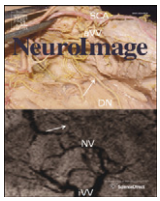
NeuroImage 56 (2011) 1134–1144

Contents lists available at ScienceDirect

NeuroImage

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



Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease

Juha Koikkalainen [a,*], Jyrki Lötjönen [a], Lennart Thurfjell [b], Daniel Rueckert [c], Gunhild Waldemar [d], Hilkka Soininen [e] and the Alzheimer's Disease Neuroimaging Initiative [1]

[a] VTT Technical Research Centre of Finland, Tampere, Finland
[b] GE Healthcare, Medical Diagnostics R&D, Uppsala, Sweden
[c] Imperial College London, London, UK
[d] Department of Neurology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[e] Department of Neurology, University of Eastern Finland, Kuopio Finland

ARTICLE INFO

Article history:
Received 4 October 2010
Revised 4 March 2011
Accepted 10 March 2011
Available online 16 March 2011

Keywords:
Tensor-based morphometry
Multi-template
Multi-atlas
Data classification
Alzheimer's disease

ABSTRACT

In this paper methods for using multiple templates in tensor-based morphometry (TBM) are presented and compared to the conventional single-template approach. TBM analysis requires non-rigid registrations which are often subject to registration errors. When using multiple templates and, therefore, multiple registrations, it can be assumed that the registration errors are averaged and eventually compensated. Four different methods are proposed for multi-template TBM. The methods were evaluated using magnetic resonance (MR) images of healthy controls, patients with stable or progressive mild cognitive impairment (MCI), and patients with Alzheimer's disease (AD) from the ADNI database ($N = 772$). The performance of TBM features in classifying images was evaluated both quantitatively and qualitatively. Classification results show that the multi-template methods are statistically significantly better than the single-template method. The overall classification accuracy was 86.0% for the classification of control and AD subjects, and 72.1% for the classification of stable and progressive MCI subjects. The statistical group-level difference maps produced using multi-template TBM were smoother, formed larger continuous regions, and had larger t-values than the maps obtained with single-template TBM.

Introduction

Morphometric techniques are widely utilized in computational neuroanatomy to study differences in the anatomy of the brain across populations, for example, to identify the effects of disease or changes due to aging. In addition, morphometry can be used in decision support to characterize and diagnose a single patient.

Various morphometric methods exist. Voxel-based morphometry (VBM) is a commonly known technique where the density or concentration of gray-matter is measured locally after accounting for global differences in anatomy (Ashburner and Friston, 2000). A high-resolution voxel-based morphometry method based on RAVENS maps and HAMMER elastic registration was proposed in (Shen and Davatzikos, 2003). An alternative approach is to characterize differences in brain shape using deformation- or tensor-based morphometry (DBM and TBM). In DBM and TBM, images are registered to a common reference space, and the analysis is done by comparing the parameters of resulting deformation fields or measures derived from them (Ashburner et al., 1998; Chung et al., 2001). In TBM, the most often used measure is the determinant of the Jacobian matrix of the deformation field, often referred to as *Jacobian*, which measures local volume change. Numerous studies have been published in which morphometry methods have been used, for example, to study Alzheimer's disease (Teipel et al., 2007; Hua et al., 2008a,b).

The conventional way to perform morphometry analysis is to use one template image to which all study images are registered. Usually this template is an image of a single subject (Chung et al., 2001; Leporé et al., 2008a), a generally available average template, or an average template specifically generated for the particular application and data (Teipel et al., 2007; Leporé et al., 2007; Hua et al., 2008a,b). As registrations between images are never perfect, the use of different templates leads to different results. The characteristics of a template may cause either false-positive or false-negative findings in the resulting parameter maps.

The bias caused by imperfect registration is a commonly known problem in atlas-based segmentation. Multi-atlas segmentation has

* Corresponding author at: VTT, P.O. Box 1300, FIN-33101 Tampere, Finland. Fax: +358 20 722 3499.
E-mail address: juha.koikkalainen@vtt.fi (J. Koikkalainen).

[1] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI Manuscript Citations.pdf).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.03.029

been presented as a solution to this problem: several atlases are registered to the patient data and the propagated segmentations are combined by techniques, such as, classifier fusion. It has been reported that segmentation accuracy (Heckemann et al., 2006; Klein et al., 2005; Rohlfing et al., 2004; Warfield et al., 2004; Lötjönen et al., 2010) and the classifications derived from the segmentations (Aljabar et al., 2008; Chou et al., 2008) can be improved significantly by using multi-atlas segmentation instead of single-atlas segmentation.

Recently, multi-atlas approach has been adopted in TBM analysis (Leporé et al., 2008b). Each study image was registered to nine template images, and the resulting deformation tensors were averaged in a reference space of a tenth template image to improve registration accuracy and increase the statistical power of the analysis. In (Brun et al., 2009), the average Jacobians were computed and used to study morphometry differences between twins.

Whereas the two previous multi-template TBM studies (Leporé et al., 2008b; Brun et al., 2009) have addressed only group-level analysis, this paper uses multi-template TBM methods for subject-level analysis. To our knowledge, this is the first study where a multi-template TBM approach is 1) used to classify subjects, 2) extensively compared to single-template TBM, and 3) applied to a large dataset. In addition, we present new methods to utilize multiple templates in TBM. The methods are evaluated quantitatively by extracting features from the TBM analysis and using them to classify the data, and by computing sample size estimates. Visual evaluation is performed by examining the statistical group-level differences. We use a large number of subjects ($N = 772$) and also a large number of templates ($N = 30$). The objectives of this paper are to study how the multiple templates should be utilized in the TBM analysis, what is the optimal way to compute classification features from the TBM analysis, and whether the use of multiple templates leads to more accurate and robust information for classification as compared to the single-template TBM analysis.

The methods proposed are applied to the diagnostics of Alzheimer's disease (AD). AD is a neurodegenerative disease that causes atrophy in the cerebral cortex and subcortical structures, such as the hippocampus and amygdala. In a recent paper, the International Working Group for New Research Criteria for the Diagnosis of AD (Dubois et al., 2010) propose the use of biomarkers as supporting evidence for diagnosis of AD. Medical temporal atrophy measured with structural MRI is one of the proposed biomarkers. Mild cognitive impairment (MCI) is a condition in which a patient has noticeable problems with memory, language, or other mental functions but activities of daily living are preserved (Petersen, 2004). It is a risk factor for AD, but not every MCI patient develops AD. In this paper, those subjects who finally develop AD are referred to as progressive MCI (P-MCI) and those who remain stable are referred to as stable MCI (S-MCI).

The paper is organized as follows: In Materials and methods section, the data used in the study, the methods developed for the TBM, and the evaluation methods are introduced. The Results section summarizes the results obtained. Finally, the methods and results are discussed in Discussion section, and the conclusions are drawn in Conclusions section.

Material and methods

Data

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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 non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. The determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research — approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years.

The study group consisted of T1-weighted 1.5 T MR images of 772 subjects from the ADNI database (Table 1a). Only the baseline images were used. The follow-up data of ADNI database were used to determine the MCI subjects who had converted to AD (P-MCI) and who had remained stable (S-MCI) during the study. For the P-MCI subjects, the mean time and its standard deviation from the baseline to the moment when the dementia threshold was reached was 18.1 ± 8.9 months. Detailed information on the conversions is presented in Table 1a. The scanners used in the ADNI study are from General Electric (GE) Healthcare, Philips Medical Systems, and Siemens Medical Solutions. The images used were sagittal 3D MP-RAGE images with resolutions ranging from 0.9 mm × 0.9 mm × 1.20 mm to 1.3 mm × 1.3 mm × 1.20 mm. The 30 templates were obtained from the ADNI database, too. The template images were 1.5 T T1-weighted MRIs of 10 control subjects, 10 MCI subjects, and 10 AD subjects (Table 1b). The subjects were randomly selected from the ADNI database for each group.

Skull-stripping and intensity inhomogeneity correction was performed on both study and template images using an unpublished in-house tool. To guarantee successful skull-stripping, the results were manually checked, and the failed results were run with new parameters until all the results were acceptable.

Methods

The conventional singe-template and the multi-template TBM approaches used in this study are summarized in Fig. 1. The single-template (ST) TBM consists of the following steps:

1. *Registrations*: A template image is non-rigidly registered to each study image. Registrations produce for each voxel of the template image the deformations recquired to map it to the corresponding anatomical

Table 1
Demographic data for a) the study subjects (and the number of AD conversion in each follow-up visit) and b) the template subjects.

a

Study subjects	n	Age	% females	P-MCI conversions	n
Controls	215	76 ± 5	48	6 months	22
S-MCI	215	75 ± 8	33	12 months	45
P-MCI	154	75 ± 7	40	18 months	33
AD	188	75 ± 7	48	24 months	35
				30 months	0
				36 months	19

b

Template subjects	n	Age	% females
Controls	10	80 ± 5	50
MCI	10	74 ± 10	40
AD	10	77 ± 6	50

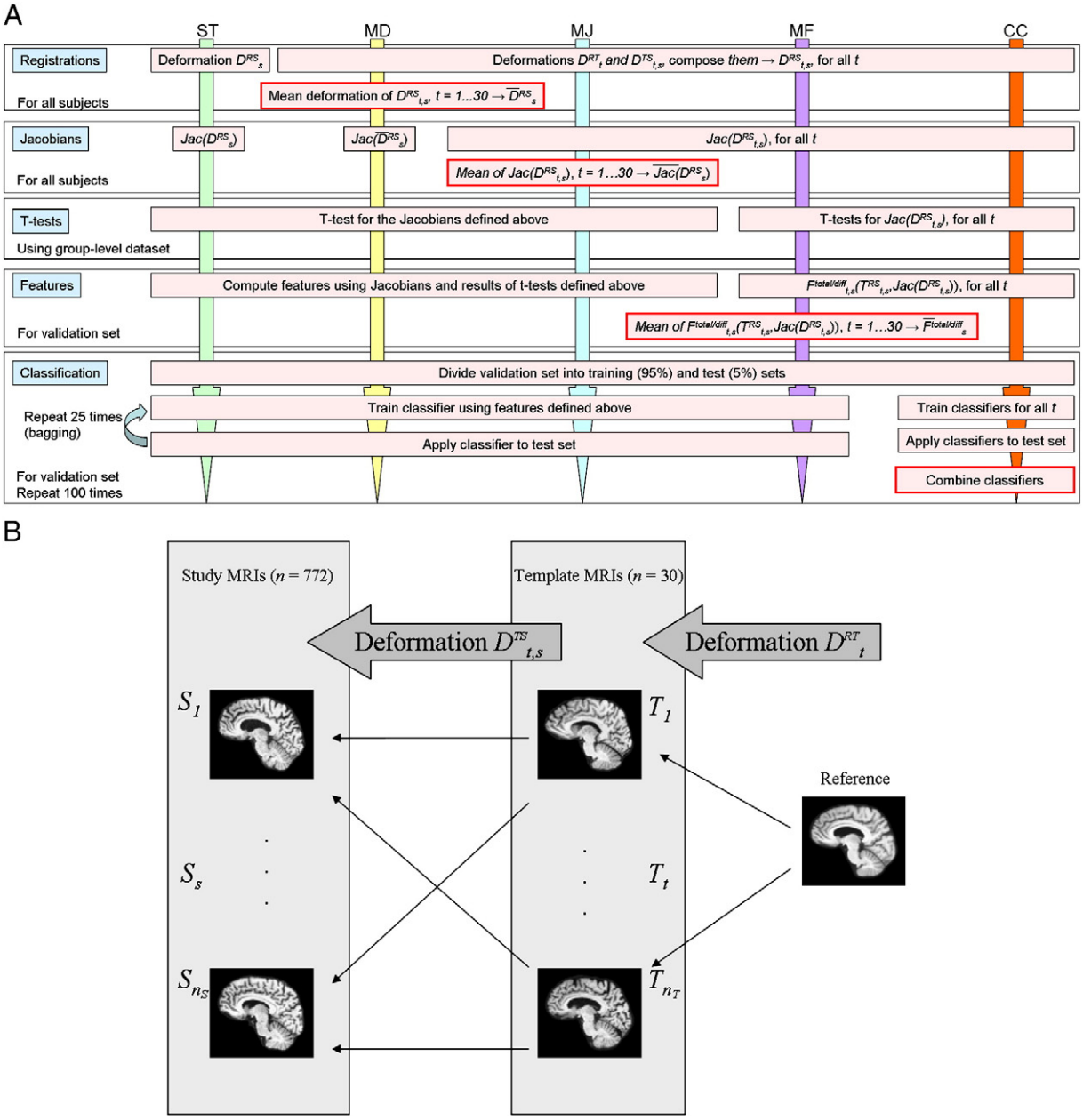


Fig. 1. A) The classification procedure of the TBM methods studied. The boxes with the red borders show the locations where the multiple templates are combined. B) The flowchart of the registration procedure.

location in each study image. The study images are typically collected from two study groups whose differences are studied.

2. *Jacobians*: A scalar value quantifying the deformation is computed for each voxel of each subject. The determinant of the Jacobian matrix was used in this study.
3. *T-tests*: The group-level differences of two study groups are computed using a voxel-wise statistical test. A *t*-test was used in this study.
4. *Features*: If region of interest (ROI) based analysis is performed, the scalar values are combined providing a single feature value for each ROI.
5. *Classification*: Data classification is performed using the feature values.

In multi-template TBM, an extra step is needed to combine the information from multiple templates. The combination can be accomplished at different stages of the single-template procedure (highlighted with red borders in Fig. 1A):

1. *Mean deformation (MD)*: combination of the deformations (after Registrations step),
2. *Mean of Jacobians (MJ)*: combination of the Jacobians (after Jacobians step),
3. *Mean of features (MF)*: combination of the feature values (after t-tests and Features steps), and
4. *Combination of classifiers (CC)*: combination of the classification results (after Classification step).

The TBM steps defined in Fig. 1A are described in detail in the Registrations up to the Classification section, and after that the different multi-template methods are presented in Multi-template TBM methods section.

Registrations

In the standard TBM, a template image is registered with all the study images. Each of these non-rigid registrations contains errors that decrease the accuracy of the TBM analysis. The idea of multi-template TBM is to reduce these errors by averaging the result over several templates. In multi-template TBM, all templates are registered with all the study images (Fig. 1B). As information is fused from the registrations of multiple templates, a common reference space is needed to spatially normalize the results from multiple templates, to normalize the values of the morphometric analysis, and to establish a reference space in which the results are presented.

The reference image used in this study is a mean anatomical template (MAT) generated from all the template images as proposed by Guimond et al. (2000):

1. one template image (floating image) was registered to each template image,
2. the mean deformation was computed,
3. all the template images were deformed to the floating image using the inverse transformations,
4. the mean intensity was computed for each voxel from the deformed templates, and
5. the mean deformation was applied to the mean intensity image.

The mean anatomical template used in this study is shown in Fig. 2.

For the conventional single-template TBM using MAT as the reference image, the MAT was registered to each study image. The resulting deformations are denoted as D_s^{RS}, $s=1,\ldots,n_S$, where n_S is the number of study subjects. The registration procedure of the multi-template TBM contained two steps shown in Fig. 1B: First, all the template images, T_t, $t=1,\ldots,n_T$, where n_T is the number of templates, were non-rigidly registered to each study image, S_s, $s=1,\ldots,n_S$ (deformations $D_{t,s}^{TS}$). Second, the reference image (MAT) was non-rigidly registered to each template image (deformations D_t^{RT}). In single-template TBM, the morphometic information of the study subjects are contained in the deformations D_s^{RS}. In the multi-template methods, this information exists in the deformations $D_{t,s}^{TS}$. To make the TBM results of all the templates comparable, this information has to be normalized to a common reference space. In practice, this was obtained by composing the deformations D_t^{RT} and $D_{t,s}^{TS}$ for each pair (t,s), $t=1,\ldots,n_T$, $s=1,\ldots,n_S$. This resulted in the deformation $D_{t,s}^{RS}$ of the MAT to the study image S_s computed via the template image T_t. Consequently, it turned out that in multi-template TBM the reference image is registered n_T times to each study image, and each time the registration is done via a different template image.

In this study, we are not interested in the differences in the global brain size. Therefore, global differences in the pose and scale were removed from the data by registering the study images and the template images to the reference space using 9-parameter affine transformation. Consequently, all the images in Fig. 1B were in the same space and no further affine registrations were required. The registration methods used are described in detail in (Lötjönen et al., 2010).

Jacobians

The determinant of the Jacobian matrix of the deformations, the Jacobian, was selected as the measure of local morphometry in this study. The Jacobian of the deformation $D(\vec{p})$, where $\vec{p}=(x,y,z)$ gives the coordinates of the voxel studied, is computed as:

$$J(\vec{p}) = Jac\left(D(\vec{p})\right) = \begin{pmatrix} \frac{\partial D^x(\vec{p})}{\partial x} & \frac{\partial D^x(\vec{p})}{\partial y} & \frac{\partial D^x(\vec{p})}{\partial z} \\ \frac{\partial D^y(\vec{p})}{\partial x} & \frac{\partial D^y(\vec{p})}{\partial y} & \frac{\partial D^y(\vec{p})}{\partial z} \\ \frac{\partial D^z(\vec{p})}{\partial x} & \frac{\partial D^z(\vec{p})}{\partial y} & \frac{\partial D^z(\vec{p})}{\partial z} \end{pmatrix}, \quad (1)$$

where $D^{x/y/z}$ denotes the deformation in $x/y/z$-direction. Flowchart in Fig. 1A shows the actual deformations for which the Jacobians are computed using the different TBM methods.

T-tests

TBM analysis is often used to search for statistically significant morphometric differences between two study groups. The Jacobians of the subjects of two study groups were compared voxel by voxel using a t-test. In order to make the distribution of data more Gaussian, a logarithmic transformation was applied to the Jacobians prior the t-test. False discovery rate (FDR) correction was performed to compensate for multiple comparisons (Genovese et al., 2002).

In our classification study, the results of the t-tests were used to provide prior information on the group-level differences between study groups. We used the results without the FDR-correction in the classification studies. The flowchart in Fig. 1A shows the data for which the t-tests were performed in different TBM methods.

Features

In addition to the group-wise analysis of size and shape differences, morphometic information can be used to classify unseen images into one of the study groups (Vemuri et al., 2008; Fan et al., 2008; Misra et al., 2009; Klöppel et al., 2008). For this purpose, a set of features was computed from the Jacobians both globally and for 83 structures obtained from an atlas defined in the reference space (Heckemann et al., 2006).

A commonly used method to analyze the morphometric properties of structures is to compute their volumes. As the Jacobian measures local volume change, a feature related to the total volume of a structure is obtained by averaging the Jacobians within a structure:

$$F_s^{total}(R) = \frac{\sum_{\vec{p}\in R} \log J_s^c(\vec{p})}{\sum_{\vec{p}\in R} 1}, \quad (2)$$

where J_s^c is the Jacobian value (defined below) of subject s and R denotes the structure studied.

However, a structure may contain regions that dilate and those that shrink in AD. If the average is computed over the whole structure, these regions may cancel each other and the feature obtained does not provide good classifications. Therefore, we computed the mean of Jacobians from the dilating voxels only and from the shrinking voxels only, and used the difference of the two mean values as a classification feature:

$$F_s^{diff}(R) = \frac{\sum_{\vec{p}\in R\cap T(\vec{p})>0} W(\vec{p}) \cdot \log J_s^c(\vec{p})}{\sum_{\vec{p}\in R\cap T(\vec{p})>0} W(\vec{p})} - \frac{\sum_{\vec{p}\in R\cap T(\vec{p})<0} W(\vec{p}) \cdot \log J_s^c(\vec{p})}{\sum_{\vec{p}\in R\cap T(\vec{p})<0} W(\vec{p})}, \quad (3)$$

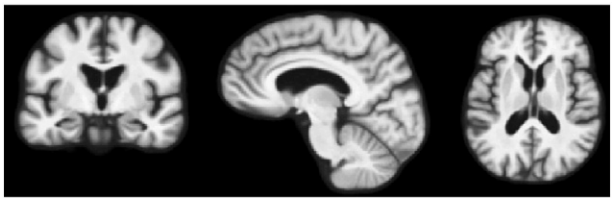


Fig. 2. Three orthogonal slices from the mean anatomical template (MAT) used as the reference image.

where $T\left(\vec{p}\right)$ is the t-value from a group-level t-test and $W\left(\vec{p}\right)$ a weighting function defined as

$$W\left(\vec{p}\right) = \frac{\log(p_{max}) - \log\left(P\left(\vec{p}\right)\right)}{\log(p_{max}) - \log(p_{min})}, \quad (4)$$

where $P\left(\vec{p}\right)$ is the p-value, $p_{max} = 0.05$ and $p_{min} = 0.000001$ are user-defined parameters, and the p-values are constrained to the interval defined by p_{min} and p_{max}. The weighting $W\left(\vec{p}\right)$ was used to focus the computations on the voxels that have statistically significant group-level differences. It gives a voxel a larger weight the smaller the p-value $P\left(\vec{p}\right)$ is, and no weight is given to the statistically non-significant ($P\left(\vec{p}\right) > 0.05$) voxels. The threshold p_{min} was used to avoid situations where just a few highly significant voxels would have a too large impact on a feature value. Other types of weightings based on the t- or p-values were studied, but no major differences were found.

Flowchart in Fig. 1A summarizes the data for which the feature values were computed in each TBM method.

Classification

A regression-based classifier was used in all classification studies. A label −1 was given to the subjects of one study group and label 1 to the subjects of the other study group. Then, linear regression model parameters were optimized using a training set, and the parameters were applied to test set data. A test set subject was classified to the first group if the regression value obtained was negative and otherwise to the second group. Finally, the classification performance was computed from the results of the test set.

We used a bagging strategy to improve the robustness of the classifiers (Breiman, 1996; Bauer and Kohavi, 1999): the training set was sampled 25 times randomly and each time a new classifier was trained and applied to the test set. Then, the classification results were combined by computing the mean of the regression values, and the class of a test set subject was inferred. The sampling of the training set was performed so that an equal number of samples (the size of the larger study group) were chosen from both study groups. Sampling was done with replacement, so the same subject could appear multiple times in a training set.

All the classification studies were implemented using Matlab (Matlab R2007b, The MathWorks Inc.).

Multi-template TBM methods

In the *single-template (ST) TBM*, the MAT was used as the template image and the analysis was performed using the Jacobians of the deformations D_s^{RS}, $J_s^c\left(\vec{p}\right) = Jac\left(D_s^{RS}\left(\vec{p}\right)\right)$. Group-level statistical analysis was performed by applying the t-test to the Jacobians of two groups studied.

The first multi-template method used *mean deformations* (MD). The hypothesis behind the multi-template approach was that the registration errors are compensated by averaging a set of registrations. Therefore, a mean deformation was computed as:

$$\overline{D}_s^{RS}\left(\vec{p}\right) = \frac{1}{n_T}\sum_{t=1}^{n_T} D_{t,s}^{RS}\left(\vec{p}\right). \quad (5)$$

We decided to use the Euclidean space instead of a log-Euclidean framework (Arsigny et al., 2006) to reduce the computational burden as much as possible. The Jacobians $J_s^c\left(\vec{p}\right)$ were then computed from the resulting mean deformation,

$$J_s^c\left(\vec{p}\right) = Jac\left(\overline{D}_s^{RS}\left(\vec{p}\right)\right), \quad (6)$$

and used both to classify data and to study group-level differences.

In the second method, the *mean of Jacobians* (MJ) were computed for each voxel. In this case, the Jacobians used in classification were computed as follows:

$$J_s^c\left(\vec{p}\right) = \left(\prod_t^{n_T} J_{t,s}\left(\vec{p}\right)\right)^{1/n_T}. \quad (7)$$

These Jacobians were used to study group-level differences as well. The method is equivalent to the method used by Brun et al. (2009).

In the third multi-template method, the feature values were separately computed for each template, $F_{s,t}^{total/diff}(R)$, and then the *mean feature values* (MF) were computed as:

$$\overline{F}_s^{total\,/\,diff}(R) = \frac{1}{n_T}\sum_{t=1}^{nT} F_{s,t}^{total\,/\,diff}(R). \quad (8)$$

The last method based on multiple templates used each template individually in classification, and then combined the classification results. The *combination of classifiers* (CC) was performed by averaging the regression values of all the templates, and the subject was classified based on the mean regression value.

The last two methods could not be used to qualitatively study group-level differences and were used only in classification.

Evaluation

Evaluation was performed by comparing control subjects with Alzheimer's disease subjects (controls vs. AD comparison) and stable MCI subjects with progressive MCI subjects (S-MCI vs. P-MCI comparison). The comparison of S-MCI vs. P-MCI is especially important considering the need for having a method for the detection of which subjects with mild memory problems will progress to AD.

The overall evaluation procedure is shown in Fig. 1A. For the evaluation, the dataset was divided randomly into two sets: 100 subjects were selected from each group as a group-level dataset which was used to establish group-level statistical differences needed for the qualitative evaluation and for the computation of the feature values in Eqs. (3)–(4), and the remaining subjects (115 controls, 115 S-MCIs, 54 P-MCIs, and 88 ADs) established the validation set that was used to evaluate the classification performance. The 30 template subjects were used in the group-level dataset so that they were not used to evaluate the classification accuracies of the methods.

For the visual analysis of group-level differences, the whole brain t-maps were visualized using color overlays in three orthogonal directions. The same color scales were used for each TBM method (controls vs. AD had different scale than S-MCI vs. P-MCI) so that the absolute values and statistical power of each method could be compared.

Evaluation of classification accuracies was performed using a cross-validation technique for the validation set as follows: we randomly chose 5% of the subjects that were excluded from the validation set and were classified using the remaining subjects as a training set. This was repeated 100 times.

The first objective was to study how different multi-template methods compare to the single-template method when each structure is used one by one in classification (one-dimensional classifier). The comparison was performed statistically by comparing pair-wise the classification accuracies of two methods using the non-parametric Wilcoxon signed rank test. If there were no statistically significant voxels within a structure, the feature F_s^{diff} could not be computed. Consequently, the number of structures for which the feature values exists varied between TBM methods. To make the comparison fair and simple, only those structures that were available from each method were used in this comparison.

Next, the overall classification accuracy that can be reached using all the information available was evaluated. We first tried multi-dimensional regression analysis combined with stepwise feature

selection for all the 84 structures. However, we noticed that the size of the dataset was too small for such a high-dimensional classification problem, and the classification accuracies did not improve from the results of the best single structures. Therefore, we decided to use the same method as in the CC method for the combination of templates, i.e., the average regression value computed over a set of ROIs was used as the classification feature. If all the ROIs were used, the results of the worst ROIs would have decreased the classification accuracy. Therefore, we used the group-level dataset to determine an optimal set of ROIs: 1) the regression values of the n ROIs producing the best single-structure classification accuracies were averaged and the classification accuracy was computed using the average regression value as the classification feature, 2) the value n was varied between one to 84, and 3) the n giving the optimal classification accuracy was searched. The same classifiers and cross-validation methods were used for the group-level dataset as for the validation set. However, as the group-level dataset was used to compute the t-tests for the computation of feature F_s^{diff} the optimal sets of ROIs was biased and could not be applied to the validation set. Therefore, the optimal set of ROIs defined for the feature F_s^{total} was used also for the feature F_s^{diff}.

We also computed the sample size that is required to hypothetically detect a change in feature values. The sample size was estimated from

$$n = 2\frac{\sigma^2\left(z_{1-\alpha/2} + z_{1-\beta}\right)}{\delta^2}, \quad (9)$$

where σ is the standard deviation of the data, δ is the deviation to be detected, α is the significance level (here $\alpha = 0.05$), $1-\beta$ is the power (here $1-\beta = 0.8$), and z is the standard normal probability distribution.

Results

Visual evaluation of t-maps

Fig. 3 shows the t-maps of statistically significant regions for both controls vs. AD and S-MCI vs. P-MCI comparisons. From the TBM methods studied such voxel-wise analysis can be performed only for the single-template (ST) TBM, mean deformation (MD), and mean of Jacobians (MJ) methods. In Fig. 3, red is used to show the regions of smaller values of the Jacobians (atrophy) in the AD/P-MCI group and blue is used to show the regions with larger Jacobians (dilation) in the AD/P-MCI group. FDR-corrected results are shown only for the controls vs. AD comparison because in the S-MCI vs. P-MCI comparison only a few voxels survived from the FDR-correction.

The results show increased atrophy in temporal and parietal lobes in both controls vs. AD and S-MCI vs. P-MCI comparisons. Multi-template methods produce larger t-values and notable less noisier maps with larger continuous regions of significant morphometry differences than the single-template method. The two multi-template methods give nearly identical results.

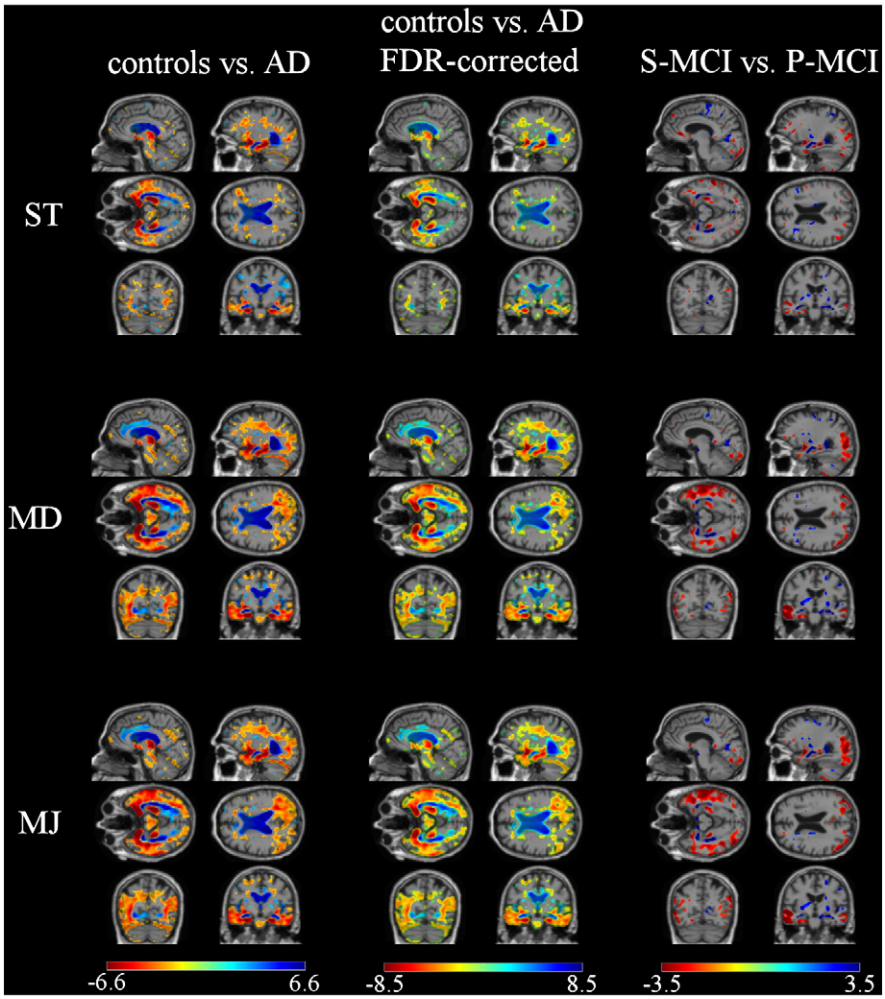


Fig. 3. T-maps of group-level dataset for three methods both for controls vs. AD and S-MCI vs. P-MCI comparisons. Only the voxels with statistically significant differences ($p<0.05$) are shown.

Data classification

The results for single-structure classifications are shown in Figs. 4 and 5. The results are presented for each TBM method and structure, and for each feature. The figures show the results both structure by structure and in an ordered way in which the results of each TBM method are ordered separately. The structures used are listed in Appendix. The best classification accuracies are 82.0% for controls vs. AD comparison and 68.4% for S-MCI vs. P-MCI comparison. The ordered graphs show clearly that the multi-template methods yield better classifications than the single-template method, especially when the feature F_s^{diff} utilizing prior group-level statistical information is used. For example, when studying the 30th best ROIs of each TBM method, the classification accuracy of the single-template TBM was the worst one in each graph.

Table 2 reports the p-values obtained by comparing pair-wise the classification accuracies of two TBM methods. When comparing the average of the single-structure classification accuracies in Figs. 4 and 5 (data not shown), the combination of classifiers method gave the best results, and in three out of four cases the single-template method was the worst method. The combination of classifiers was statistically significantly better than the MD and MJ methods for feature F_s^{diff}, but for the feature F_s^{total} the differences were non-significant. Tables 3 and 4 show the classification accuracies for structures producing the best results.

The overall classification accuracies for different TBM methods are shown in Table 5. The best classification accuracies are 86.0% for the controls vs. AD comparison and 72.1% for the S-MCI vs. P-MCI comparison. About 4 − 6% improvement (statistically significant, $p<0.05$) was obtained with the best multi-template method as compared to the single-template method.

Sample size estimates

In the sample size calculation we set δ to be 0.25 times the difference in the mean values of the two study groups and the standard deviation of the AD or P-MCI group was used as the σ. The whole validation set was used without cross-validation. The sample size estimates were not computed for the CC method because the features used in it require a separate training set, and therefore, the way how the validation set is divided into training and test sets affects the sample size estimate. The results for the individual ROIs with the smallest sample sizes are presented in Table 6. The results are consistent with the classification results: the sample size estimates of the multi-template methods are smaller than the ones of the single-template method, and the feature F_s^{diff} yields smaller sample sizes than the feature F_s^{total}.

Discussion

Various aspects of multi-template tensor-based morphometry were studied and compared to conventional single-template TBM. We studied different features for TBM and different methods to implement multi-template approach. The performance of the methods was

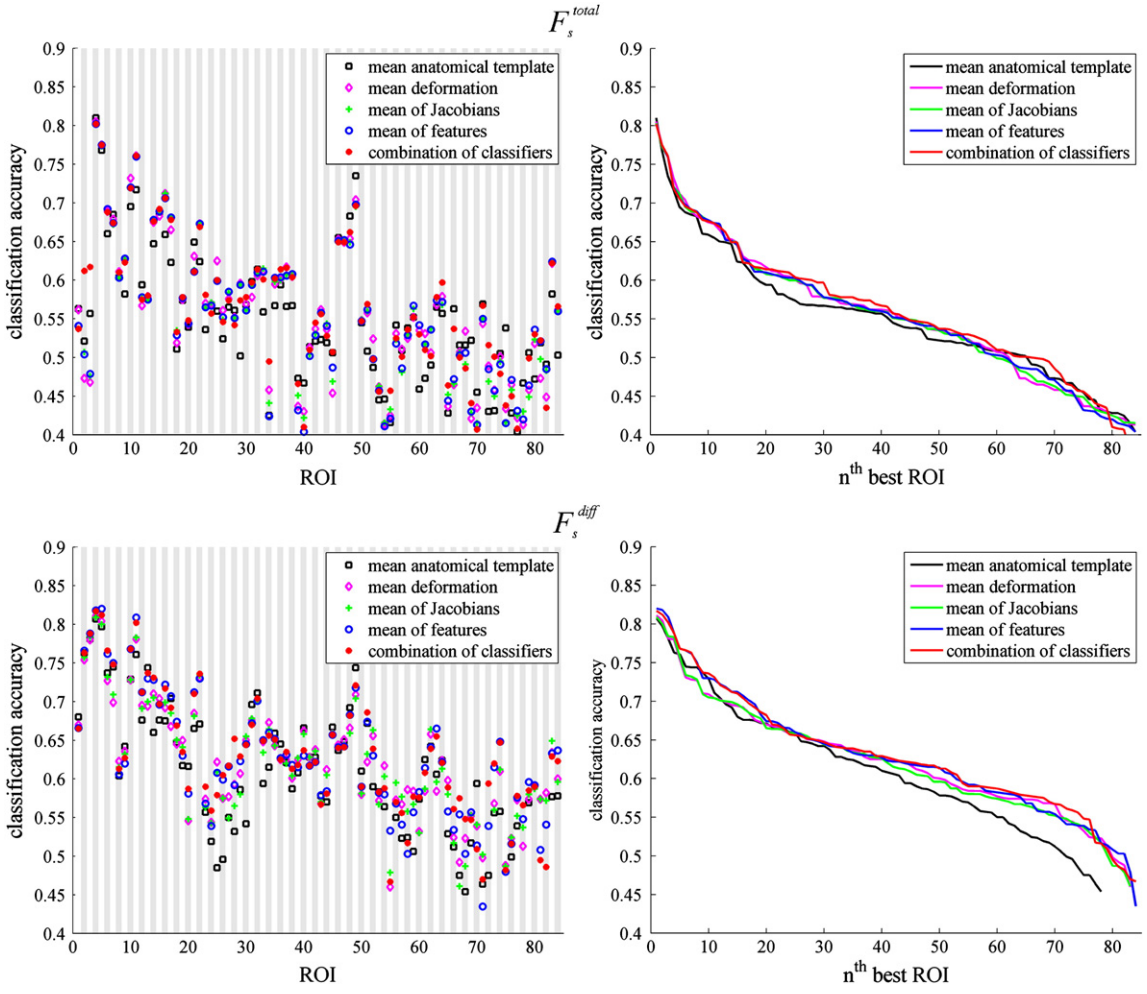


Fig. 4. Controls vs. AD comparison: single-structure classification accuracies for different features and methods. Left: ROIs ordered as listed in the Appendix. Right: ROIs ordered based on the classification accuracy.

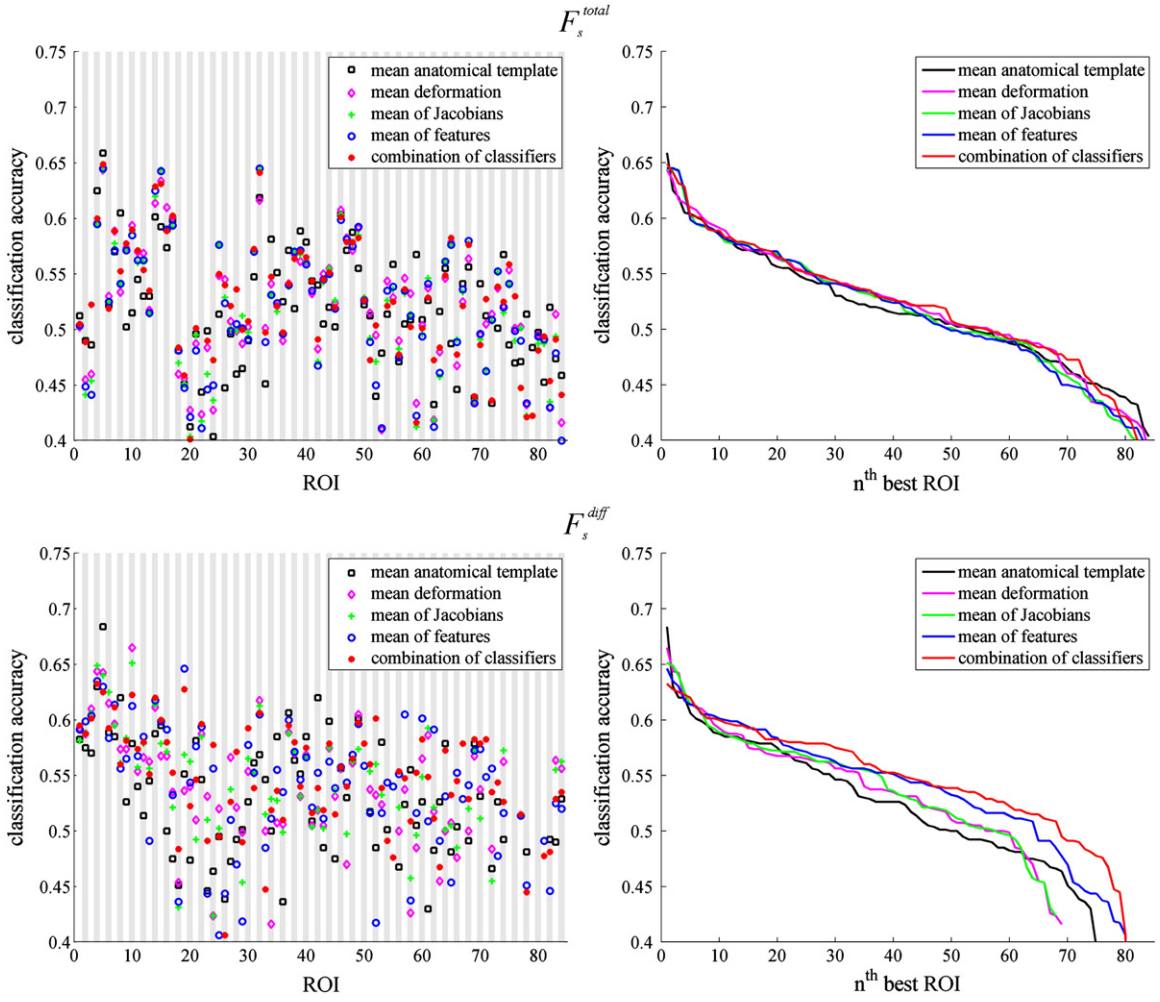


Fig. 5. S-MCI vs. P-MCI comparison: single-structure classification accuracies for different features and methods. Left: ROIs ordered as listed in the Appendix. Right: ROIs ordered based on the classification accuracy.

evaluated in the diagnostics of Alzheimer's disease, i.e., we compared the classification accuracy between healthy controls and AD patients and between stable and progressive MCI cases. In one-dimensional (one structure) classification of controls vs. AD subjects, all multi-template methods were statistically significantly better than the single-template method when the feature F_s^{diff} was used (Table 2). In the classification of S-MCI and P-MCI subjects, the combination of classifiers method was statistically significantly better when the feature F_s^{diff} was used (Table 2). When all the structures were combined in multi-dimensional classification, the accuracy of the controls vs. AD classification was 79.6% for the single-template method and 86.0% for the best multi-template method. The corresponding figures for S-MCI vs. P-MCI classification were 68.3% and 72.1%.

The best multi-template method was the combination of classifiers, i.e., each template was used separately in classification, and the classification results were combined (here using the average of regression values). This method gave for both features studied and in both controls vs. AD and S-MCI vs. P-MCI classifications the best average single-structure classification accuracy, and when the feature F_s^{diff} was used, the differences to the MD and MJ methods were statistically significant (Table 2). This also means that the combination of classifiers method was statistically significantly better than the method previously proposed for the multi-template TBM by Brun et al. (2009).

Table 2
P-values of the pair-wise comparison of different TBM methods.

F_s^{total}	ST	MD	MJ	MF	CC	F_s^{diff}	ST	MD	MJ	MF	CC
Controls vs. AD											
ST	–	0.07	0.1	0.1	0.007	ST	–	0.002	0.02	0.0000	0.0000
MD		–	0.5	0.7	0.6	MD		–	0.04	0.08	0.008
MJ			–	1.0	0.2	MJ			–	0.006	0.001
MF				–	0.2	MF				–	0.2
CC					–	CC					–
S-MCI vs. P-MCI											
ST	–	0.3	0.8	0.9	0.2	ST	–	0.1	0.08	0.06	0.0002
MD		–	0.009	0.01	0.3	MD		–	0.3	0.4	0.008
MJ			–	0.7	0.06	MJ			–	0.7	0.009
MF				–	0.06	MF				–	0.08
CC					–	CC					–

Table 3
Controls vs. AD comparison: classification accuracies (%) for some structures with the combination of classifier method.

	F_s^{total}	F_s^{diff}
Amygdala right	80.2	81.7
Gyri parahippocampalis et ambiens left	76.1	80.2
Hippocampus left	61.7	78.8
Anterior temporal lobe, medial part right	68.8	76.6
Superior temporal gyrus, posterior part left	58.0	73.7
Insula right	66.9	73.6
Medial and inferior temporal gyri right	67.6	73.1
Lateral ventricle, temporal horn left	69.6	72.1

Table 4
S-MCI vs. P-MCI comparison: classification accuracies (%) for some structures with the combination of classifier method.

	F_s^{total}	F_s^{diff}
Amygdala left	64.9	62.5
Posterior temporal lobe right	64.1	60.6
Medial and inferior temporal gyri left	63.1	60.0
Gyri parahippocampalis et ambiens right	59.0	62.3
Anterior temporal lobe, medial part left	58.9	61.1
Hippocampus left	52.3	60.1

The best one-dimensional classification accuracies obtained were 82.0% for the controls vs. AD comparison, and 68.4% for the S-MCI vs. P-MCI comparison. For the multi-dimensional classification, the best results were 86.0% and 72.1%, respectively. The related classification accuracies presented recently in the literature have varied between 76% and 94% for controls vs. AD comparison (Vemuri et al., 2008; Fan et al., 2008; Chupin et al., 2009; Wolz et al., 2010; Teipel et al., 2007; Klöppel et al., 2008) and between 65% and 85% for S-MCI vs. P-MCI comparison (Misra et al., 2009; Chupin et al., 2009; Wolz et al., 2010; Teipel et al., 2007). Our results are in concordance with these prior results. However, the comparison of our results to the results of other studies is difficult, as many studies have been performed using single-site data (Teipel et al., 2007; Klöppel et al., 2008), scanners of a single manufacturer (Vemuri et al., 2008), notable smaller datasets (Teipel et al., 2007; Klöppel et al., 2008; Misra et al., 2009), or shorter follow-up times (Chupin et al., 2009). Also, it has to be noticed that the objective of this study was not to optimize the classification accuracy, but to compare the multi-template methods with the single-template method. Classification accuracy could possibly be improved by extracting more powerful features from the data and by using more sophisticated multi-dimensional classification and feature selection methods.

The best structures for the classification included amygdala, hippocampus and regions in the medial temporal lobe. These are well-known areas of aberrations in AD, and therefore show that the results are meaningful.

In the controls vs. AD comparison, the overall trend was that the feature F_s^{diff} outperformed the feature F_s^{total} (Tables 3 and 5). In other words, the utilization of the results of the group-level statistical analysis improved the classification accuracy. The differences were more obvious in the one-dimensional classification (Table 3) than in the multi-dimensional classification (Table 5). In the S-MCI vs. P-MCI comparison, the differences between the features F_s^{total} and F_s^{diff} were much smaller (Tables 4 and 5). In fact, in the multi-dimensional classification the feature F_s^{total} was clearly better, but this is very likely due to the way how the optimal set of ROIs was defined always from the feature values F_s^{total}. Apparently, the anatomical differences in the MCI groups are so heterogeneous that the statistical analysis with group-level dataset cannot model all those differences accurately. In the controls vs. AD comparison, the morphometry differences are so prominent and localized that the group-level information on statistical differences is robust and can be used for unseen subjects. It must be noticed that different features worked best for different structures. Therefore, it might be useful to use different features for different structures, especially in multi-dimensional classification.

Table 5
Overall classification accuracies (sensitivity/specificity) (%) when using all the structures simultaneously.

	Controls vs. AD		S-MCI vs. P-MCI	
	F_s^{total}	F_s^{diff}	F_s^{total}	F_s^{diff}
ST	78.9 (70/85)	79.6 (74/84)	68.3 (72/67)	66.4 (68/65)
MD	83.3 (78/88)	83.7 (78/88)	67.1 (69/67)	67.5 (64/69)
MJ	83.3 (76/89)	85.5 (79/91)	66.5 (65/68)	69.6 (74/68)
MF	83.4 (76/89)	86.0 (81/91)	71.0 (76/70)	63.0 (64/63)
CC	84.7 (78/90)	85.3 (79/90)	72.1 (77/71)	63.6 (65/63)

Table 6
Sample size estimates for each TBM method (minimum sample size of individual ROIs).

	Controls vs. AD		S-MCI vs. P-MCI	
	F_s^{total}	F_s^{diff}	F_s^{total}	F_s^{diff}
ST	176	143	510	412
MD	155	141	492	406
MJ	156	143	479	410
MF	156	134	479	393

The sample size estimate results (Table 6) support the findings of the classification studies. The results underline the superiority of multi-template methods over the conventional single-template method, and the feature F_s^{diff} outperformed the feature F_s^{total}. In other words, use of prior group-level information increases the statistical power of the features.

The ADNI data have been acquired with different scanners in many sites. The data used here were not divided based on the site or scanner, but only random division to the group-level dataset and validation set. For comparison, we divided the dataset so that the subjects were first ordered based on the imaging site and after that the subjects were divided in the group-level dataset and validation set. In other words, all the subjects from one site were either in the group-level dataset or in the validation set. This mimics the actual clinical situation, where the dataset is continuously enlarged from the images acquired from the particular clinical site. Using this data division, the best classification accuracy for the S-MCI vs. P-MCI classification was 77.1%, i.e., 5% unit improvement as compared to the fully random data division was obtained. Therefore, it can be concluded that there are notable differences in the MCI data between different imaging sites, and the data used in classification should be as uniform as possible. In the controls vs. AD comparison, the differences between the study groups are so large that the inter-site differences do not have noticeable effects on the results.

The methods were evaluated visually using the statistical maps of group-differences (Fig. 3). It was observed that the maps obtained with the multi-template methods were smoother, established larger continuous regions, and had larger t-values. The main application of this study was the classification of subjects, where the group-level statistical maps were utilized to provide prior information. A specific study should be performed to evaluate the usefulness of the statistical group-level difference maps in detecting anatomical differences.

The major drawback of utilizing multiple templates in TBM analysis is the increase in the computation time needed to make the registrations. The increase is linear as a function of the number of templates used. The reference method, single-template TBM using a mean anatomical template, requires only one registration for a target subject, which makes the computations faster and no catenation of registrations is needed. In multi-atlas segmentation, atlas selection methods have been shown to be useful (Aljabar et al., 2009; Lötjönen et al., 2010). In these methods, only a subset of atlases, those that are the most similar to the study image, are registered with the study image. Consequently, the computation time is decreased, and at the same time the accuracy is improved. A similar selection of templates in the TBM analysis using some intelligent heuristics would decrease the computation time notably, and would probably improve the classification results, too. However, the registration method used in this study enables relatively fast computation of TBM analysis even with a large number of templates, as a single non-rigid registration takes under one minute, and with multi-core computers the registrations can be performed in parallel (Lötjönen et al., 2010).

In the multi-dimensional classification, optimal sets of ROIs were defined from the group-level dataset. However, this could be done only for the features that did not utilize the results of group-level

statistical analyses. Consequently, the multi-dimensional classification results of the feature F_s^{diff} would likely get better if a larger dataset that enables efficient multi-dimensional classifications were available.

FDR-correction for multiple comparisons was used when the group-level differences were studied. However, it was not used in the classification application because it would have limited the analysis on only the most significant ROIs in the temporal lobe and it would have made the analysis of the MCI subjects impossible. In the future studies, it should be studied if the utilization of FDR-correction improves the classification accuracy of the features computed from the temporal lobe.

This study is part of an EU funded research project PredictAD (www.predictad.eu) aiming at developing a standardized and objective solution that would enable an earlier diagnosis of Alzheimer's disease, improved monitoring of treatment efficacy and enhanced cost-effectiveness of diagnostic protocols. In order to be clinically feasible, analysis methods have to be computationally efficient. Therefore, to make computations fast, mean deformations were computed in the Euclidean space. However, it is known that the deformations do not live in the Euclidean space, but to obtain more accurate approximations of the mean deformations the log-Euclidean framework presented by Arsigny et al. (2006) should be used instead. The main objective of our paper was to study if multi-template methods are able to improve the results of TBM analysis. Even with the mean deformations computed in the Euclidean space this was proven. In the future studies it should be studied if the utilization of the log-Euclidean framework still improves the results.

The methods presented are not limited to specific registration algorithms, brain MRIs, or AD but can be applied with any registration algorithm to any images and study groups.

Conclusions

We presented methods to utilize multiple templates in tensor-based morphometry as opposed to the conventionally used single template. Both visual and quantitative results showed that the multi-template methods produce clear improvement to the single-template method: the statistical maps are smoother with the multi-template method, and the classification accuracies are statistically significantly better and the sample size estimates are smaller for the features from the multi-template TBM methods.

Acknowledgments

The work has been partially funded under the 7th Framework Programme by the European Commission (http://cordis.europa.eu/ist) and Tekes - Finnish Funding Agency for Technology and Innovation (www.tekes.fi).

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, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI 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.

Appendix. List of structures

1 Global
2 Hippocampus right
3 Hippocampus left
4 Amygdala right
5 Amygdala left
6 Anterior temporal lobe, medial part right
7 Anterior temporal lobe, medial part left
8 Anterior temporal lobe, lateral part right
9 Anterior temporal lobe, lateral part left
10 Gyri parahippocampalis et ambiens right
11 Gyri parahippocampalis et ambiens left
12 Superior temporal gyrus, posterior part right
13 Superior temporal gyrus, posterior part left
14 Medial and inferior temporal gyri right
15 Medial and inferior temporal gyri left
16 Lateral occipitotemporal gyrus, gyrus fusiformis right
17 Lateral occipitotemporal gyrus, gyrus fusiformis left
18 Cerebellum right
19 Cerebellum left
20 Brainstem, spans the midline
21 Insula left
22 Insula right
23 Occipital lobe left
24 Occipital lobe right
25 Cingulate gyrus, anterior part left
26 Cingulate gyrus, anterior part right
27 Cingulate gyrus, posterior part left
28 Cingulate gyrus, posterior part right
29 Frontal lobe left, becomes middle frontal gyrus after subdivision of frontal lobe
30 Frontal lobe right, becomes middle frontal gyrus after subdivision of frontal lobe
31 Posterior temporal lobe left
32 Posterior temporal lobe right
33 Parietal lobe left
34 Parietal lobe right
35 Caudate nucleus left
36 Caudate nucleus right
37 Nucleus accumbens left
38 Nucleus accumbens right
39 Putamen left
40 Putamen right
41 Thalamus left
42 Thalamus right
43 Pallidum, globus pallidus left
44 Pallidum, globus pallidus right
45 Corpus callosum
46 Lateral ventricle, frontal horn, central part and occipital horn right
47 Lateral ventricle, frontal horn, central part and occipital horn left
48 Lateral ventricle, temporal horn right
49 Lateral ventricle, temporal horn left
50 Third ventricle
51 Precentral gyrus left
52 Precentral gyrus right
53 Straight gyrus, gyrus rectus left
54 Straight gyrus, gyrus rectus right
55 Anterior orbital gyrus left
56 Anterior orbital gyrus right
57 Inferior frontal gyrus left
58 Inferior frontal gyrus right
59 Superior frontal gyrus left

60 Superior frontal gyrus right
61 Postcentral gyrus left
62 Postcentral gyrus right
63 Superior parietal gyrus left
64 Superior parietal gyrus right
65 Lingual gyrus left
66 Lingual gyrus right
67 Cuneus left
68 Cuneus right
69 Medial orbital gyrus left
70 Medial orbital gyrus right
71 Lateral orbital gyrus left
72 Lateral orbital gyrus right
73 Posterior orbital gyrus left
74 Posterior orbital gyrus right
75 Substantia nigra left
76 Substantia nigra right
77 Subgenual frontal cortex left
78 Subgenual frontal cortex right
79 Subcallosal area left
80 Subcallosal area right
81 Pre-subgenual frontal cortex left
82 Pre-subgenual frontal cortex right
83 Superior temporal gyrus, anterior part left
84 Superior temporal gyrus, anterior part right

References

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46 (3), 726–738.

Aljabar, P., Rueckert, D., Crum, W., 2008. Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. Neuroimage 43 (2), 225–235.

Arsigny, V., Commowick, O., Pennec, X., Ayache, N., 2006. A log-Euclidean framework for statistics on diffeomorphisms. Proc. Medical Image Computing & Computer-Assisted Intervention (MICCAI'06), pp. 924–931.

Ashburner, J., Friston, K., 2000. Voxel-based morphometry — the methods. Neuroimage 11 (6), 805–821.

Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., 1998. Identifying global anatomical differences: deformation-based morphometry. Hum. Brain Mapp. 6 (5–6), 348–357.

Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Mach. Learn. 36, 105–139.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Brun, C., Lepore, N., Pennec, X., Lee, A., Barysheva, M., Madsen, S., Avedissian, C., Chou, Y.-Y., de Zubicaray, G., McMahon, K., Wright, M., Toga, A., Thompson, P., 2009. Mapping the regional influence of genetics on brain structure variability — a tensor-based morphometry study. Neuroimage 48 (1), 37–49.

Chou, Y., Leporé, N., de Zubicaray, G., Carmichael, O., Becker, J., Toga, A., Thompson, P., 2008. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. Neuroimage 40 (2), 615–630.

Chung, M., Worsley, K., Paus, T., Cherif, C., Collins, D., Giedd, J., Rapoport, J., Evans, A.C., 2001. A unified statistical approach to deformation-based morphometry. Neuroimage 14 (3), 595–606.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, 2009. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19 (6), 579–587.

Dubois, B., Feldman, H., Jacova, C., Cummings, J., DeKosky, S., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, Gauthier, D.G.S., Hampel, H., Jicha, G., Meguro, K., O'Brien, J., Pascuier, F., Robert, P., Rossor, M., Salloway, S., Sarazin, M., de Souza, L., Stern, Y., Visser, P., Scheltens, P., 2010. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 9 (11), 1118–1127.

Fan, Y., Batmanghelich, N., Clark, C., Davatzikos, C., the Alzheimer's Disease Neuroimaging Initiative, 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39 (4), 1731–1743.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15 (4), 870–878.

Guimond, A., Meunier, J., Thirion, J.-P., 2000. Average brain models. A convergence study. Comput. Vis. Image Underst. 77 (2), 192–210.

Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33 (1), 115–126.

Hua, X., Leow, A., Lee, S., Klunder, A., Toga, A., Lepore, N., Chou, Y.-Y., Brun, C., Chiang, M.-C., 2008a. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. Neuroimage 41 (1), 19–34.

Hua, X., Leow, A., Parikshak, N., Lee, S., Chiang, M.-C., Toga, A., Jack Jr., C.R., M. W., Thompson, P., The Alzheimer's Disease Neuroimaging Initiative, 2008b. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects. Neuroimage 43 (3), 458–469.

Klein, A., Mensh, B., Ghosh, S., Tourville, J., Hirsch, J., 2005. Mindboggle: Automated brain labeling with multiple atlases. BMC Med. Imaging 5 (7).

Klöppel, S., Stonnington, C., Chu, C., Draganski, B., Scahill, R., Rohrer, J., Fox Jr., N., C. J., Ashburner, J., Frackowiak, R., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681–689.

Leporé, N., Brun, C., Chou, Y.-Y., Chiang, M.-C., Dutton, R., Hayashi, K., Luders, E., Lopez, O., Aizenstein, H., Toga, A., Becker, J., Thompson, P., 2008a. Generalized tensor-based morphometry of hiv/aids using multivariate statistics on deformation tensors. IEEE Trans. Med. Imaging 27 (1), 129–141.

Leporé, N., Brun, C., Chou, Y.-Y., Lee, A., Barysheva, M., de Zubicaray, G., Meredith, M., McMahon, K., Wright, M., Toga, A., Thompson, P., 2008b. Multi-atlas tensor-based morphometry and its application to a genetic study of 92 twins. Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2008). Sep. 6.

Leporé, N., Brun, C., Pennec, X., Chou, Y.-Y., Lopez, O., Aizenstein, H., Becker, J., Toga, A., Thompson, P., 2007. Mean template for tensor-based morphometry using deformation tensors. Proc. of the 10th International Conference on Medical Image Computation and Computer Assisted Intervention (MICCAI-2007), Part II, pp. 826–833.

Lötjönen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., The Alzheimer's Disease Neuroimaging Initiative, 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 49 (3), 2352–2365.

Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. Neuroimage 44 (4), 1415–1422.

Petersen, R., 2004. Mild cognitive impairment as a diagnostic entity. J. Intern. Med. 256 (3), 183–194.

Rohlfing, T., Russakoff, D., Brandt, R., Menzel, R., Maurer Jr., C., 2004. Performance-based multi-classifier decision fusion for atlas-based segmentation of biomedical images. Proc. 2004 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'04). Arlington, USA, pp. 404–407.

Shen, D., Davatzikos, C., 2003. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. Neuroimage 18 (1), 28–41.

Teipel, S., Born, C., Ewers, M., Bokde, A., Reiser, M., Möller, H.-J., Hampe, H., 2007. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 38 (1), 13–24.

Vemuri, P., Gunter, J., Senjem, M., Whitwell, J., Kantarci, K., Knopman, D., Boeve, B., Petersen, R., Jack Jr., C., 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 39 (3), 1186–1197.

Warfield, S., Zou, K., Wells, W., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23 (7), 903–921.

Wolz, R., Heckemann, R., Aljabar, P., Hajnal, J., Hammers, A., Lötjönen, J., Rueckert, D., 2010. Measuring atrophy by simultaneous segmentation of serial mr images using 4d graph-cuts. Proc. of the 7th International Symposium on Biomedical Imaging: From Nano to Macro, pp. 960–963. ISBI-2010.