SharpMean: Groupwise registration guided by sharp mean image and tree-based registration

Guorong Wu [a], Hongjun Jia [a], Qian Wang [a,b], Dinggang Shen [a,*]

[a] *Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA*
[b] *Department of Computer Science, University of North Carolina at Chapel Hill, USA*

ARTICLE INFO

ABSTRACT

Groupwise registration has become more and more popular due to its attractiveness for unbiased analysis of population data. One of the most popular approaches for groupwise registration is to iteratively calculate the group mean image and then register all subject images towards the latest estimated group mean image. However, its performance might be undermined by the fuzzy mean image estimated in the very beginning of groupwise registration procedure, because all subject images are far from being well-aligned at that moment. In this paper, we first point out the significance of always keeping the group mean image sharp and clear throughout the entire groupwise registration procedure, which is intuitively important but has not been explored in the literature yet. To achieve this, we resort to developing the robust mean-image estimator by the adaptive weighting strategy, where the weights are adaptive across not only the individual subject images but also all spatial locations in the image domain. On the other hand, we notice that some subjects might have large anatomical variations from the group mean image, which challenges most of the state-of-the-art registration algorithms. To ensure good registration results in each iteration, we explore the manifold of subject images and build a minimal spanning tree (MST) with the group mean image as the root of the MST. Therefore, each subject image is only registered to its parent node often with similar shapes, and its overall transformation to the group mean image space is obtained by concatenating all deformations along the paths connecting itself to the root of the MST (the group mean image). As a result, all the subjects will be well aligned to the group mean image adaptively. Our method has been evaluated in both real and simulated datasets. In all experiments, our method outperforms the conventional algorithm which generally produces a fuzzy group mean image throughout the entire groupwise registration.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Groupwise registration has become more and more popular in recent years due to its attractiveness for unbiased analysis of population data (Crum et al., 2004; Maintz and Viergever, 1998; Zitová and Flusser, 2003). Compared to the pairwise registration, groupwise registration can simultaneously estimate the transformation fields for all subjects without explicitly specifying an individual subject as a template, thus avoiding any possible bias (introduced by template selection) in the subsequent data analysis.

One of the most popular groupwise registrations was proposed by Joshi et al. (2004) in 2004. In their method, the groupwise registration is implemented by iteratively constructing the group mean image and estimating the transformation fields of all subjects towards the estimated tentative group mean image. However, this method has several limitations. One major drawback of this method is that it equally treats all subjects during the construction of the group mean image, which can be very fuzzy especially in the beginning of groupwise registration, since all subjects at that moment are not well aligned. As a result, the fuzzy group mean image fails to provide clear guidance to the subsequent pairwise registrations and leads to 1) the loss of anatomical details which can be hardly recovered from the initial fuzzy group mean image; 2) the degradation of the alignment in each round of groupwise registration due to the difficulty of establishing reliable correspondences between sharp subject images and the fuzzy group mean image during the iterative registration procedure; and 3) the slow convergence of the groupwise registration due to the lack of clear and consistent information from the fuzzy group mean image to guide the registration.

Several follow-up papers (Fletcher et al., 2009; Ma et al., 2008) have been published recently. To be robust with the outlier subjects, Fletcher et al. (2009) extended Joshi's method to the Riemannian manifold and proposed to use the geometric median of the group to handle the possible outliers that may deviate the Fréchet mean far away from the real population center. Rather than averaging on image intensity, Ma et al. (2008) proposed a Bayesian-based approach to

* Corresponding author. Fax: +1 919 843 2641.
E-mail addresses: grwu@med.unc.edu (G. Wu), jiahj@med.unc.edu (H. Jia), qianwang@cs.unc.edu (Q. Wang), dgshen@med.unc.edu (D. Shen).

estimate the atlas by iteratively deforming an initial guess to the center of the population. In this way, the final atlas is sharp but at the expense of the bias in choosing an individual subject as the start point. Generally, these methods still used an equal weight for all subjects to build the group mean image. Also, their goal is the robust estimation of the group mean image, not the groupwise registration. However, the importance of keeping the sharpness of the group mean image during the groupwise registration is still not addressed.

Another limitation of the conventional groupwise registration method lies in the pairwise registrations used in each round of groupwise registration, used to align each subject to the latest estimated group mean image. As well known, good pairwise registration helps estimate the clear group mean image which can be used to better guide the next round of groupwise registration. However, in case of large inter-subject variations, it is difficult to register two faraway subjects with different anatomical structures (Hamm et al., 2009; Tang et al., 2009). In light of this, several methods have been proposed to improve the registration. For example, Tang et al. (Kim et al., 2010; Tang et al., 2009) proposed to generate an intermediate template sufficiently similar to each subject for improving the registration. Since the deformation from the template to any intermediate template is pre-known in the training stage, the final registration can be completed by compositing the deformation from the template to the intermediate template with the deformation from the intermediate template to the subject under registration.

In addition, Hamm et al. (2009) presented the tree-based method for groupwise registration. In their method, the pseudo-geodesic median image is selected as the root template after learning the intrinsic manifold of the whole data set. Since a fixed image (i.e., the root image) is used as the final template to register all other subjects, the bias is inevitably introduced in this scenario due to the possible discrepancy between the selected template and the real population center. However, this method benefits from the tree-based organization of images for registration. Specifically, since the subjects are organized in the tree-based hierarchy by considering each subject image as the node, any pair of neighboring images in the tree is similar to each other. Therefore, even for the subject having large anatomical differences from the root subject, it can be still well-aligned through the registration path from its node to the root.

The goal of this paper is to outline the limitations of the conventional groupwise registration (Joshi et al., 2004) and then propose solutions to improve its performance. Specifically, we will first point out the importance of always keeping a sharp group mean image during the groupwise registration. To achieve the sharp group mean image without introducing bias, we generalize the conventional groupwise registration method by defining a new objective function. Specifically, we treat each subject adaptively throughout the registration. Only the registered subjects that are close enough to the tentatively-estimated group mean image will be involved in updating the group mean image, since equally treating the subjects in the early stage will lead to the irreversible loss of structural details especially when most subjects are not well aligned in the beginning. As the registration progresses, subjects are more likely to agglomerate to the population center. Then, more subjects will be allowed to participate into the construction of the group mean image and their contributions will gradually become similar to others. Working under this scenario, the group mean image in our method will gradually approach to the population center, as will be shown by our experimental results on simulated dataset.

The other contribution of this paper is that we improve the registration accuracy of each subject to the group mean image by a tree-based registration. In our method, we do not assume small anatomical differences between individual brain images. As mentioned above, registering two subjects with large differences usually challenges the state-of-the-art registration algorithms, and the bad registration results will undermine the sharpness of the group mean image. We

resort to employing a tree-based hierarchical registration method (Hamm et al., 2009; Jia et al., 2010b) to register each subject to the latest estimated group mean image. Specifically, considering each subject as a node in the tree, we can set the sharp group mean image, instead of any individual subject, as the root of the minimal spanning tree (built upon the image distances of each possible pair of subjects). The benefit of using the tree-based registration is that the registration accuracy can be greatly improved, since each subject will be registered only to its nearby subject (with similar anatomy) and its complete registration to the group mean image can be estimated by compositing all deformation fields along its path to the group mean image (the root of tree).

In our previous work, we proposed a robust groupwise registration method, called ABSORB, in (Jia et al., 2010a). To overcome the limitation of fuzzy group mean in the conventional method (Joshi et al., 2004), ABSORB perform the registration in the conservative way: each subject will only deform w.r.t. its neighboring subjects within a learned image manifold. The advantage is that the global structure, or the distribution of subject images in their intrinsic high-dimensional space, is always preserved during the groupwise registration procedure. However, the computation time is much longer than the conventional method. In this paper, we attack the problem of fuzzy mean image in a different way where we replace the fuzzy group mean image with the sharp one to deterministically guide the registration. Therefore, we are able to achieve much faster groupwise registration than ABSORB and even with better registration performance as will be shown in the experimental results.

In experiments, we demonstrate the advantage of our method by integrating with the Diffeomorphic Demons registration algorithm (Vercauteren et al., 2009), and also compare its performance with the tree-based registration method (Hamm et al., 2009), the conventional groupwise registration method, and ABSORB (Jia et al., 2010a). Extensive experiments on 16 NIREP Data (Christensen et al., 2006) and 40 LONI LPBA40 data (Shattuck et al., 2008a, 2008b) show that our proposed method outperforms the tree-based method, the conventional groupwise registration method, and ABSORB in terms of both registration accuracy and consistency.

In the following, we will present our improved method for groupwise registration in Method section. After that, our proposed method will be intensively evaluated and further compared with other three groupwise registration methods in Experiments section. We conclude in Conclusion section.

Method

In the framework of unbiased groupwise registration (Joshi et al., 2004), the deformation fields are estimated by iteratively registering N subjects to the latest estimated group mean image. In the t-th round of registration ($t=1,...,T$), the group mean image M^t is generated by averaging upon the intensity of the current warped subjects $I^t = \{I_s^t(x) | I_s^t(x) = I_s^0(g_s^t(x)), s = 1,...,N, t = 1,...,T, x \in \mathcal{R}^3\}$ w.r.t. the current estimated transformation fields $G^t = \{g_s^t | s = 1,...,N, t = 1,...,T\}$. Each g_s^t is calculated by considering M^{t-1} as the template and I_s^0 (affine aligned subject) as the moving images. It is worth noting that G^0 are only the initial affine transformations.

Fig. 1 illustrates the framework of the conventional unbiased groupwise registration (Joshi et al., 2004). The dashed arrows denote the pairwise registrations in each round of registration, which are used to deform each subject to the common space of the current group mean image. The solid arrows indicate the evolution of the group mean image with the progress of registration, where the group mean image gradually marches from a gray rectangle (in the first round of groupwise registration) to a final pink circle (in the later round of groupwise registration), as shown in Fig. 1. Here, the red star denotes the real population center. The convergence of this groupwise registration is guaranteed according to the convexity of the distance function (Joshi et al., 2004). However, this algorithm is not robust to
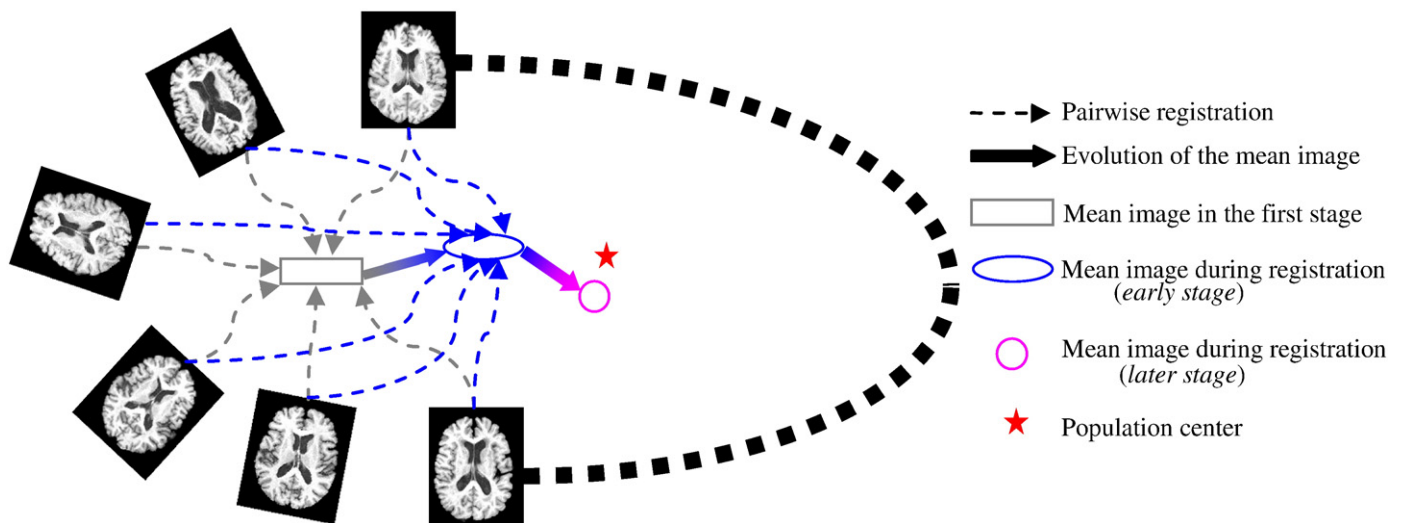
Fig. 1. The schematic illustration of the unbiased groupwise registration algorithm (Joshi et al., 2004). In each round of groupwise registration, all subject images will be registered onto the latest estimated group mean image (shown by dashed arrows), to obtain a new updated group mean image. With the evolution of the group mean image (shown by gray rectangle in the beginning to the red circle in the final), all subjects will be warped closely to the population center.

the outlier subjects (i.e., less registered subjects, especially in the beginning of groupwise registration), although its finally estimated group mean image (the pink circle) might be close to the real population center. We will explain this clearly in the following.

In the conventional groupwise registration (Joshi et al., 2004), the problem of estimating the group mean image M is formulated as a statistical estimation problem:

$$\left\{\widehat{G},\widehat{M}\right\} = arg \min_{G^t, M^t} \sum_{s=1}^{N} \left\{ \sum_x \|I_s^0\left(g_s^t(x)\right) - M^t(x)\|^2 + dist\left(e, g_s^t\right)\right\}, \quad (1)$$

where the term $dist(e,g_s^t)$ is the distance between the identity transformation e and g_s^t (Avants and Gee, 2004; Miller, 2004; Miller et al., 2002; Vercauteren et al., 2009). By fixing the transformation field g_s^t, the estimation of the group mean image is obtained as $M^t(x) = \frac{1}{N}\sum_{s=1}^{N} I_s^{t-1}(x)$, which is the simple average of the warped images according to the current-estimated deformations. It is obvious that this kind of method assumes only one center in the population. The work dealing with multiple centers can be found in (Blezek and Miller, 2007; Sabuncu et al., 2009; Wang et al., 2010).

As pointed in (Fletcher et al., 2009), the major drawback of this method is the poor robustness to the outliers because the contributions (or the weights) are the same for *not only* all subjects in the population, *but also* all voxels in each subject. Therefore, even single outlier image will dramatically mislead the estimation and lead to the fuzzy group mean image. In the following, we will first point out the Importance of always keeping sharpness of the group mean image during registration in Motivation section. Then, we will propose an improved objective function in Objective function section, and explain the solution in Optimization scheme section. Finally, we will summarize our proposed groupwise registration method in Summary section.

Motivation

Unbiased groupwise registration method (Joshi et al., 2004) seeks to iteratively estimate the group mean image and register each subject to the tentative group mean image. However, the initial group mean image $M^1 = \frac{1}{N}\sum_{s=1}^{N} I_s^0(x)$, generated right after the linear alignment, is generally very fuzzy, since the subjects I_s^0 are not well aligned in the beginning of registration. According to our knowledge,

few articles have addressed the importance of keeping the sharpness of the group mean image during the registration. Indeed, the fuzzy group mean image would undermine the groupwise registration performance in two ways: 1) it is difficult to register an individual subject with clear anatomical structures to the group mean image with fuzzy structures; 2) a fuzzy group mean image will challenge the convergence of optimization since it might not provide sufficient anatomical information to guide the registration. The importance of the sharp group mean image in groupwise registration is demonstrated in Fig. 2 by 61 toy images, which are distributed in three branches. Each branch (with 20 images) is generated from the same base image with a red box in Fig. 2(a), to represent one type of cortical folding. For clarity, only three images (i.e., 1st, 10th, and 20th) in each branch are shown as examples. Before registration, the group mean image is very blurry (as shown in Fig. 2(b)). If the groupwise registration starts from this fuzzy group mean image, it will result in an unsatisfactory group mean image (as shown in Fig. 2(b)), since the fuzzy group mean image is not able to informatively guide the groupwise registration of individual images.

This toy example shows the importance of always keeping the sharpness of the group mean image during the groupwise registration. The fundamental reason why the conventional method fails is illustrated in Fig. 3. For easy explanation, we project all 61 images in Fig. 2(a) onto a two-dimensional (2D) space with PCA, as shown by blue circles in Fig. 3, where each branch represents one type of folding pattern. In the conventional method (Joshi et al., 2004), all subjects are equally treated in building the group mean image, regardless of their anatomical differences. This equal weighing strategy, throughout the whole registration, is shown as flat green meshes in Figs. 3(a)–(c). As a result, the group mean image obtained in the first round of registration (shown in Fig. 2(b)) is not only fuzzy but also located far away from the real population center (shown in Figs. 3(b) and (c)), where the red triangles denote for the warped results of all subjects w.r.t. this fuzzy group mean image after projecting them onto the same 2D PCA space. Since the detailed anatomical information has already been lost in the first round of registration, it can be hardly recovered and moved back to the real population center. Fig. 3(c) shows the final groupwise registration result by the conventional method, where the warped subjects are not close to each other and also far away from the real population center.

Instead of equally weighting all subjects, we propose to adaptively weight subjects based on their distances to the latest estimated group
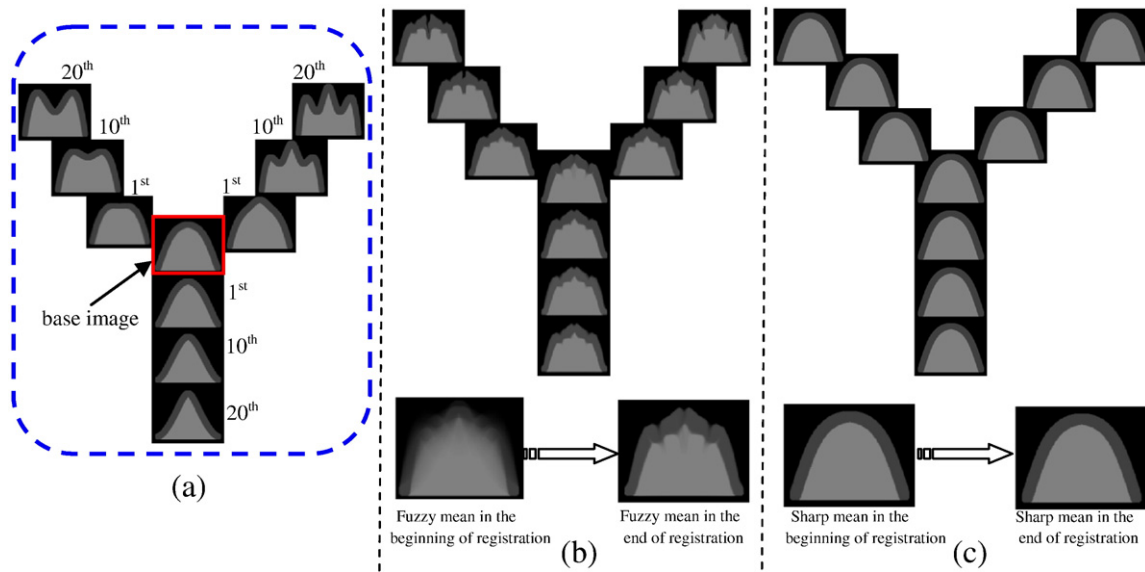
Fig. 2. A toy example demonstrating the importance of always keeping the sharpness of the group mean image during the groupwise registration. The synthetic data are shown in (a). (b) shows the groupwise registration results starting with a fuzzy group mean image, which was achieved by equally weighting all warped subject images in the population. (c) demonstrates the results by the proposed method which starts with the sharp group mean image and always keeps its sharpness during the registration by adaptively weighting each warped subject image for building the group mean image (see the method in Objective function section).

mean image. In the above example (Fig. 3), the distance metric between two images is the sum of squared differences on image intensity. Our idea has been demonstrated through Figs. 3(d)–(f). Specifically, given an estimated group mean image M^{t-1} from the previous round of registration, the contribution of each subject in constructing a new group mean image M^t should be dynamically

determined according to its tentative warping result I_s^{t-1}. For each warped subject I_s^{t-1}, the closer of its distance to M^{t-1} is, the larger weight it should have to contribute in the construction of M^t. To estimate M^t, adaptive weightings for all subjects can be determined based on the distances of those warped subjects to the previous group mean image M^{t-1}. However, the way on how to obtain M^0 in the first
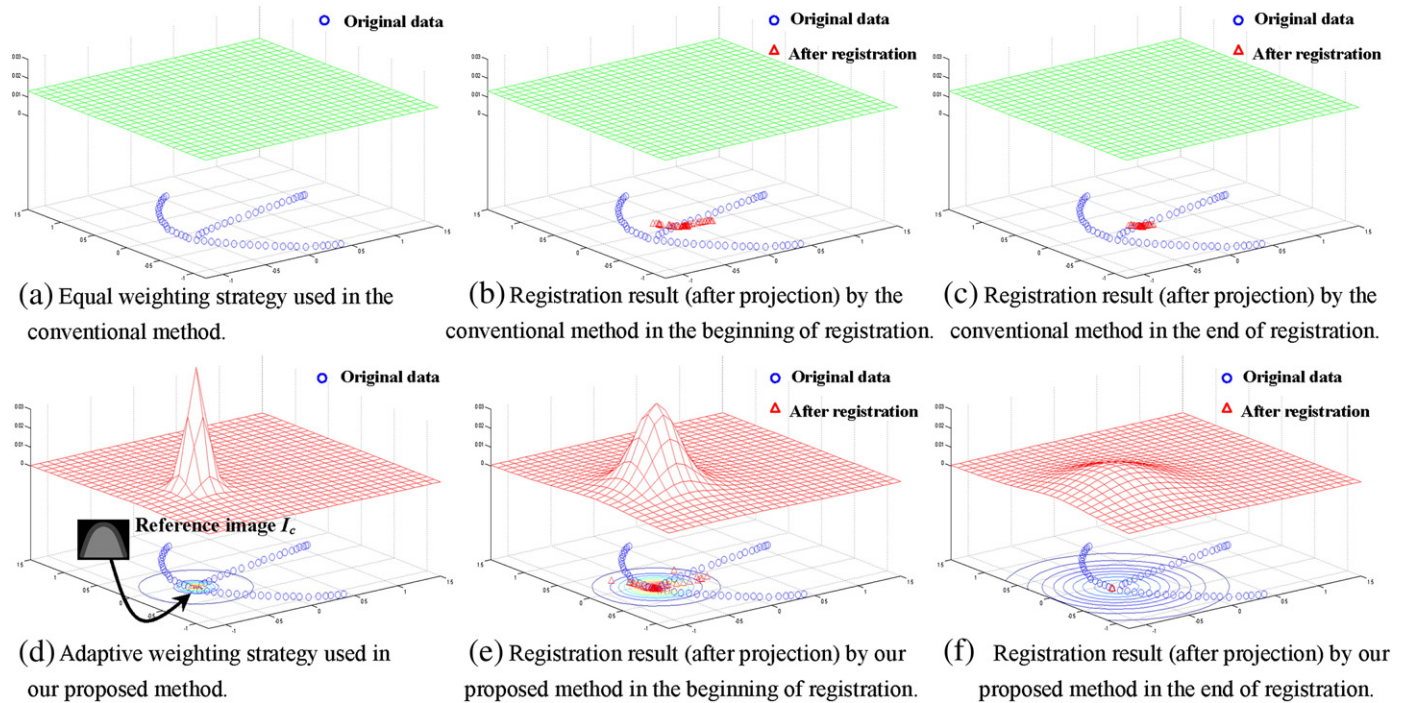

Fig. 3. Demonstration of the advantage of our proposed method. The blue circles in (a)–(f) denote the projection of 61 toy images of Fig. 2 in the 2D PCA space. (a–c) and (d–f) show different weighting strategies and registration results by the conventional method (with equal weighting) and our method (with adaptive weighting), respectively. In our method, we propose assigning large weights to the subjects nearby the group mean image, while small weights to those faraway subjects. The weights for each image throughout the whole registration are displayed with red meshes in (d)–(f), where the shape of the red meshes changes from sharp to flat. The evolutions of groupwise registration results by the conventional method and our proposed method are specially shown in (b–c) and (e–f), respectively, where the red triangles denote the warped subjects in the same 2D PCA space.

round of registration is still unclear. Since we do not have much knowledge on the population center in the beginning, we first select the median image I_c among I^0 as the M^0 by performing the following operations:

$$M^0 = I_c, \quad c = arg\ \min_i \sum_{j=1, j\neq i}^{N} d\left(I_i^0, I_j^0\right),$$ (2)

where d is the distance measurement between two images. Although the advanced geodesic distance can be used here, we turn to use the Euclidian distance due to its simplicity. It is clear that I_c has the minimum sum of distances to all other subjects. Setting M^0 as the median image I_c has the advantage of closeness to the population center. Therefore, we use I_c as the reference image to calculate the weight for each I_s^0 in the first round of groupwise registration according to $d(I_c, I_s^0)$. Fig. 3(d) shows the reference image I_c for 61 toy images in Fig. 2(a), which is very close to the baseline image that we used to generate three branches of images with different cortical foldings. As shown in Fig. 3(d), only the subjects close to I_c will have high weights in constructing M^1, while the faraway subjects have very low weights. The weights used in our method are displayed by the red meshes in Figs. 3(d)–(f), where the shapes of these meshes changes from very sharp in the beginning to very flat in the end of registration. As displayed in Fig. 2(c), the performance of groupwise registration has been significantly improved with our adaptively weighted M^1, compared to that by the conventional method in Fig. 2(b). Similarly, we project the warped results in the first and the final rounds of registration onto the same 2D PCA space, as shown in Figs. 3(e) and (f), respectively. Obviously, our group mean image is more reasonable and much clearer than that by the conventional method, and importantly all subjects have agglomerated very closely to the real population center. In the following section, we will detail our algorithm for achieving the sharp group mean image and improving the overall registration performance.

Objective function

As demonstrated in Motivation section, different subjects should have different dynamic weights, instead of equal and fixed weights, during the groupwise registration. On the other hand, each anatomical region in each subject may have its own difference from the respective region in the group mean image. Therefore, applying the same weight (generally obtained from the entire subject) to all anatomical regions of the same subject may lead to different fuzziness across different regions of the group mean image. To mathematically formulate these problems and solve them together, we propose two strategies below.

First, we propose a distance measurement for each voxel x w.r.t. the current group mean image M^t and the warped subject image I_s^t as:

$$D\left(I_s^t, M^t, x, b\right) = \sum_{y\in P_b(x)} ||I_s^t(y) - M^t(y)||^2,$$ (3)

where $\mathcal{P}_b$ denotes a local image patch centered at voxel x, with the neighborhood size b. The term D measures the overall intensity difference between the corresponding local image patches (centered at x) in the images I_s^t and M^t, by visiting all voxels y in the neighborhood $P_b(x)$. Although other advanced geodesic distance (defined in the manifold) can be employed, the simple intensity difference is used here for easy computation. On the other hand, recall that $I_s^t = I_s^0(g_{ss}^t(x))$, thus the distance measurement in our method is also the function of transformation field g_{ss}^t, albeit implicit. It is worth noting that D approaches to the image distance between two entire images when b is large enough, while D becomes voxel-wise difference in Eq. (3) when b converges to 1. Since the registration results are usually refined from global shape to local shape, the value of b is large in the initial round of registration and then gradually decreases with the progress of registration. We will describe the way to dynamically change the size of b in Eq. (8) of Optimization scheme section.

Second, to treat each subject differently, we introduce a hidden variable $\omega_s^t(x)$ ($s.t. \sum_{s=1}^{N} \omega_s^t(x) = 1, \forall x\in R^3$) to indicate the contribution of each warped subject I_s^t in the construction of the group mean image at a particular voxel x. In the initial rounds of registration, all subjects are not well aligned (especially right after affine registration). If we equally weight all subjects (i.e., the entropy of the weighting set $\{\omega_s^t(x)|s=1,…,N\}$ is high), it will lead to a fuzzy mean image as demonstrated in Fig. 2. To keep the sharp group mean image throughout registration, only those warped subjects I_s^t that are close enough to the currently estimated group mean image M^{t-1} are qualified to have large weights $\omega_s^t(x)$, while other subjects are penalized with small weights $\omega_s^t(x)$. With the progress of registration, all subjects will likely agglomerate to the population center. At that moment, all I_s^ts, as long as they are close enough to the group mean image, will contribute almost equally to the construction of the group mean image, thus providing opportunity to estimate an unbiased group mean image. Since all subjects are already well aligned, very little fuzziness will be introduced to the group mean image. In this paper, the dynamic changes of weights $\{\omega_s^t(x)|s=1,…,N\}$ from strictly binary to loosely uniform are controlled by requiring the entropy of the weighting set $\{\omega_s^t(x)|s=1,…,N\}$ to increase with the progress of registration.

By replacing the distance measurement with Eq. (3) and adding the dynamic control of weights of subjects to the objective function, we arrive at a new objective function for groupwise registration as:

$$F\left(\omega^t, M^t, G^t\right) = \sum_{s=1}^{N} \sum_x \left\{ \left[\omega_s^t(x)\cdot D\left(I_s^t, M^t, x, b\right) + r\cdot\omega_s^t(x)\cdot log\left(\omega_s^t(x)\right)\right] + \frac{1}{\sigma_g} dist\left(e, g_s^t\right)\right\},$$

$$subject\ to : \forall x\in R^3, \sum_{s=1}^{N} \omega_s^t(x) = 1.$$

(4)

Where the scalar r controls the penalty of large distance from I_s^t to M^t. σ_g is used to balance the strength of regularization in the objective function. Compared with the objective function in the conventional method (Eq. (1)), our formulation generalizes the objective function in (Joshi et al., 2004) by introducing the adaptive weights ω for not only each subject but also each spatial location in the common space. We will show its important role when explaining the solution to Eq. (4) in Optimization scheme section.

Optimization scheme

It is non-trivial to simultaneously optimize F in Eq. (4) with many parameters. Instead, we resort to decoupling it with two sub-problems, i.e., alternatively (Sub-Problem 1, SP$_1$) estimate the group mean image and (Sub-Problem 2, SP$_2$) estimate the transformation fields for all subjects. First, given the warping results I^{t-1} (w.r.t. G^{t-1}) in the last round of registration, compute the optimal ω^t. Then the mean image M^t in the current round can be estimated according to ω^t. Second, using G^{t-1} as the initialization, we employ the pairwise registration algorithm (e.g., Diffeomorphic Demons (Vercauteren et al., 2009)) to calculate the transformation field G^t for each subject I_s^0 towards M^t. As a result, new warped results I^t are ready for the next round of registration. Therefore, with the progress of groupwise registration, all I_s^t become closer and closer to the group mean image M^t gradually.

SP$_1$: Estimate the adaptively weighted group mean image

After registering each I^0 with M^{t-1} in the previous round of registration, we are able to obtain the transformation field G^{t-1} as well as the warped subjects I^{t-1}. Discarding all variables irrelevant with ω^t and M^t, the objective function F turns out to be:

$$F_1\left(\omega^t, M^t\right) = \sum_{s=1}^{N} \sum_x \left\{ \omega_s^t(x)\cdot D\left(I_s^{t-1}, M^{t-1}, x, b\right) + r\cdot\omega_s^t(x)\cdot log\left(\omega_s^t(x)\right)\right\}.$$

(5)

The optimal solution to $\omega_s^t(x)$ in Eq. (5) can be immediately calculated by setting $\partial F_1/\partial \omega_s^t(x) = 0$. After removing the constant, the estimation of weight $\omega_s^t(x)$ on each voxel x is given as:

$$\widehat{\omega}_s^t(x) = \exp\left(-\frac{D\left(I_s^{t-1}, M^{t-1}, x, b\right)}{r}\right), \widehat{\omega}_s^t(x) \leftarrow \frac{\widehat{\omega}_s^t(x)}{\sum_{s=1}^{N}\widehat{\omega}_s^t(x)} \tag{6}$$

where $\widehat{\omega}_s^t$ at location x is highly related with two parameters: the temperature r and the neighborhood size b. Here the parameter r is used to control the fuzzyness of the mean image, acting as the inverse temperature in the annealing system based on our observations in Motivation section. Initially, the degree of r is low, i.e., the contribution of particular subject in constructing the group mean image will decrease in the exponential way unless it is very close to the current group mean image. This strategy is used to keep the group mean image sharp. With the progress of groupwise registration, all subjects become closer to the population center; at that time, the temperature r will be increased to encourage equal weighting of all warped subjects I_s^t. Given the overall number of iterations T, the temperature in the t-th round is determined as:

$$r(t) = r_0 + \Delta_r \cdot \frac{t}{T}, \tag{7}$$

where r_0 is the initial temperature and Δ_r denotes the step length of temperature increase. In all of our experiments, we set $r_0 = 1$ and $\Delta_r = max_s\, d\left(I_s^0, I_c\right)$, where I_c is the median image among I^0.

To cater for the possible misalignment, especially in the initial rounds of groupwise registration, we use the neighborhood size b to adaptively control the scale, i.e., from global to local, in measuring the image similarity. Therefore, the weight $\widehat{\omega}_s^t(x)$ used for averaging warped images is different across not only the subjects but also spatial locations in the group mean image space. The value of b in the t-th round of registration is given as:

$$b(t) = b_0 \cdot \left(1 - \frac{t}{T}\right) + 1, \tag{8}$$

where b_0 is the whole image size. When the iteration t increases to T, $b(T) = 1$ and D in Eq. (3) becomes the voxelwise intensity difference.

After determining the weight $\widehat{\omega}_s^t(x)$ for each subject I_s at each location x by Eq. (6), the calculation of the group mean image $M^t(x)$ at each location x is straightforward by setting $\partial F_1/\partial M^t(x) = 0$ (with the derivation given in the Appendix A):

$$\widehat{M}^t(x) = \sum_{s=1}^{N}\psi_s^t(x) \cdot I_s^{t-1}(x), \quad \psi_s^t(x) = \frac{1}{|P_b(x)|}\sum_{y \in P_b(x)}\widehat{\omega}_s^t(y) \tag{9}$$

where $|P_b(x)|$ is the total number of pixels in the local image patch $P_b(x)$. It is clear that (1) the group mean image is the weighted average of all currently registered subjects in the common space; (2) the weight ψ_s^t is adaptive to each subject according to its contribution $\widehat{\omega}_s^t$ calculated in Eq. (6); (3) the weight ψ_s^t is also locally adaptive since the $\widehat{\omega}_s^t$ is computed from the local patch discrepancy.

It is worth noting that M^1 in the first round of registration is the equally weighted average of affine aligned subjects in the conventional groupwise registration (Joshi et al., 2004), which could be very fuzzy (see Fig. 2(b)). In our method, however, M^1 is the weighed mean image of all warped subjects, where the weights are calculated by taking the median image I_c as the reference. Moreover, no bias is introduced to our method for the following reasons: 1) I_c is only used as the reference to calculate the contribution of each subject, instead of using it as a direct template for registration; 2) we gradually increase the temperature r to ensure that all warped subjects have the opportunity to equally contribute to the construction of the group

mean image as long as they are well registered. This argument will be further supported by the experimental results below.

SP$_2$: Estimate the transformation field

By fixing $\widehat{\omega}_s^t$ and $\widehat{M}^t$ obtained in SP$_1$, the objective function in this step becomes (with the derivation given in Appendix B):

$$F_2\left(G^t\right) = \sum_{s=1}^{N} F_2\left(g_s^t\right), \; F_2\left(g_s^t\right) = \sum_x \frac{1}{\sigma_s^t(x)}\|I_s^0\left(g_s^t(x)\right) - \widehat{M}^t(x)\|^2 + \frac{1}{\sigma_g}dist\left(e, g_s^t\right), \tag{10}$$

where $\sigma_s^t(x) = 1/\sum_{y \in P_b(x)}\widehat{\omega}_s^t(y)$. Here we assume each g_s^t is independent. Therefore, the solution to each g_s^t is a well-known optimization problem of quadratic objective function as discussed in many gradient-based registration algorithms (Vercauteren et al., 2008, 2009). The regularization term $dist(e, g_s^t)$ is related with the pairwise registration algorithm used in our method. For example, $dist(e, g_s^t)$ denotes the geodesic distance in LDDMM method (Beg et al., 2005) whose velocity is the function of time starting from template to subject. In this paper, we use Diffeomorphic Demons (Vercauteren et al., 2009) as the pairwise registration component. Therefore, the regularization term on each transformation field g_s^t is defined as the geodesic distance but with constant velocity.

Usually the transformation field $\widehat{g}_s^t$ is independently solved by taking M^t as the template in registration. By taking those 61 toy images in Fig. 2 as examples, the procedure of registering each I_s^{t-1} to M^t is shown in Fig. 4(a). However, there might exist large anatomical variations among the subjects. Assuming all subjects reside on the manifold, it is relatively easy for the pairwise registration algorithm to align two nearby subjects than two faraway subjects. In light of this, we employ a tree-based registration method (Fig. 4(b)), where each subject will only register to its nearby subject on the manifold. To achieve it, we first define the measurement of distance between two subjects I_A and I_B as:

$$d(I_A, I_B) = \sum_x D(I_A, I_B, x, b), \tag{11}$$

where $D(.)$ is defined in Eq. (3) with the neighborhood size b. Let $Q = \{\{I_s^{t-1}\}, M^t\}$ as a set of images under registration in Step 2 (SP_2) of the t-th round of groupwise registration. The distance between any two subjects in Q can be calculated by Eq. (11). Then, a fully connected graph can be built by considering each subject as the node and the manifold distance as the edge weight. Next, the minimal spanning tree (MST) is extracted from the graph, using Kruskal's algorithm (Kruskal, 1956), where M^t is set as the root node since the goal of step 2 (SP_2) is to calculate the transformation field from each subject I_s^{t-1} to M^t. In this way, all images can be organized into a tree structure where only similar images are connected. The advantage of using the tree-based registration is obvious: It allows the subjects to be aligned more robustly and accurately to the current group mean image, especially for the subjects far away from the population center.

Fig. 4(b) demonstrates the procedure of MST-based registration in our method. In registering each subject I_s^{t-1} with M^t, the pairwise registration algorithm will be employed between I_s^{t-1} and M^t if I_s^{t-1} is directly connected with M^t (root node). Otherwise, the registration is performed sequentially along the path determined during the construction of the MST. For clearly describing SP$_2$ in our method, we briefly summarize it as follows:

1. Calculate the distance between any pair of images in the set Q according to Eq. (11);
2. Construct the fully connected graph;
3. Extract the MST by setting M^t as the root node;
4. For each subject $I_s^{t-1} \in I^{t-1}, s = 1, .., N$:
   (a) If the transformation field from I_s^{t-1} to M^t is already computed, then exit;
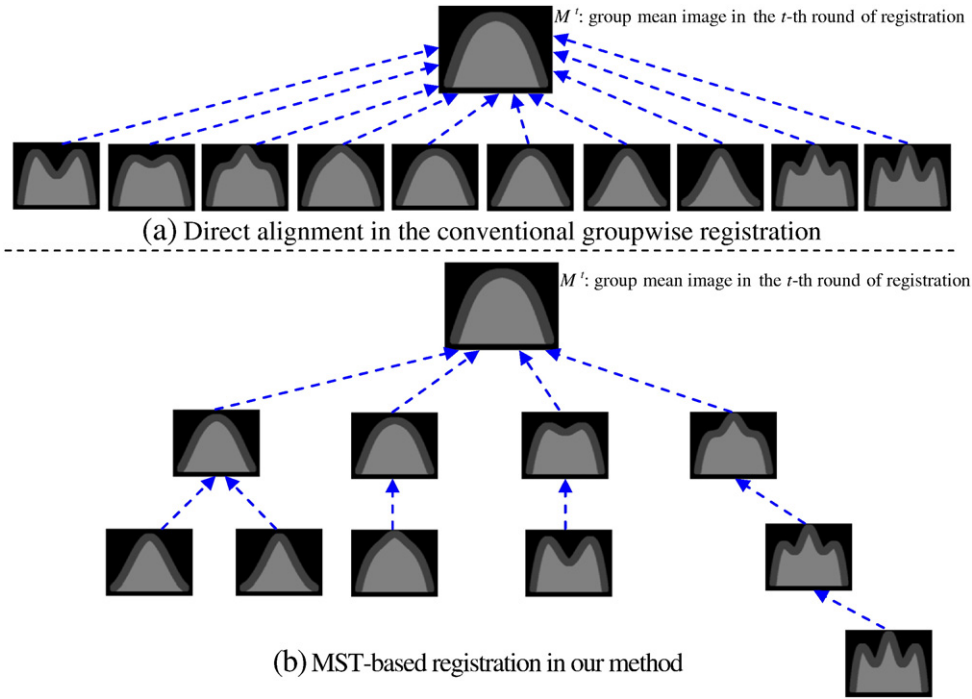
Fig. 4. (a) Pairwise registration is independently performed for each subject, regardless of its difference to the current group mean image. (b) Our MST-based registration framework is used in each round of groupwise registration, where all subjects are organized into a tree-based hierarchy by considering each subject as the tree node. In our registration, each subject will be sequentially registered to its parent nodes one by one until it reaches the root node (i.e., the group mean image).

(b) Otherwise, transverse from node I_s^{t-1} to M^t by composing all transformation fields from the current node to its parent node and then the next level of parent node until reaching the root node M^t.

(c) Obtain the overall transformation field for I_s^{t-1} by composing all transformation fields along the path from I_s^{t-1} to M^t. This overall transformation field will be used as a good initialization to estimate more accurate transformation field between I_s^{t-1} and M^t.

Summary

Compared to the conventional method the advantages of our method are: 1) the group mean image is a weighted average of aligned subjects (Eq. (9)), instead of a simple arithmetic average. Weights are adaptively determined not only for each aligned subject I_s^t, but also for each image location x; 2) the contribution of each subject is dynamically adjusted throughout the groupwise registration in the annealing scenario; 3) the sharpness of the group mean image is always preserved throughout the registration; 4) the registration accuracy in each round of groupwise registration has been improved by using the MST-based hierarchical registration framework.

For comparison, we first summarize the conventional method as follows:

Assume I^0 is the set of affine aligned subjects.

1. Compute the group mean image M^1 in the first round of registration by averaging all I_s^0s with equal weight;
2. Set $t=1$;
3. Register each subject I_s^0 to the currently estimated M^t by Diffeomorphic Demons (Vercauteren et al., 2009) and get the registration result I_s^t;
4. Compute the group mean image M^{t+1} by averaging all I_s^ts with equal weight;
5. Set $t \leftarrow t+1$;
6. Loop from step 3 to step 5 until convergence or reaching the abortion criterion ($t=T$).

Our new groupwise registration method is summarized as follows:

1. Calculate the median image I_c among the population by Eq. (2);
2. Set $t=1, r_0=1, \Delta_r = max_s d(I_s^0, I_c)$, and b_0 the whole image size;
3. Initialize each g_s^0 with only affine transformation to the median image I_c;
4. Set $M^0 = I_c$ and calculate the weight of each I_s^0 by Eq. (6);
5. Compute the group mean image M^1 in the first round of registration by Eq. (9);
6. Build the MST on the image set $Q = \{\{I_s^{t-1}\}, M^t\}$;
7. Register each subject I_s^0 to the currently estimated M^t by the MST-tree-based registration (Diffeomorphic Demons (Vercauteren et al., 2009) is used here as the pairwise registration component);
8. Warp the original subject I_s^0 to the group-mean-image space by g_s^t, thus obtaining I_s^t for the next round of registration;
9. Calculate the weight of each current-warped subject I_s^t by Eq. (6);
10. Compute the group mean image M^{t+1} by Eq. (9);
11. Set $t \leftarrow t+1$;
12. Loop from step 6 to step 11 until convergence or reaching the abortion criterion ($t=T$).

The diagrams of the conventional method and our groupwise registration method are shown in Figs. 5(a) and (b), respectively. As summarized in the conventional method, the construction of group mean image and the registration of each subject to the group mean image are integrated to help each other for reaching the goal of groupwise registration. Our method improves each of these two steps, i.e., building the sharp mean image by dynamic and adaptive weighting of warped subjects and also performing registration of each subject to the group mean image by the tree-based method. Thus, our method can produce better results as detailed below.

Experiments

In our experiments, we have extensively evaluated the performance of our groupwise registration method in atlas construction and
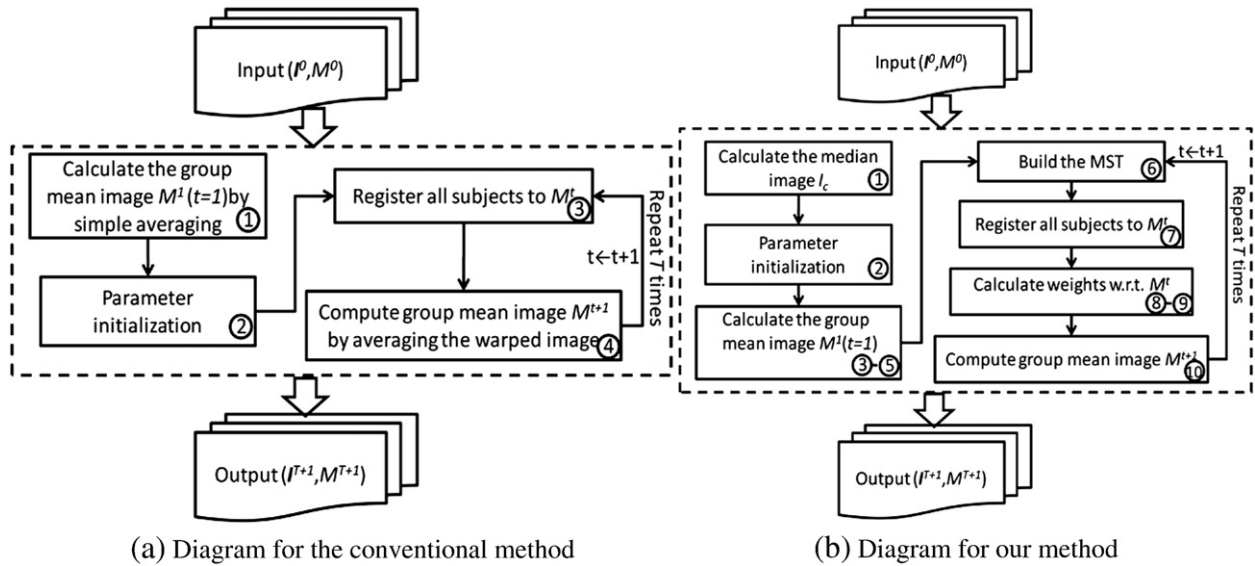
Fig. 5. Diagrams for the conventional method (a) and our method (b). The step numbers in the figure are consistent to those used in Summary section for the two methods.

ROI labeling. For comparison, we set the tree-based registration method (Hamm et al., 2009), the conventional groupwise registration method (Joshi et al., 2004), and the ABSORB (Jia et al., 2010a) as the baseline methods. It is worth noting that the tree-based method is not the groupwise registration method, since it needs to first set an individual image as the root and then register all other subjects to the root image. The latter two methods are both groupwise registration methods. To be fair, Diffeomorphic Demons algorithm (Vercauteren et al., 2009) is used as the pairwise registration component in all these four methods (including our method). Also, we use the same iteration number T for both the conventional method and our method.

We first evaluate the group mean image and the overlap ratios on WM, GM, and ventricle (VN) for 18 elderly subjects. Then, we compare the overlap ratios of aligned ROIs using NIREP dataset (http://www.nirep.org/) and LPBA40 dataset (http://www.loni.ucla.edu/Atlases/LPBA40), which have 32 and 54 manually delineated labels for each brain image, respectively.

The overlap ratio of ROIs is an important criterion to judge the registration performance. Here we use the Jaccard Coefficient metric (Jaccard, 1912) to measure the alignment of two regions with same label. For the two registered regions A and B, the Jaccard Coefficient is defined as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|},$$ (12)

where $|\cdot|$ denotes the volume of the underlying region. Since there is no template with labeling information in the groupwise registration,

we need to vote a reference image before calculating the overlap ratio on each ROI. Here, we vote the tissue assignment of each voxel with the majority of all tissue labels at the same location from all the aligned subjects. Then, the Jaccard Coefficient between each registered subject image and the voted reference will be calculated. It is worth noting that we use this procedure to evaluate the overlap ratio for all four registration methods. In the following experiments, we report the average score of Jaccard Coefficients as the overlap ratio for each tissue label and ROI.

Experiment on 18 elderly brains

Fig. 6 shows a group of 18 elderly brain images. Each image has the size of $256 \times 256 \times 124$ and the resolution of $0.9375 \times 0.9375 \times 1.5 mm^3$. It can be observed that the anatomical structures vary a lot across different subjects, especially for the ventricle and cortex.

Conventional groupwise registration method starts from a very fuzzy group mean (with its 3D rendering shown in blue box of Fig. 7). On the contrary, our method begins with a clear group mean image (as shown in red box of Fig. 7), which is close to the population center. The evolutions of the group mean image by the conventional and our registration methods are provided in the top and bottom rows, respectively. It is clear that, although the final group mean images are similar by both methods, ours is much sharper than that by the conventional method. Also, this experiment has demonstrated that our method will not introduce bias in the final group mean image since the group mean images by two methods (ours and the conventional unbiased registration method) are very similar. To
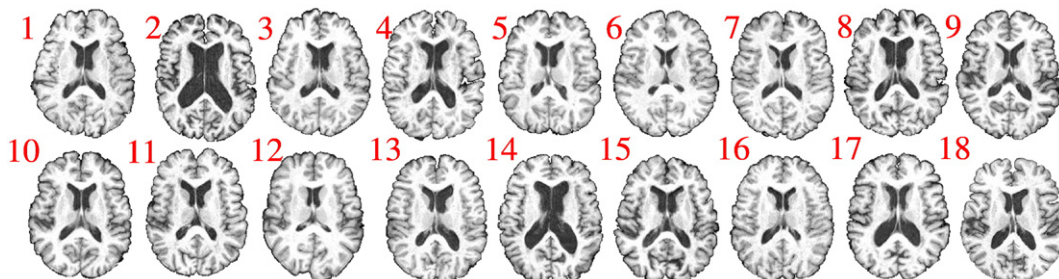

Fig. 6. 18 elderly brain images used for the evaluation of the tree-based registration method and three groupwise registration methods.
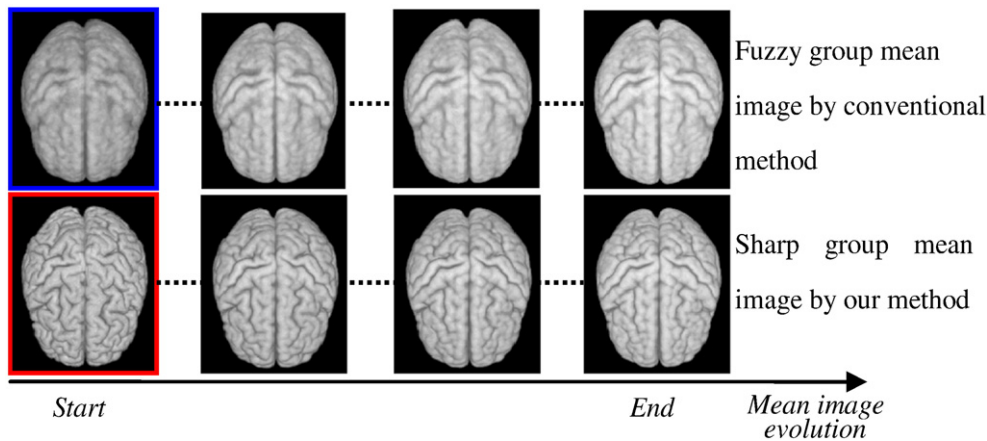
Fig. 7. The evolution of the group mean image. The evolutions of the conventional groupwise registration starting with a fuzzy mean (in blue box) and our method with a sharp mean (in red box) are displayed in the top and bottom rows, respectively.

better evaluate the group mean images by these two methods, we show six transverse views in Fig. 8.

To quantitatively measure the registration accuracy by overlap ratio, we first vote a reference image based on the warped results of all (tissue-segmented) subject images, as mentioned in the beginning of this section. Then, the overlap ratio is calculated between each warped subject image and the reference image one by one. Table 1 reports the overlap ratio on white matter (WM), gray matter (GM), and ventricle (VN) by the tree-based method and three groupwise registration methods, respectively. The average overlap ratio on three tissues (WM, GM, and CSF) is 68.12% by the tree-based method, which is lower than any of the groupwise registration method. This demonstrates the advantage of groupwise registration over the tree-based registration method which has bias in selection of template and thus may affect registration of a group of subjects. For the groupwise registration methods, the overlap ratio is 78.07% by our method and 70.84% by the conventional method, indicating a 7.23% improvement. Compared to the ABSORB method, the overlap ratio by our method is also slightly better, with 2% improvement. The standard deviations on

three tissues by these four methods are also reported in Table 1. Accordingly, we show the maps of overlap on WM, GM, and VN by the tree-based method, the conventional method, the ABSORB method, and our method in Figs. 9(a)–(d), respectively.

Although the evolution of the group mean image can be visually inspected in Fig. 7, we further quantitatively plot the evolution curves of the tissue overlap ratio during the groupwise registration by the conventional method and our method. To demonstrate the advantage of employing the adaptive and dynamic weighting strategy (controlled by the local distance measurement D and the gradually decreased temperature r in Eq. (6)), we also set the median image as the initial group mean and then allows the conventional groupwise registration method to perform the rest of registration. We call this method as "pseudo sharp mean" method. Fig. 10 shows the evolution curves of overlap ratio of white matter (Fig. 10(a)), gray matter (Fig. 10(b)), and ventricle (Fig. 10(c)) with respect to the iteration number. It can be observed that the performance of the conventional method and the "pseudo sharp mean" method is comparable and their overlap ratios are much lower than ours, not only in the end of registration but also during

(a) by conventional method
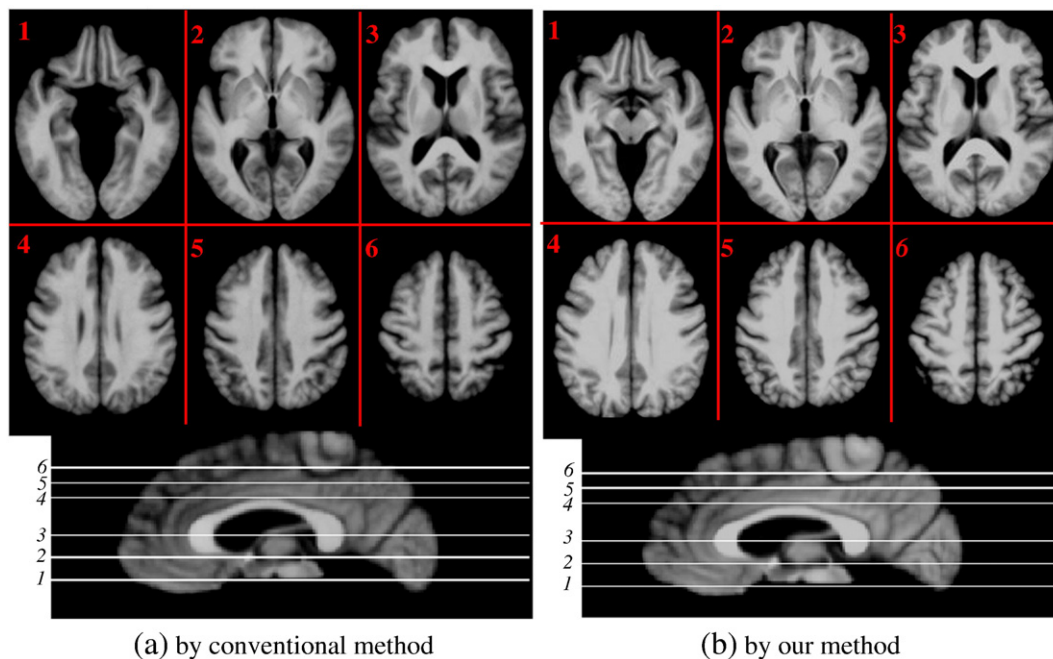
(b) by our method

Fig. 8. Six transverse views of the group mean images by the conventional method and our method.

Table 1
Overall overlap ratios and standard deviation of WM, GM, and VN by the tree-based registration method, the conventional groupwise registration method, ABSORB algorithm, and our registration method.

	WM	GM	VN	Overall
Tree-based method	68.15% ($\pm$1.96%)	59.74% ($\pm$3.51%)	76.51% ($\pm$3.08%)	68.12% ($\pm$2.85%)
Conventional method	73.88% ($\pm$1.52%)	60.51% ($\pm$2.28%)	78.14% ($\pm$3.53%)	70.84% ($\pm$2.44%)
ABSORB	79.01% ($\pm$1.27%)	66.82% ($\pm$2.60%)	82.33% ($\pm$2.24%)	76.05% ($\pm$2.04%)
Our method	81.36% ($\pm$1.11%)	70.29% ($\pm$3.03%)	81.72% ($\pm$2.36%)	78.07% ($\pm$2.16%)

the whole registration procedure. Specifically, the "pseudo sharp mean" method still produce worse results, compared to ours, although it has the same performance in the first iteration as our method's. The reason is that, even using the median image as the initial group mean image in the "pseudo sharp mean" method, the subsequently constructed group mean images by simple averaging the warped subjects are still fuzzy, thus suffering the same problem as the conventional method.

Besides providing the evaluation on overlap ratios, we further quantitatively evaluate the registration consistency according to the entropy of tissue probability on each brain voxel across 18 aligned subjects. Note that lower entropy indicates more consistent registration results across individual subjects. The average entropy value is 0.72 after affine registration. After non-rigid registration, the average entropy value is 0.54 by the tree-based method, 0.51 by the conventional method, 0.42 by ABSORB, and 0.34 by our method. Again, our groupwise registration method achieves the best performance in terms of registration consistency.

Experiment on NIREP data

In this experiment, we align 16 brain subjects in NIREP dataset (Christensen et al., 2006) by the tree-based method and all three groupwise registration methods, respectively. The NIREP dataset consists of 16 3D MR images of 8 normal male adults and 8 normal female adults, each with 32 manually-delineated gray matter ROIs. All 16 MR images have been aligned according to the anterior and posterior commissures (AC and PC). The image size is $256 \times 300 \times 256$, and the voxel dimension is $0.7 \times 0.7 \times 0.7 mm^3$.

After calculating the overlap ratio for each ROI after the registration of 16 images, the overlap ratio is 61.69% by the tree-based method, 61.25% by the conventional groupwise registration method, 65.31% by ABSORB, and 66.66% by our groupwise

registration method (which achieves the highest overlap ratio among all four methods). Fig. 11 shows the overlap ratio in each ROI by the three groupwise registration methods with blue for the conventional method, green for ABSORB, and red for our method, respectively. Furthermore, we use the blue "†" and red "*" to designate the significant improvement of overlap ratio (with the *p*-value less than 0.05 in *t*-test) on ROIs by our groupwise registration method, compared to the conventional method and ABSORB, respectively. It can be observed from Fig. 9 that our method outperforms the conventional method in all ROIs, demonstrating the importance of always keeping the sharpness of the group mean image during the entire registration procedure. Compared with ABSORB, our method has obtained significant improvements on 20 out of 32 ROIs.

Experiment on LONI LPBA40 data

In this experiment, we use the LONI LPBA40 dataset (Shattuck et al., 2008a, 2008b) with 40 brain images and 54 manually labeled ROIs in each brain image. Similarly, we employ the tree-based registration method and three groupwise registration methods to align these 40 brain images, respectively. The overall overlap ratio is 66.95% by the tree-based method, 66.36% by the conventional method, 69.50% by ABSORB, and 70.36% by our method. Fig. 12 shows the overlap ratios in all 54 ROIs by the three groupwise registration methods, with blue for the conventional method, green for ABSORB, and red for our method, respectively. Also, the blue "†" is used to denote the significant improvement on ROIs achieved by our method over the conventional method, while the red "*" is used to denote the significant improvement by our method over ABSORB. It can be observed that our method outperforms the conventional method on most ROIs. On the other hand, although our method produces slightly better overlap ratio than ABSORB, the improvement
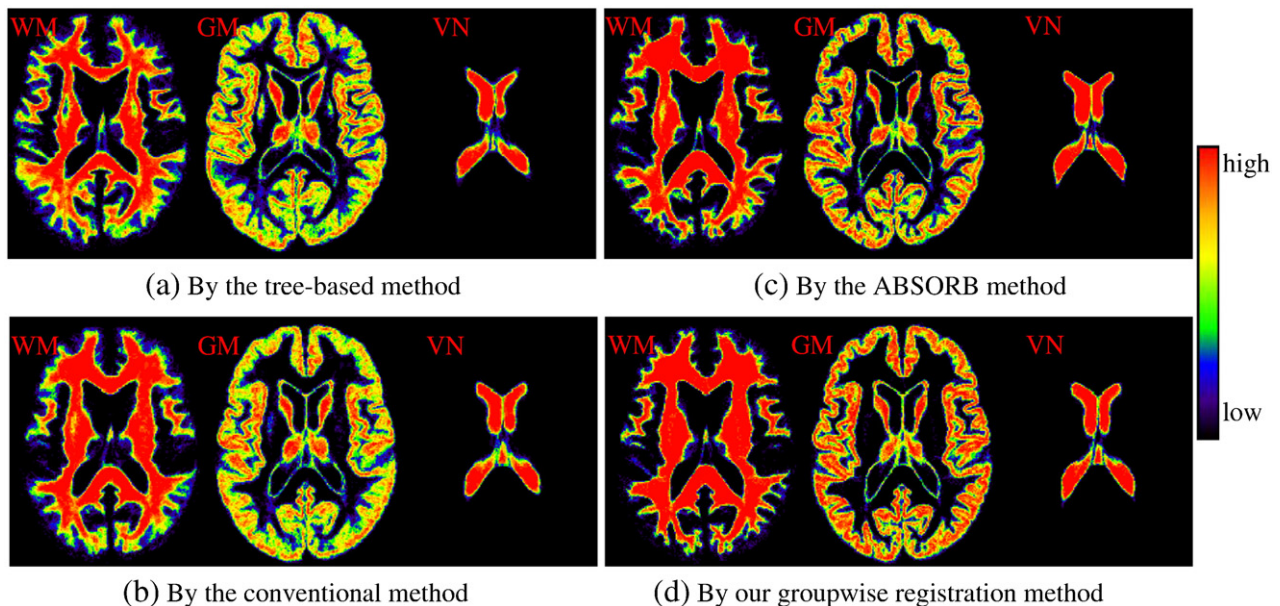

(a) By the tree-based method

(c) By the ABSORB method

(b) By the conventional method

(d) By our groupwise registration method

Fig. 9. The overlap of VN, GM, and WM by the tree-based method, the conventional method, the ABSORB method, and our groupwise registration method.

(a) Evolution curves of overlap ratio on white matter

(b) Evolution curves of overlap ratio on gray matter

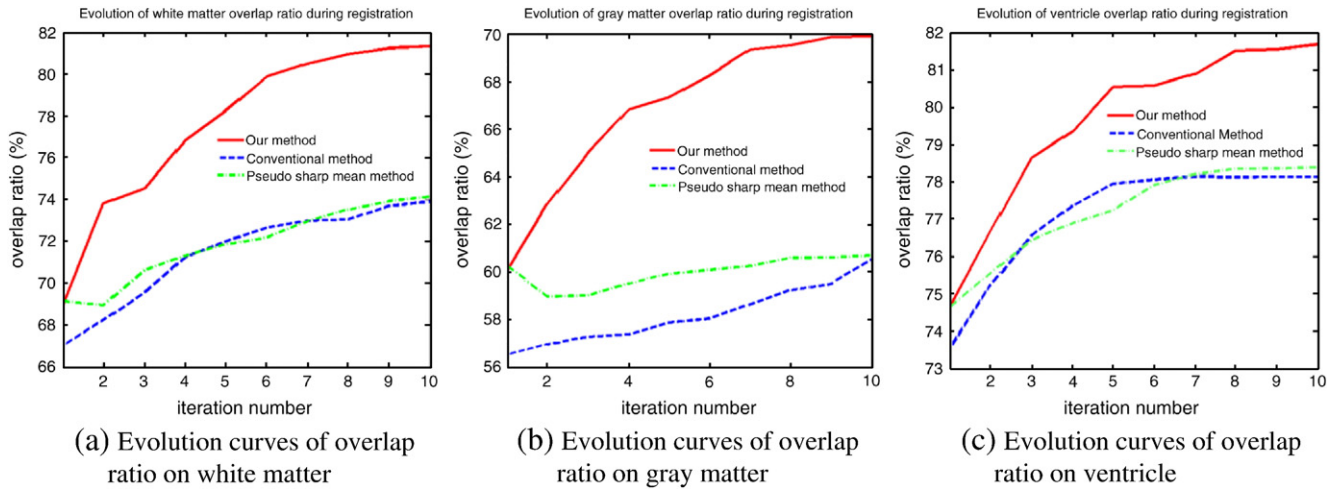(c) Evolution curves of overlap ratio on ventricle

Fig. 10. The evolution curves on tissue overlap ratio during the registration. From left to right shows the evolution curve of overlap ratio on white matter, gray matter, and ventricle, respectively. In each figure, the evolution curves by the conventional method, pseudo sharp mean method, and our method are displayed in blue, green, and red, respectively.

on each ROI is not significant after performing t-test. However, as we will show below, our method is faster; it uses only around 1/3 computation time of ABSORB.

Computation time

We perform all of the above experiments on Dell workstation (with 8 Xeon CPU@2.66 GHz and 32 G DDR memory). The computation time of the tree-based method, the conventional method, ABSORB, and our method is provided in Table 2. It is worth noting that we set the same

number of iterations (10 rounds) for the conventional method and our method. For ABSORB, we choose its optimal parameters. Our method is faster than ABSORB, but slightly slower than the conventional method, and much slower than the tree-based method since it performs only 1 round of registration for the whole group.

Discussion

MST-based strategy is used in each round of our groupwise registration method to improve the robustness of registration results.

✤ Our method achieves significant improvement over the conventional method

✱ Our method achieves significant improvement over the ABSORB method
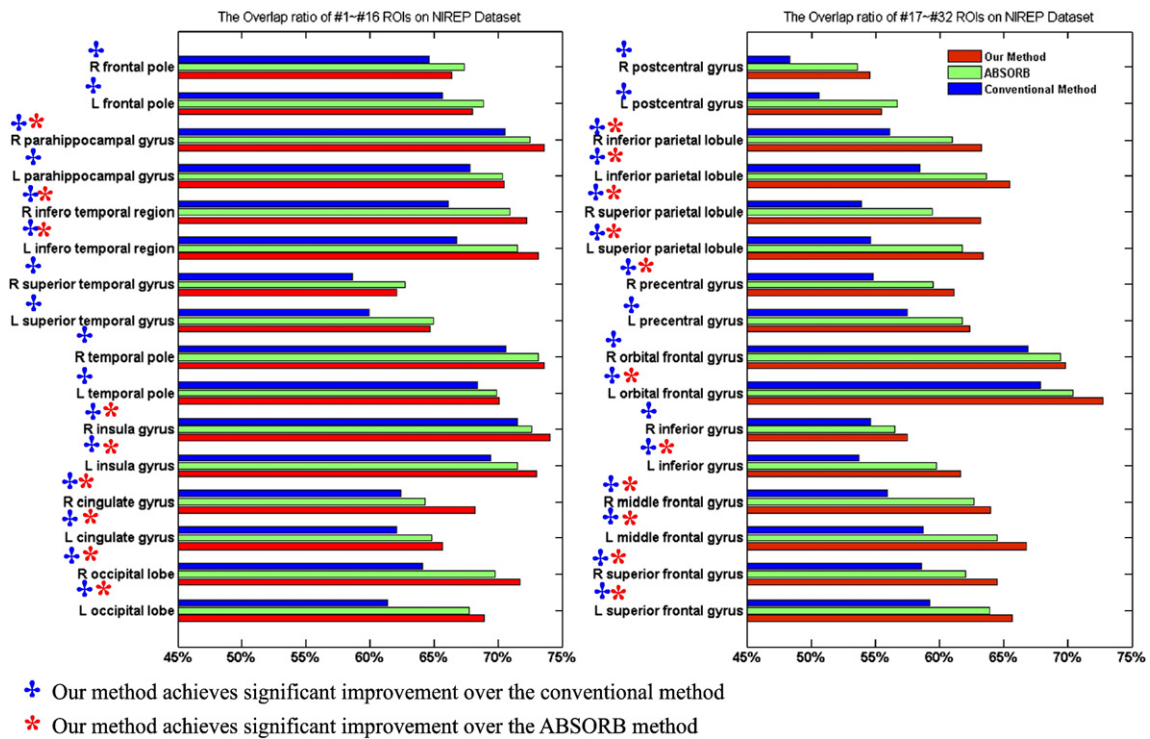
Fig. 11. The overlap ratios of 32 manually delineated labels in 16 NIREP brain images. The overlap ratios by the conventional groupwise registration method, ABSORB, and our groupwise registration method are shown in blue, green, and red, respectively. It can be observed that our method achieves much better alignment result in all ROI regions than the conventional method. Compared with ABSORB, our registration results outperform in most ROIs. The overall overlap ratio is 61.25% by the conventional method, 65.31% by ABSORB, and 66.66% by our method. For each ROI, we use blue "†" and red "*" to denote the significant improvement of overlap ratio (with p-value less than 0.05) by our method over the conventional method and ABSORB, respectively.
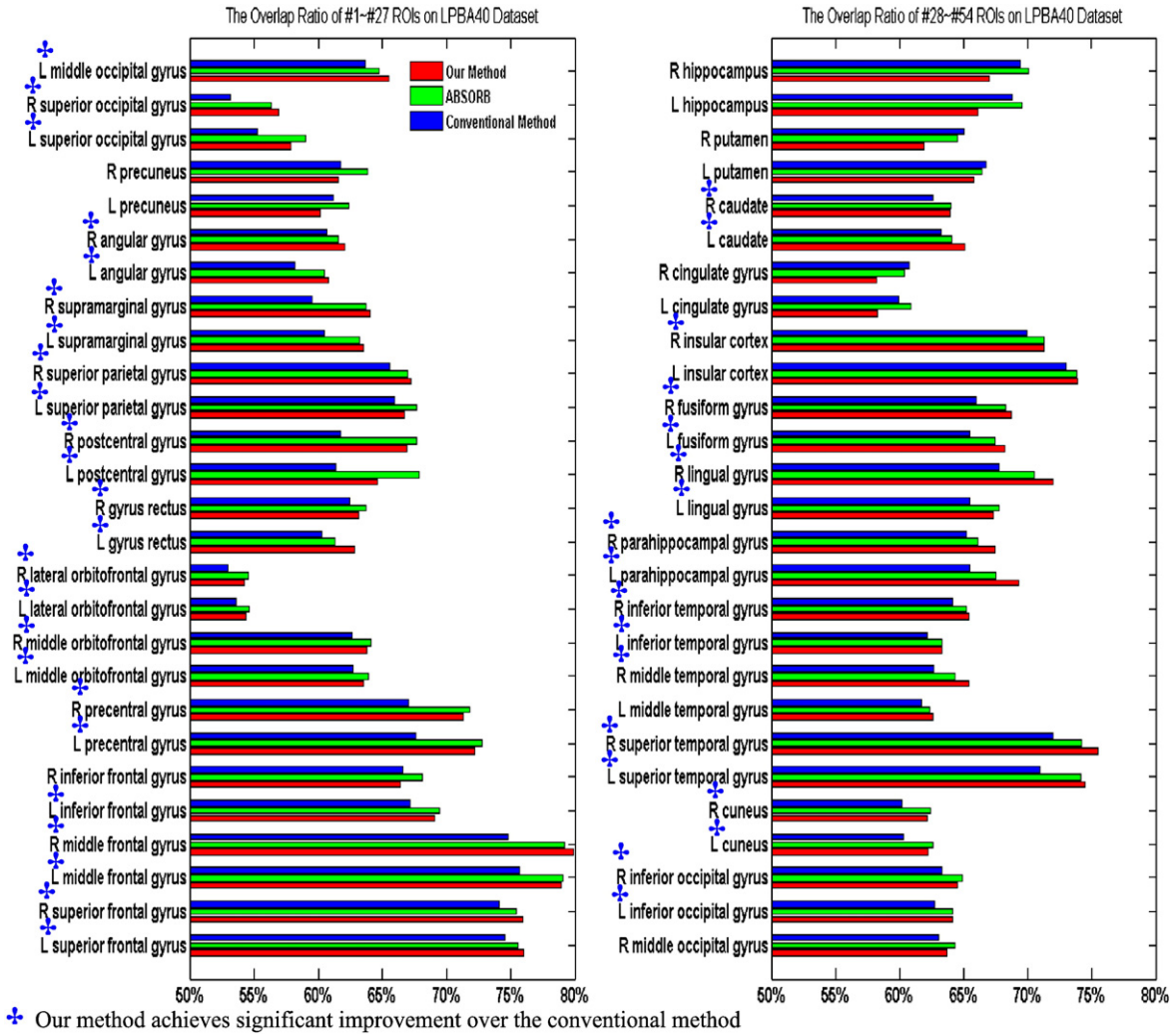
Fig. 12. The overlap ratios of 54 manually delineated labels in 40 LONI LPBA40 dataset. The overlap ratios by the conventional groupwise registration method, ABSORB, and our groupwise registration method on each ROI are shown in blue, green, and red, respectively. For each ROI, we use blue "†" to denote the significant improvement of overlap ratio (with p-value less than 0.05) by our method over the conventional method.

In order to evaluate the contribution of MST in our method, we apply our groupwise registration method, with and without MST (by keeping all other same parameters), to 18 elderly brains, NIREP data set, and LONI LPBA 40 data set, respectively. Table 3 shows their overall average overlap ratio and the standard deviation. Obviously our groupwise registration method with MST consistently achieves better registration accuracy than the counterpart without MST, demonstrating the advantage of employing MST in our complete method.

Particularly, we notice that the improvement on registration accuracy is much more in the data set of 18 elderly brains than the

other two data sets (NIREP and LONI LBPA40). One explanation is that the anatomical shape variations in 18 elderly brains (see in Fig. 6) are much larger than those in NIREP and LBPA40 data sets. Our method can fully take the advantage of MST to hierarchically register each subject to the group mean image by compositing multiple segments of transformation fields from the underlying node (individual subject) to the root of the tree (group mean image). The minimal spanning tree used in the first round of our groupwise registration is shown in Fig. 13. Obviously each pair of neighboring nodes in the tree has similar anatomical shape. In brief summary, this experiment strongly demonstrates the importance of using MST-based registration in each round of groupwise registration, especially in case of large anatomical variations in the data set.

Table 2
Computation time of the tree-based method, the conventional method, ABSORB, and our method.

Dataset	Num.	Image size	Tree-based method	Conventional method	ABSORB	Our method
Elderly brains	18	256×256×124	2.5 h	4.0 h	6.0 h	5.2 h
NIREP	16	256×300×256	4 h	9.4 h	39.0 h	11.1 h
LONI LPBA40	40	181×271×181	4.5 h	20.5 h	60.0 h	22.6 h

Computation environment: 8 Xeon CPU @2.66 GHz, 32 G DDR memory.

Table 3
The overall overlap ratio and standard deviation by our method with and without MST in three data sets.

Method	18 elderly brains	NIREP	LONI LPBA40
Our method with MST	78.07% (±2.16%)	66.66% (±3.64%)	70.36% (±3.73%)
Our method without MST	75.28% (±2.38%)	66.05% (±3.34%)	69.15% (±4.26%)
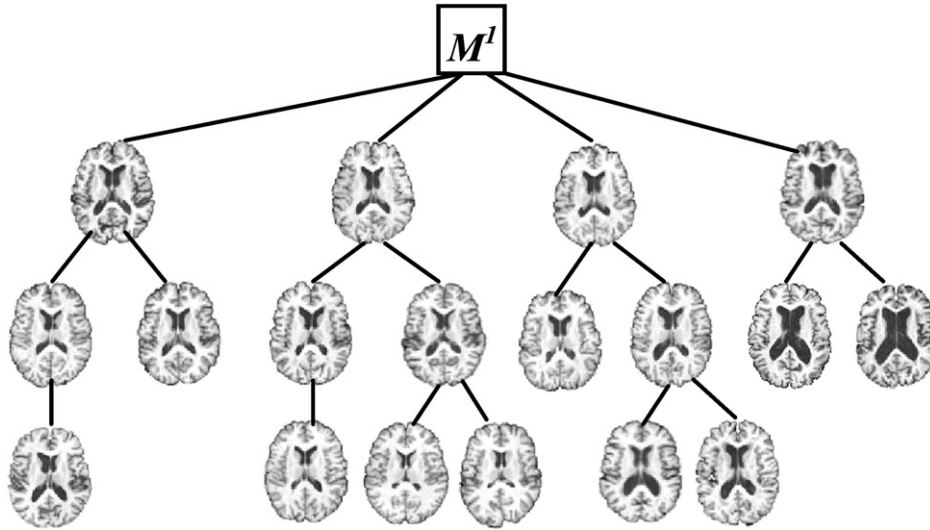
Fig. 13. The minimal spanning tree (MST) built for the 18 elderly brains in the first iteration ($t=1$). Each node in the MST is corresponding with a subject image shown in Fig. 6.

Conclusion

We have demonstrated the importance of always keeping a sharp group mean image during the entire groupwise registration procedure. In order to improve the overall registration performance of the whole population, we generalize a popular unbiased groupwise registration method in this paper. In our new framework, different subjects are adaptively weighed to construct the sharp group mean image, according to their similarities to the previous-estimated group mean image. Also, we utilize a tree-based registration to improve the registration quality in each round of groupwise registration. Based on these two innovative formulations, our groupwise registration method has achieved much better results than other state-of-the-art groupwise registration methods, and the tree-based registration method, using simulated and real data.

In the current method, we only use the intensity-based registration method, i.e., Diffeomorphic Demons (Vercauteren et al., 2009), as a pairwise registration component in each round of groupwise registration. However, minimizing the intensity difference does not necessarily mean the good anatomical correspondences across subjects. In our future work, we will first integrate other feature-based registration methods, e.g., HAMMER in (Shen, 2007), into our groupwise registration framework. Second, since it is very important to process the large population data efficiently in many clinical applications, we will improve the overall registration speed by optimizing the program and using the learning-based method for fast predication of deformation field as we did for our pairwise registration method (Kim et al., 2010). Finally, we will also apply our groupwise registration method to various neuroscience studies, such as atlas building for early human brain development (Kazemi et al., 2007) and Alzheimer's Disease study using ADNI data (ADNI, 2004).

Appendix A. Estimation of group mean image

After calculating the weight $\widehat{\omega}_s^t(x)$ for each subject s at each location x and removing all un-necessary terms with M^t, the objective function F_1 in SP_1 becomes:

$$F_1\left(M^t\right) = \sum_{s=1}^N \sum_x \left\{ \widehat{\omega}_s^t(x) \cdot D\left(I_s^{t-1}, M^t, x, b\right) \right\}$$
$$= \sum_{s=1}^N \sum_x \left\{ \widehat{\omega}_s^t(x) \cdot \sum_{y \in P_b(x)} \|I_s^{t-1}(y) - M^t(y)\|^2 \right\}.$$

Here x and y denote the arbitrary spatial location in the common space. P_b denotes a local (spherical) image patch, thus $\forall x \in P_b(y) \Rightarrow y \in P_b(x)$. D is the overall similarity in the local image patch. The estimation of M^t can be calculated by requiring the derivative $\frac{\partial F_1}{\partial M^t(y)} = -2 \cdot \sum_{s=1}^N \sum_{x \in P_b(y)} \widehat{\omega}_s^t(x) \cdot \left(I_s^{t-1}(y) - M^t(y)\right) = 0$. As the result, we get:

$$\sum_{s=1}^N \sum_{x \in P_b(y)} \widehat{\omega}_s^t(x) \cdot I_s^{t-1}(y) = \sum_{x \in P_b(y)} \left[\sum_{s=1}^N \widehat{\omega}_s^t(x) \right] M^t(y)$$
$$\Rightarrow \sum_{s=1}^N \sum_{x \in P_b(y)} \widehat{\omega}_s^t(x) \cdot I_s^{t-1}(y) = \sum_{x \in P_b(y)} M^t(y) \ \left(since \ \sum_{s=1}^N \widehat{\omega}_s^t(x) = 1 \right)$$
$$\Rightarrow \sum_{s=1}^N \sum_{x \in P_b(y)} \widehat{\omega}_s^t(x) \cdot I_s^{t-1}(y) = |P_b(y)| \cdot M^t(y)$$

($|P_b(y)|$ is the total number of points in the neighborhood $P_b(y)$)

$$\Rightarrow M(y) = \frac{1}{|P_b(y)|} \cdot \sum_{s=1}^N \sum_{x \in P_b(y)} \widehat{\omega}_s^t(x) \cdot I_s^{t-1}(y)$$
$$= \sum_{s=1}^N \left[\frac{1}{|P_b(y)|} \sum_{x \in P_b(y)} \widehat{\omega}_s^t(x) \right] I_s^{t-1}(y).$$

By swapping the variable x and y, we obtain the Eq. (9) in the paper to compute the mean image at each point x.

Appendix B. The objective function in step 2

Recall that $I_s^t = I_s^0(g_s^t(x))$ is the deformed I_s^0 w.r.t. the transformation field g_s^t. Then, after determining the group mean image $\widehat{M}^t$ and weights $\widehat{\omega}_s^t(x)$, the overall objective function in Eq. (4) turns to be the function of transformation fields G:

$$F_2(G) = \sum_{s=1}^N \left\{ \sum_x \widehat{\omega}_s^t \cdot D\left(I_s^t, \widehat{M}^t, x, b\right) + \frac{1}{\sigma_g} dist\left(e, g_s^t\right) \right\}$$
$$= \sum_{s=1}^N \left\{ \sum_x \widehat{\omega}_s^t(x) \cdot \sum_{y \in P_b(x)} \|I_s^t(y) - \widehat{M}^t(y)\|^2 + \frac{1}{\sigma_g} dist\left(e, g_s^t\right) \right\}$$
$$= \sum_{s=1}^N \left\{ \sum_x \left[\sum_{y \in P_b(x)} \widehat{\omega}_s^t(y) \right] \cdot \|I_s^t(x) - \widehat{M}^t(x)\|^2 + \frac{1}{\sigma_g} dist\left(e, g_s^t\right) \right\}$$
$$= \sum_{s=1}^N \left\{ \sum_x \frac{1}{\sigma_s^t(x)} \cdot \|I_s^t(x) - \widehat{M}^t(x)\|^2 + \frac{1}{\sigma_g} dist\left(e, g_s^t\right) \right\}.$$