

Groupwise Registration Based on Hierarchical Image Clustering and Atlas Synthesis

Qian Wang,^{1,2} Liya Chen,³ Pew-Thian Yap,² Guorong Wu,²
and Dinggang Shen^{2*}

¹Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

²Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

³Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China

Abstract: Groupwise registration has recently been proposed for simultaneous and consistent registration of all images in a group. Since many deformation parameters need to be optimized for each image under registration, the number of images that can be effectively handled by conventional groupwise registration methods is limited. Moreover, the robustness of registration is at stake due to significant intersubject variability. To overcome these problems, we present a groupwise registration framework, which is based on a hierarchical image clustering and atlas synthesis strategy. The basic idea is to decompose a large-scale groupwise registration problem into a series of small-scale problems, each of which is relatively easy to solve using a general computer. In particular, we employ a method called affinity propagation, which is designed for fast and robust clustering, to hierarchically cluster images into a pyramid of classes. Intra-class registration is then performed to register all images within individual classes, resulting in a representative center image for each class. These center images of different classes are further registered, from the bottom to the top in the pyramid. Once the registration reaches the summit of the pyramid, a single center image, or an atlas, is synthesized. Utilizing this strategy, we can efficiently and effectively register a large image group, construct their atlas, and, at the same time, establish shape correspondences between each image and the atlas. We have evaluated our framework using real and simulated data, and the results indicate that our framework achieves better robustness and registration accuracy compared to conventional methods. *Hum Brain Mapp* 31:1128–1140, 2010. © 2010 Wiley-Liss, Inc.

Key words: groupwise registration; image clustering; hierarchical registration

INTRODUCTION

Medical image registration has been an important topic in research and clinical studies for decades, due to its wide spread applications in aiding alignment and comparison of longitudinal and cross-sectional data, thus facilitating diagnosis, guiding treatments, and monitoring disease progression [Woods et al., 1992; Hajnal et al., 1995; Maintz and Viergever, 1998; Holden et al., 2000; Hill et al., 2001; Zitová and Flusser, 2003]. For example, to monitor tumor growth using the patient images obtained at different time

*Correspondence to: Dinggang Shen, MRI Building, 106 Mason Farm Road, Chapel Hill, NC 27599-7515. E-mail: dgshen@med.unc.edu

Received for publication 7 May 2009; Revised 8 September 2009; Accepted 9 September 2009

DOI: 10.1002/hbm.20923

Published online 8 January 2010 in Wiley InterScience (www.interscience.wiley.com).

points, we can first use a mechanical model to simulate the tumor growth and then use a registration method to estimate the remaining deformations as we did in Mohamed et al. [2006]. Besides techniques for single-modality image registration, multimodality registration techniques are required to fuse information across modalities for better insight and understanding of the anatomy and functions of a specific organ [Hill et al., 1991; Meltzer et al., 1990]. To this end, methods using information theoretic metrics, such as mutual information (MI) and normalized mutual information (NMI), both of which are capable of measuring multimodality similarity, have been developed and successfully applied [Luan et al., 2008; Maes et al., 1997; Pluim et al., 2003; Studholme et al., 1999; Wells et al., 1996]. For accurate registration, transformations with higher degrees of freedom (DOF) than those of rigid or affine transformations are crucial to ensure proper capturing of the subtle changes in the anatomical structures. Such transformations can be represented by a linear combination of polynomial terms [Woods et al., 1998a,b], basis functions [Friston et al., 1995], or B-Splines [Rueckert et al., 1999].

Although many techniques have been proposed for medical image registration, most of them in nature belong to the category of pairwise registration, where a floating image is mapped to the space of a fixed image. When pairwise methods are directly applied to register a group of images, one image from the group needs to be selected as a reference or template, and this may introduce bias to the subsequent image analysis. With advances in imaging and storage technologies, implementation of more efficient techniques involving registration of more than a pair of images is now more feasible. To this end, groupwise registration has been recently proposed to match correspondences and register all images in a population simultaneously [Crum et al., 2004; Toga and Thompson, 2001]. Guimond et al. [2000] proposes to model a group of images in terms of the average intensity and shape image. Pairwise registration is employed to deform all images to a reference. The average intensity and shape image is then produced by warping the average of those registered images via the mean of the deformations linking the reference to all individual images. By iteratively mapping images to their average in a multiresolution fashion, an atlas, which is the average of both intensity and geometry, can be constructed to depict the group mean, while the variation of all individual images can be captured by their estimated deformation fields to the group mean [Kovacic et al., 2004, 2005; Woods, 2003]. Seghers et al. [2004] performed pairwise registration between all possible pairs of images in the group, and constructs the atlas by voxelwise averaging of all images after mapping them to their mean morphological images. Each mean morphological image is determined by the average of transformations from a particular image to all other images in the group. Park et al. [2005] defined an image closest to the population mean geometry as a tentative template and generates

the atlas by iteratively registering all images onto the template and replacing the template with the mean of the aligned images. The mean geometry is determined with the aid of information of relative locations provided by multidimensional scaling (MDS). These methods succeed in achieving the goal of groupwise registration; however, they suffer from high computational complexity, especially when the size of the image dataset increases.

More efficient groupwise registration methods have subsequently been proposed. Joshi et al. [2004] proposed a method for atlas estimation in a large deformation diffeomorphic setting. In Zöllei et al. [2005], a gradient-based stochastic optimizer is employed to minimize an information-theoretic objective function, and an affine congealing [Learned-Miller, 2006; Miller et al., 2000] mechanism is used to drive each image to the center of the group simultaneously. This scheme does not call for any arbitrary selection of template and eventually produces an unbiased atlas. Balci et al. [2007] later used the framework used in Zöllei et al. [2005] and extended it to a nonrigid groupwise registration algorithm by incorporating free-form B-Splines to represent nonrigid deformations. For convenience, we will use the term “congealing” referring to the method reported in Balci et al. [2007] in the rest of the article.

Recently, it has been reported that a single mode is not sufficient to account for the variation of all images in a population [Allasonnière et al., 2007; Blezek and Miller, 2007; Sabuncu et al., 2008]. Moreover, independently estimating correspondences between each individual image and the atlas is not always the most efficient approach. Some images are in fact very similar and spatially close prior to groupwise registration, and it is more efficient to leverage their similarity and warp them to the atlas in subgroups rather than individually. A multiclass approach as such can be expected to be more robust and will result in more accurate deformation estimations, with, in the end, a more refined atlas.

In this article, we propose a novel groupwise registration framework based on a hierarchical image clustering and atlas synthesis strategy. In our framework, images are hierarchically clustered into several classes using affinity propagation [Frey and Dueck, 2007]. This clustering process will be iterated until the sizes of all child classes are under a certain threshold. After the pyramid of image classes has been established in a top-down fashion, intraclass groupwise registration will be performed bottom-up in the pyramid. In each class, images can be registered using any existing groupwise technique, producing the respective class center image, which will be then registered at higher levels in the pyramid. When the intraclass registration reaches the summit of the pyramid, the atlas of the whole image dataset will eventually be synthesized. At the same time, all images in the group will be registered onto the atlas based on their respective transformation routes. In the following, we will describe our proposed groupwise registration framework in more detail.

METHODS

A hierarchical groupwise registration framework adopting a divide-and-conquer strategy is described below. We decompose a large-scale groupwise registration problem into a series of independent small-scale problems. The small-scale problems are independently solved and their results are hierarchically synthesized, leading to a complete groupwise registration. In the section “Image Clustering,” we will discuss how to decompose a large-scale registration problem by an image clustering technique. In the section “Atlas Synthesis,” we will introduce how to produce a group mean during the registration procedure.

Image Clustering

The images in a group are clustered into different classes, each containing only images that are spatially close to each other in their intrinsic high-dimensional space. Large classes are further clustered into several smaller subclasses, and finally a pyramid of classes is constructed. Groupwise registration will be performed classwise from bottom to top of the pyramid. When intraclass registration reaches the summit of the pyramid, a single atlas is eventually synthesized.

Prior to image clustering, we spatially normalize all images by performing a coarse groupwise registration to remove some of the confounding factors that might interfere with subsequent processes. This normalization phase aligns all images using low DOF transformation so that the clustering algorithm can more accurately capture the image spatial distribution. Because of processing speed concern, we perform only a very coarse groupwise registration. Specifically, the method reported in Zöllei et al. [2005] is used for linear alignment, while the nonlinear part is handled by the congealing method [Balci et al., 2007], where the density of B-Spline control points (corresponding to the DOF of transformations) is set very low (four control points in each dimension). The transformation for each image estimated in this phase is stored and will be utilized later when establishing correspondences between images and the synthesized atlas.

A common approach for clustering is to locate a set of class centers, which could minimize certain distance measure such as the sum of Euclidean distances between the class centers and their constituent subjects. The conventional k -means clustering technique [MacQueen, 1967], for instance, iteratively updates assignments of the subjects to k classes based on the specific distance metric and calculates the mean of the subjects in each class as tentative center. Nevertheless, this approach is quite sensitive to the initial condition, where k subjects are randomly selected as the initial class centers. Sometimes, no decent solution could be obtained even if a huge number of trials with different initial conditions are used. Furthermore, even some seemingly reasonable solutions could introduce unexpected bias to the clustering results and hence lead to an

incorrect representation of the true distribution of the subject images [Jain et al., 1999].

In this study, we employ a recently proposed method called *affinity propagation* [Frey and Dueck, 2007] for fast and robust clustering. For each detected class by affinity propagation, a single subject image is determined as the exemplar to represent the whole class. In contrast to conventional clustering techniques, affinity propagation considers all images as potential exemplars without the need of arbitrary preselection. By relating each image in the group with a node in a network/graph, affinity propagation tries to locate an optimal set of exemplars and their corresponding classes via a message-passing mechanism along the edges between each pair of nodes. The message-passing mechanism updates the “responsibility” and “availability” of each node by exchanging messages between the nodes, and the two attributes will finally determine the likelihood of a certain node in becoming an exemplar. To initialize the responsibility and availability values, affinity propagation takes as inputs real-valued similarities between each pair of input images. The similarity measure $s(j,k)$ between images I_j and I_k indicates how close spatially the two images are. A higher $s(j,k)$ value implies that, if image I_j is determined as an exemplar image, the other image I_k will have more tendencies to stay within the class represented by I_j . On the contrary, if the value of $s(j,k)$ is lower, I_k will have fewer chances of being classified as belonging to the class represented by I_j . Obviously, $s(j,k)$ can be defined in a variety of ways, though it is commonly defined as the negative Euclidean distance, i.e., $s(j,k) = -\|I_j - I_k\|^2$.

In our case, we have adopted mutual information (MI) [Maes et al., 1997; Pluim et al., 2003; Wells et al., 1996] as the similarity measure. It is worth noting that other similarity measures, such as least square error (LSE) and normalized mutual information (NMI) [Studholme et al., 1999], can also be used. NMI, for example, may better reflect the geometrical distance than MI, especially in the case where the overlapping regions of two images are relatively tiny. However, since all images in our following experiments are of the same modality and their spatial homogeneity is enforced by the normalization phase prior to image clustering, MI performs sufficiently well in measuring the similarity for any pairwise combination of images.

Suppose that there are n images, $n(n-1)/2$ pairwise MI values need to be provided for clustering using affinity propagation. Therefore, it is crucial that the calculation of MI be sufficiently fast so that computation time of the image clustering stage can be kept at a competent level. To this end, several approaches are implemented:

- A simple but effective way is to lower the sampling rate of the number of voxels contributing to the calculation of individual histograms as well as the joint histogram. In our case, MI is estimated from only 25–50% of the sampled voxels used in regular MI estimation.

- Parzen windowing is commonly used to interpolate the (joint) histograms for more precise estimation of intensity probabilities and entropies [Mattes et al., 1993; Viola and Wells, 1997]. To speed up computation, this step is removed, and the probability for any intensity (pair) is extracted directly from a single bin in the discrete (joint) histogram.
- The time-consuming calculation of logarithms in entropy estimation is circumvented by computing in advance a large logarithm lookup table (LUT), which is a reasonable approach since all possible probabilities take a finite set of discrete values.

Although this simplified estimation of MI is relatively coarse and might potentially influence the clustering results, one should keep in mind that the ultimate objective of image clustering in our framework is to aid the decomposition of a complicated registration problem into a number of small relatively easy-to-solve problems, rather than to accurately determine the exact image spatial distribution. In fact, we found that the speedup approaches delineated above do not have significant effects on the clustering results as well as the final constructed atlas.

In affinity propagation, the “preference” value, or the “self-similarity” of each node, determines the likelihood of a particular image becoming an exemplar and needs to be specified before clustering. The number of estimated exemplars, which equals the number of classes, is strongly dependent on the “preference” values [Frey and Dueck, 2007]. Since no single image is more likely than others in becoming an exemplar, the “preferences” for all images are set to a common value, i.e., the median of all pairwise MI similarity values.

After a single round of clustering, there might still be too many images in certain classes, exceeding the capacity that can be efficiently handled by a particular groupwise registration algorithm. Clustering is hence performed again within any class, which is still too large, to divide it further into smaller subclasses. Extra bookkeeping is needed to ensure that the number of child classes belonging to the same parent class should not be too large. Otherwise, it will be challenging and time-consuming to align all the center images of the child classes and to estimate the center image representing the parent class itself. For this purpose, a threshold based on our experience is selected, and the size of any class at any time is restricted to be no larger than this threshold so that intraclass registration can be performed smoothly. The size of each class should be as close as possible, to avoid the redundant levels inside the pyramid and also to keep the number of classes as small as possible. Since all center images need to be further registered at subsequent higher levels in the pyramid, we expect that by keeping the class sizes more uniformly distributed, the center images of all classes will have similar image quality and thus a good performance of registration on these center images can be achieved. According to this design, our image clustering procedure can produce a pyramid of image classes as illustrated in Figure 1.

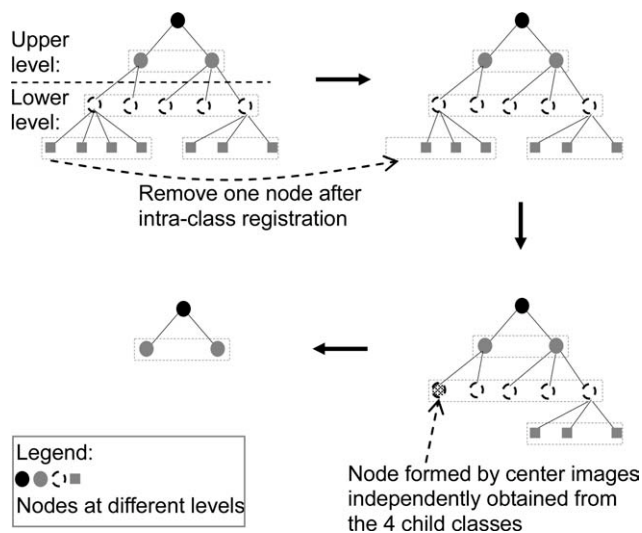


Figure 1.

Illustration of a hierarchical pyramid formed by classes of images. Each node in the pyramid denotes an identified class containing a number of images. The edge between two nodes at a lower level and an upper level indicates that the class at the lower level belongs to the class at the upper level. The progression of the bottom-up groupwise registration scheme is illustrated by the arrows.

Atlas Synthesis

Based on the image pyramid established through the image clustering stage, we gradually perform intraclass groupwise registration in a bottom-up fashion and build an atlas space for the whole image dataset hierarchically. Images in each class can be groupwise registered via any existing method to produce the center image. Center images from different classes belonging to the same parent node are then groupwise registered to form the center image for the parent node higher up in the pyramid. Upon reaching the summit of the pyramid, an atlas for the whole dataset can eventually be synthesized. We have selected the congealing method [Balci et al., 2007] to perform intraclass registration. It is an open-source algorithm built upon ITK (Insight Toolkit) and thus makes evaluation of our framework more convenient and fair. In the congealing method, voxels sampled from the same location of different images form a stack, and the sum of intensity entropies from various stacks forms the stack entropy. Stack entropy serves as the objective function and is minimized by a gradient-based optimizer, which simultaneously estimates individual transformations for different images and finally warps all images to the atlas space. A multiresolution registration strategy is also deployed in the congealing method. Utilizing deformation fields regulated by different densities of B-Spline control points, the multiresolution approach could mitigate influences of local minima and increase the robustness of registration.

For each individual node at the lowermost level of the multilevel pyramid, intraclass groupwise registration is performed and a center image is produced by averaging all the registered images. The node is then temporarily removed from the pyramid, and the center image representing this class is added as a new member image to its corresponding parent node (cf. Fig. 1). By repeating the intraclass registration and the node removal steps, the pyramid can gradually be depleted, and the atlas will eventually be synthesized when the summit of the pyramid is reached. The key ideas are summarized in Figure 1, with the atlas-synthesizing process indicated by the solid arrows.

In any node other than those at the lowermost level, the synthesis process is more challenged, since the constituent members inevitably consist of blurred center images from its child nodes. We illustrate this in Figure 2a, where we want to construct the atlas from two image classes, each of which contains five subjects. Intraclass registration is performed independently to the two classes, and it is not surprising that both center images, shown in the right panels of the figure, are relatively blurred. The quality of the atlas, synthesized from the center images, is hence compromised, and structural details are lost. Especially when the DOF of the transformation is high, the optimized solution will become unreasonable if it is estimated upon blurred anatomical information. As a remedy to this potential problem, we introduce a refinement phase to the atlas synthesis stage, when registration is running at high DOF.

In refinement, we modify the objective function used in the congealing method to incorporate information not only from the to-be-registered center images representing child classes but also information from the original images in the population. To achieve this, we first note that, as the registration progresses up the pyramid, each image is progressively deformed following edges connecting nodes inside the pyramid. Therefore, for each center image, we essentially have a number of corresponding images deformed to the space defined by the center image, which we call support images for easy reference. Since support images generally retain much more structural information than their center images, we exploit the information they contain for stack entropy estimation. Stack entropy is now formulated using the support images of each center image with respect to the other center images. Given the intensity $v_i(x)$ at location x of the i th center image and $v_{i,m}(x)$ the intensity of the m th support image for the i th center image, the stack entropy f_H in the refinement phase can be written as

$$\left\{ \begin{array}{l} f_H \propto \sum_x \sum_{i=1}^N \log \frac{1}{N} \sum_{j=1}^N d(x; i, j) \\ d(x; i, j) = \frac{1}{S_i \sigma} \sum_{m=1}^{S_i} G_\sigma(\|v_{i,m}(T_i(x)) - v_j(T_j(x))\|) \end{array} \right. \quad (1)$$

where N is the number of to-be-registered center images, S_i the number of support images for the i th center image,

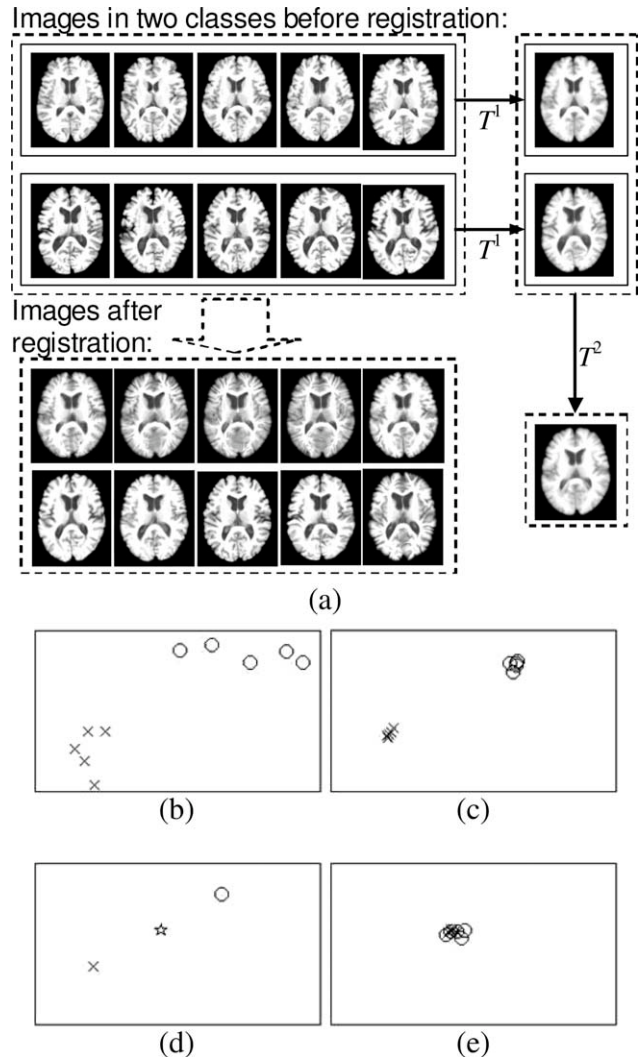


Figure 2.

Illustration of atlas synthesis. (a) Ten images are clustered into two classes, and these images are groupwise registered via our proposed framework. (b) Distributions of the 10 images before registration, where for visualization purpose ISOMAP [Tenenbaum et al., 2000] is used to map the 10 images into a low-dimensional space and preserving their relative distances. The two classes are represented by circles and crosses, respectively. (c) Distributions of the 10 images after independent intraclass registration. (d) Center images of the two classes are registered to the final constructed atlas as indicated by the star. (e) Distributions of the 10 images after being mapped into the space of the atlas, using their respective concatenated deformation fields.

and $G_\sigma(\cdot)$ a Gaussian kernel of variance σ^2 used for entropy estimation. By optimizing f_H , the transformations $\{T_i(\cdot), i = 1, \dots, N\}$ for the center images to their parent class center can be estimated. Note that we only need to

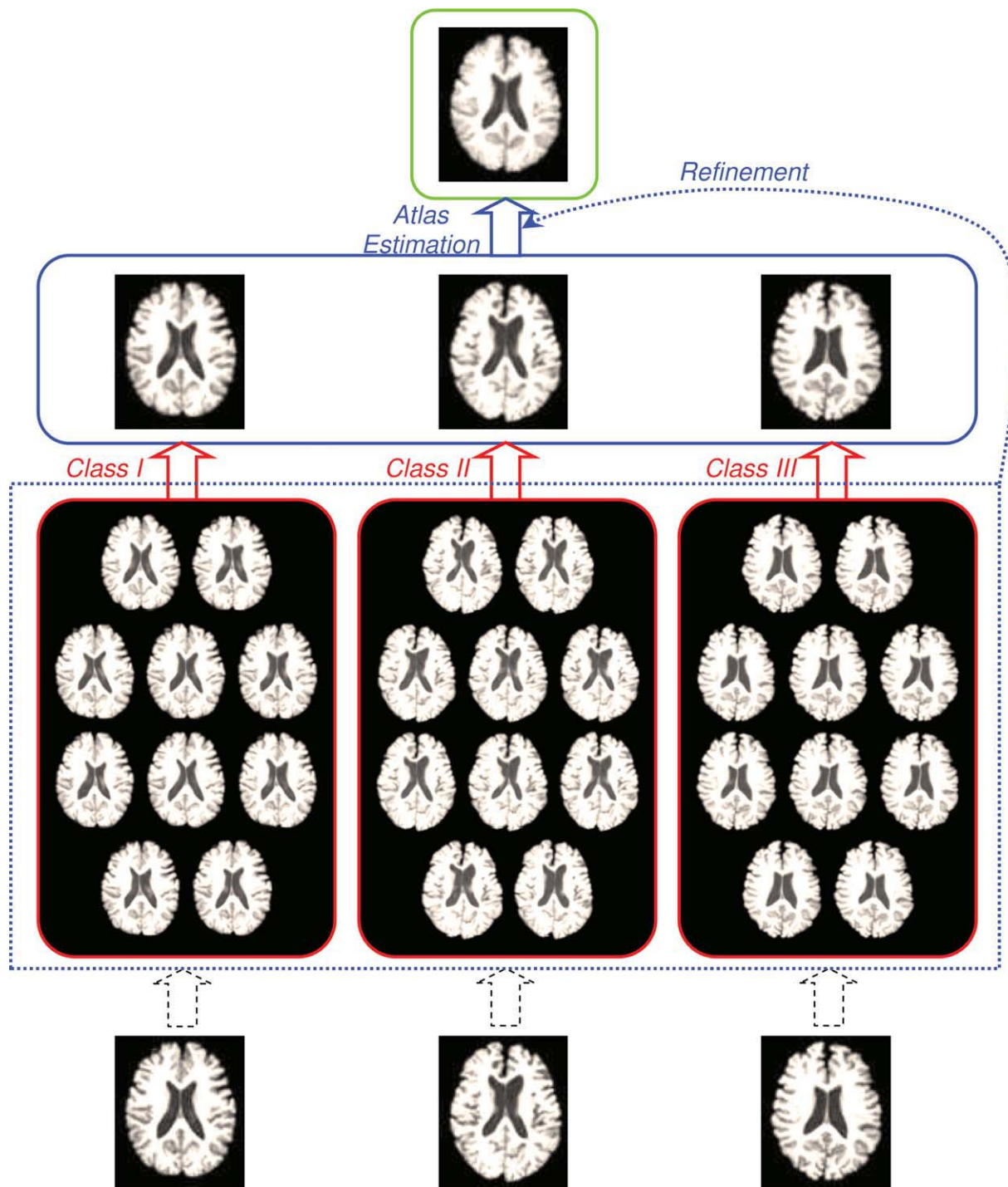


Figure 3.

Each of the three very similar brain images in the bottom is used to generate 10 simulated images, which are shown within the red boxes, by applying zero-mean random deformations. Intra-class groupwise registration is performed for each class (red arrows) resulting in three center images (blue box). These

three images are further groupwise registered (solid blue arrow) and refined (dashed blue arrow) to synthesize the final atlas (green box) for the 30 images. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

TABLE I. Comparison of computation time costs

	Time cost (s)		
	Congealing framework	RG framework	Proposed framework
Image clustering (inclusive of normalization phase)	—	—	201
Atlas synthesis			
Class/Group I	—	931	862
Class/Group II	—	924	851
Class/Group III	—	922	858
Atlas estimation	4,148	285	301
Overall	4,148	3,062	3,073

Images and computation environment: 30 images ($83 \times 98 \times 75$ voxels); Intel Q6600 (4 Cores, 2.4 GHz) CPU.

estimate N rather than $\sum_i S_i$ transformations. The refinement phase will hence not greatly increase the computation cost of the whole registration process. At the same time, the robustness and accuracy of the registration can be expected to improve, since images with more structural details are utilized in the optimization.

For a particular image in the population, the transformation between itself and the final atlas can be estimated by concatenating the transformations involved in different paths of the route traversed by the image in reaching the summit of the pyramid. As an illustration, the two classes shown in Figure 2a are groupwise registered, yielding two different center images through transformations $\{T_{ji}^1 | j = 1, 2; i = 1, \dots, 5\}$, where j denotes a class, i an image in the class, and T^1 the transformations of the images at the bottom level of the pyramid to their center images. The two centers are further groupwise registered, with refinement, to form the final atlas through transformations $\{T_j^2 | j = 1, 2\}$. Concatenating the individual overall transformations $T_j^2 \circ T_{ji}^1 \circ T_{ji}^0$, where T_{ji}^0 indicates the transformation produced in the normalization phase prior to clustering, each image can be mapped from its original space into the atlas space, as demonstrated in Figure 2a.

To provide better insight and illustrate the dynamics of the proposed groupwise registration framework, we have employed a dimensionality reduction method called “ISOMAP” [Tenenbaum et al., 2000] to provide a visualization of the spatial distribution of images before and after registration. Since ISOMAP approximately preserves the relative distance (negative similarity) between any two images, the two-dimensional representation preserves the spatial relationship of the images in their high dimensional space. MI is used here to measure the similarity between any two images. Figure 2b shows the distribution of the original ten images before registration. Two different symbols, i.e., cross and circle, are used to represent the two classes identified in the clustering stage. After these two classes have been groupwise registered, we can observe

from Figure 2c that the distribution of images in each class becomes much tighter. Further, the two center images are groupwise registered, with refinement, to the final atlas, and they are all provided in Figure 2d. The distribution of all the images registered to the estimated atlas space is shown in Figure 2e.

RESULTS

In the following, we will use a series of experiments to evaluate the performance of our proposed groupwise registration framework. In particular, we will demonstrate that the proposed framework performs sufficiently fast and yields improved accuracy for both real and simulated brain images. We will show that our framework has the ability to capture subtle morphological changes in brain images, which is essential for clinical applications, such as identifying subtle neurological disorder-related atrophy.

Running Time

To compare the running time of different registration frameworks, we selected 3 very similar brain images as seeds and separately produced 10 simulated images from each of them by applying a set of zero-mean random deformations. These seed images are acquired on a GE Signa 1.5 Tesla scanner using a high-resolution volumetric spoiled-grass (SPGR) axial series (TR = 35 ms, TE = 5 ms, FOV = 24 cm, flip angle = 45° , matrix = 256×256 , NEX = 1, voxel dimensions $0.9375 \times 0.9375 \times 1.5 \text{ mm}^3$ slice thickness). The random deformation is regulated by B-Spline transformation model, in which 11 control points along each axis are used. The number of control points here is a prime number and is different from the number of control points used in our hierarchical groupwise registration (where 8, 16, and 32 control points are used in different resolutions). This is designed for making the registration problem difficult, e.g., the simulated deformations cannot be well represented by a deformation model used in the registration method. Furthermore, while the mean shift of all control points is constrained to zero, the standard deviation of their shifts is set to 10 mm so that large brain variations can be simulated. All simulated images are visually inspected to ensure that the deformations are realistic, reflecting biologically plausible anatomical variation.

The three seed brain images are displayed at the bottom of Figure 3, and the corresponding simulated images are shown within the three red boxes in the same figure. Although these 30 simulated images were visually very similar, our image clustering scheme successfully clustered them into three classes, which matched the original classification exactly. After groupwise registration of all images in each of these three classes, three center images were produced. These three center images were groupwise

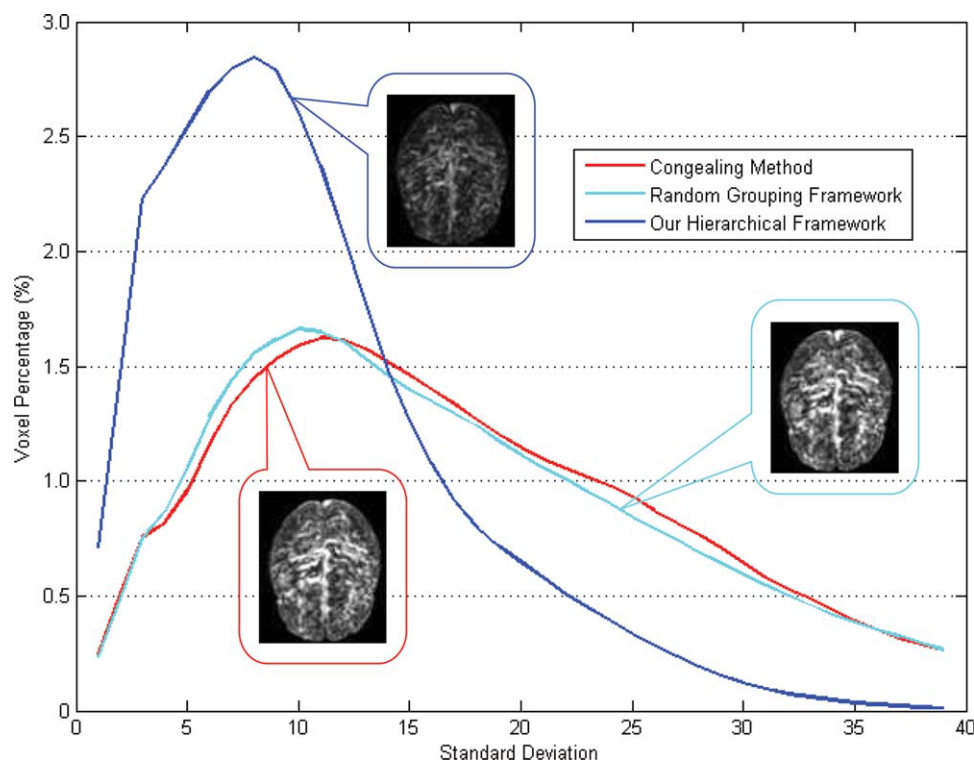


Figure 4.

Histograms of the standard deviation (STD) maps of images registered via different frameworks in comparison. 3D renderings of the STD maps are provided for visual inspection. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

registered with refinement using all 30 registered images as support images to synthesize the final atlas. The process is illustrated by the blue arrows in Figure 3.

For evaluation of running time, we compared our framework with a random grouping (RG) framework, which uses a hierarchical registration scheme similar to ours. However, in the RG framework, all images are randomly assigned into different groups, instead of being clustered based on their spatial distribution. In this case, images in each group would have similar distribution with that of the original dataset. This is significantly different from our framework where images in a class are spatially close and have a distribution that is quite different from that of the whole dataset. For fair comparison, we used an image pyramid with structure (nodes and branches) identical to the clustered result of our framework for the RG framework.

We used the same multiresolution groupwise registration strategy for the congealing framework,¹ the RG framework, as well as our framework, to assure fairness of comparison. The numbers of iterations in different resolutions are automatically determined, to assure reasonable

registration accuracy with no redundant iterations. In Table I, we have recorded the time costs for different processing phases in individual frameworks. The following observations can be obtained:

- Given the same number of images, the processing time is highly dependent on the scale of the registration problem, as well as the complexity of the spatial distribution of the input images. To register images that are more similar costs much less computation time. For example, the intraclass registration for a single class containing 10 images costs only ~860 s via our framework, while the registration to a single group in the RG framework, where images are less similar, costs > 920 s.
- The stage of image clustering, which includes the coarse normalization, incurs additional computation overhead, though it is relatively insignificant. The bulk of the computation happens in the stage of atlas synthesis, where a series of groupwise registration is performed. Overall, the hierarchical strategy proves its ability in decreasing running time, as both the proposed framework and the RG framework lead by a large margin in terms of computation time cost. Our framework only falls a little behind the RG framework, mainly due to the extra time needed in the image clustering stage.

¹Direct application of the congealing algorithm to the whole set of images.

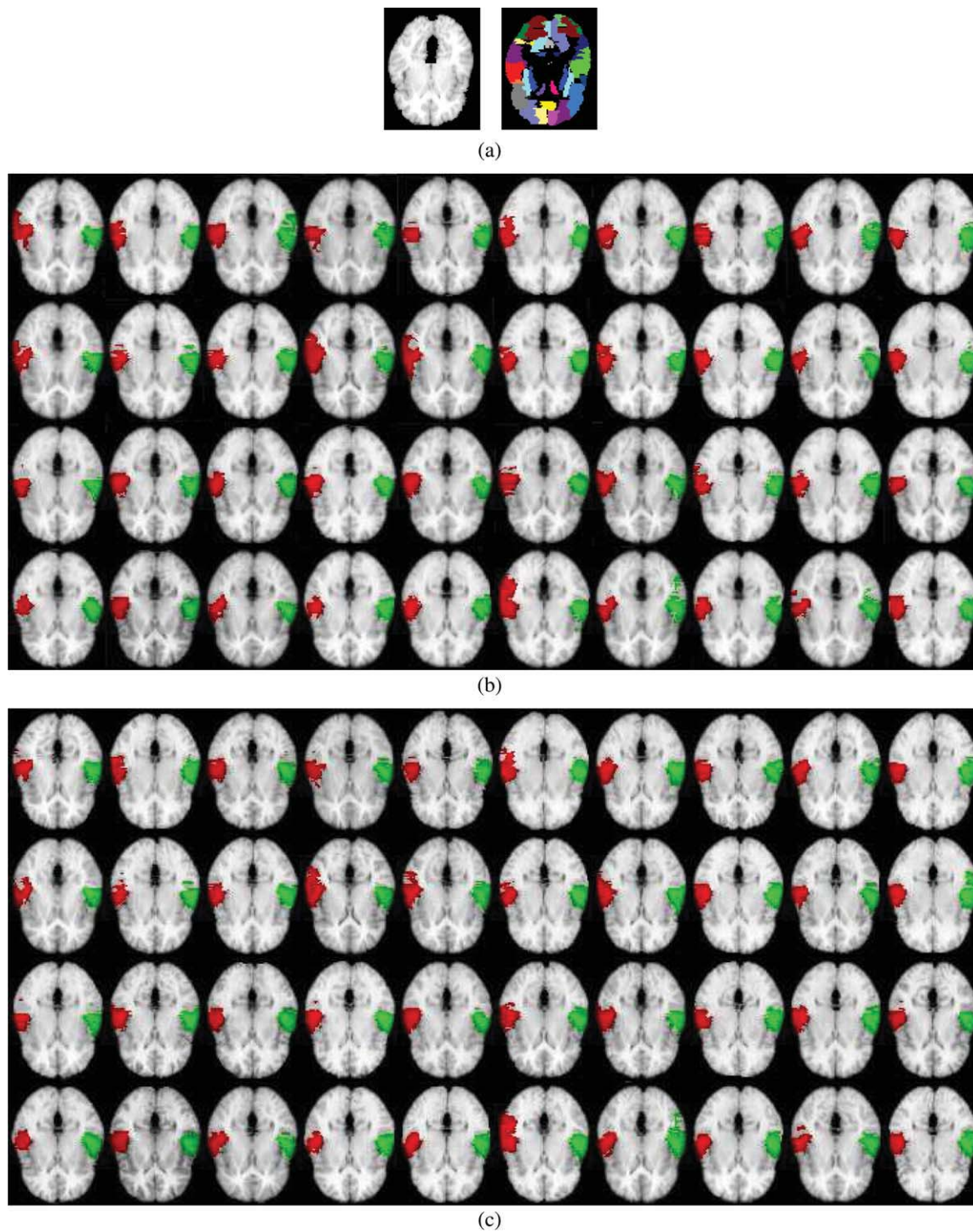


Figure 5.

Visual comparison of registration results yielded via both the congealing method and our method on LPBA40. (a) A volume from LPBA40, along with its respective manual labels in different colors. (b) Forty brain images in the LPBA40 dataset after registration via the congealing method; the left and the right superior temporal gyri

(L/R STGs) are colored in red and green, respectively. (c) The corresponding brain images after groupwise registration via our method. For all cases, the same slice is taken. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

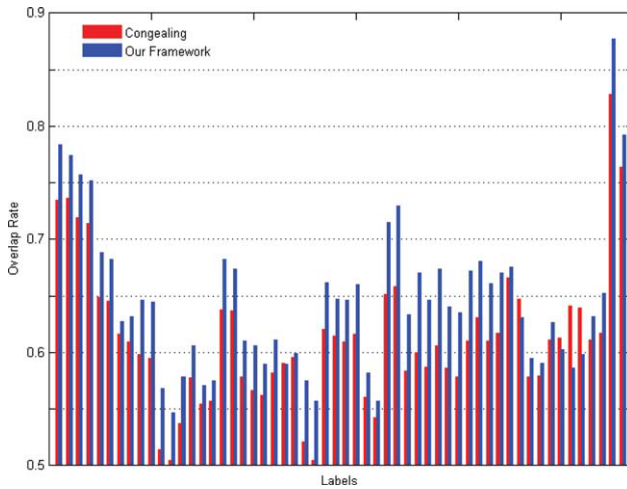


Figure 6.

Jaccard Coefficient improvements gained in all 56 manually labeled anatomical regions across the 40 images. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Registration Accuracy

Computation speed is an important aspect in groupwise registration, but maintaining registration accuracy is even more crucial. After groupwise registration, we can quantitatively evaluate the registration accuracy, since all images are now in a common atlas space. Specifically, for each given location in the image domain, the standard deviation of voxel intensities from all aligned images can be calculated. Then, we can compose those standard deviation values of different locations into a single image, called as a standard deviation map. The histogram of this standard deviation map reflects quantitatively the residual distributions after image registration. Figure 4 shows the results from several different registration frameworks. Obviously, our framework (blue curve) produces a distribution, which is more tightly concentrated at the low deviation end of the spectrum as in Figure 4, indicating less residual errors for the 30 aligned images and hence better registration quality. The congealing framework and the RG framework produce similar results. The top views of the standard deviation maps are adjusted to the same contrast level and the 3D renderings are provided in Figure 4. Through visual inspection, the result produced by our method is significantly darker than other two volumes.

We employed the LONI Probabilistic Brain Atlas (LPBA40) [Shattuck et al., 2008] to further evaluate the registration accuracy of our proposed framework. In LPBA40, there are 40 brain MR images with 56 manual labels for different anatomical structures, and all of these 40 images have already been linearly aligned to a common space. A typical slice extracted from one of the images and its corresponding label maps are shown in Figure 5a. By utilizing the congeal-

ing framework and our framework, these 40 images were groupwise registered, with their labels warped to the respective atlas spaces following the same deformation estimated upon image intensities. For fair comparison, computation costs for both methods were kept the same.

For visual inspection, the same slices of the 40 groupwise registered brain images, via both the congealing framework and our framework, are shown in Figure 5b,c. In particular, the left and the right superior temporal gyri (L/R STGs) are colored in red and green, respectively. It is, however, challenging to visually determine which method yields more accurate anatomical alignment. To solve this problem, we describe in the following a statistical approach to quantitatively evaluate the registration accuracy.

The Jaccard Coefficient metric [Jaccard, 1912], also known as the Tanimoto Coefficient, is closely related to the Dice overlap measure and is often used to measure the similarity or overlap between a specific region of two registered images (V and U) [Postelnicu et al., 2009]. For given regional label L , it is defined as

$$J(V, U; L) = \frac{|V_L \cap U_L|}{|V_L \cup U_L|}. \quad (2)$$

For quantitative comparison of the registration accuracy across more than a pair of images in a certain dataset, the metric needs to be further extended. We first create a reference image, which summarizes the common labels from a total of $N = 40$ images through a voting process. The Jaccard Coefficient could then be measured for each image with respect to the reference.

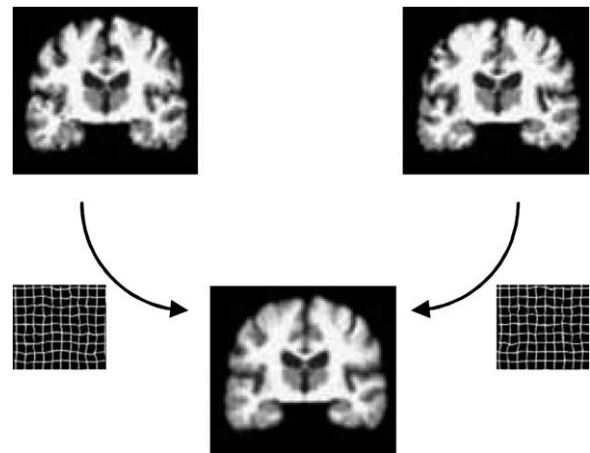


Figure 7.

A visual illustration of how morphological information of individual images (top) can be captured by the deformation fields resulting from their transformations to the atlas space (bottom). The estimated deformation fields can be used to compare the morphological differences between the images.

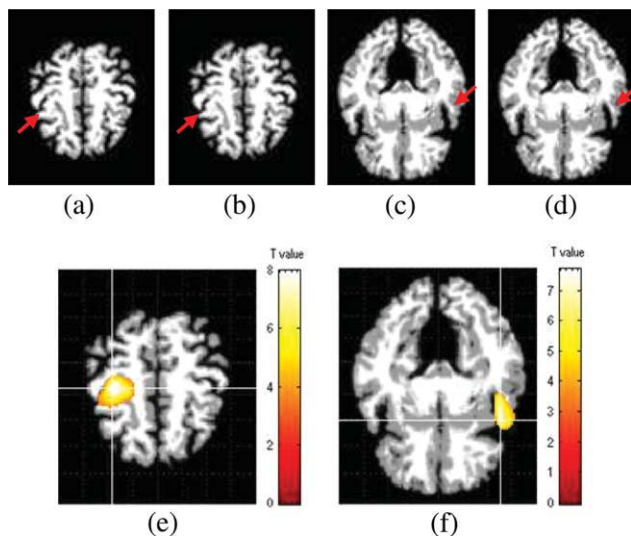


Figure 8.

Original MR images, (a) and (c), and their corresponding images with simulated atrophies in (b) the precentral gyrus (PCG) and (d) the superior temporal gyrus (STG), respectively. In (e) and (f), distributions of t -values are provided. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

For any selected label, the Jaccard Coefficient of each image registered to the atlas is computed. In Figure 6, we show the average values of the Jaccard Coefficients of different images after groupwise registration, yielded via the congealing method and our framework, for all the 56 manually labeled anatomical regions. It is obvious that both methods can effectively improve the matching of most anatomical regions. However, in most regions (51 out of 56), our framework achieves higher scores than the congealing framework. Taking into consideration the volume sizes for different anatomical labels, the improvement of the weighted average overlap rate across the whole brain is 3.966% by our framework, compared to the congealing framework. The experimental results indicate that our method yields higher registration accuracy.

Ability to Capture Morphological Changes

The estimated transformation for each image defines the anatomical correspondence and captures the relative morphological distances [Baloch et al. 2007; Davatzikos et al. 1996] between the original image and the final atlas. Therefore, the individual transformation can be used to quantitatively characterize pathological abnormalities [Ashburner and Friston, 1999; Christensen et al., 1993; Miller and Younes, 2001; Miller et al., 1997; Shen and Davatzikos, 2002]. This is illustrated in Figure 7, where two images are warped into the atlas space utilizing their respective estimated transformations. By reversing their respective deformation fields, morphological differences

TABLE II. Paired t -test results of simulated brain atrophy detection

Locations	Congealing method			Our framework		
	P_{PWE}	P_{FDR}	t -value	P_{PWE}	P_{FDR}	t -value
STG	0.036	0.007	6.98	0.004	0.001	7.97
PCG	0.055	0.007	6.75	0.007	0.001	7.67

between them can be measured in a common datum defined by the atlas.

We used two sets of images, each of which contained 12 images, as the testing subjects. The first set was composed of 12 images, and the second set was simulated from the first set by inducing simulated atrophies in the superior temporal gyrus (STG) and the precentral gyrus (PCG) [Davatzikos et al., 2001; Xue et al., 2006]. Sample slices are shown in Figure 8a–d. For each simulated image, we knew exactly the amount of atrophies (around 10% of the original volume), as indicated by red arrows in Figure 8a–d. After registration via either the congealing method or our framework, the deformation field between each image and the final estimated atlas was obtained. Based on the inverse of the deformation field, the Jacobian determinant map was computed for each image to measure the relative voxelwise spatial change in relation to the atlas. In the atlas space, we performed paired t -tests using SPM (<http://www.fil.ion.ucl.ac.uk/spm/>) on the Jacobian determinant maps to compare the group differences between the two sets of images. Paired t -tests can discard random variations and identify significant group differences. In Figure 8e,f, we show the distribution of t -values produced by the paired t -tests. The cluster of high t -values indicates the location where the simulated atrophies can be detected. Detailed values of the paired t -tests are shown in Table II. Higher t -values indicate that the detected group difference is more significant. Hence, we expect to see the higher t -values in STG and PCG where simulated atrophies

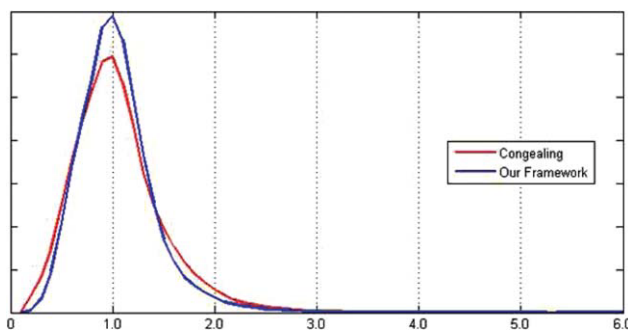


Figure 9.

Comparison of histograms of Jacobian determinants yielded via the congealing framework and our framework. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

are located. As in Table II, the t -values in both STG and PCG show that our method gives significant improvement over the congealing method, considering the nonlinearity of t -values.

As an endnote, it is also important to note that a smooth deformation field is important to ensure preservation of the anatomical topology. We compare the histograms of Jacobian determinants of the deformation fields produced via both the congealing method and our framework. From Figure 9, we could obviously observe that the Jacobian determinants of the deformation fields obtained via our framework distribute more tightly around 1.0 than those of the congealing method. This proves the capability of our framework in producing less noisy transformations, since a more concentrated distribution of Jacobian determinants around 1.0 implies more smoothness of the deformation field.

CONCLUSION

In this article, a novel groupwise registration framework, utilizing a hierarchical *image clustering* and *atlas synthesis* strategy, is proposed. A large-scale groupwise registration problem is decomposed hierarchically into a series of small-scale groupwise registration problems, each of which can be solved with relatively less effort and at increased accuracy. We show that image cluster information can be employed to guide the groupwise registration and to effectively increase its robustness, since within-class transformations can be estimated with more certainty. Experimental results indicate that higher registration accuracy can be achieved with the proposed framework, within an acceptable length of computation time. In the future, we will further validate the effectiveness of our framework by applying it to various large-scale clinical studies.

REFERENCES

- Allasonnière S, Amit Y, Trouvé A (2007): Towards a coherent statistical framework for dense deformable template estimation. *J R Stat Soc Series B Stat Methodol* 69:3–29.
- Ashburner J, Friston KJ (1999): Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* 7:254–266.
- Balci SK, Golland P, Wells WM (2007): Non-rigid groupwise registration using b-spline deformation model. In: *Proceedings of MICCAI 2007 Workshop on Open-Source and Open-Data*. pp 106–122.
- Baloch S, Verma R, Davatzikos C (2007): An anatomical equivalence class based joint transformation-residual descriptor for morphological analysis. *Inf Process Med Imaging* 20:594–606.
- Blezek DJ, Miller JV (2007): Atlas stratification. *Med Image Anal* 11:443–457.
- Christensen GE, Rabbit RD, Miller MI (1993): A deformable neuroanatomy textbook based on viscous fluid mechanics. In: *Proceedings of CISS 1993*. pp 211–216.
- Crum WR, Hartkens T, Hill DLG (2004): Non-rigid image registration: Theory and practice. *Br J Radiol* 77:S140–S153.
- Davatzikos C, Vaillant M, Resnick S, Prince J, Letovsky S, Bryan R (1996): A computerized approach for morphological analysis of the corpus callosum. *J Comput Assist Tomogr* 20:88–97.
- Davatzikos C, Genc A, Xu D, Resnick SM (2001): Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated longitudinal atrophy. *Neuroimage* 14: 1361–1369.
- Frey BJ, Dueck D (2007): Clustering by passing messages between data points. *Science* 315:972–976.
- Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RSJ (1995): Spatial registration and normalization of images. *Hum Brain Mapp* 2:165–189.
- Guimond A, Meunier J, Thirion JP (2000): Average brain models: A convergence study. *Comput Vis Image Underst* 77:192–210.
- Hajnal JV, Saeed N, Soar EJ, Oatridge A, Young IR, Bydder GM (1995): A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *J Comput Assist Tomogr* 19:289–296.
- Hill DLG, Hawkes DJ, Crossman JE, Gleeson MJ, Cox TC, Bracey EE, Strong AJ, Graves P (1991): Registration of MR and CT images for skull base surgery using point-like anatomical features. *Br J Radiol* 64:1030–1035.
- Hill DLG, Batchelor PG, Holden M, Hawkes DJ (2001): Medical image registration. *Phys Med Biol* 46:R1–R45.
- Holden M, Hill DLG, Denton ERE, Jarosz JM, Cox TCS, Rohlfing T, Goodey J, Hawkes DJ (2000): Voxel similarity measures for 3-D serial MR brain image registration. *IEEE Trans Med Imaging* 19:94–102.
- Jaccard P (1912): The distribution of flora in the alpine zone. *New Phytol* 11:37–50.
- Jain AK, Murty MN, Flynn PJ (1999): Data clustering: a review. *ACM Comput Surv* 31:264–323.
- Joshi S, Davis B, Jomier M, Gerig G (2004): Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 23:S151–S160.
- Kovacevic N, Chen J, Sled JG, Henderson J, Henkelman M (2004): Deformation based representation of groupwise average and variability. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*. Berlin: Springer. LNCS 3216, pp 615–622.
- Kovacevic N, Henderson JT, Chan E, Lifshitz N, Bishop J, Evans AC, Henkelman RM, Chen XJ (2005): A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability. *Cereb Cortex* 15:639–645.
- Learned-Miller E (2006): Data driven image models through continuous joint alignment. *IEEE Trans Pattern Anal Mach Intell* 28:236–250.
- Luan H, Qi F, Xue Z, Chen L, Shen D (2008): Multimodality image registration by maximization of quantitative–qualitative measure of mutual information. *Pattern Recognit* 41:285–298.
- MacQueen JB (1967): Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley: University of California Press. pp 281–297.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997): Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 16:187–198.
- Maintz JBA, Viergever MA (1998): A survey of medical image registration. *Med Image Anal* 2:1–36.
- Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W (1993): PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imaging* 22:120–128.

- Meltzer CC, Bryan RN, Holcomb HH, Kimball AW, Mayberg HS, Sadzot B, Leal JP, Wagner HN Jr, Frost JJ (1990): Anatomical localization for PET using MR imaging. *J Comput Assist Tomogr* 14:418–426.
- Members and Collaborators of the Wellcome Trust Centre for Neuroimaging (2009): SPM—Statistical Parametric Mapping. <http://www.fil.ion.ucl.ac.uk/spm/>. Last modified 2 April 2009.
- Miller M, Younes L (2001): Group actions, homeomorphisms, and matching: A general framework. *Int J Comput Vis* 41: 61–84.
- Miller MI, Banerjee A, Christensen GE, Joshi SC, Khaneja N, Grenander U, Matejic L (1997): Statistical methods in computational anatomy. *Stat Methods Med Res* 6:267–299.
- Miller E, Matsakis N, Viola P (2000): Learning from one example through shared densities on transforms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2000, Vol. 1. pp 464–471.
- Mohamed A, Zacharaki EI, Shen D, Davatzikos C (2006): Deformable registration of brain tumor images via a statistical model of tumor-induced deformation. *Med Image Anal* 10:752–763.
- Park H, Bland PH, Hero AO, Meyer CR (2005): Least biased target selection in probabilistic atlas construction. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005*. Berlin: Springer. LNCS 3750, pp 419–426.
- Pluim JPW, Maintz JBA, Viergever MA (2003): Mutual information based registration of medical images: A survey. *IEEE Trans Med Imaging* 22:986–1004.
- Postelnicu G, Zöllei L, Fischl B (2009): Combined volumetric and surface registration. *IEEE Trans Med Imaging* 28:508–522.
- Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999): Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans Med Imaging* 18:712–721.
- Sabuncu MR, Balci SK, Golland P (2008): Discovering modes of an image population through mixture modeling. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*. Berlin: Springer. LNCS 5242, pp 381–389.
- Seghers D, D’Agostino E, Maes F, Vandermeulen D, Suetens P (2004): Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*. Berlin: Springer. LNCS 3216, pp 696–703.
- Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW (2008): Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39:1064–1080.
- Shen D, Davatzikos C (2002): HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans Med Imaging* 21:1421–1439.
- Studholme C, Hill DLG, Hawkes DJ (1999): An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit* 32:71–86.
- Tenenbaum JB, de Silva V, Langford JC (2000): A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323.
- Toga AW, Thompson PM (2001): The role of image registration in brain mapping. *Image Vis Comput* 19:3–24.
- Viola P, Wells WM (1997): Alignment by maximization of mutual information. *Int J Comput Vis* 24:137–154.
- Wells WM, Viola P, Atsumi H, Nakajima S, Kikinis R (1996): Multi-modal volume registration by maximization of mutual information. *Med Image Anal* 1:35–51.
- Woods RP (2003): Characterizing volume and surface deformations in an atlas framework: Theory, applications, and implementation. *Neuroimage* 18:769–788.
- Woods RP, Cherry SR, Mazziotta JC (1992): Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr* 16:620–633.
- Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC (1998a) Automated image registration: I. General methods and intrasubject, intramodality validation. *J Comput Assist Tomogr* 22:139–152.
- Woods RP, Grafton ST, Watson JD, Sicotte NL, Mazziotta JC (1998b) Automated image registration: II. Intersubject validation of linear and nonlinear models. *J Comput Assist Tomogr* 22:153–165.
- Xue Z, Shen D, Davatzikos C (2006): Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. *Med Image Anal* 10:740–751.
- Zitová B, Flusser J (2003): Image registration methods: A survey. *Image Vis Comput* 21:977–1000.
- Zöllei L, Learned-Miller E, Grimson E, Wells W (2005): Efficient population registration of 3D data. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2005*. Berlin: Springer. LNCS 3765, pp 291–301.