supervision, what supports the hypothesis that AD is indeed linked to hippocampal changes. However, in the difference maps of the atlases, we observe slightly different patterns. For instance, it can be seen that the semisupervised case seems slightly more driven by the hippocampus and the inferior horn of the lateral ventricle, compared to the unsupervised case where the lateral ventricle (in general) seems to play a more important role in driving the clustering process. Furthermore, the difference maps of the atlases obtained from SPARC (both unsupervised



(a) Normal (b) Normal (c) AD (d) AD

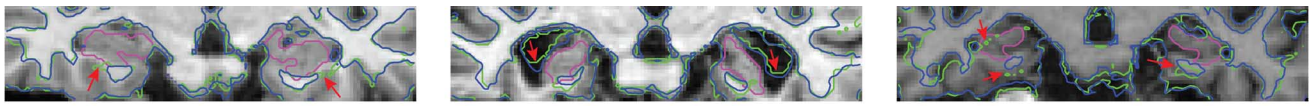
(e) AD - CN (f) AD - CN (g) AD - CN (h) AD - CN

Fig. 13. Exp. 4: ADNI images with GM segmentations obtained from SPARC (blue) and hippocampus segmentation available from ADNI (green): (a) a healthy subject (clinical diagnosis) belonging to the "normal" cluster, (c) an AD subject (clinical diagnosis) belonging to the "AD" cluster, (e)–(g) two AD subjects (clinical diagnosis) belonging to the "normal" cluster although the hippocampi are small. (b), (d), (f), (h): the voxelwise cluster memberships for each image, rescaled based on the estimated cluster membership of the image, i.e., in (16), for visualization purposes. The rescaled voxelwise cluster memberships highlight the voxels that want to drive the clustering even more towards a subgroup than the estimated image-wise cluster memberships. Red indicates driving more to the normal cluster, while blue indicates driving more to the AD cluster. We observe (for this slice) that in (f) voxels in the hippocampal region and the inferior horn of the lateral ventricles are rather driving to the AD cluster, while voxels in the cortex are pointing to the normal cluster. In (h) the voxels in the hippocampal region are rather driving to the normal cluster, while voxels in/around the central part of the lateral ventricles are driving to the AD cluster.

and semisupervised) show larger differences in morphology than those obtained using SPARC-c (supervised). Finally, the atlases obtained using SPARC are visually sharper, which is confirmed by their smaller self-information (entropy) of the atlases $-\sum_k A_{kt} \log A_{kt}$, which equals resp. 0.1618 and 0.1666 for SPARC and 0.1629 and 0.1678 for SPARC-u, compared to 0.1808 and 0.1747 for SPARC-c. This indicates that our image-based clustering is more appropriate to expose morphological features between both groups than a pure clinically based clustering. The fact that SPARC based on the hippocampal region results in sharper atlases than SPARC-u might indicate that the hippocampal region drives the clustering to a more global optima. This amplifies the advantage of our semisupervised model where we can include prior knowledge about the major changes in brain morphology.

We studied the general cluster memberships and the corresponding groupwise morphological cluster-specific patterns obtained from our semisupervised model. We now analyze the individual images of the data set. Fig. 13(a) and (c) show similar coronal slices through the hippocampus of the MR images of a normal and a AD subject, which are correctly classified as such by SPARC. Fig. 13(e) and (g) show MR images of two AD subjects with relatively small hippocampus volume, but classified by SPARC as belonging to the normal cluster. The corresponding voxelwise cluster memberships for all these images are shown in Fig. 13(b), (d), (f), and (h). These voxelwise cluster memberships highlight the regions that want to drive the clustering further towards one of both subgroups compared to the image-wise cluster membership. As the hippocampal region determines the prior on the cluster memberships for 90%, and has therefore a large impact on the image-wise cluster memberships, voxels in this region will only be highlighted if the cluster memberships of voxels in other regions (which determine the prior on the cluster membership for 10%) will be relatively different. Other highlighted voxels show that these regions can be even more important the hippocampal region to guide the clustering process between those two clusters. Studying these images, visualizing the driving voxels gives an indication of why the images shown in Fig. 13(a) and (c) are similarly classified by our algorithm than by the clinical diagnosis, while the images of Fig. 13(e) and (g) are differently classified by our algorithm than by the clinical diagnosis.

Finally, the contours of the binary GM segmentation, obtained by assigning each voxel of the probabilistic segmentation maps generated by SPARC to its most likely tissue class, are overlaid in blue in Fig. 13. It can be seen here that the obtained GM segmentations are overall quite accurate, but include some non-brain voxels, as is also apparent from the atlases shown in Fig. 9. This can be explained by the fact that the brain masks available from ADNI are sometimes inaccurate (as they were only constructed to perform bias field correction [63]), while for the experiments in this paper only three brain tissue classes were considered (and no class for the remaining is included). Furthermore, the segmentation in the hippocampal region, i.e.,



(a) (b) (c)

Fig. 14. Exp. 4: Prior models $\sum_Z P(L|Z,A,R)$ for GM segmentations in the neighborhood of the hippocampus obtained from SPARC: resp. based on cluster-specific atlases (SPARC) (blue), and based on disease-specific atlases (SPARC-c) (green). Hippocampus segmentations available from ADNI are given in pink. The shown images correspond to the images of Fig. 13 where we also showed the obtained segmentations by SPARC: from left to right resp. the images corresponding to Fig. 13(a) a healthy subject (clinical diagnosis) belonging to the "normal" cluster, Fig. 13(c) an AD subject (clinical diagnosis) belonging to the "AD" cluster, and to Fig. 13(g) an AD subjects (clinical diagnosis) belonging to the "normal" cluster although the hippocampus is small. Red arrows indicate regions were the prior based on the cluster-specific atlases seem to be better adapted towards the true segmentation compared to the disease-specific atlases.

the region that majorally determines the clustering, is studied more into detail. The segmentation is determined by the intensity model and a prior based on the constructed atlases. Fig. 14 shows the prior on the segmentations $\sum_Z P(L|Z,A,R)$ obtained from SPARC and SPARC-c for three cases of the images presented in Fig. 13. For the image assigned to the normal cluster by both SPARC and the clinical diagnosis, we obtain similar results for SPARC and SPARC-c, although the prior of SPARC-c seems to segment some GM to WM in the region below the hippocampus. For the image assigned to the AD cluster by both SPARC and the clinical diagnosis, we observe that both underestimate the seriously enlarged inferior horn of the lateral ventricle, but that the underestimation is worse for the SPARC-c prior. For the image of the AD person (clinical diagnosis) that is assigned to the "normal" cluster by SPARC, we observe the largest differences between the priors of SPARC and SPARC-c. In particular, SPARC-c seems to expect that the hippocampus is located lower than visualized by the image, what is indicated by the WM region below the hippocampus, as it is estimated by SPARC-c below the expected true segmentation. Although, the differences between the SPARC and SPARC-c priors are subtle, it can in general be observed that the prior based on the cluster-specific atlases (SPARC) seems closer to true segmentation for all images, than the prior based on the disease-specific atlases (SPARC-c).

We now analyze the segmentation of the AD image of Fig. 13(c) [also Fig. 15(a)], that shows seriously enlarged ventricles and leukoaraiosis (WM deficits), more into detail. Hereby, we will illustrate how the analysis of such images benefits from our framework for combined segmentation and atlas stratification, compared to techniques based on separate and sequential atlas construction and segmentation (i.e., SPM8b, SPARC-a-spm, and SPARC-a-Wang). Firstly, we segment the image using SPM8b and SPARC-a-spm, which make use of an *a priori* constructed fuzzy probabilistic atlas for WM, GM, CSF, constructed using a global registration algorithm from a data set of normal images. We also segment the image using SPARC-a-Wang, which makes use of a nonrigidly constructed probabilistic atlas for WM, GM, and CSF obtained from a data set of normal images. The segmentation results obtained from these techniques for the AD image [Fig. 15(a)] are shown in Fig. 15. It can be seen that the use of an *a priori* constructed fuzzy atlas (SPM8b, SPARC-a-spm) results in inconsistent incorporation of the leukoaraiosis in the tissue classes, i.e., different parts of the leukoaraiosis are segmented as different tissue classes (WM, GM and CSF). This is likely to lead to inconsistenties between the leukoaraiosis segmentations of the different images. Moreover, it also has an impact on the GM segmentation as the deep GM structures are underestimated. The use of an *a priori* constructed sharp nonrigid atlas (SPARC-a-Wang) results in an underestimation of CSF and GM as the large difference in morphology between the AD image and the sharp atlas, constructed from normal images, could not be captured by the model. Segmentations determined based on a fixed prior atlas can therefore result in a poor extraction of disease-specific image features, e.g., when analyzing the GM volume to quantify GM atrophy. In Fig. 16, the results obtained from SPARC are shown for the same slice. It is clear that the SPARC atlas is sharp, but can adapt towards the image as it describes the corresponding cluster-specific morphology, resulting in a more accurate segmentation. The leukoaraiosis is consistently included into the GM tissue class, which indicates that leukoaraiosis occurs in multiple images in more or less similar places (and is not merely an outlier occurring in a single image). Volumetric studies on the GM tissue class to extract imaging features can therefore also be problematic when using our model. However, our model extracts cluster-specific features directly based on the segmentations in a voxelwise way, i.e., by performing voxelwise clustering. As the leukoaraiosis is consistently detected between the images, it can possibly be picked up as feature. Our model indicates indeed that leukoaraiosis is more likely to occur in AD patients as these regions are highlighted [Fig. 16(b) and (d)]. Hence, our method for combined segmentation and atlas stratification provides segmentations and atlases adapted to the morphological pattern, which results in more consistent segmentations over all images of the data set (i.e., in contrast to the methods relying on *a priori* constructed atlases, handling an individual image), such that the clustering and feature detection can benefit from this specific segmentation. A more profound analysis using more clusters, i.e., to separate AD and normals with and without leukoaraiosis, will provide extra insights and might also contribute to the analysis of the differences between the clinical diagnosis and our morphological clustering of this heterogeneous data set.

V. Discussion

In this paper, we have presented a unified probabilistic framework, called SPARC, that simultaneously segments a set of images in tissue classes and clusters them into different subpopulations without the need for prior knowledge. The method automatically generates nonrigid probabilistic atlases for each subpopulation and reveals the location of cluster-specific morphological features for each image. The unified framework makes

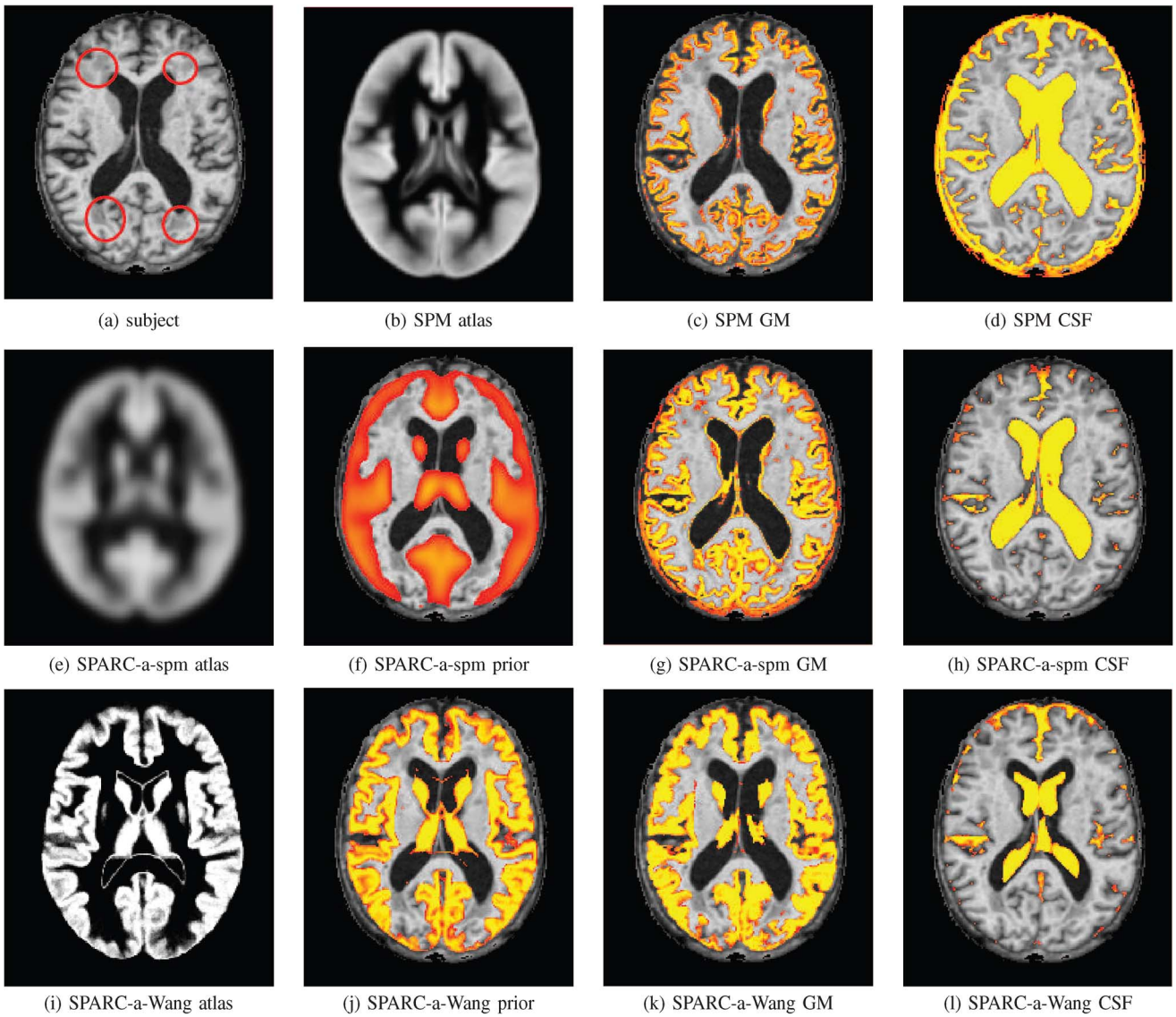


Fig. 15. Exp. 4: Illustration of the need for appropriate atlases for segmentation and analysis of a subject. Atlas-driven tissue class segmentation of a brain MR image of an AD patient with seriously enlarged ventricles and leukoaraiosis (red circles). (a) Subject, (b)–(d) the SPM8b atlas and corresponding segmentations obtained with SPM8b [12], (e)–(l): the *a priori* atlases (see text for details), corresponding deformed atlases and obtained segmentations using SPARC-a. Colors [red to yellow] indicate probabilistic values [0.5–1] in all images. It is clear that a fuzzy atlas [(b) and (e)] results in a random incorporation of the leukoaraiosis (i.e., outliers) in the WM or GM (or CSF) segmentation, and in an underestimation of the deep GM structures. The segmentation based on the sharp atlas (i) is not influenced by the leukoaraiosis, but is not able to cope with the large changes in brain morphology caused by AD.

that all these aspects can benefit from each other. The combination of segmentation, registration, atlas construction and (unsupervised) clustering, while exposing the morphological features between clusters, makes the proposed method a comprehensive image-driven population analysis framework.

We explained in Section II-C how our framework relates to state-of-the-art segmentation techniques for label fusion [19], [37]–[41] including local atlas selection and the use of an intensity model to further improve the segmentation performance as we deal with homogeneous regions as in [18], [19]. Contrary to the methods of [18], [19], the parameters of the intensity model used for atlas-guided segmentation and of the atlas-to-image registration are iteratively and jointly updated by our method as in [12], [14], [15], which is shown to be beneficial over performing segmentation and registration sequentially. However, the models in [12], [14], [15] are only capable of handling a single atlas, while our algorithm makes use of multiple atlases.

A second major difference to all previously mentioned methods is that our method does not assume the atlases to be given *a priori*, but constructs these during the segmentation process itself. Combining atlas construction and segmentation has already been shown to be beneficial over performing both algorithms separately and sequentially [43], [44] as the same registration flexibility is used for atlas construction and atlas-guided segmentation [28]–[30] and as the atlases are better adapted to the images. In case a single cluster is used, SPARC reduces to an unbiased probabilistic atlas estimation algorithm, as discussed in Section II-F, similar to the ones proposed in [21], [25], [27].

However, the main issue addressed in this paper is that a single template is not sufficient to summarize the variability in a

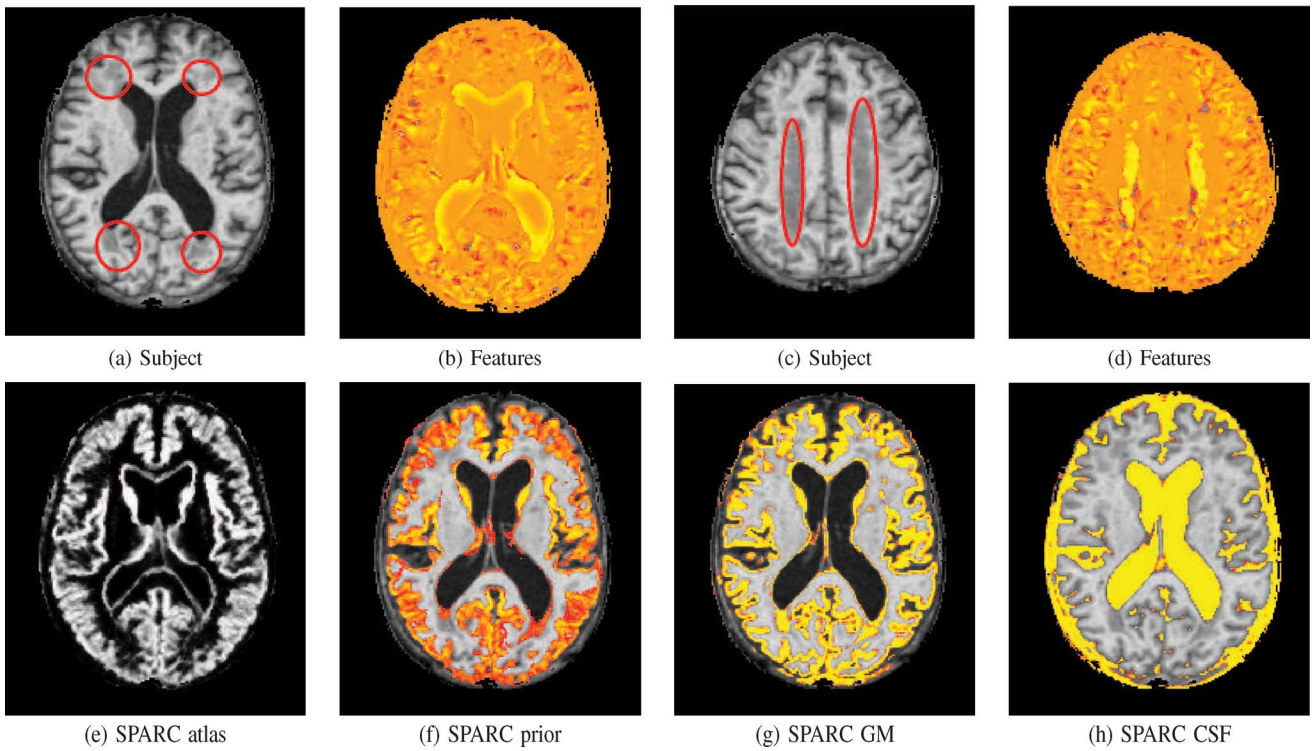


Fig. 16. Exp. 4: Illustration of the advantages of combined segmentation and atlas stratification using SPARC for the analysis of an individual subject, compared to state-of-the-art segmentation techniques (Fig. 15). Atlas-driven tissue class segmentation of a brain MR image of an AD patient with seriously enlarged ventricles and leukoaraiosis (red circles). (a) and (c) Two slices of the same subject. (b) and (d) The corresponding voxelwise cluster memberships to belong to AD for these slices. For the first slice, we also show: (e) AD atlas for GM, (f) deformed atlas, (g)–(h) probabilistic segmentations of GM and CSF ([red yellow] $= [0.5\,1]$). The atlas is sharp, but can clearly adapt towards the subject to be segmented. The leukoaraiosis is systematically segmented as GM, indicating that it systematically occurs in the data set, in particular in the AD subjects, in the same spatial location. This becomes also clear from the voxelwise clustering highlighting the leukoaraiosis regions as more likely to belong to AD.

large and heterogeneous population of images as also argued by [31], [32]. Hence, SPARC adopts a multi-atlas stratification and clustering strategy, in line with recent work [50]–[52]. The main difference with these methods is, however, that SPARC models the heterogeneity of the population explicitly by creating probabilistic atlases and uses this setting immediately to segment the set of images and to find the cluster-specific morphological features based on these segmentations. The cluster-specific features are obtained both groupwise, by comparing the atlases, and individually, as the relevant features are highlighted in each image separately by the voxelwise cluster memberships.

The use of segmentations for atlas stratification avoids the need for intensity normalization and overcomes problems with image artifacts, compared to the stratification/clustering approaches proposed in [50]–[52]. Handling both segmentation and atlas stratification in a single framework allows these techniques to cooperate. The stratification method embedded in the segmentation procedure makes that more appropriate image-specific prior knowledge is driving the segmentation process than in case a single atlas was constructed. This can be explained by the fact that the use of *a priori* defined subgroups, e.g., based on a clinically diagnosis, does not necessarily result in an optimal separation (and atlas construction) according to morphological patterns, while our stratification explicitly focuses on a discrimination based on the morphology. In addition, the local cluster memberships defined in SPARC, allow local selection of images for atlas construction and subsequently an appropriate local selection of atlases for segmentation, based on the same criterion. In contrast, performing combined segmentation and atlas construction based on clinically defined subgroups, boils down to a global selection. The use of local weights for atlas stratification not only allows a locally adaptive and hence more optimal atlas selection, but also makes that more images contribute to the construction of the atlases (and therefore to the segmentations) in case the inter-cluster differences are small, as discussed in Section II-E. Both these facts, i.e., cluster-specific atlases rely on more information (i.e., due to local selection) and directly rely on the morphology of the data set (i.e., based on the optimal similarity between atlases and segmentations of the images), makes that segmentation based on cluster-specific atlases is likely to outperform segmentation based on disease-specific atlases. This second argument is also stated by [52]. Furthermore, the statement is further amplified as by combining the clustering and the atlas construction, the same criterion is used to assign the images to a subgroup for atlas construction as to select the most appropriate atlas for image segmentation.

In turn, the atlas stratification also benefits from the combination with the segmentation algorithm. From previous discussion, it follows that the obtained segmentations are a more accurate representation of the specific morphological patterns of the data set, compared to segmentations obtained based on *a priori* constructed atlases. As such the morphological patterns are more distinct what can contribute to an improved stratification.

The fact that our method explicitly models the heterogeneity of the population by detecting relevant image clusters and exposing the cluster-specific image features, makes our method capable to perform cross-sectional studies. The major difference with state-of-the-art cross-sectional studies (e.g., [3], [6]) is that our algorithm performs an unsupervised clustering, while most state-of-the-art methods are based on a clinical hypothesis and perform therefore supervised analyses. The advantage is that our method can contribute to the discovery of new subgroups, e.g., characterizing new subtypes of a disease, and new morphological features leading to new research questions. However, different types of prior information can be taken into account in the semisupervised clustering scheme described in Section II-E, including prior information about regional features and/or clinical information, to study specific clinical hypotheses.

The experiments demonstrate the use of SPARC in multiple settings. In the first experiment, a proof-of-concept of the framework is given based on synthetic images. The second experiment demonstrates that the segmentations obtained with SPARC are more accurate than those obtained with conventional segmentation methods. The experiment also confirms that combined segmentation and atlas construction results in improved segmentation results over the use of a predefined probabilistic atlas for segmentation. The improvement is, however, limited in this experiment as the data set is homogeneous and consists of healthy controls, while predefined atlases, such as used by e.g., SPM, are also often constructed from images of healthy controls. A greater improvement can therefore be expected when segmenting images showing large morphological changes, e.g., caused by a neurodegenerative disorder, compared to normal images.

The third and fourth experiment explore our algorithm in a heterogeneous setting, namely normal controls versus HD and AD patients respectively. The clustering obtained with SPARC largely agrees with the clinical classification for both experiments. This not only provides evidence that our algorithm is capable of distinguishing subgroups in a population, as the major variability in morphology within the population is expected to be caused by the disease, but also shows that the obtained subgroups are clinically relevant. As such the voxelwise cluster memberships estimated by our method allow to localize disease-specific features for each image individually, while also the features at cluster level can be obtained (e.g., by comparing the cluster-specific atlases generated by our method). Hence, our method reveals for each subject local features that conform or do not conform with the overall pattern of each of the clusters. This could be exploited in further research to develop novel strategies for testing specific hypotheses regarding an individual subject, for instance for early diagnosis of subtle disease patterns.

Furthermore, experiment 4 shows that our clustering approach results in cluster-specific atlases with more pronounced differences than in case atlases are constructed based on clinical knowledge. This indicates that our approach for atlas stratification results in a more optimal separation of morphological patterns and in atlases that better capture the variability within the population than using clinical defined subgroups. Hence, our method could be capable of detecting more subtle morphological variations and could contribute to the discovery of new subgroups, possibly differing in unexpected ways. Moreover, the improved capturing of the morphological variability by the atlases supports our expectation that segmentation based on cluster-specific atlases corresponding to the image-modes in a population, is likely to outperform segmentation based on disease-specific atlases. This expectation is further amplified in the experiment by the priors of the segmentations obtained from disease-specific and cluster-specific atlases, where the priors based on cluster-specific atlases seem closer to the true segmentations than those based on the disease-specific atlases. Finally, experiment 4 also provides an explicit example of the benefits of combined segmentation and atlas stratification using a data set of normal and AD patients, where some images show a WM deficit (leukoaraiosis). The example shows that in sequential techniques, prior information used for the segmentation is often not appropriate, e.g., coming from normal patients to segment a diseased patient and as such introducing bias towards the normal population. Furthermore, in sequential techniques, they must define specific guidelines for specific morphological patterns, such as these WM deficits, to obtain consistencies within and between the images, while they are automatically handled using our method. In particular, the experiment shows how some morphological patterns deviating from the normal tissue structures (e.g., leukoaraiosis) can be picked up by our combined method when they arise in multiple images, while they might be discarded when performing both processes separately and sequentially. The detection of these novel morphological patterns has shown to improve the segmentation performance and allows the detection of novel cluster-specific features.

Although, we have shown the advantages of combined segmentation and atlas stratification (i.e., less bias and detection new morphological patterns), we must notice that the use of prior information about the segmentations might still be helpful. In our algorithm, we assume that all information can be learned from the images. However, this assumption might not be valid, e.g., structural differences within a tissue might lead to poor image contrast visible in all brain MR images. Furthermore, the current algorithm still suffers from a few limitations. Firstly, the method requires atlas-to-image and image-to-atlas registration (Section II-H and Appendix B) in each iteration between all atlases and all images. This makes the method computationally intensive. Also, the registrations are updated based on the optimization of the expected loglikelihood function (analogously to [15]). However, as there is no closed solution for the expected loglikelihood in terms of the registrations R, an improved registration might be obtained by optimizing the loglikelihood function directly (analogously to [12]), what also results in more consistency between registration and segmentations. Secondly, the method as implemented here requires accurate brain masks to overcome that non-brain tissue is included in brain segmentation (i.e., as shown in experiment 4). The inclusion of non-brain tissue will also induce a bias in the Gaussian mixture model parameters. Moreover, these incorrectly segmented voxels can influence the cluster memberships. Adding an extra tissue class for non-brain tissue can resolve this problem and including constraints as proposed in [16] could further improve the segmentation of non-brain versus brain tissue. A third limitation is that

the initialization can have an impact on the clustering, although the basic version of the method is assumed to be unsupervised. This requires a more profound analysis. In this paper, the impact of the initialization on the clustering was only determined by the initialization of the atlases, i.e., as a specific subset of images was selected for their initial construction (Sections II-G and III). However, experiment 3 indicates that the influence is restricted as a "random" initialization is used (Section III), while the obtained results for the clustering follow our expectations. Moreover, often a specific research question is defined and/or clinical knowledge is available. This helps to select an more optimal set of images for the initial construction of the atlases (e.g., experiment 4). A last limitation concerns the parameter ϵ_{jt}^2 which need to be defined in advance. In experiment 1, we concluded based on the artificial images that the influence of this parameter on the final results is limited for a quite large range of values. However, locally adapting the prior, e.g., allowing a larger variance in regions where the inter-subject variability is known to be larger such as the cortex, might further improve the clustering. Also, varying the parameter for the different subgroups, as it can be expected that the variability is larger in a "diseased" cluster compared to the normal cluster, might further improve our algorithm.

This paper focuses on the segmentation accuracy and the clustering performance of our algorithm. Further research will focus on the use of the constructed atlases to segment new unseen images as well as on the classification of new images. Also, not discussed in this paper, the framework is implemented to handle multi-modal images. As the method is a generic framework, it can be used in all types of settings. Different prior distributions and models could be plugged in our framework, e.g., non-Gaussian models for the intensity [64] or alternative models for the clustering and registration prior. This enlarges the generalisability of our framework even more. Finally, we assumed that the number of clusters is specified in advance. In case of a specific research question or clinical hypothesis, the optimal number of clusters is typically known, e.g., when we want to investigate how AD patients can be distinguished from normal controls. However, the proposed framework can also be very useful in cases where the optimal number of clusters is unknown, e.g., in spectrum disorders, the optimal number of stages in an aging study, etc. Thereto, further research could focus on the development of a model to determine this optimal number. Different methods for automatically detecting the optimal number of clusters in a data set have already been proposed in literature [65], [66]. Such techniques are often based on systematically optimizing the loglikelihood for a different number of clusters, while including a cost term for this number. Evaluating the cluster memberships generated by our method could also provide insight in the required number of clusters for a particular data set: if multiple images are assigned with equal probability to two different clusters, these images are likely to form a new subgroup.

VI. Conclusion

We have presented a unified probabilistic framework, for the analysis of large heterogeneous sets of brain MR images, that simultaneously performs image segmentation, clustering in subgroups, and groupwise registration. The framework has been illustrated to be a generalization of different state-of-the-art techniques for the individual aspects. However, it's setup differs in four major ways: 1) atlas-guided segmentation and atlas stratification are performed simultaneously, 2) atlas stratification is performed locally, 3) cluster-specific morphological patterns are automatically exposed in a groupwise way as well as in each individual image, 4) clinical prior knowledge can be included directly to complement information extracted from the brain morphology.

The framework is applied to multiple data sets, including BrainWeb and ADNI, to illustrate its feasibility and evaluate its performance. It is shown that the segmentation benefits from the combination with atlas stratification. Furthermore, the experiments demonstrate that our algorithm is capable of finding the major modes of variability in the population, while the constructed atlases clearly represent the cluster-specific brain morphology. Finally, it is illustrated how new morphological features might be picked up by our combined method over handling the different aspects separately and sequentially.

On a methodological side, future work will focus on improving the registration and automatically determining the optimal number of clusters. However, as important will be exploring the potential contribution of the presented method in multiple applications. Our major interest goes here to the impact of regrouping images, based on the morphology compared to clinical subgroups, for the detection of novel imaging biomarkers. Furthermore, the method might contribute to the development of proper representations of spectrum diseases constructing morphology-based subgroups, or to the extraction of endo-phenotypes for genetic association studies by separating environmental factors by those correlated with genetics.

Appendix A

In this appendix, we derive the equations to update the atlases in the maximization step of the SPARC algorithm. We have to determine

$$\frac{\partial Q}{\partial A_{ukt}} = 0, \text{ subject to } \quad \sum_{k=1}^{N_k} A_{ukt} = 1 \tag{30}$$

with u indexing the voxels in the atlas space. Therefore, we rewrite the part of the Q-function dependent on the atlas, in terms of voxels in the atlas space

$$\begin{aligned}
&\sum_i \sum_j \sum_k \sum_t \rho_{ijt} p_{ijkt} \log\left[A_{kt}\left(R_{it}(x_{ij})\right)\right] \\
&\quad \propto \sum_{i,k,t} \int_{\Omega_j} \rho_{it}(x_{ij}) p_{ikt}(x_{ij}) \log\left[A_{kt}\left(R_{it}(x_{ij})\right)\right] dj \\
&\quad = \sum_{i,k,t} \int_{\Omega_u} \rho_{it}\left(H_{it}(u)\right) p_{ikt}\left(H_{it}(u)\right) \log\left[A_{kt}(u)\right] \cdot \\
&\qquad \left|\mathrm{Jac}\left(H_{it}(u)\right)\right| du \\
&\quad \propto \sum_{i,k,t,u} \rho_{it}\left(H_{it}(u)\right) p_{ikt}\left(H_{it}(u)\right) \log\left[A_{kt}(u)\right] \cdot \\
&\qquad \left|\mathrm{Jac}\left(H_{it}(u)\right)\right|
\end{aligned}$$

with $H_{it}(u) = R_{it}^{-1}(u)$, with $|\cdot|$ the determinant of a matrix and $\mathrm{Jac}(H_{it}(u))$ the Jacobian matrix of H_{it} containing the partial derivatives of the warp field with respect to the coordinates. Furthermore, Ω_j is a continuous and compact subset of $\mathbb{R}^3$ that covers the discrete set $j = \{1, \ldots, N_J\}$. Thus, the Q-function is rewritten in terms of the voxels in the atlas space using the substitution $u = R_{it}(x_{ij})$ and by interpreting the discrete lattice temporarily as a continuum using an integral instead of a summation. Solving (30) comes down to the computation of the saddle point of the following Lagrangian with Lagrange multipliers λ_{ut}:

$$\min_{\lambda_{ut}} \max_{A_{ukt}} \sum_i \sum_k \sum_t \sum_u \rho_{it}\left(R_{it}^{-1}(u)\right) p_{ikt}\left(R_{it}^{-1}(u)\right) \log\left[A_{ukt}\right] \cdot \left|\mathrm{Jac}\left(R_{it}^{-1}(u)\right)\right| - \lambda_{ut}\left(\sum_k A_{ukt} - 1\right)$$

resulting in

$$A_{kt}(u) = \frac{\sum_i \rho_{it}\left(R_{it}^{-1}(u)\right) p_{ikt}\left(R_{it}^{-1}(u)\right) \left|\mathrm{Jac}\left(R_{it}^{-1}(u)\right)\right|}{\sum_i \rho_{it}\left(R_{it}^{-1}(u)\right) \left|\mathrm{Jac}\left(R_{it}^{-1}(u)\right)\right|}$$

The atlas is thus constructed as the weighted sum of the segmented images deformed towards the atlas space. The weights equal the cluster membership probabilities (voxelwise). Moreover, the atlas construction implies a modulation step, i.e., a scaling of the deformed segmentation maps with the Jacobian determinant to preserve their local probabilistic volumes.

Appendix B

In this appendix, we discuss the computation of the inverse registration of R to determine the image-to-atlas registration for atlas construction given by (17) and detailed in Section II-H. An exact derivation of the inverse is possible as we constrain all registrations R to be diffeomorphic. However, the atlas-to-image registrations and image-to-atlas registrations are not exactly each others inverses as simultaneously updating the atlases and the atlas-to-image registrations is unfeasible. In practice the atlases A and registrations R are not updated at the same time by the M-step, but sequentially, i.e., the atlases are reestimated first before updating the atlas-to-image registrations. Accordingly, forward- and inverse registrations transform (to) different estimations of the atlases. This becomes also clear from the schematic presentation of our framework (Fig. 1). It shows that first the segmentations are estimated. Subsequently, an estimate of the image-to-atlas registrations is required to compute a new estimate for the atlases. Lastly, the atlas-to-image registrations are computed based on the new estimates of the atlases and as such a new prior is obtained to update the segmentations. It is clear that the required image-to-atlas registrations will only equal the inverse atlas-to-image registrations when the algorithm has converged, i.e., when the estimated segmentations and atlases do not change anymore. Therefore, we prefer to replace the inverse registration R^{-1} by a forward image-to-atlas registration, called H, updating a similarity measure based on the previous estimation of the atlas. As such, we allow a more adapted registration per iteration, what avoids ending up in local optima of the overall EM algorithm. The registration H must converge to the inverse registration R^{-1} when new updates of the image segmentations and atlases become minimal and when the deformed images exactly match the atlases and vice versa. The last argument states that the assumption of a diffeomorphic relationship between the atlas and the image must be valid.

From the Q-function it follows that the following similarity measure need to be maximized to update the registration H:

$$\rho_{it}(H_{it}(u)) \left[\left(\sum_k p_{ikt}\left(H_{it}(u)\right) \log \frac{A_{ukt}}{p_{ikt}\left(H_{it}(u)\right)} \right) + \log\left(\frac{1}{\sqrt{2\pi\gamma_{ut}^2}} \exp\left(-\frac{1}{2} \frac{(H_{iut} - M_{ut})^2}{\gamma_{ut}^2} \right) \right) \right] \quad (31)$$

with M_{ut} the mean deformation in u from the image towards the atlas space and γ^2 the variance. The first part of the similarity measure described in (31) tries to match the probability maps of the image and the atlas (for cluster t). This part equals almost exactly the expected likelihood under the substitution $u = R_{it}(j)$ (considering the grid temporally as a continuum). However, the posterior of the expected likelihood is not fixed anymore during the maximization step, i.e., when updating the model parameters. Therefore, the second term obtained in Jensen's inequality [see (9)] is not constant anymore and should be taken into account when updating H_{it}. Under the assumption that ρ_{ijt} is locally constant, the first part of (31) is obtained. Remark that this part is equal to the similarity measure used in [27] to construct a probabilistic atlas. Moreover, under the assumption that ρ_{ijt} is constant, the measure has the same optimum as $\sum_k A_{ukt} \log(p_{ikt}(H_{it}(u)))$, which is symmetric to the similarity measure used to update the forward atlas-to-image registration.

The second part of (31) describes a similar constraint on the registration than the one described on the forward atlas-to-image registration, although it has a slightly different interpretation. It is assumed that the voxels in the image spaces corresponding to the same voxel u in the atlas space are Gaussian distributed with mean voxel $M_t(u)$. Therefore, the registration H_{it} must stay close to a groupwise image-to-atlas registration M_t. Here, M_{ut} equals the identity transformation as we reestimate the atlas each time such that G_{tj} equals the identical transformation (see implementation: Section II-H) and M_{ut} can be seen as the inverse of G_{tj}. The variance γ^2 in the image space equals then

$$\gamma_t^2(u) = \epsilon_t^2\left(M_t(j)\right) = \epsilon_t^2(j). \quad (32)$$

The force field which drives the image-to-atlas registration H_{it} is now as follows:

$$F_{iut} = \rho_{it}\left(H_{it}(u)\right) \left[\left[-\frac{(H_{iut} - M_{ut})}{\gamma_{ut}^2} \right] + \left[\sum_k \left(\log\left(\frac{A_{ukt}}{p_{ikt}\left(H_{it}(u)\right)} \right) - 1 \right) \nabla_{H_{it}(u)} p_{ikt}\left(H_{it}(u)\right) \right] \right]. \quad (33)$$

Both terms in the force field are multiplied with $\rho_{it}(H_{it}(u))$, forcing the force field to have a larger impact on the construction of the deformation field in case a voxel with a larger probability to belong to the atlas is reached. For simplicity and stability, the

transformed cluster memberships $\rho_{it}(H_{it}(u))$ are kept constant in each iteration step, when updating the registrations, i.e., no derivative to update the registration is computed towards this term.

ACKNOWLEDGMENT

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data, but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

REFERENCES

[1] V. Valcour, K. Masaki, J. Curb, and P. Blanchette, "The detection of dementia in the primary care setting," *Arch. Internal Med.*, vol. 160, no. 19, pp. 2964–2968, Oct. 2000.

[2] J. Vonsattel and M. Difiglia, "Huntington disease," *J. Neuropath. Exp. Neurol.*, vol. 57, no. 5, pp. 369–384, 1998.

[3] D. Chan, N. Fox, R. Scahill, W. Crum, J. Whitwell, G. Leschziner, A. Rossor, J. Stevens, L. Cipolotti, and M. Rossor, "Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease," *Ann. Neurol.*, vol. 49, no. 4, pp. 433–442, Apr. 2001.

[4] N. Fox and J. Schott, "Imaging cerebral atrophy: Normal ageing to Alzheimer's disease," *Lancet*, vol. 363, no. 9406, pp. 392–394, 2004.

[5] V. Molina, S. Reig, J. Sanz, C. Benito, J. Pascau, F. Collazos, F. Sarramea, J. Artaloytia, J. Gispert, R. Luque, T. Palomo, C. Arango, and M. Desco, "Association between relative temporal and prefrontal sulcal cerebrospinal fluid and illness duration in schizophrenia," *Schizophrenia Res.*, vol. 58, no. 2–3, pp. 305–312, Dec. 2002.

[6] J. Barnes, J. Whitwell, C. Frost, K. Josephs, M. Rossor, and N. Fox, "Measurements of the amygdala and hippocampus in pathologically confirmed alzheimer disease and frontotemporal lobar degeneration," *Arch. Neurol.*, vol. 63, no. 10, pp. 1434–1439, Oct. 2006.

[7] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, J. A. Clifford, R. Jack, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, Feb. 2008.

[8] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based bias field correction of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 885–896, Oct. 1999.

[9] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.

[10] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, Jan. 2002.

[11] B. Fischl, D. H. Salat, N. A. J. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale, "Sequence-independent segmentation of magnetic resonance images," *NeuroImage*, vol. 23, pp. S69–S84, 2004.

[12] J. Ashburner and K. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.

[13] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[14] K. Pohl, J. Fisher, W. Grimson, R. Kikinis, and W. Wells, "A Bayesian model for joint segmentation and registration," *NeuroImage*, vol. 31, no. 1, pp. 228–239, 2006.

[15] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A unified framework for atlas based brain image segmentation and registration," in *Biomedical Image Registration*. New York: Springer, 2006, vol. 4057, pp. 136–143.

[16] M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin, "Load: A locally adaptive cortical segmentation algorithm," *NeuroImage*, vol. 56, no. 3, pp. 1386–1397, Jun. 2011.

[17] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "An information theoretic approach for non-rigid image registration using voxel class probabilities," *Med. Image Anal.*, vol. 10, no. 3, pp. 413–431, 2006.

[18] F. van der Lijn, T. den Heijer, M. M. Breteler, and W. J. Niessen, "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts," *NeuroImage*, vol. 43, no. 4, pp. 708–720, Dec. 2008.

[19] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, D. Rueckert, and T. A. D. N. Initiative, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage*, vol. 49, no. 3, pp. 2352–2365, Feb. 2010.

[20] Statistical parameter mapping [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm

[21] A. Guimond, J. Meunier, and J.-P. Thirion, "Average brain models: A convergence study," *Comput. Vis. Image Understand.*, vol. 77, no. 2, pp. 192–210, 2000.

[22] D. E. Rex, J. Q. Ma, and A. W. Toga, "The LONI pipeline processing environment," *NeuroImage*, vol. 19, no. 3, pp. 1033–1048, 2003.

[23] K. Bhatia, J. Hajnal, B. Puri, A. Edwards, and D. Rueckert, "Consistent groupwise non-rigid registration for atlas construction," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2004, vol. 1, pp. 908–911.

[24] B. Avants and J. C. Gee, "Geodesic estimation for large deformation anatomical shape averaging and interpolation," *NeuroImage*, vol. 23, pp. 139–150, 2004.

[25] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. 23, no. 1, pp. 151–160, 2004.

[26] Q. Wang, E. D'Agostino, D. Seghers, F. Maes, D. Vandermeulen, P. Suetens, and A. Hammers, "Construction and validation of mean shape atlas templates for atlas-based brain image segmentation," in *Inf. Proc. in Med. Imag.*. New York: Springer, 2005, vol. 3565, LNCS, pp. 689–700.

[27] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi, "Multi-modal image set registration and atlas formation," *Med. Image Anal.*, vol. 10, no. 3, pp. 440–451, Jun. 2006.

[28] K. Van Leemput, "Encoding probabilistic brain atlases using Bayesian inference," *IEEE Trans. Med. Imag.*, vol. 28, no. 6, pp. 822–837, Jun. 2009.

[29] L. Zöllei, M. Shenton, W. Wells, and K. Pohl, "The impact of atlas formation methods on atlas-guided brain segmentation," in *Proc. MICCAI*, 2007, vol. 1, pp. 39–46.

[30] T. Yeo, M. R. Sabuncu, R. Desikan, B. Fischl, and P. Golland, "Effects of registration regularization and atlas sharpness on segmentation accuracy," *Med. Image Anal.*, vol. 12, no. 5, pp. 603–615, Oct. 2008.

[31] P. Thompson, R. Woods, M. Mega, and A. Toga, "Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain," *Human Brain Mapp.*, vol. 9, no. 2, pp. 81–92, Feb. 2000.

[32] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins, The Brain Development Cooperative Group, "Unbiased average age-appopriate atlases for pediatric studies," *NeuroImage*, vol. 49, no. 2, pp. 1601–1611, Jan. 2010.

[33] C. Studholme and V. Cardenas, "A template free approach to volumetric spatial normalization of brain anatomy," *Pattern Recognit. Lett.*, vol. 25, no. 10, pp. 1191–1202, Jul. 2004.

[34] L. Zöllei, E. Learned-Miller, E. Grimson, and W. Wells, "Efficient population registration of 3-D data," in *Int. Conf. Computer Vision*. New York: Springer, 2005, vol. 3765, LNCS, pp. 291–301.

[35] G. Wu, P.-T. Yap, Q. Wang, and D. Shen, "Groupwise registration from exemplar to group mean: Extending HAMMER to groupwise registration," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 396–399.

[36] T. Rohlfing, R. Brandtb, R. Menzelb, and C. Maurer, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.

[37] M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein, "Optimum template selection for atlas-based segmentation," *NeuroImage*, vol. 34, no. 4, pp. 1612–1618, Feb. 2007.

[38] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, Jul. 2009.

[39] X. Artaechevarria, A. Munoz-Barrutia, and C. O. de Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1266–1277, Aug. 2009.

[40] E. M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. A. Viergever, J. P. Pluim, and B. van Ginneken, "Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus," *Med. Image Anal.*, vol. 14, no. 1, pp. 39–49, Feb. 2010.

[41] M. R. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Trans. Med. Imag.*, vol. 29, no. 10, pp. 1714–1729, Oct. 2010.

[42] V. S. Petrovic, T. F. Cootes, C. J. Twining, and C. Taylor, "Automatic framework for medical image registration, segmentation and modeling," *Proc. Med. Image Understand. Anal.*, vol. 2, pp. 141–145, 2006.

[43] K. K. Bhatia, P. Aljabar, J. P. Boardman, L. Srinivasan, M. Murgasova, S. J. Counsell, M. A. Rutherford, J. Hajnal, A. D. Edwards, and D. Rueckert, "Groupwise combined segmentation and registration for atlas construction," in *MICCAI 2007*. New York: Springer, 2007, vol. 10, LNCS, pp. 532–540.

[44] T. Riklin-Raviv, K. Van Leemput, B. H. Menze, W. M. Wells, and P. Golland, "Segmentation of image ensembles via latent atlases," *Med. Image Anal.*, vol. 14, no. 5, pp. 654–665, Oct. 2010.

[45] J. Ashburner and K. J. Friston, "Voxel-based morphometry: The methods," *NeuroImage*, vol. 11, pp. 805–821, Jun. 2000.

[46] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C. S. Carter, V. A. Stenger, S. Davis, H. Aizenstein, J. Becker, O. Lopez, and C. Meltzer, "Discriminative MR image feature analysis for automatic schizophrenia and Alzheimer's disease classification," in *MICCAI*. New York: Springer, 2004, vol. 3216, LNCS, pp. 393–401.

[47] Y. Fan, D. Shen, and C. Davatzikos, "Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM," in *MICCAI*. New York: Springer, 2005, LNCS, pp. 1–8.

[48] S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G. B. Frisoni, and D. L. Collins, "MRI-based automated computer classification of probable AD versus normal controls," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 509–520, Apr. 2008.

[49] K. M. Pohl and M. R. Sabuncu, "A unified framework for MR based disease classification," in *Inf. Proc. in Med. Imag*. New York: Springer, 2009, vol. 5636, LNCS, pp. 300–313.

[50] D. Blezek and J. MIller, "Atlas stratification," *Med. Image Anal.*, vol. 11, no. 5, pp. 443–457, Oct. 2007.

[51] S. Allassonniere, Y. Amit, and A. Trouvé, "Towards a coherent statistical framework for dense deformable template estimation," *J. R. Stat. Soc., B, Stat. Methodol.*, vol. 69, pp. 3–29, 2007.

[52] M. R. Sabuncu, S. K. Balci, M. E. Shenton, and P. Golland, "Image-driven population analysis through mixture modeling," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1473–1487, Sep. 2009.

[53] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. s61–s72, Mar. 2009.

[54] S. Basu, M. Bilenko, A. Banerjee, and R. Mooney, *Semi-Supervised Learning, Chapter Probabilistic Semi-Supervised Clustering With Constraints*, O. Chapelle, B. Schölkopf, and A. Zien, Eds. Cambridge, MA: MIT Press, 2006.

[55] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2570–2583, Oct. 1992.

[56] D. Langan, K. Molnar, J. Modestino, and J. Zhang, "Use of the mean-field approximation in an EM-based approach to unsupervised stochastic model-based image segmentation," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1992, vol. 3, pp. 57–60.

[57] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.

[58] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[59] L. Collins, "3-D model-based segmentation of individual brain structures from magnetic resonance imaging data," Ph.D. dissertation, McGill Univ., Montreal, Canada, Oct. 1994.

[60] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapp.*, vol. 17, pp. 143–155, 2002.

[61] B. Aubert-Broche, M. Griffin, G. Pike, A. Evans, and D. Collins, "Twenty new digital brain phantoms for creation of validation image data bases," *IEEE Trans. Med. Imag.* vol. 25, no. 11, pp. 1410–1416, Nov. 2006.

[62] W. Vandenberghe, P. Demaerel, R. Dom, and F. Maes, "Diffusion-weighted versus volumetric imaging of the striatum in early symptomatic Huntington disease," *J. Neurol.*, vol. 256, no. 1, pp. 109–114, Jan. 2009.

[63] C. J. Jack, M. Bernstein, and N. F. *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, Apr. 2008.

[64] S. Roy, A. Carass, P.-L. Bazin, and J. L. Prince, "A rician mixture model classification algorithm for magnetic resonance images," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 406–409.

[65] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory*, 1973, pp. 267–281.

[66] M. A. Figueiredo and A. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–395, Mar. 2002.