

Published in final edited form as:

Neuroimage. 2012 January 2; 59(1): 422–430. doi:10.1016/j.neuroimage.2011.07.036.

Iterative Multi-Atlas-Based Multi-Image Segmentation with Tree-Based Registration

Hongjun Jia, Pew-Thian Yap, and Dinggang Shen

Department of Radiology and BRIC University of North Carolina at Chapel Hill, NC 27599, U.S.A.

Abstract

In this paper, we present a multi-atlas-based framework for accurate, consistent and simultaneous segmentation of a group of target images. Multi-atlas-based segmentation algorithms consider concurrently complementary information from multiple atlases to produce optimal segmentation outcomes. However, the accuracy of these algorithms relies heavily on the precise alignment of the atlases with the target image. In particular, the commonly used pairwise registration may result in inaccurate alignment especially between images with large shape differences. Additionally, when segmenting a group of target images, most current methods consider these images independently with disregard of their correlation, thus resulting in inconsistent segmentations of the same structures across different target images. We propose two novel strategies to address these limitations: 1) a novel tree-based groupwise registration method for concurrent alignment of both the atlases and the target images, and 2) an iterative groupwise segmentation method for simultaneous consideration of segmentation information propagated from all available images, including the atlases and other newly segmented target images. Evaluation based on various datasets indicates that the proposed multi-atlas-based multi-image segmentation (MABMIS) framework yields substantial improvements in terms of consistency and accuracy over methods that do not consider the group of target images holistically.

Keywords

Multiple atlases; groupwise registration; groupwise segmentation; intermediate template

1 Introduction

Accurate delineation of different anatomical structures is a critical prerequisite for precise quantification of brain growth and pathology. Manual labeling by human experts is time consuming and may fall short in terms of inter-rater consistency even after sufficient training, making it unsuitable for labeling large datasets. Automatic segmentation methods (Beucher and Meyer, 1992; Klein et al., 2008; Sethian, 1999), which minimizes human intervention, are hence highly desirable.

Atlas-based segmentation methods are highly efficient and capable of yielding reproducible results with accuracy comparable to the manual segmentation (Klein et al., 2008). The basic assumption of atlas-based segmentation methods is that the knowledge regarding the atlas

© 2010 Elsevier Inc. All rights reserved.

{jiahj, ptyap, dgshen}@med.unc.edu .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

can be carried forward to the target image, modulated by some form of structural similarity measure that reflects anatomical resemblance (Langerak et al., 2010). This will however be problematic if the atlas does not match the target image very well. In fact, recent studies have shown that segmentation accuracy can be significantly improved if a suitable atlas catering specifically for the target image is employed. The optimal atlas for the target image (Rohlfing et al., 2004) can be selected among a set of atlases based on either image similarity (Aljabar et al., 2009) or prior knowledge, such as demographic information (Xue et al., 2007). The importance of utilizing an optimal atlas has been demonstrated in a number of hippocampus and cortical/sub-cortical structures related studies (Avants et al., 2010; Wu et al., 2007).

Instead of using a single ‘optimal’ atlas, more recent approaches employ a multi-atlas paradigm that typically involves registering all atlases to the target image, and then fuse all segmentation information from the atlases to produce a final segmentation result (Aljabar et al., 2009; Artaechevarria et al., 2009; Isgum et al., 2009; Khan et al., 2011; Langerak et al., 2010; Lötjönen et al., 2010; Sabuncu et al., 2010; Shi et al., 2010). The potential bias introduced by using only a single atlas can thus be compensated to some extent by combining information from multiple sources. The atlases can be weighted globally with a constant weight for all voxels of a particular atlas (Aljabar et al., 2009), locally with specific weights assigned to local patches (Isgum et al., 2009; Khan et al., 2011), or with an adaptive local-global combined weighting strategy to achieve greater segmentation accuracy (Artaechevarria et al., 2009; Shi et al., 2010). Extensive performance comparison of single-atlas- and multi-atlas-based segmentation methods (Collins and Pruessner, 2010) confirms that the multi-atlas approach yields higher accuracy. Studies further show that the selection of a specific subset of atlases for each target image (Lötjönen et al., 2010), can provide more accurate segmentation results (Aljabar et al., 2009), compared with the case when all, or a randomly selected subset of, atlases are used. For further improvement, iterative atlas selection strategies (Langerak et al., 2010) can be employed for refining atlas selection.

State-of-the-art multi-atlas-based segmentation methods are, however, limited due to the following reasons: 1) The commonly used pairwise registration (Christensen and Johnson, 2001; Klein et al., 2009; Shen and Davatzikos, 2002; Vercauteren et al., 2009) is problematic when a pair of images under registration have large shape differences; 2) The registration between atlases and the target images is usually carried out independently, instead of taking groupwise registration approaches, which are typically more effective in achieving better registration results (Balci et al., 2007a; Balci et al., 2007b; Hamm et al., 2010; Jia et al., 2010; Jia et al., 2011; Joshi et al., 2004; Munsell et al., 2009; Wu et al., 2011); 3) Segmentation is typically performed one image at a time, which may lead to segmentation inconsistency across images; and 4) Many current implementations lack a feed-back loop for further segmentation refinement.

In this paper, we present an iterative multi-atlas-based multi-image segmentation (MABMIS) algorithm for concurrent and consistent segmentation of a group of target images. Two novel strategies are employed to address the limitations discussed above: 1) a novel tree-based groupwise registration method for simultaneously registration of both the atlases and the target images, and 2) an iterative groupwise segmentation strategy for simultaneously segmentation of multiple target images for improved accuracy and across-image consistency. Experimental results show that the new tree-based registration framework can significantly improve registration accuracy, especially for target images with large anatomical differences compared with the atlases, and the iterative groupwise segmentation can dramatically improve segmentation consistency over a group of target images. We will discuss next the details of the MABMIS algorithm.

2. Methods

In this section, the overall framework of MABMIS is summarized in Fig. 1. All atlases are first registered to a common space, facilitated by a combinative tree that is constructed based on an expanded atlas population comprising both the original and the simulated atlases. The images are represented as nodes on the tree, and the deformation fields relating the connected nodes are stored for further use. For each of the target images that needs to be segmented, we locate its most similar image on the tree, so that all atlases can be warped to the target image based on the stored deformation fields. An initial segmentation of the target image is obtained via label fusion. Finally, the segmentation results of all target images are refined with an iterative groupwise labeling strategy until the results converge. Each component of this proposed framework is detailed in the following sections.

2.1. Combinative-tree-based atlas registration

Most existing multi-atlas-based segmentation methods are focused on improving segmentation precision after all atlases have been aligned to the target image by, e.g., using label fusion strategies (Artaechevarria et al., 2009; Khan et al., 2011; Sabuncu et al., 2010) or optimal atlas(es) selection (Aljabar et al., 2009; Langerak et al., 2010; Wu et al., 2007). However, it is worth noting that the accurate and reliable registration also plays an indispensable role in the whole atlas-based segmentation framework and places limits on the achievable segmentation accuracy.

Intermediate templates guided registration—It is generally difficult to obtain accurate registration between images with large shape differences. Good initialization of the spatial transformation can help relieve this difficulty by bringing the images close enough to avoid local minima. Recently, several intermediate templates (IT) guided registration methods have been demonstrated to be effective in the registration of brain structural images (Hamm et al., 2010; Kim et al., 2010; Tang et al., 2009), diffusion tensor images (Jia et al., 2011), 2D shapes (Munsell et al., 2009), and 3D cortical surfaces (Dalal et al., 2010). The key concept is to decompose a large deformation, with the help of intermediate templates, into several smaller ones that can be estimated with higher reliability. According to the method by which that the intermediate templates are constructed, intermediate templates based methods can be classified into two categories: 1) intermediate templates generation (ITG) (Kim et al., 2010; Tang et al., 2009) and 2) intermediate templates selection (ITS) (Hamm et al., 2009; Jia et al., 2011; Munsell et al., 2009).

ITG methods attempt to construct simulated images that are more similar to the target images by using statistical deformation models learned from a training dataset (Fig. 2, left). In (Tang et al., 2009), principal component analysis (PCA) is utilized to learn the variations of the training deformation fields and generate a set of intermediate templates. In this method, the training images are first registered to the template to generate a set of training samples of deformation fields, which are then used to build a statistical model with PCA. This model can be used to approximate the distribution of deformation fields between the individual samples and the template. By sampling the PCA space constructed by the top eigen-vectors, a number of simulated deformation fields are generated for warping the template to obtain a set of intermediate templates. On the other hand, support vector regression (SVR) can be also applied in (Kim et al., 2010) to correlate image appearances with the deformation coefficients for the effective prediction of an initial deformation field for a target image. These methods correspond to a simple 1-level tree structure, in which all paths in the top level are represented by the known deformation fields relating the intermediate templates with the final template (see red dashed arrows in Fig. 2). In this case, only the residual deformation field between the target image and the intermediate template

needs to be estimated. In contrast to ITG, ITS methods aim to select the intermediate templates from a set of real images, and more than one intermediate template may be traversed by a target image before reaching the final template (Fig. 2, right). A minimum spanning tree (MST) can be constructed to relate all target images to the final template (Jia et al., 2011; Munsell et al., 2009), and the registration is performed by warping each target image progressively towards the final template through a series of intermediate templates. Note that the differences between the neighboring images on the tree are relatively small, and thus the deformations between them can be estimated easily with higher accuracy.

We present a novel groupwise image registration framework that combines the key ideas of ITG and ITS - we construct a combinative tree comprising both the original and the simulated images. The combinative tree is capable of growing incrementally, allowing new target images to be appended to it. The three major steps of the combinative-tree-based registration framework are illustrated in Fig. 3. First, one atlas, determined manually or automatically from a population of atlases, is denoted as the root of the tree. A set of simulated images are then generated based on the root atlas/template using a learned statistical model of deformation fields. Finally, a combinative tree based on the expanded image set is constructed. At the same time, we estimate the spatial transformations between all atlases. In the following, we describe in detail the steps involved in Fig. 3.

Generation of simulated images and construction of combinative tree—To minimize the potential bias, the selected root of the tree should be reasonably representative of the atlas population. We adopt the method described in (Hamm et al., 2010; Jia et al., 2011) and select the geometric median image as the root. We assume that the population of atlases, $\mathbf{A}=\mathbf{I} \cup \mathbf{L}$, consist of N intensity images $\mathbf{I}=\{I_0, I_1, \dots, I_{N-1}\}$ and their corresponding label images $\mathbf{L}=\{L_0, L_1, \dots, L_{N-1}\}$. We denote as $\mathbf{S}=\{S_0, S_1, \dots, S_{M-1}\}$ the target images that we want to segment. Our goal is to accurately and consistently segment all images in \mathbf{S} and give segmentation results $\mathbf{U}=\{U_0, U_1, \dots, U_{M-1}\}$ based on the label information provided in \mathbf{A} . Without loss of generality, the root atlas is denoted as I_0 . All other atlases $\{I_1, I_2, \dots, I_{N-1}\}$ are registered to I_0 to obtain a set of deformation fields, which will be used for training. A statistical model, e.g., the PCA based model in (Tang et al., 2009) or the perturbation model in (Kim et al., 2010), can then be adopted to characterize the variability of the deformation fields. In (Kim et al., 2010; Tang et al., 2009), thousands of simulated images are generated to ensure the dense sampling of the space of deformation fields. This approach, however, will cause the expanded set to be significantly biased towards the simulated images and the root template. Moreover, due to dense sampling, many of the simulated images are closely similar to each other or the training image, and are hence redundant. As a remedy, we employ a simulated image selection scheme to reduce the redundancy in the expanded set. For each simulated image warped to the root image $\hat{I}_i \in \{\hat{I}_1, \hat{I}_2, \dots, \hat{I}_W\}$, its shortest distance to the atlases is determined by

$$d_i = \min_{j=0, \dots, N-1} \text{dist}(\hat{I}_i, I_j), i=1, \dots, W, \quad (1)$$

where $\text{dist}(\hat{I}_i, I_j)$ is a distance metric between a pair of images, and W is the number of simulated images prior to selection. To reduce computational load, we define the distance metric as the mean squared difference of the intensity differences after affine registration. Only images with distances larger than the threshold (d_{th}) are kept and used to form the expanded training set together with the original atlases. A total of K simulated images $\{I_N, I_{N+1}, \dots, I_{N+K-1}\}$ are selected, along with their deformation fields $\{G_N, G_{N+1}, \dots, G_{N+K-1}\}$ with respect to the root I_0 . The threshold d_{th} is set so that the number of simulated images is comparable to that of the original atlases; we set $K=2N$.

After generating and selecting simulated images, an expanded training set is formed. This training set provides a better sampling of the image space than the real images (Hamm et al., 2010; Jia et al., 2011) or simulated images (Kim et al., 2010; Tang et al., 2009) alone. We construct a combinative tree with this expanded training set by using the MST algorithm described in (Jia et al., 2011; Munsell et al., 2009). In MST-based approaches, a fully-connected graph is first constructed by calculating the pairwise distances of all pairs of images. The same distance measure defined above is also used for this purpose. To further approximate the empirical manifold of the image space, a k -NN sub-graph is extracted by keeping only the connections between one image and its k nearest neighbors, where k is set to be able to give a connected k -NN sub-graph. And then, the pairwise distance between two images on the sub-graph is updated by adopting Dijkstra's algorithm (Dijkstra, 1959). The MST is finally obtained by applying Kruskal's algorithm (Kruskal, 1956).

On the MST, the relationship of each image to the root template I_0 is indicated by the respective path on the tree as shown in Fig. 4. Each image that is connected to the root template via a red dashed arrow is a simulated image with known deformation field generated from the deformable model. The deformation field of a pair of images connected by a blue solid arrow is estimated using a deformable registration algorithm. It is possible that the target or the atlas images are connected to the root template with more than one path. For example, there are two possible paths for connecting the leftmost atlas of Fig. 4 to the template, since this atlas is directly connected with a simulated image. In this case, the shortest path, which has the least number of segments on the path, will be selected for registration and segmentation. In summary, we can register all atlases to I_0 by following their respective paths on the graph for obtaining the deformation fields $\{G_1, G_2, \dots, G_{N-1}\}$.

Registration of the population of atlases to a target image—Each atlas needs to be aligned to the target image before the segmentation information can be carried over. However, most existing methods, as illustrated in Fig. 5a, register the atlases independently with disregard of their correlation. A better strategy is shown in Fig. 5b (Avants et al., 2010). In this strategy, for each target image, the best-matching atlas is first located, and since this atlas is related to all other atlases on the tree, the stored deformations can be used to warp all atlases through the located atlas to the target image. For this to work, we need to first estimate the deformation fields between every pair of atlases, which can be done by composing two existing deformation fields by using I_0 as a bridge. Note that the deformation field used for warping each atlas to the root template is calculated by following the respective path on the tree. Specifically, the deformation field for warping I_t to I_s can be computed as $G_t \circ G_s^{-1}$ where G_t (or G_s) denotes the deformation field estimated from I_t (or I_s) to the root template I_0 and \circ is the deformation composition operator. Since all deformations of connected images on the tree are known, the registration problem is now reduced to the registration of the target image with the best-matching atlas. Since the target image and its best-matching atlas are relatively similar, the risk of being trapped by local minima is significantly reduced, and the registration accuracy can thus be improved, as indicated by the experimental results below. All deformation fields are stored, together with the tree structure, for use in subsequent steps.

2.2. Initial groupwise labeling

To take advantage of the stored deformation fields between atlases and also the additional similarity information among all target images, we design a sequential registration and groupwise segmentation solution, as illustrated in Fig. 6. Two major steps are involved: 1) All target images are first sequentially attached to the current combinative tree, and 2) All atlases are registered, using the information provided by the updated tree, to each target image for initial segmentation.

Specifically, we first calculate the distance from each new target image to the current tree, which is defined as the minimum distance from each target image to any image (simulated or real image) on the current tree. We then sort all target images according to their distances to the current tree and select one giving the minimum distance, and then attach the selected image to the tree as a child node of its respective best-matching image on the tree, as shown in Fig. 4. Note that if the best-matching image is the root template, the target image is attached directly to the root template and set it as a child node of the root template. This process is iterated until all target images have been attached. This expands the tree to accommodate more images without the need to re-perform all registration. Next, based on the information given by the tree, we will align all atlases to each of the target images for initial segmentation.

For example, given a target image, S_j , $0 \leq j < M$, if its parent image is an atlas, we follow the strategy illustrated in Fig. 5b to warp all atlases to S_j . If the parent image is a simulated image, we trace back to the root template I_0 to find the nearest atlas and use the same strategy in Fig. 5b to align all atlases to S_j . After aligning the atlases with the new target image S_j , an initial segmentation can be obtained through label fusion by concurrent consideration of information from all warped atlases.

During the sequential registration, an initial segmentation of each new image can be obtained. To segment S_j into different tissue classes or functional ROIs, a multi-atlas-based segmentation approach is taken. We determine the label of each voxel in the target space by weighting its local-patch similarity with the aligned atlases in a manner similar to (Isgum et al., 2009). Specifically, for each atlas I_i , $i = 0, 1, \dots, N - 1$, its intensity and label images aligned to S_j are denoted as \tilde{I}_i^j and \tilde{L}_i^j , respectively. To label the c^{th} voxel of S_j , we calculate the similarity of a (cubic) patch centered at this voxel with each of the aligned atlases.

Denoting the patch intensity difference as $d_c(S_j, \tilde{I}_i^j)$, the weight signifying the contribution by the atlas I_i , is given as

$$w_c(S_j, I_i) = \omega(d_c(S_j, \tilde{I}_i^j)), \quad (2)$$

where $\omega(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{x^2}{2\sigma^2}}$ is the Gaussian function. The standard deviation σ is set to take value of the median of $\{d_c(S_j, \tilde{I}_i^j), i=0, 1, \dots, N-1\}$. The Gaussian kernel function is adopted here to emphasize patches that are more similar to that of the target image, while the images dissimilar to the target image will be given very small weights. By this way, we circumvent the need to determine the optimal number of selected atlases (Lötjönen et al., 2010). After obtaining the initial segmentation results of all target images in \mathcal{S} , the segmentation step with label fusion are repeated to further refine the segmentation results as described below.

2.3. Iterative update of segmentation results

In most multi-atlas-based segmentation algorithms, the segmentation is performed once without a feedback loop for correcting inconsistent labeling of the same anatomical structure among different target images. To further improve segmentation consistency, we adopt an iterative labeling approach where, in each iteration, we update the segmentation result of each target by using not only the information from the warped atlases, but also the newly segmented target images.

For the new target images $\mathcal{S} = \{S_0, S_1, \dots, S_{M-1}\}$, we denote the initial segmentation results as $\mathbf{U}^1 = \{U_0^1, U_1^1, \dots, U_{M-1}^1\}$. In the t^{th} iteration, image S_j 's labeling result U_j^t will be updated to U_j^{t+1} by a weighted label fusion step based on all current segmentation outputs and the warped atlases. If S_j differs significantly from all other target images in \mathcal{S} , its segmentation result is less affected by them and is thus derived mainly from the atlases. Therefore, its segmentation result will not be significantly affected by further iterations. On the other hand, if S_j is close to some target images in \mathcal{S} , its segmentation result will be influenced and updated with further iterations. Iterative updating is stopped either when the maximum number of iterations is reached, or when the average difference of the overlap ratio between two consecutive iterations is below a pre-defined threshold. In practice, typically only a few iterations are needed for convergence.

3. Experiments

We evaluate the proposed MABMIS framework on real brain images, by comparing it with state-of-the-art algorithms. All images are first affine-registered to a common space, and the image distance metric is defined to be the mean squared difference of the intensity differences between a pair of images after affine registration. We use PCA to learn the statistical model and construct simulated images following the approach described in (Tang et al., 2009). Diffeomorphic demons (Vercauteren et al., 2009) is used for pairwise registration. The MST algorithm applied to build the combinative tree is implemented as described in (Jia et al., 2011; Munsell et al., 2009). Two data sets are used to evaluate different aspects of the proposed MABMIS method: 1) a subset of 100 images from the ADNI dataset for extensively assessment of the combinative and incremental tree based registration method, and 2) LONI LPBA40 data set with 40 manually labeled images to judge the segmentation accuracy and consistency based on different ROI labels.

3.1. ADNI Dataset

Registration accuracy is fundamentally important in the atlas-based segmentation methods. One can reliably propagate atlas information only when the registration is sufficiently accurate. Thus, we first evaluate the registration accuracy of the proposed MABMIS framework based on the ADNI dataset (ADNI, 2004).

We apply the proposed method to a subset of the ADNI dataset, consisting of totally 100 images randomly selected with 50 normal controls and 50 mild cognitive impairment (MCI) patients. Only the preprocessed images in NIFTI format of 1.5T baseline scans for each subject were used. To reduce sensitivity to intensity variation, the following pre-processing steps were performed:

1. AC (anterior commissure) - PC (posterior commissure) correction was performed on all images, which were then resampled to have a common dimension.
2. The N3 algorithm (Sled et al., 1998) was applied with default parameters to correct for intensity inhomogeneity.
3. Skull stripping was then performed by applying both Brain Surface Extractor (BSE, Shattuck and Leahy, 2002) and Brain Extraction Tool (BET, Smith, 2002), followed by further manual editing for clean results.
4. Histogram matching was performed on all images for overall intensity normalization.
5. FLIRT was applied for affine registration with default parameters (e.g., 12 degrees of freedom). The template image was determined based on the root of a tree built

from the dataset. The distance metric was defined as the intensity differences between two images after pairwise affine registration.

To segment the brains into different tissue types (GM, WM, and CSF), FAST from the FSL package (Smith et al., 2004; Zhang et al., 2001) was applied after the images were skull-stripped. Several samples are shown in Fig. 7 to demonstrate the large anatomical differences within the dataset. Half of the images in each category are randomly chosen for training, and the rest of the images are used as testing images. To generate the simulated images, we first locate, from the training images, the median image that gives the minimum average distance to all other images (Hamm et al., 2010; Jia et al., 2011), and then assign it as the root (I_0) of the tree. All other images in the training set, which will be used as atlases, are registered, based on their paths to the root of the tree, onto the root image and the resulting deformation fields are used to train the statistical model. Using the PCA model (Tang et al., 2009), we select the four eigen-vectors with the largest eigen-values, and sample along each eigenvector's direction according to a Gaussian distribution to get 4 vectors with different scales. Thus, a total of $4^4 = 256$ simulated deformation fields are generated by summing up the eigen-vectors weighted by different scales, and the corresponding simulated images are also obtained by warping the root image. 100 simulated images are finally selected to be combined with the original images to build the combinative tree.

On the training images (atlases)—We first evaluate the registration accuracy by measuring the intensity difference over the registered training image group as atlases. The intensity differences between each warped atlas and the root image is calculated after registration using the pairwise registration method (Vercauteren et al., 2009), the statistical model based registration method (Tang et al., 2009), the tree-based registration method described in (Hamm et al., 2010; Jia et al., 2011), and the proposed combinative tree based registration. The distributions of intensity differences, plotted in Fig. 8, indicate that the proposed method yields smaller alignment discrepancy. Tissue overlap rate evaluation is also performed using Dice ratio (Dice, 1945) defined as $D(U, V) = 2 \times |U \cap V| / (|U| + |V|)$, where U and V are two regions of the same tissue type in two different images, and $|\cdot|$ denotes the volume of a region. The average tissue overlap rates of all four tissues, i.e., WM, GM, Ventricle (VN), and CSF, are calculated. With the combinative tree based registration, the average overlap rate is 83.1%, which is higher than that by the pairwise registration (80.5%), the statistical model based registration (81.1%), and the tree-based registration (80.9%). The paired t -test shows that the improvement of the proposed method over the conventional tree-based registration is significant: $p < 0.005$ for intensity differences, and $p < 0.01$ for overlap rates.

On the test images—We evaluate the registration accuracy of the testing images using various measures. The results, listed in Table 1, indicate that the proposed method achieves the best performance among all methods in comparison. In addition to intensity difference and entropy, segmentation accuracy is also evaluated by computing the average overlap rate between the ground-truth and the estimated segmentation of each testing image to be segmented. Table 1 shows the average tissue overlap rates of all four tissues on all registered test images, again showing the good performance of our method. In addition, as mentioned in Section 2.3, the initial segmentations could be iteratively refined using additional information from the newly segmented images. We stop the iteration when the difference in average overlap rates in two consecutive iterations is less than 0.1%. Over all test images, our method (MABMIS) gives an initial average overlap rate of 81.0%, and further increases to 81.9% after the iterative updating, which is much higher than direct pairwise registration (75.9%), statistical model based registration (78.5%), and the traditional tree-based registration (78.6%). A paired t -test on the average overlap ratios shows that the

segmentation accuracy is significantly improved by MABMIS in comparison with all other methods ($p < 0.005$).

It is worth noting that the running time of our method is only 60 minutes for registration of 50 images, which is significantly lower than the pairwise registration (158 min), the statistical model based registration (85 min), and the traditional tree-based registration (142 min). Our method can hence achieve a better registration performance in a shorter amount of time. The training stage in our framework takes about 2 hours including statistical model building, simulated image generation, and training data registration to the root template, and additional 12 hours to achieve the registrations between all pairs of training images. This part of the computation, fortunately, needs to perform once and the results can be stored for subsequent processing. In Table 1, WM gives a very high overlap ratio, only second to VN. The ventricular region usually has the best contrast, and thus the registration accuracy is significantly higher. Since the overlap ratio for WM is already quite high in the affine-registered images, it is more difficult to further improve its registration accuracy, compared with GM and CSF.

3.2. LONI LPBA40 dataset

The 40 brain images in the LONI LPBA40 dataset (Shattuck et al., 2008) are manually labeled with a total of 54 ROIs on different cortex and sub-cortex regions. Since the LONI LPBA40 dataset has a fixed number of subjects, more training subjects means less testing images, and vice versa. If more subjects are chosen as atlases (i.e., the training subjects), the registration accuracy of the atlases could increase. But the performance improvement due to the iterative update will become limited. On the other hand, if fewer subjects are chosen as atlases, alignment accuracy among images could be affected while segmentation consistency could be improved as more testing images are involved in iterative update. To balance these, we select an equal number of images for training and testing. We randomly select 20 images for training and the other 20 images for testing. We generate 40 simulated images to be used together with the atlases for constructing the tree. The test images are sequentially attached to the combinative tree and registered to the template with guidance by their best matches on the tree. The registration accuracy of the 20 test images is measured by the average overlap rate of the labels of the aligned images. Our method achieves an average overlap rate of 80.5%, which is higher than the pairwise registration (77.0%), the statistical model based registration (78.3%), and the traditional tree-based registration (78.7%). The improvement given by our method over the other methods is statistically significant at $p < 0.001$ (with paired t -test). The computation time-cost of our method is about 47 min, which is less than the pairwise registration (96 min), the statistical model based registration (52 min), and the tree-based registration (80 min).

In Table 2, we list the Dice ratio, averaged over all ROIs, between the estimated labels and the corresponding ground-truth. Using multi-atlas-based labeling, the proposed method outperforms the other methods using various registration schemes. The result for the single-atlas-based segmentation using the best-matching atlas is included as a reference. The average overlap rates for all ROIs given by the three *groupwise registration* based segmentation methods are shown in Fig. 9, where our method, MABMIS, tops in 46 out of 54 ROIs. Fig. 10 shows results for qualitative comparison of the segmentation accuracy given by different methods with the close-up views of two regions provided for demonstrating the improvement.

To measure segmentation consistency across all target images, the segmentations of all target images are warped to a common space, which is the root template space in our case. We then calculate the segmentation overlap rates between all image pairs, and the results are shown using the box-and-whisker plot in Fig. 11. MABMIS gives the best accuracy, with

average overlap rates of 83.8% and 85.0% before and after iterative updating, whereas the other three label fusion based methods give overlap rates of 78.9%, 80.3%, and 80.9%, respectively. We have also evaluated the performance of our algorithm by repeating 10 runs of experiments on 10 different sets of training and testing subjects randomly selected from the LONI LPBA40 dataset. For each run of experiment, we calculated the segmentation accuracy based on the test images. Thus, the mean and standard deviation across runs can be measured as given in Table 3, which indicates again that the MABMIS algorithm achieved much better results compared with other methods.

3.3. Discussions

MABMIS consistently achieves higher segmentation accuracy when compared with state-of-the-art methods. Two key factors contribute to this improvement: 1) high registration accuracy gained through the combinative tree based registration, and 2) utilization of additional information given by the target images in a multi-atlas segmentation setting. In particular, more accurate registration generally translates to better guide of label propagation, especially for complex anatomical structures. By using other labeled target images, the pool of candidate labels for voting is expanded, and more importantly, more candidate labels are coming from the most similar images. So the voting results could match the real label with more chance, which is consistent with previously reported results (Lötjönen et al., 2010). It is also noticed that multi-atlas-based label fusion solution is much better than single-atlas-based label propagation, even if the atlas is optimally chosen among the atlas population.

It is worth noting that, in Fig. 10, the labeling result of the proposed MABMIS method shows a smoother boundary than other methods. There are two possible reasons. First, improved registration accuracy may increase the smoothness of the boundary after weighted-voting based segmentation. With better registration accuracy, the boundaries in the same region of multiple atlases are more consistent, and the voted common boundary will hence be smoother. This can be observed from Fig. 10. We further note that, from the same figure, it can be observed that the boundary becomes increasingly smoother as registration accuracy increases. For example, the result given by the pairwise registration is the least smooth, while the tree-based and the statistical model based registration methods gives slightly smoother results since they produce better registration accuracy. The proposed method achieved the best registration and hence the smoothest boundary. Second, in the iterative updating step performed in MABMIS, we include not only the original atlases, but also incorporate other target images into the final atlas set to help segment each target image. Since there are now an increased number of atlases, the outcome necessarily becomes smoother with more images taking part in deciding the boundaries.

Note that MABMIS is a general framework capable of incorporating different groupwise registration and segmentation algorithms. It is also flexible enough to accommodate more atlases and target images without discarding any existing computation results as long as the selected root atlas/template can reasonably represent the whole dataset. In the case where the root atlas is no longer representative of the image population as progressively more images are augmented, the root image can simply be updated using the augmented image population. With more images being added to the image population, the combinative tree becomes increasingly larger. To efficiently determine, for a target image, the best match on the tree, a simplified distance metric based on down-sampled images, or a fast retrieval solution (Jacobs et al., 1995), can be adopted.

Although the proposed MABMIS method can significantly improve the labeling results on multiple images, it takes a significant amount of time for training and requires a large amount of memory for storing the deformation fields. For example, approximately 14 hours

are required to train using the ADNI dataset (50 training images with dimensions $216 \times 216 \times 128$) and 8 hours using LONI LPBA40 (20 training images with dimensions $220 \times 220 \times 184$) on our servers. To make our method applicable to the large-scale dataset, it is important to reduce the training time and storage requirement. Possible solutions include optimizing the number of simulated images for the underlying dataset to reduce the training time and finding more efficient representation of the dense deformation fields to save the storage space and memory. Future work will be directed to address these issues.

4. Conclusions and future work

In this paper, we present a novel multi-atlas-based multi-image segmentation (MABMIS) framework for achieving accurate and consistent groupwise image segmentation. In contrast to other atlas-based segmentation methods, we focus on the problem of simultaneously labeling a group of images. To this end, two new strategies are proposed: 1) A new image registration framework based on a combinative and incremental tree for better registration between atlases and new target images, and 2) a multi-atlas-based segmentation scheme using an iterative groupwise labeling strategy for improving segmentation accuracy and consistency. The proposed MABMIS framework was validated using various sets of in vivo brain images and the results indicate that significant improvement over the state-of-the-art methods can be achieved. Future work entails evaluation of different fusion schemes and different groupwise registration techniques for further improving the labeling accuracy.

References

- ADNI. 2004. <http://www.loni.ucla.edu/ADNI/>
- Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*. 2009; 46:726–738. [PubMed: 19245840]
- Arteachevarria X, Muñoz-Barrutia A, Ortiz-de-Solórzano C. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *IEEE Transactions on Medical Imaging*. 2009; 28:1266–1277. [PubMed: 19228554]
- Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, Gee JC. The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*. 2010; 49:2457–2466. [PubMed: 19818860]
- Balci SK, Golland P, Shenton M, Wells WM. Free-Form B-spline Deformation Model for Groupwise Registration. *MICCAI*. 2007a:2007:23–30.
- Balci SK, Golland P, Wells W. Non-rigid Groupwise Registration using B-Spline Deformation Model. Open Source and Open Data for MICCAI. 2007b:105–121.
- Beucher, S.; Meyer, F. The Morphological Approach to Segmentation: The Watershed Transformation.. In: Dougherty, E., editor. *Mathematical Morphology in Image Processing*. Marcel Dekker; New York: 1992.
- Christensen GE, Johnson HJ. Consistent Image Registration. *IEEE Transactions on Medical Imaging*. 2001; 20:568–582. [PubMed: 11465464]
- Collins DL, Pruessner JC. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *NeuroImage*. 2010; 52:1355–1366. [PubMed: 20441794]
- Dalal, P.; Shi, F.; Shen, D.; Wang, S. *MICCAI 2010*. Beijing, China: 2010. Multiple cortical surface correspondence using pairwise shape similarity.; p. 349-356.
- Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945; 26:297–302.
- Dijkstra EW. A note on two problems in connexion with graphs. *Numerische Mathematik*. 1959; 1:269–271.
- Hamm, J.; Davatzikos, C.; Verma, R. *MICCAI 2009*. London, UK: 2009. Efficient large deformation registration via geodesics on a learned manifold of images.; p. 680-687.

- Hamm J, Ye DH, Verma R, Davatzikos C. GRAM: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis*. 2010; 14:633–642. [PubMed: 20580597]
- Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, Ginneken B.v. Multi-Atlas-Based Segmentation With Local Decision Fusion - Application to Cardiac and Aortic Segmentation in CT Scans. *IEEE Transactions on Medical Imaging*. 2009; 28:1000–1010. [PubMed: 19131298]
- Jacobs CE, Finkelstein A, Salesin DH. Fast multi-resolution image querying. *SIGGRAPH*. 1995:277–286.
- Jia H, Wu G, Wang Q, Shen D. ABSORB: Atlas Building by Self-Organized Registration and Bundling. *NeuroImage*. 2010; 51:1057–1070. [PubMed: 20226255]
- Jia H, Yap P-T, Wu G, Wang Q, Shen D. Intermediate Templates Guided Groupwise Registration of Diffusion Tensor Images. *NeuroImage*. 2011; 54:928–939. [PubMed: 20851197]
- Joshi S, Davis B, Jomier M, Gerig G. Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. *NeuroImage*. 2004; 23:S151–S160. [PubMed: 15501084]
- Khan AR, Cherbuin N, Wen W, Anstey KJ, Sachdev P, Beg MF. Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): Validation on hippocampus segmentation. *NeuroImage*. 2011 accepted.
- Kim, M.; Wu, G.; Yap, P-T.; Shen, D. MICCAI 2010. Beijing, China: 2010. A Generalized Learning Based Framework for Fast Brain Image Registration.; p. 306-314.
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*. 2009; 46:786–802. [PubMed: 19195496]
- Klein S, Heide U.A.v.d. Lips IM, Vulpen M.v. Staring M, Pluim JPW. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Medical Physics*. 2008; 35:1407–1417. [PubMed: 18491536]
- Kruskal JB. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*. 1956; 7:48–50.
- Langerak TR, Heide U.A.v.d. Kotte ANTJ, Viergever MA, Vulpen M.v. Pluim JPW. Label Fusion in Atlas-Based Segmentation Using a Selective and Iterative Method for Performance Level Estimation (SIMPLE). *IEEE Transactions on Medical Imaging*. 2010; 29:2000–2008. [PubMed: 20667809]
- Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*. 2010; 49:2352–2365. [PubMed: 19857578]
- Munsell BC, Temlyakov A, Wang S. Fast Multiple Shape Correspondence by Pre-Organizing Shape Instances. *IEEE Conference on CVPR*. 2009:840–847.
- Rohlfing T, Brandt R, Menzel R, Maurer CR. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*. 2004; 21:1428–1442. [PubMed: 15050568]
- Sabuncu MR, Yeo BTT, Leemput KV, Fischl B, Golland P. A Generative Model for Image Segmentation Based on Label Fusion. *IEEE Transactions on Medical Imaging*. 2010; 29:1714–1729. [PubMed: 20562040]
- Sethian, J. *Level Set Methods and Fast Marching Methods*. Cambridge University Press; Cambridge, UK: 1999.
- Shattuck DW, Leahy RM. BrainSuite: An Automated Cortical Surface Identification Tool. *Medical Image Analysis*. 2002; 8:129–142. [PubMed: 12045000]
- Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*. 2008; 39:1064–1080. [PubMed: 18037310]
- Shen D, Davatzikos C. HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. *IEEE Trans. on Medical Imaging*. 2002; 21:1421–1439.
- Shi F, Yap P-T, Fan Y, Gilmore JH, Lin W, Shen D. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *NeuroImage*. 2010; 51:684–693. [PubMed: 20171290]

- Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *Medical Imaging, IEEE Transactions on*. 1998; 17:87–97.
- Smith SM. Fast robust automated brain extraction. *Human Brain Mapping*. 2002; 17:143–155. [PubMed: 12391568]
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, Luca MD, Drobnjak I, Flitney DE, Niazy R, Saunders J, Vickers J, Zhang Y, Stefano ND, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. 2004; 23:208–219.
- Tang S, Fan Y, Shen D. RABBIT: Rapid Alignment of Brains by Building Intermediate Templates. *NeuroImage*. 2009; 47:1277–1287. [PubMed: 19285145]
- Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*. 2009; 45:S61–72. [PubMed: 19041946]
- Wu G, Wang Q, Jia H, Shen D. Feature-based groupwise registration by hierarchical anatomical correspondence detection. *Human Brain Mapping*. 2011 accepted.
- Wu M, Rosano C, Lopez-Garcia P, Carter CS, Aizenstein HJ. Optimum template selection for atlas-based segmentation. *NeuroImage*. 2007; 34:1612–1618. [PubMed: 17188896]
- Xue H, Srinivasan L, Jiang S, Rutherford M, Edwards AD, Rueckert D, Hajnal JV. Automatic segmentation and reconstruction of the cortex from neonatal MRI. *NeuroImage*. 2007; 38:461–477. [PubMed: 17888685]
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. on Medical Imaging*. 2001; 20:45–57.

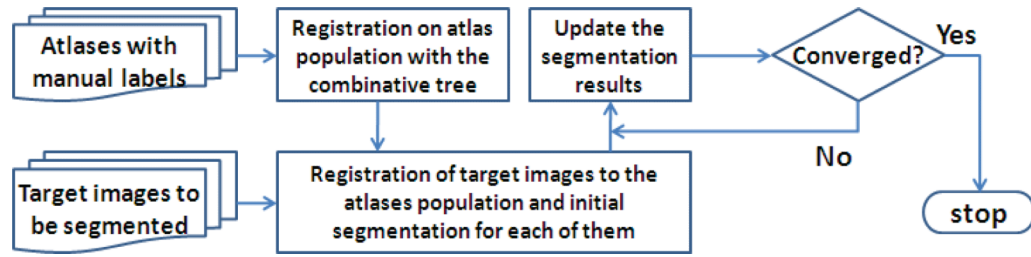


Fig. 1.
The overall framework of MABMIS. The three key components are represented by the rectangular boxes.

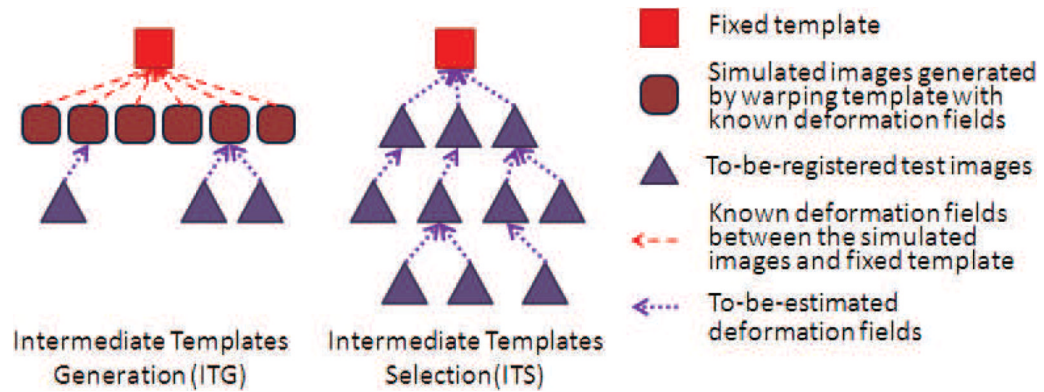


Fig. 2.

Two intermediate templates guided registration schemes: intermediate templates generation (ITG, left) and intermediate templates selection (ITS, right). In ITG, one image is generated, by warping the template with a known deformation field, as the intermediate template for each test image. This intermediate template is more similar to the test image than the original template; therefore, the residual deformation field between each test image and the intermediate template can be estimated with higher accuracy. In ITS, on the other hand, all test images are organized into a tree structure with the template as the root. The deformation field between each test image and the template is thus the combination of several smaller ones, thus joining several intermediate templates along the path traversed by the test image to reach the template.

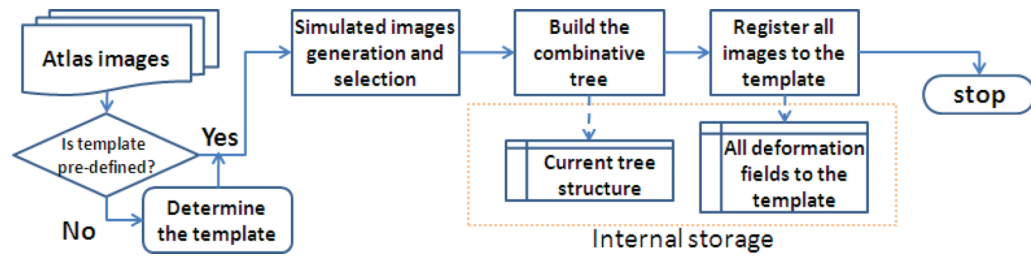


Fig. 3. Flowchart illustrating the registration based on the combinative tree.

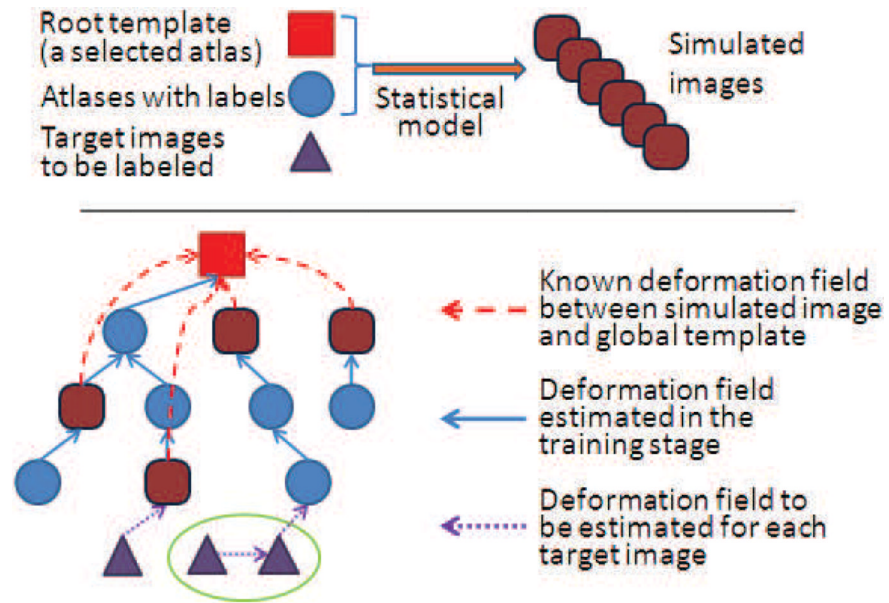


Fig. 4. Illustration of a combinative tree, which can be expanded by appending the target images. For each target image, its best-matching image on the current tree can be an original atlas, a simulated atlas, or even another target image as shown within the green circle. Some images (e.g., the left-most atlas) may be connected to the root template via more than one path.

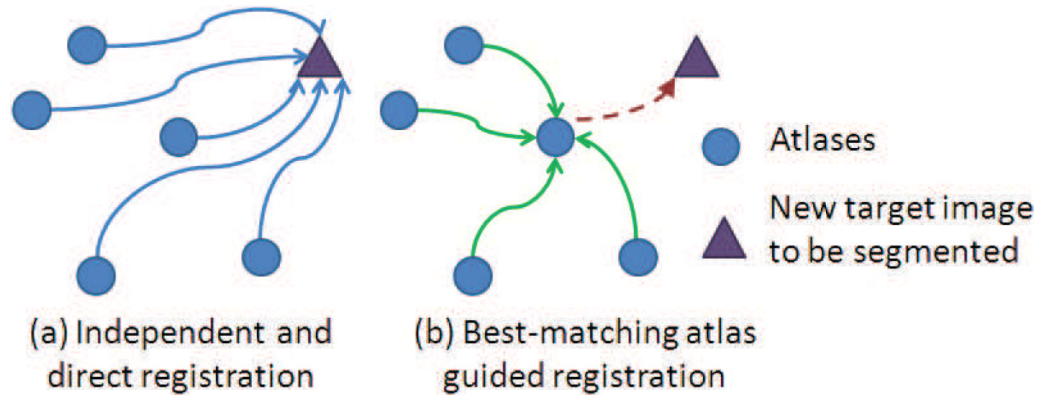


Fig. 5.
Two different ways of aligning the atlases to the target image.

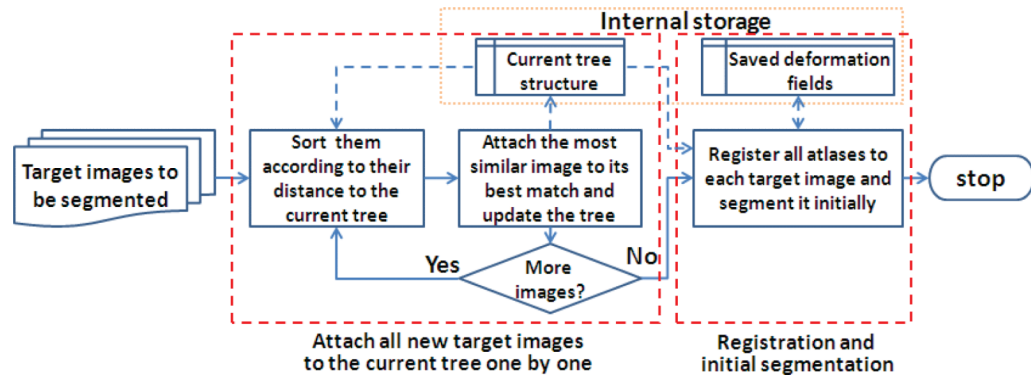


Fig. 6. Sequential registration and segmentation of the target images.

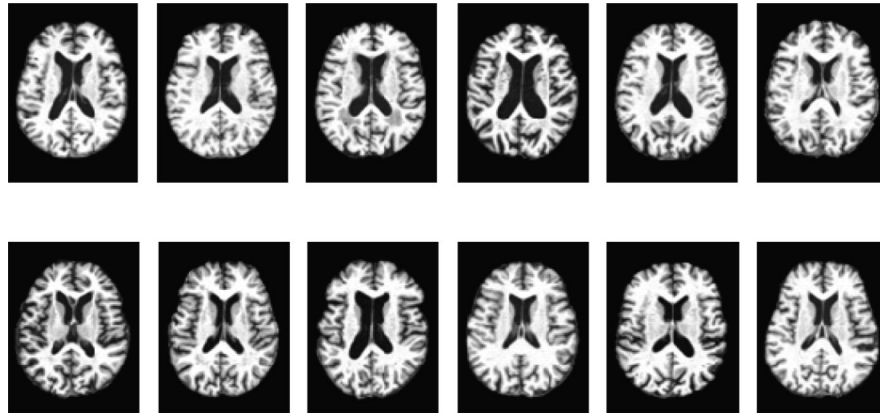


Fig. 7. Sample images from the ADNI dataset. Large anatomical differences remain even after affine registration: MCI patients (top row) and normal controls (bottom row).

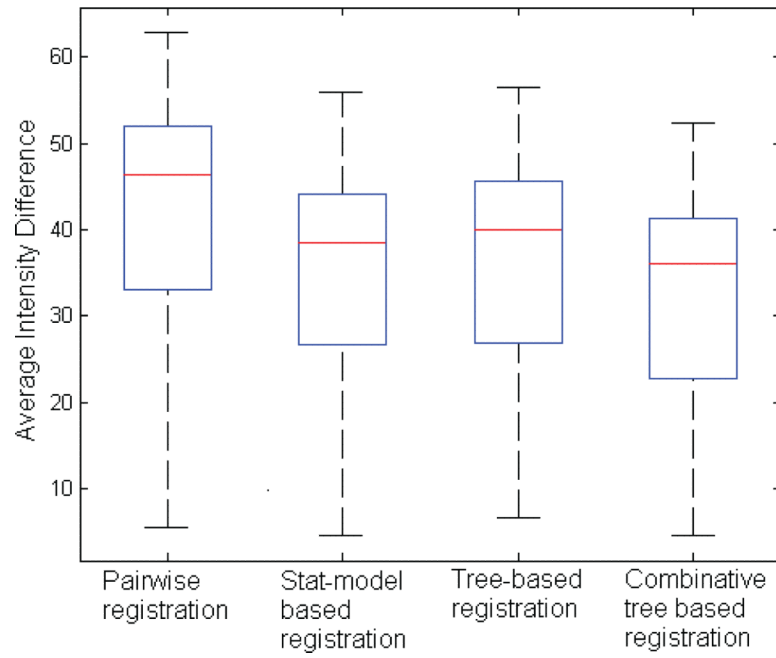


Fig. 8. The distribution of intensity differences of all atlases after registration by four different methods on ADNI dataset.

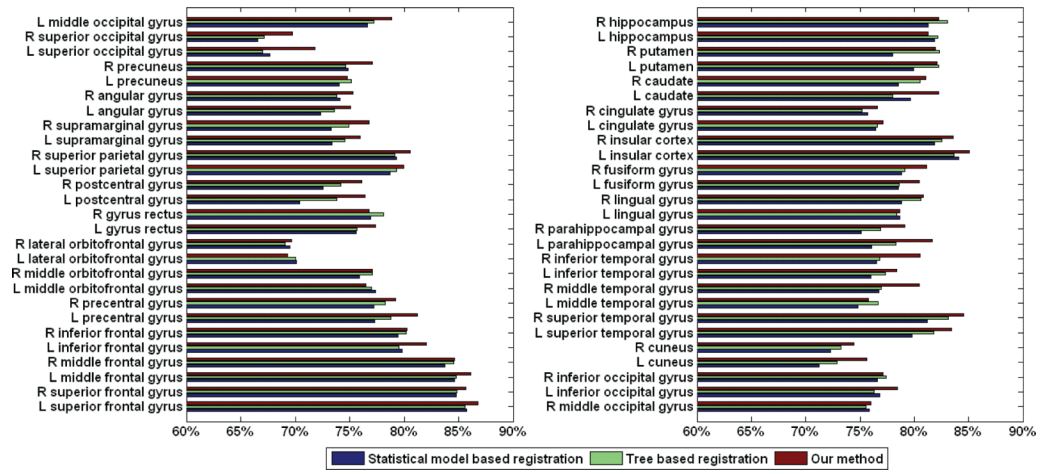


Fig. 9. The proposed method (MABMIS) achieves the best overlap rates for 46 out of 54 ROIs.

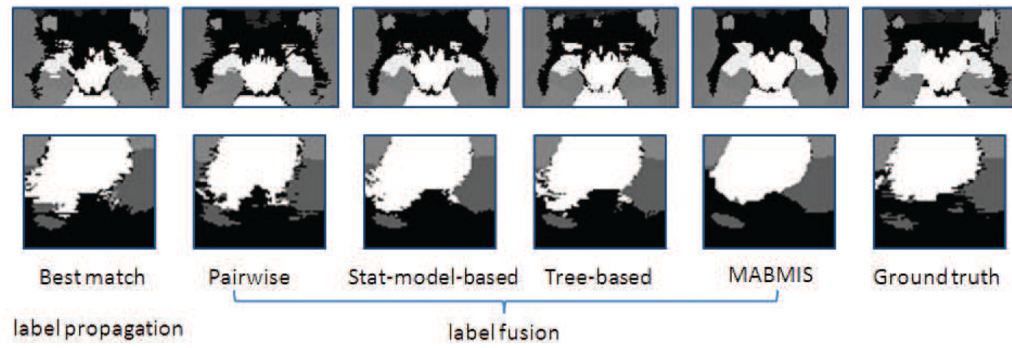


Fig. 10.

Comparison of segmentation accuracy. From left to right: (1) the label propagation result using a best-matching atlas, (2) the label fusion results with the pairwise registration, the statistical model based registration, the tree-based registration, and MABMIS with iterative updating, and (3) manual labels (ground-truth).

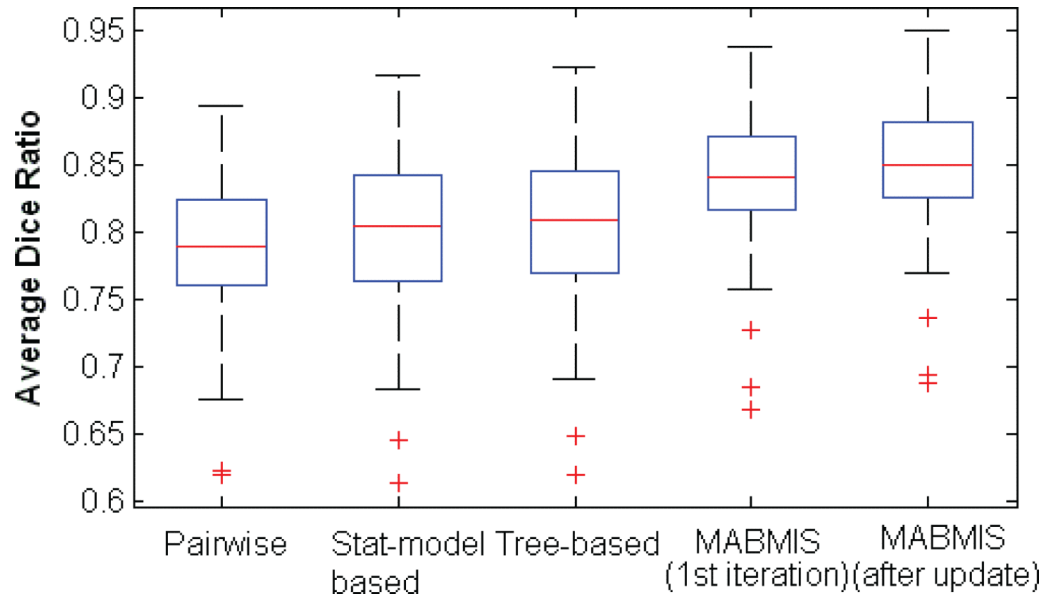


Fig. 11. Segmentation consistency measured by the average overlap rates given by different multi-atlas-based image segmentation methods.

Table 1

The average overlap rates, entropies, and intensity differences of the registered testing images given by different methods.

	Overlap rate			Entropy	Intensity difference
	GM	WM	VN		
Original data	59.3%	70.9%	76.7%	0.788	59.36
Pairwise registration	65.9%	78.6%	87.9%	0.617	44.37
Statistical model based	67.2%	81.0%	88.2%	0.568	39.52
Traditional tree based	67.1%	80.9%	89.3%	0.570	39.63
Combinative tree based	* 69.3%	* 81.9%	* 90.7%	* 0.547	* 36.81

* indicates that the difference with respect to the conventional tree-based registration is statistically significant at $p < 0.01$.

Table 2

Average overlap rates given by different methods on segmentation accuracy.

	Multi-atlas based segmentation with label fusion					
	Single-atlas	Statistical model based registration	Tree-based registration	MABMIS (first round)	MABMIS (after update)	
Segmentation accuracy	Best matched atlas with label propagation 75.5%	Pairwise registration 77.4%	78.5%	78.9%	81.7%	82.8%

Segmentation accuracy and consistency of different methods on the LONI LPBA40 dataset, obtained using ten runs of experiments with random grouping of training and testing subjects.

Table 3

	Pairwise registration	Statistical model based registration	Tree-based registration	MABMIS (1 st round)	MABMIS (after update)
Segmentation accuracy	76.9%±3.2%	78.9%±2.6%	79.1%±2.3%	82.1%±2.0%	83.3%±2.0%
Segmentation consistency	78.3%±2.9%	79.9%±2.2%	80.5%±2.4%	83.6%±1.9%	85.2%±1.8%