Registration of Brain MR Images in Large-Scale Populations

Qian Wang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2013

Approved by:

Dinggang Shen, Advisor

Marc Niethammer, Reader

Pew-Thian Yap, Reader

Leonard McMillan, Committee Member

Martin Styner, Committee Member

ABSTRACT
QIAN WANG: Registration of Brain MR Images in Large-Scale Populations.
(Under the direction of Dinggang Shen.)

Non-rigid image registration is fundamentally important in analyzing large-scale population of medical images, e.g., T1-weighted brain MRI data. Conventional pairwise registration methods involve only two images, as the moving subject image is deformed towards the space of the template for the maximization of their in-between similarity. The population information, however, is mostly ignored, with individual images in the population registered independently with the arbitrarily selected template. By contrast, this dissertation investigates the contributions of the entire population to image registration.

- First, the population can provide guidance to the pairwise registration between a certain subject and the template. If the subject and an intermediate image in the same population are similar in appearances, the subject shares a similar deformation field with the intermediate image. Thus, the guidance from the intermediate image can be beneficial to the subject, in that the pre-estimated deformation field of the intermediate image initiates the estimation of the subject deformation field when the two images are registered with the identical template.

- Second, all images in the population can be registered towards the common space of the population using the groupwise technique. Groupwise registration differs from the traditional design of pairwise registration in that no template is predetermined. Instead, all images agglomerate to the common space of the population simultaneously. Moreover, the common space is revealed spontaneously

during image registration, without introducing any bias towards the subsequent analyses and applications.

This dissertation shows that population information can contribute to both pairwise registration and groupwise registration. In particular, by utilizing the guidance from the intermediate images in the population, the pairwise registration is more robust and accurate compared to the direct pairwise registration between the subject and the template. Also, for groupwise registration, all images in the population can be aligned more accurately in the common space, although the complexity of groupwise registration increases substantially.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Dinggang Shen, for his excellent guidance, caring, inspiration, and financial support for my research. Dr. Pew-Thian Yap co-authored several of my publications and helped me to improve my scientific thinking and writing skills a lot. I would also like to thank Dr. Marc Niethammer, who chaired my committee and provided assistance to me even before I arrived at Chapel Hill. Special thanks go to Dr. Martin Styner and Dr. Leonard McMillan, who served on my committee and devoted a lot of effort to my study.

Also I would like to thank the co-authors and collaborators of my various publications, including Dr. Guorong Wu, Dr. Min-Jeong Kim, Dr. Hongjun Jia and all the people who ever worked in the IDEA lab of BRIC. In particular, Guorong and Hongjun contributed significantly to the part on groupwise image registration in this dissertation. We shared a lot of discussions and even "quarrels" sometimes.

My fellow students in Computer Science and several other departments also helped me a lot in getting through these years. To name a few of them, Liang Shan, Chen-Rui Chou, Tian Cao, Yi Hong, etc. And special thanks to Weibo Wang, who was my high school classmate and incredibly became my classmate again at UNC; also Cihat Eldeniz, who was affiliated with Biomedical Engineering and served as my "tutor" of imaging techniques.

Finally I would like to thank my family members. My parents steadily stood behind me and supported every decision I made. My wife sacrificed her promising career and

stayed with me. My son, little Maxwell, came into this world a few months ago and has brought me a lot of joy as well as encouragement to complete the last stage of my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
ADNI	Alzheimers Disease Neuroimaging Initiative
AP	affinity propagation
CSF	corticospinal fluid
GM	grey matter
HAMMER	hierarchical attribute matching mechanism for elastic registration
ING	iterative neighborhood graph
MDL	minimum description length
MI	mutual information
MR	magnetic resonance
MRI	magnetic resonance imaging
MST	minimum spanning tree
NCC	normalized cross correlation
PDF	probability density function
PDFs	probability density functions
RBF	radial basis function
RBFs	radial basis functions
ROI	region of interest
ROIs	regions of interest
PCA	principal component analysis
PCG	pre-central gyrus
SNR	signal-to-noise ratio
SSD	sum of squared differences

STG superior temporal gyrus

TPS thin-plate splines

WM whiter matter

Chapter 1

Introduction

1.1 Significance of Image Registration

Without introducing hazardous ionizing radiation, magnetic resonance imaging (MRI) offers the capability to visualize internal structures of the human body in a non-invasive manner. The technique has thus been widely applied to numerous clinical and research works ever since it was invented decades ago. Brain magnetic resonance (MR) images, in particular, facilitate the diagnoses and the treatments for many neurological diseases thanks to the well rendered contrasts of brain tissues. Large-scale studies are enabled for investigating brain development (Casey *et al.*, 2000; Giedd *et al.*, 1999; Lenroot and Giedd, 2006; Sowell *et al.*, 1999b; Thompson *et al.*, 2000), maturation (Holland *et al.*, 1986; Sowell *et al.*, 1999a; Paus *et al.*, 1999, 2001), aging (Resnick *et al.*, 2000; Raz *et al.*, 2005), disease-induced anomalies (Frisoni *et al.*, 2010; Laakso *et al.*, 2000; Polman *et al.*, 2011; Thompson *et al.*, 1998), and for monitoring the effects of pharmacological intervention over treatment time (Jack *et al.*, 2004; Mulnard *et al.*, 2000; Resnick and Maki, 2001).

Computer-aided methods with the capability of analyzing brain MR images automatically are highly desired for related applications. Many diseases, e.g., major neuropsychiatric and substance abuse disorders, can cause very complex spatiotempo-

ral changes in brain tissue patterns that are difficult to identify visually. Meanwhile, technique advances help reduce the costs of MRI, thus large-scale populations of MR images are often employed in recent studies. To manually process and analyze large-scale populations of brain MR images, however, can be made difficult by (1) the rapidly increasing number of acquisitions; (2) the large size (or the dimensionality) of the image data; (3) the subtle changes of image appearances; (4) the difficulty and the cost in recruiting human experts. As a result, many sophisticated computer-aided medical image analysis methods have been proposed in the literature, including image registration (Glocker *et al.*, 2011; Hajnal and Hill, 2010; Oliveira and Tavares, 2012; Rueckert and Schnabel, 2011), segmentation (Balafar *et al.*, 2010; Heimann and Meinzer, 2009; Pham *et al.*, 2000), disease-oriented classification (Aggarwal *et al.*, 2012; Ecker *et al.*, 2010; Klöppel *et al.*, 2008; Zhang *et al.*, 2011), etc.

Image registration acts as one of the most fundamentally important tools in the area of medical image analysis. The task of registration aims to align a certain moving image, which is often termed as the *subject*, with another fixed image often termed as the *template*. After registration, the subject and the template are regarded as being within the same image space, thus the two images can be quantitatively compared. By registering longitudinal scans of a certain patient to the common template space, for instance, the intra-subject temporal changes can be accurately captured (Holland and Dale, 2011; Li *et al.*, 2012; Smith *et al.*, 2001). Similarly, the inter-subject variation can also be modeled by the means of registration, after all subjects under consideration are spatially aligned. By registering a population of images, researchers are also able to construct an atlas that reflects the common information of the population. And inter-population comparisons can then be conducted (Learned-Miller, 2006; Joshi *et al.*, 2004; Zhang *et al.*, 2013). Furthermore, the understanding of a certain image (i.e., anatomical parcellation) can be easily propagated to other images that have been registered to the

same template space, thus significantly reducing the costs in processing and analyzing large-scale image populations (Aljabar *et al.*, 2009; Cuadra *et al.*, 2004; Vemuri *et al.*, 2003).

1.2 Brain MR Image Registration

The registration of brain MR images has drawn abundant interest and also intensive investigations due to its importance in the area of medical image analysis. Most conventional methods regard the task of registration as a typical optimization problem, where a pair of moving *subject* and fixed *template* are often involved. The classical model of pairwise registration is illustrated by Figure 1.1, as the optimizer estimates the transformation, following which the subject is mapped to the space of the template in order to maximize the similarity between the two images. Typically, the solution to the transformation is attained in an iterative manner. In general, the optimization related to image registration tries to maximize the following objective function:

$$\text{Objective}(\phi) = \text{Similarity}(T, S \circ \phi) + \text{Regularization}(\phi). \tag{1.1}$$

In (1.1), variables S, T, and ϕ represent the subject, the template, and the transformation, respectively. The term $S \circ \phi$ indicates the transformed subject image. This formulation, as well as its variants, will be revisited in the dissertation frequently. Moreover, several components in the model of Figure 1.1 need to be addressed carefully, including:

- **Subject and Template.** The input subject and template images may belong to either a single or multiple modalities, although multi-modal registration can be much more complicated due to the significantly different appearances of the two images under consideration. This dissertation, however, will focus on the mono-modal registration of T1-weighted brain MR images only, meaning both

the subject and the template are required to be mono-modal. Discussions of other images modalities and organs are outside the scope of this dissertation.

- **Similarity Metric.** The similarity (or the distance) between the (transformed) subject and the template can be computed in various ways, especially considering the fact that brain MR images are high-dimensional data. Though there are a huge number of voxels in each image and the anatomical variation of the images may be high, the similarity can be simply computed from the intensities of two images, i.e., as the inverse of the sum of squared differences (SSD) of intensities. More sophisticated measures and advanced features of image contexts can also contribute to the evaluation of the image similarity.

- **Transformation Model.** The transformation is a mapping between the image spaces of the subject and the template. The mapping of the rigid transformation, for example, allows rotation and translation of the subject. By incorporating scalings and shearings in addition, the model becomes the affine transformation. Non-rigid transformation affords unprecedented freedoms in deforming the subject. This dissertation will focus on image registration that is related to non-rigid deformation. Further, the Lagrangian framework is adopted as the convention to denote the transformation in this dissertation. In particular, the transformation $\phi(\cdot)$ is applied to register the moving subject S with the fixed template T. Then, $\phi(x) : \Omega_T \rightarrow \Omega_S$ maps the template point $x \in \Omega_T$ to the subject point $\phi(x) \in \Omega_S$. Similarly, the inverse transformation $\phi^{-1}(y) : \Omega_S \rightarrow \Omega_T$ maps the subject point $y \in \Omega_S$ to the template point $\phi^{-1}(y) \in \Omega_T$.

- **Optimizer and Correspondence Detection.** The optimization aims to find the optimal transformation that maximizes the similarity of the deformed subject and the template under a certain regularization. From an alternative perspective,

the optimizer can also be perceived as a correspondence detector for the two images. The correspondence is defined as (1) two underlying points that are similar in appearances; (2) the two points that signify similar neuroanatomies. The optimizer estimates the transformation, following which individual points in the subject should be mapped to the coordinate locations of their correspondences in the template space. In this way, the optimizer is able to maximize the similarity between the transformed subject image and the template image.

- **Multi-Resolution Hierarchy.** Registration is regarded as an ill-posed optimization problem and suffers severely from the notorious curse of high dimensionality. Brain MR images are intrinsically high dimensional due to their large data size and the very abundant anatomical variations. The non-rigid transformation, or deformation, also requires a huge number of parameters for the sake of its representation. Therefore, to alleviate the aforementioned concerns, a multi-resolution strategy is often introduced by solving the registration problem hierarchically. For example, both the subject and the template can be down-sampled first, while a low-resolution transformation is computed. Then, the low-resolution transformation is up-sampled to be further optimized at the higher level, where high-resolution images should participate. The multi-resolution scheme allows more abundant image information (corresponding to varying resolutions) to be used by image registration, and results in a more robust estimation of the transformation.

1.2.1 Intensity-Based Methods

Brain MR images provide clinically usable contrasts between tissues, as image intensities carry important appearance information. For the sake of image registration, the

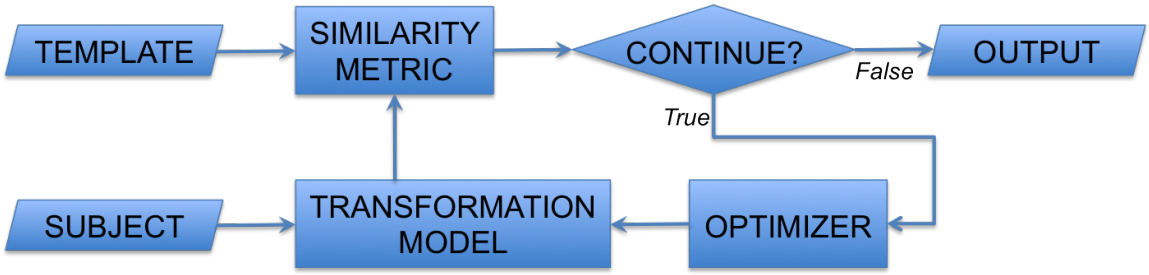


Figure 1.1: The classical model of pairwise image registration as a typical optimization problem.

similarity between the subject and the template can be calculated from image intensities directly. Though simple and straightforward, the inverse of the SSD of image intensities gives an effective image similarity measure in the Euclidean space. SSD adopts the assumption that noises in images are Gaussian distributed, and is successfully applied in many state-of-the-art registration methods (Oliveira and Tavares, 2012). Several other intensity-based metrics are also available in the literature, including the statistical measure of normalized cross correlation (NCC) (Avants *et al.*, 2008), the information-theory-based mutual information (MI) (Maes *et al.*, 1997; Pluim *et al.*, 2003; Viola and Wells III, 1997), minimum description length (MDL) (Davies *et al.*, 2002), etc.

Diffeomorphic Demons (Vercauteren *et al.*, 2009) is among one of the most popular intensity-based registration methods that are capable to handle brain MR images. The image similarity term in (1.1) is simply calculated as $\text{Similarity}(T, S \circ \phi) = \|T - S \circ \phi\|^2$ by Demons. After being combined with the regularization upon ϕ, the derivative $\partial \cdot \text{Objective}/\partial\phi$ can be computed to indicate the steepest ascent direction of the objective function. Details of the regularization of ϕ will be discussed later in Section 1.2.3. Then, gradient-based optimization methods can be plugged in to optimize the transformation and to maximize the objective function. The final solution of the transformation is attained in an iterative and multi-resolution manner.

1.2.2 Feature-Based Methods

Correspondence detection, which is conveyed by image registration, can be challenged by high ambiguities in brain MR images if only intensity information is considered. Given two points that share the same intensity, it is risky to assert that they are correspondences to each other. In fact, though brain MR images offer good contrast between grey matter (GM) and whiter matter (WM), the intensities within a certain tissue can be very similar. To this end, it is often impossible to identify respective correspondences of neighboring points by using intensity information only. As an alternative, it is well known in computer vision research that correspondences can be better established by using sophisticated context features. The rule also applies in the case of brain MR image registration, to which correspondences between points (Maurer Jr *et al.*, 1997), curves (Lyu *et al.*, 2013), and surfaces (Jiang *et al.*, 1992; Pelizzari *et al.*, 1989) may contribute.

The method of hierarchical attribute matching mechanism for elastic registration (HAMMER) (Shen and Davatzikos, 2002) is a typical example of the feature-based category. In HAMMER, each point in the brain MR image is signified by its own attribute vector. Then, a certain point in the template can detect its correspondence in the subject image space, by searching for the nearby subject point with the most similar attribute vector. Specifically, both the subject and the template images are segmented into different tissue types including GM, WM, and corticospinal fluid (CSF), respectively. The attribute vector for each point is then computed to incorporate the geometric moments (Lo and Don, 1989) of all tissue types in the nearby area of the point. Furthermore, a set of key points are selected, as correspondence detection is performed upon each selected key point. The key points are usually located in tissue transitions, e.g., crowns of gyri and roots of sulci. Thus their abundant context features can help reduce the ambiguities in correspondence detection, while their anatomically impor-

tant locations are crucial for driving the correct alignment of corresponding anatomies of the two brains. For the sake of optimization, the selection of key points and the correspondence detection are further arranged into an iterative and multi-resolution scheme. With the progress of registration, more key points are devoted to the process as the correspondence detection becomes more accurate. The deformation field, which can be interpolated from the key points and their correspondences, thus approaches the desired solution gradually.

1.2.3 Deformation and Its Smoothness

The non-rigid transformation, or the deformation field, is usually expected to be physiologically meaningful and smooth in the registration of brain MR images. A simple criterion requires that the Jacobian determinant of the deformation field is positive everywhere. In this way, though the deformation allows free movement of individual points in the image space, it is guaranteed that no folding of brain tissues could possibly be introduced during the warping of the subject. Meanwhile, the smoothness of the deformation can be implemented in several different ways, e.g., via low-pass filtering of the deformation field.

- The way to enforce the smoothness regularization is tightly related to the way the deformation field is represented, which can generally be categorized into non-parametric and parametric models. The non-parametric modeling regards the entire deformation as a whole and offers flexibility in describing respective deformations of individual grid points. The smoothness of the deformation field can be easily achieved by the low-pass filtering technique (Myronenko and Song, 2010). Simple isotropic Gaussian smoothing, for instance, is often applied to ensure that the resulting deformation field is smooth. Moreover, the low-pass filter can be specifically designed to cater to varying demands of the smoothness requirement

of the deformation field.

- The deformation can also be parameterized as a collection of basis functions. B-Splines, for instance, are capable of representing the free-form deformation (Rueckert *et al.*, 1999). In particular, for a certain *control point* in the image space, a B-Spline is attached to describe its contribution to the entire deformation field. The set of control points should cover the whole brain volume, and often needs to be decided carefully. The deformation field, which is spanned across the entire image space, can be interpolated by integrating contributions from all control points. Other choices for the parametric modeling of the deformation field include but are not restricted to thin-plate splines (TPS) (Bookstein, 1989; Chui and Rangarajan, 2003), Fourier basis functions (Amit *et al.*, 1991; Ashburner and Friston, 1999), and wavelets (Amit, 1994). Compared to non-parametric modeling, parametric modeling is able to reduce the number of parameters that are needed for the representation of the deformation field. Thus, the optimization of the registration usually becomes less challenging. Also, most parametric modeling provides subtle controls upon the smoothness when the deformation field is interpolated.

- The deformation can be described by both the small deformation model and the large deformation model (Christensen *et al.*, 1996). In the small deformation model, the smoothness requirement is attained in the elastic style. That is, the smoothness regulation is imposed on the deformation field directly, e.g., by applying the Gaussian filter to the deformation. By contrast, the large deformation model functions in the fluid style while the smoothness requirement is typically imposed on the incremental field of the deformation. Mathematically, the large deformation model rewrites the regularization in (1.1) to Regularization($\Delta\phi$), while ϕ can be self-updated following the rule $\phi \leftarrow \phi \circ \Delta\phi$. Though the elastic

and the fluid styles are different, they can sometimes be combined and utilized as a hybrid. In particular, the incremental of the deformation estimated in every iteration complies with the smoothness constraint in the fluid style, while the iteratively integrated overall deformation can be regulated by the elastic constraint as well.

1.2.4 Performance Evaluation

It is necessary to evaluate the performances of brain MR image registration, before the registration outputs can be potentially used by the following clinical applications and research studies. The lack of prior knowledge or the ground truth, however, leads to the dilemma that the performance evaluation could be non-trivial. The evaluation often starts from the assessment of the robustness of the registration results. Experts with minimal training can easily conclude whether a registration is satisfactory by visually inspecting the similarity between the deformed subject and the template images. The robustness of image registration is important in that the registration method is expected to handle a large-scale population of images automatically. Manual inspection obviously interrupts the pipeline in processing images and requires additional cost.

The quantitative evaluation is more difficult in the absence of the ground truth. To this end, prior knowledge of the deformation field is highly desired. For example, West *et al.* (1997) use skull-implanted markers to derive the rigid transformations between multi-modal images. Thus the outputs from different registration methods can be compared with respect to displacements of external markers. However, it might be difficult to generate a non-rigid transformation from external markers. Therefore, most methods prefer to simulate physically plausible deformations that are comparable to realistic scenarios. The simulation then allows the researcher to evaluate the accuracy of the registration as well as the accompanying statistical analyses, e.g., for the studies

of the Alzheimer's Disease (AD) (Camara *et al.*, 2008).

The annotations of anatomical structures provide a convenient way to evaluate registration accuracy. Note that the purpose of image registration is to correctly align corresponding anatomies. Thus, the consistency or overlapping of identical anatomical structures after registering the subject to the template space is a natural indicator of the registration quality. For instance, by (manually) labeling regions of the same anatomical role but from the deformed subject and the template, respectively, the overlap of the two regions can be captured by the Dice ratio

$$R_D(\text{Region}_1, \text{Region}_2) = \frac{2 \times \text{Vol}(\text{Region}_1 \cap \text{Region}_2)}{\text{Vol}(\text{Region}_1) + \text{Vol}(\text{Region}_2)}. \tag{1.2}$$

Here, the operator $\text{Vol}(\cdot)$ computes the volume of the underlying anatomical region. Besides the Dice ratio, other measures (e.g., the Jaccard ratio) may be applicable as well. The registration accuracy metric can also extend to incorporate more corresponding anatomical structures, including landmarks, curves, surfaces, etc.

To end the retrospect of the literature, a list of typical registration methods for brain MR images is provided in Table 1.1. Also, it is worth noting that a comprehensive investigation of performances of individual registration methods is reported in Klein *et al.* (2009). Systematic understanding and surveys of the current status of medical image registration are also available in the literature (Hajnal and Hill, 2010; Oliveira and Tavares, 2012; Rueckert and Schnabel, 2011).

1.3 Limitations of Conventional Methods

As discussed in the above, most conventional registration methods belong to the pairwise category, which is designed to handle a pair of subject and template images only.

Method	Similarity Metric	Transformation and Regularization
ART	NCC	Non-parametric deformation, Low-pass filtering
AIR	Mean squared difference	Polynomial warps
Diffeomorphic Demons	SSD	Non-parametric deformation, low-pass filtering
FLIRT	SSD, MI, NCC, etc.	Linear/rigid/affine transformation
FNIRT	SSD	B-Splines
HAMMER	Feature matching	Non-parametric deformation, low-pass filtering
IRTK	Normalized MI	B-Splines
LDDMM	SSD	Non-parametric deformation, low-pas filtering
S-HAMMER	Feature matching	Symmetric deformation, TPS, bending energy
SPM	Mean squared difference, etc.	Discrete cosine transform, bending energy
SyN	NCC	Non-parametric and symmetric deformation, low-pass filtering

Table 1.1: Summary of state-of-the-art pairwise registration methods for brain MR images. Only exemplar methods are reported in this table, including: ART (Ardekani and Bachman, 2009), AIR (Woods *et al.*, 1998b,a), Diffeomorphic Demons (Vercauteren *et al.*, 2009), FLIRT (Jenkinson and Smith, 2001; Jenkinson *et al.*, 2002), FNIRT (Andersson *et al.*, 2008), HAMMER (Shen and Davatzikos, 2002), IRTK (Schnabel *et al.*, 2001), LDDMM (Beg *et al.*, 2005), S-HAMMER (Wu *et al.*, 2012d), SPM (Friston *et al.*, 2011), and SyN (Avants *et al.*, 2008).

Contrarily, image registration is more frequently applied to large-scale populations of images nowadays. Analyses based on large-scale image populations, whose prevalence is partially due to the introduction of automatic image analysis techniques and the lowered costs of image acquisitions, are statistically more convincing and attractive. When applying pairwise image registration methods to a large-scale population of images, researchers typically need to designate a certain template to which all subject images are registered. The template can be randomly selected from the population, or determined as the specific image that best represents the population. Standard image

spaces, which are not necessarily related to the image population under consideration, may also provide convenient alternatives.

Images in the population can be regarded as individual subjects and handled by pairwise registration in either the round-robin or parallelized manner. The registration tasks of individual subject images are essentially independent of each other. However, it is known that two similar subject images (i.e., denoted by S_1 and S_2), when registered with the same template, should have similar deformation fields. This property is unfortunately ignored when applying pairwise registration to a large-scale population of images. Even though S_1 and S_2 are very similar in appearances and deformations, their registration with the template is independently considered. However, it would be desirable for the two subject images to share certain information during the registration process. For example, the registration of S_1 (or S_2) might be a good approximation of the registration of S_2 (or S_1). In other words, to register one of them could well initialize the registration of the other.

Though a template is inevitably necessary for applying pairwise registration to large-scale image populations, the determination of the proper template can be highly complex and sensitive. The template implies the common image space where all images in the population are deformed to and then compared. Thus, the template can easily introduce bias towards the statistical analyses following image registration. For example, given a population that consists of both normal controls and disease-affected subjects, the influence of using a normal or abnormal image as the template would be very hard to predict. To this end, a better solution of image registration would get rid of the concern over the determination of the template, while all subject images could still be registered to the common space for facilitating analyses and evaluations upon them.

1.4 Thesis

Thesis: *Registration of brain MR images in large-scale populations benefits from utilizing information from the entire population, instead of from the subject and the template only.*

This dissertation investigates ways of incorporating information contributed by the entire image population into solving the registration problem. In particular, two specific aims are proposed to improve brain MR image registration.

- ***Specific Aim 1:*** *Pairwise registration can be improved by utilizing the guidance from other images in the population.* Instead of registering the subject with the template directly in the conventional style of pairwise registration, the subject can initiate its registration by using guidance provided by other images (namely the *intermediate images*) in the population. The guidance of the intermediate images helps the subject to identify its deformation towards the template more easily. Thus the registration of the subject with the template becomes more robust and accurate in the end, compared to direct registration of the two images.

- ***Specific Aim 2:*** *The large-scale population of brain MR images can be registered towards their common space in a groupwise manner.* Groupwise registration significantly differs from the traditional pairwise registration, in that the designation of the template, as well as its bias, is completely avoided. However, by iteratively optimizing the coherence of the population, groupwise registration is capable of deforming all images to the common space of the population. Meanwhile, the similarity between each pair of warped subjects is maximized, as all images are registered with each other with respect to the common space.

In order to support the thesis and the specific aims in the above, the detailed contributions of this dissertation include:

- Two types of guidance from the intermediate images, namely the *image-scale guidance* and the *patch-scale guidance*, are investigated for guiding the pairwise registration of a certain subject with the template and improving registration performance;

- The image-scale guidance allows the subject to seek help from the intermediate image in estimating its own deformation towards the template, in that the intermediate image is required to be similar to the subject and thus shares similar deformations with respect to the template;

- In the setting of patch-scale guidance, individual patches in the subject flexibly predict their associated deformations with respect to the template, in accordance with the guidance provided by different intermediate images and their corresponding patches;

- The importance and advantages of groupwise registration are addressed, as in groupwise registration all subject images in the large-scale population are warped to an unknown common space where the atlas of the population can be constructed;

- With a mean image to represent the common space, groupwise registration is feasibly implemented by a set of pairwise registrations of individual subjects and the mean, while the *sharpness* of the group-mean is proven to be significantly important in terms of the quality of groupwise registration;

- The explicit estimation of the group-mean image is not always necessary; instead, a certain subject is still able to locate the common space of the population, simply by following the deformation pathway combined from its pairwise registration towards other similar images in the population;

- With an objective function to capture the variation of all images with respect to the unknown common space, groupwise registration is able to directly optimize the objective function by simultaneously estimating the deformation fields for all subjects; meanwhile, correspondences of points from individual subjects are established via the implicit common space;

- By incorporating information of the entire population, overall registration quality is improved, thus benefiting other related applications, e.g., to propagate anatomical labeling from the known images to the unknown images more precisely after all images are registered.

1.5 Overview of Chapters

The remaining chapters of this dissertation are organized as follows.

- Chapter 2 investigates the utilization of the guidance, provided by the intermediate images, in pairwise registration of the subject with the template. The scales of the guidance, i.e., at the image or the patch scale, are further addressed and compared. The guidance from the intermediate images provides good initialization for the subject to estimate its deformation towards the template. The initialized deformation can be refined efficiently and effectively, and thus enhance the accuracy of pairwise registration significantly.

- Chapter 3 systematically overviews typical implementations of groupwise registration, which aims to register all images in the population towards the common space of the population simultaneously. In particular, groupwise registration gets rid of the template that is arbitrarily designated in pairwise registration, implying that the bias of the template is completely avoided. Moreover, the images can

be aligned with each other more accurately in the common space compared to pairwise registration, though the problem of groupwise registration is much more complicated due to the population size.

- Chapter 4 demonstrates that improved registration, i.e., by incorporating information from the entire population, can benefit the multi-atlas labeling technique to generate more reliable segmentation outcomes. Multi-atlas labeling, a state-of-the-art method, is capable of propagating anatomical labels from the known atlas to other unknown images for the sake of automatic segmentation. In that the atlases and the to-be-segmented images are registered more precisely, the resulting segmentation provided by multi-atlas labeling is also improved in the context of brain MR images.

- Chapter 5 concludes the entire dissertation. In general, this dissertation investigates the contributions of large-scale populations towards image registration. The entire population benefits the conventional pairwise registration between two images, due to the guidance from other intermediate images in the population. Moreover, from the perspective of groupwise registration, all images in the population can be registered simultaneously and accurately.

1.6 Summary

Image registration is a complicated yet important problem in the area of medical image analysis. Though much effort has been devoted to solving it, the problem is still open to further investigation. In particular, it is necessary to introduce large-scale population information into registration, since most studies nowadays rely on huge numbers of acquisitions. As outlined in this chapter, the image population contributes

to registration in at least two ways: (1) the conventional pairwise registration benefits from the guidance provided by other intermediate images in the population; (2) the pairwise limitation can be broken by using the groupwise registration technique for simultaneous warps of all images in the population. The following chapters in this dissertation will discuss these two aspects in detail, and will also demonstrate that better registration leads to better applications, e.g., segmenting brain MR images into individual anatomical areas.

Chapter 2

Population Guidance in Pairwise Registration

2.1 Overview

When applying the conventional pairwise registration methods to a large-scale population of brain MR images, the typical scenario requires researchers to specify a certain template *first* and *then* register all images with the template. Regardless of the order in handling subject images, the registration of each individual subject is generally independent of the others, as only a specific subject and the template are involved in any particular pairwise registration task. The performance of this straightforward scheme, namely the *direct* registration of the subject, could be undermined by the complexities in (1) evaluating the similarity of high-dimensional image data and (2) optimizing a huge number of parameters to represent the deformation field. The direct registration could even fail if the subject and the template differ significantly in appearances. To this end, the desire for high-performance registration has inspired enormous efforts for the improvement of existing registration methods.

Among all the blueprints proposed for better brain MR image registration, *guidance*

is utilized by a family of methods to initiate optimization and produce better registration outputs in the end. Specifically, both the template and the to-be-registered subject are incorporated by a population; other images in the population are termed as the *intermediate* images, whose registration with the template are presumed to be known already. Then, for the subject image under consideration, the guidance towards its registration with the template originates from the intermediate images in the population. In particular, two images that are similar in appearances should have similar deformations when registered with the same template. Thus, after identifying a certain intermediate image from the population that is similar with the subject in appearances, the subject initiates its own registration with the template immediately, by using the deformation field that has been pre-estimated for that specific intermediate image.

The contribution from the intermediate image, which is categorized as the *image-scale guidance* here, is illustrated by Figure 2.1. Given a large-scale population of brain MR images, a high-dimensional manifold is often instantiated to characterize the distribution of the entire population (Aljabar *et al.*, 2012). Images that are similar in appearances are spatially close to each other within the manifold. Meanwhile, the geodesic along the manifold connects a certain pair of images, and represents the deformation pathway for the registration between them. Figure 2.1(a) illustrates an image manifold and enumerates several exemplar images that are annotated by individual nodes (i.e., red for the template, purple for the subject, blue for the intermediate images). The geodesic arrow denotes the deformation pathway, following which a non-template image is registered with the template. The intermediate images are pre-registered with the template, thus the deformation pathways in blue are known already.

The direct estimation of the deformation in purple that registers the subject S_1 with the template T is frequently adopted by most state-of-the-art pairwise methods (c.f. Figure 2.1(a)). As an alternative, in Figure 2.1(b), S_1 utilizes the intermediate

image M_1 to initialize its own registration. That is, S_1 *first* registers with the intermediate image M_1, and *then* follows the existing pathway from M_1 to T to complete its registration with the template. The overall deformation of S_1 is thus derived from the integration of two deformation segments that correspond to the two registration sub-tasks. It is worth noting that the intermediate image M_1 is required to be similar to S_1 in this case. Thus the deformation segment from S_1 to M_1 is relatively short, while the segment from M_1 to T contributes to the majority of the overall deformation for S_1. In other words, the deformation that registers M_1 with T can be regarded as a well-functioning approximation (as well as initialization) to registering S_1 with T.

The *image-scale* guidance contributed by the intermediate image M_1 is able to alleviate difficulties in registering the subject S_1 with the template T directly. Given an existing deformation that registers M_1 to T reliably, the remaining task for S_1 is to find the deformation segment from S_1 to M_1 only. The length of the to-be-determined segment of the entire deformation pathway is shortened significantly. A short pathway usually implies that the registration could be handled more easily, complying with the observation that the registration between similar images faces many fewer challenges in optimizing the deformation field. After integrating two segments into a single deformation, the subject is able to further refine the deformation, i.e., via state-of-the-art registration methods, in effective and efficient manners.

The guidance from the intermediate images also allows registration to assess difficulties of individual tasks and then assign priorities to register images in the population adaptively. In the case illustrated by Figure 2.1(c) for example, the direct registration of the subject S_2 with the template T could be challenging due to the high anatomical variation and the lengthy deformation pathway between them. However, with the introduction of the intermediate image M_2, the registration of S_2 is decomposed into two sub-tasks, i.e., to register S_2 with M_2 and to register M_2 with T, respectively. Both

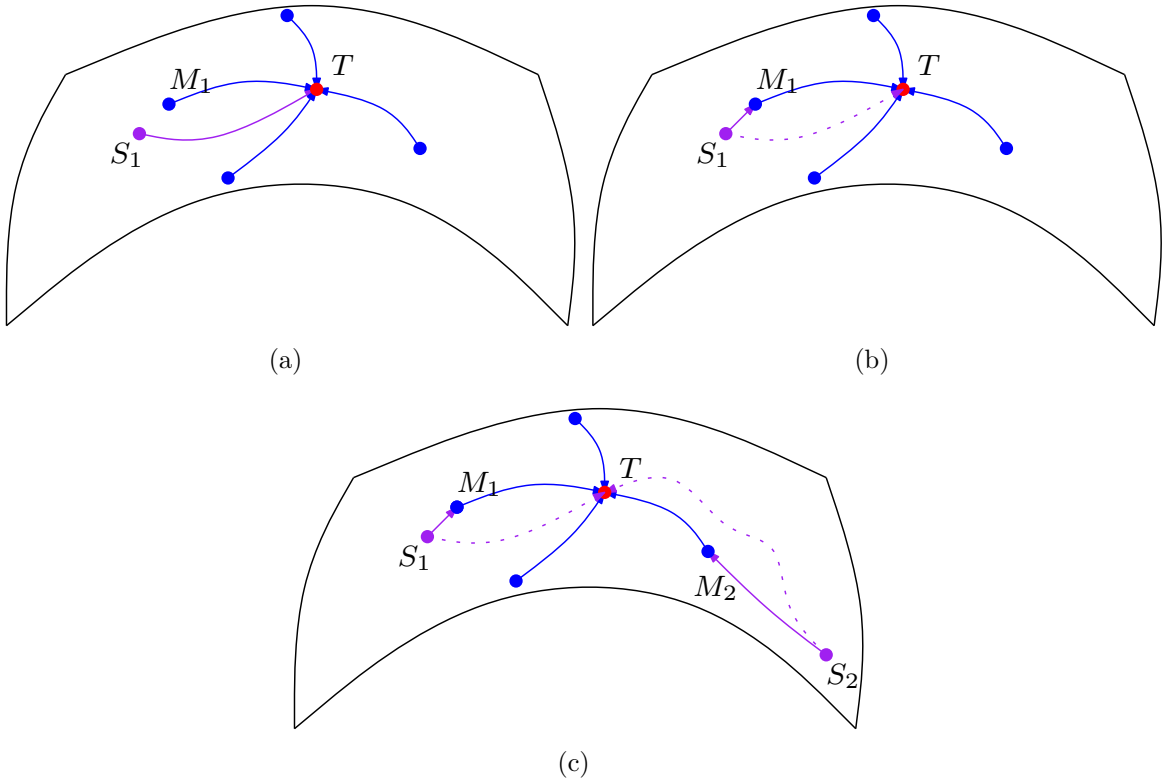


Figure 2.1: The manifold of the image population and the image-scale guidance in pairwise registration: (a) The subject S_1 is registered with the template T directly regardless of other intermediate images (including M_1) in the population; (b) Alternatively S_1 utilizes the guidance provided by M_1 to initiate its registration with T; (c) The registration of S_2 with T is decomposed into two relatively easy-to-solve problems with the introduction of M_2, i.e., from S_2 to M_2 and from M_2 to T, respectively. Red nodes correspond to the templates, blue for the intermediate images, and purple for the subjects. Curved arrows correspond to the deformation pathways that register the linked pairs of images.

sub-tasks are obviously less difficult than the direct registration of S_2 with T. To this end, the optimal strategy in handling S_2 is to defer its registration until the guidance associated with M_2 has become available. Therefore, from the perspective of the entire population, images that are more similar to the template (e.g., M_2) should be registered at higher priorities than other images that are less similar to the template (e.g., S_2). In this way, those images registered at earlier stages might be recursively utilized, as they provide guidance to initiate registration of more images at later stages. The

overall registration performance evaluated upon the entire image population can thus be improved.

Granularity of Guidance. Though above discussions are focused on *image-scale* guidance, the intermediate images are capable of contributing towards the registration of the subject image in the *patch-scale* manner. The two types of guidance differ significantly in terms of their granularities. For example, once the subject has identified the intermediate image (similar in appearances), the image-scale guidance is provided by the intermediate image within the entire image space. The rationale is that similar images (in appearances) should have similar deformation fields when registered with the same template. On the other hand, the patch-scale guidance relies on the fact that the deformations of two patches from the subject and the intermediate image are highly correlated, if the appearances of the two patches are similar (Wang *et al.*, 2013b). Therefore, for individual patches of the subject, their correspondences may come from patches belonging to different intermediate images. By reducing the granularity of the guidance from image to patch, the patch-scale guidance allows the subject to utilize all intermediate images in the population for predicting its own registration more flexibly.

Initialization and Refinement. The guidance contributed by the intermediate images yields a predicted deformation for the subject, while the prediction acts as *initialization* and needs further *refinement* to register the subject with the template completely. The refinement is typically accomplished via state-of-the-art pairwise registration methods, most of which follow the iterative optimization strategy. Moreover, the refinement is often conducted within the *high* resolution only (i.e., corresponding to the original resolution of the subject and the template). The multi-resolution framework is in general unnecessary, in that the predicted deformation is at the high resolution already. Given the initializing predicted deformation and the following effec-

tive refinement, the final deformation is able to register the subject with the template more accurately, compared to the traditional direct registration scheme.

The rest of this chapter is organized as follows. In Section 2.2, more details related to image-scale guidance are provided. Then, patch-scale guidance is proposed and elaborated in Section 2.3. A brief summary is available in Section 2.4 to conclude this chapter.

2.2 Image-Scale Guidance

Incorporating the intermediate images in the population alleviates many of the challenges encountered by the direct registration of the subject with the template. A critical problem to be resolved here is the identification of the source of the guidance, or the optimal intermediate image for a certain subject. In particular, two criteria need to be satisfied to qualify the intermediate image, which are that it (1) should be similar to the subject in appearances and (2) is registered with the template already. The image-scale guidance is often utilized in two directions (Jia *et al.*, 2012b), i.e., (1) *selecting* the intermediate image from the population to best approximate the subject, and (2) *generating* the subject-specific intermediate image via simulation.

2.2.1 Intermediate Image: Selection

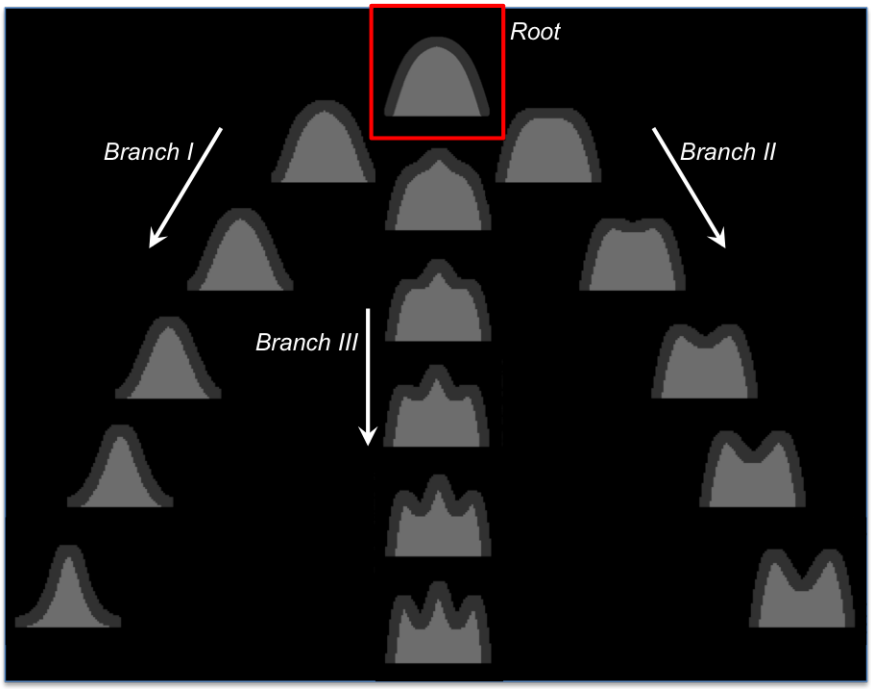
The optimal intermediate image can be cherry-picked from the image population (Dalal *et al.*, 2010; Hamm *et al.*, 2010; Jia *et al.*, 2011b; Munsell *et al.*, 2012; Wolz *et al.*, 2010). The selected intermediate image is expected to approximate the subject in appearances, such that the deformation pathways from the two images to the common template (on the manifold of the population) are also very similar to each other. In

the literature, simple metrics (e.g., SSD and MI) are often adopted to capture image similarity. More sophisticated measures associate the similarity with the deformation field between images (Beg *et al.*, 2005; Joshi *et al.*, 2004). The manifold of images is then approximated by graph-based structures, where nodes correspond to individual images and edges reflect similarities between images. Different from the full connection of all pairs of images, the graph-based structure only allows similar images to be connected and thus registered directly. The deformation pathway between images that are not directly linked is estimated by concatenating all edges of the shortest path between the two images in the graph-based structure. Thus, for a to-be-registered subject, its deformation is predictable given its path towards the template in the graph-based structure. In other words, the intermediate images, which correspond to nodes along the path from the subject to the template, provide guidance to register the subject under consideration.

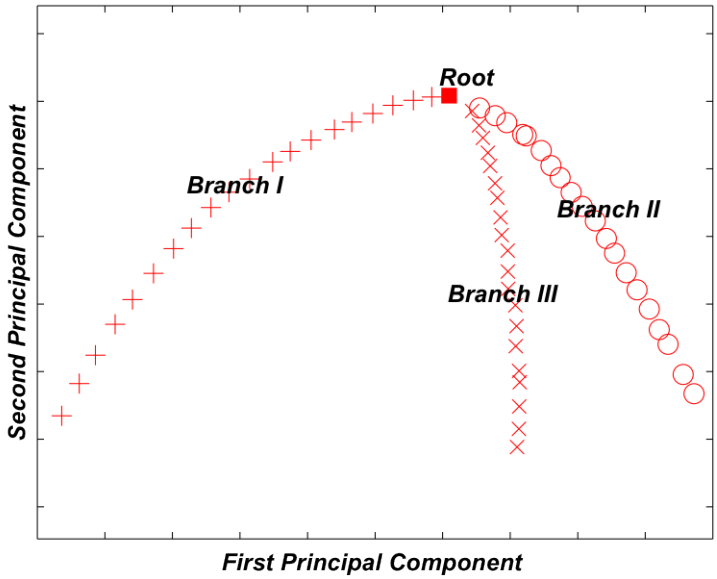
A simulated image population is used here to demonstrate the solution above. As in Figure 2.2(a), the population consists of an arbitrary root image and three branches, each of which consists of 20 images. For convenience, only a limited number of images (including the root and samples of each branch) are shown in the figure. The root image is featured with a bright area and a relatively dark layer, which mimic the appearances of WM and GM in human brains, respectively. Then, within each branch, a sequence of images is derived by consistently altering the cortical folding pattern in the root. For example, 20 images are simulated in Branch I, such that all of them share the same pattern of a single fold while their WM volumes decrease gradually. Similarly, Branches II and III consist of simulated images that are assigned with bi-folding and tri-folding patterns, respectively, while the WM volumes are also decreasing from the root to the end of each branch. The three branches generally reflect possible cortical folding patterns in human brains.

With SSD to describe the distance between each pair of images, a minimum spanning tree (MST) is instantiated in Figure 2.2(a) to model the distribution of the entire population. The root of the tree serves as the template, to which all other simulated images need to be registered. The distribution in the tree can be verified via principal component analysis (PCA). In Figure 2.2(b), all images in the population are projected on a 2D plane, which is spanned by the first two principal components identified in PCA. The root and the three branches, which lead to the structure of the tree, are clearly observable in the PCA output.

Assume that the image at the end of Branch II is the subject under consideration. The direct registration of the subject with the template (i.e., the root) can easily fail (Jia *et al.*, 2012a). Figure 2.3(b) shows the deformation yielded by the direct registration, as well as the appearance of the deformed subject. The deformation is obviously unable to describe the evolution from the root to the subject in the simulation process, such that the warped subject contains unacceptable artifacts. By contrast, with image-scale guidance from the intermediate image(s), the registration of the subject with the template can be well solved (c.f. Figure 2.3(a)). In particular, the subject identifies its optimal intermediate image as the one next to itself on Branch II. The intermediate image further seeks guidance from its own intermediate image, which is even closer to the template. Therefore, the registration of the subject with the template is essentially decomposed into a series of sub-tasks in the recursive manner, while each sub-task happens between similar images and thus is relatively easy to solve. The final deformation for the subject in Figure 2.3(a) is acquired by integrating deformation fields associated with all sub-tasks.



(a)



(b)

Figure 2.2: The population of simulated images in Section 2.2.1: (a) Three branches of images are simulated by altering the folding pattern in the root, and then organized into the tree structure; (b) The distribution of the population is verified by projecting all images on the 2D plane via PCA.

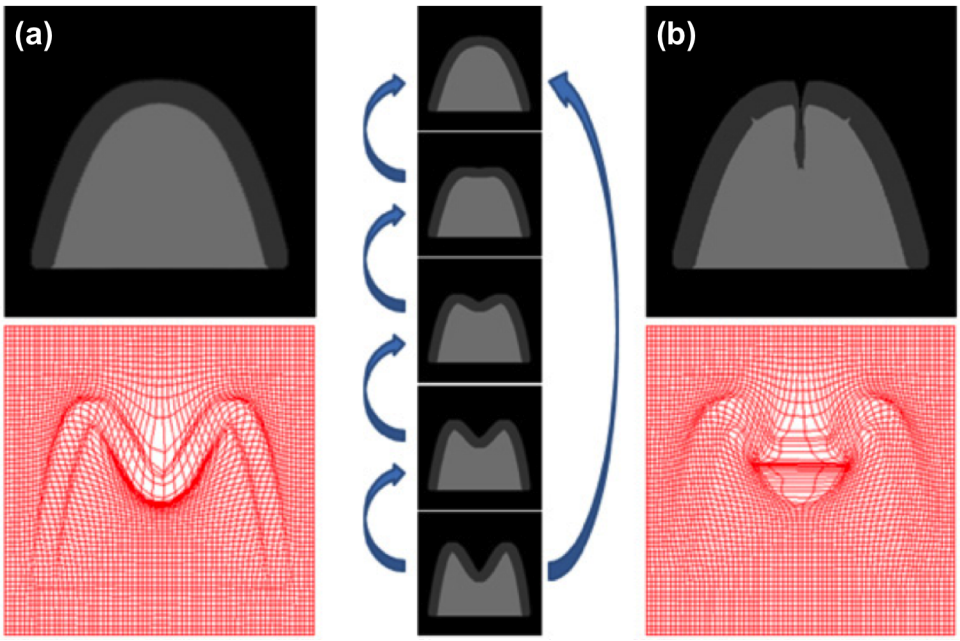


Figure 2.3: The direct registration of the subject leads to a failure in (b), while the subject essentially locates the correct deformation pathway with the guidance from the intermediate images in (b).

2.2.2 Intermediate Image: Generation

Instead of selecting the intermediate image from the population, several models are proposed to generate the intermediate image for a certain subject to be registered with the template (Chou *et al.*, 2013; Kim *et al.*, 2012; Tang *et al.*, 2009). For example, Kim *et al.* (2012) proposed to capture the correlation between appearance-derived features of training images and their deformations in registering with the common template. The correlation model is acquired via support vector regression (SVR) (Drucker *et al.*, 1997). When a new subject image comes in the testing stage, the pre-trained model is able to prompt the deformation field according to the appearance features of the subject. After inverting the deformation provided by the regression model, the intermediate image is generated by warping the template accordingly. Since the deformation pathway between the simulated intermediate image and the template is known as ground truth, the only remaining thing for the subject is to find its deformation towards the intermediate

image. Moreover, the intermediate image is highly similar to the subject in appearance, in that the deformation field for its generation encodes the appearance of the subject already. As a result, the registration of the subject to the intermediate image is easy to solve, while the entire deformation pathway from the subject to the template can be integrated and refined afterwards.

2.3 Patch-Scale Guidance

2.3.1 Motivation

It is effective to improve the performances (i.e., robustness and accuracy) in registering the subject with the template by introducing additional intermediate images in the population. The utilization of the image-scale guidance, as in Section 2.2, considers the entire image as a whole. A fundamental assumption for the image-scale guidance is that images with similar appearances should have similar deformations when registered with the same template. The assumption is closely related to the manifold instantiated for describing the large-scale population of images (e.g., Figure 2.1). Though geodesics upon the manifold are perceived to represent deformation pathways between neighboring images, the theory is challenged especially by the high-dimensional images and deformations. In fact, it is known that a tiny perturbation in the appearance of the images (and thus their mapped locations on the manifold) may possibly result in a very different deformation for registration (Modersitzki, 2004).

High anatomical variations (i.e, sulcal folding) are prevalent in brain MR images, impairing an accurate determination of the optimal intermediate image and then the proper utilization of the image-scale guidance. It is often questionable to conclude whether a certain intermediate image is similar to the subject in approximating both the appearances and the deformation field, requiring a non-trivial task to evaluate their

similarity. The subject might share common anatomical structures with the intermediate image in certain gyri or sulci, where the guidance from the intermediate image is trustworthy; at the same time, they may also differ significantly in other areas where the guidance could even be harmful. A simple illustration is provided in Figure 2.4, where a new subject is simulated (and highlighted by the blue rectangle) in addition to the dataset in Section 2.2.1 and Figure 2.2. The new subject differs significantly in appearance from any other image in the population, implying that the intermediate image can hardly be acquired for providing guidance at the scale of the entire image.

In contrast to the image-scale guidance, the patch-scale guidance (Wang *et al.*, 2013b) allows the subject to utilize multiple intermediate images at the scale of patches, though, for its own registration with the template. Meanwhile, the patch-scale guidance is interpretable in the sense of *correspondence detection*, which leads to a rigorous formulation. Specifically, if two respective patches from the subject and a certain intermediate image are similar in appearance, the center points of the two patches are generally perceived as *correspondences* to each other. Since the intermediate image is presumed to be registered with the template already, the point in the intermediate image is able to identify its correspondence in the template image space. Then, bridged by the point in the intermediate image, the correspondence between the subject point and the template point is established as both of them identify the same intermediate point as their correspondences. The subject-template correspondence gives a hint of the deformation field at the locations of specific points. With more points in the image space and their associated subject-template correspondences, the entire deformation field for registering the subject with the template is essentially predictable. In general, the collection of intermediate images contributes to the registration of the subject with the template, by providing the patch-scale guidance that bridges point-to-point correspondences between the two images. The guidance is allowed to originate from

different intermediate images for individual points in the image space.

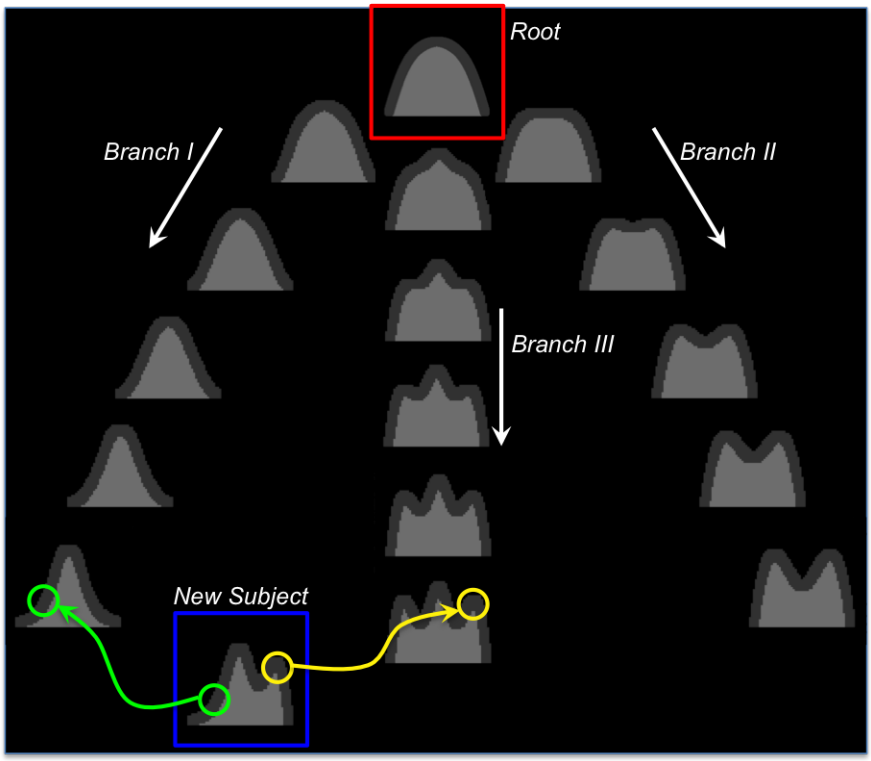
Although it is hard to utilize image-scale guidance for the specific new subject in Figure 2.4, patch-scale guidance helps its registration with the template due to the high flexibility in accessing the intermediate images. In particular, due to a certain point on the left half of the new subject, its correspondence can be easily identified from the intermediate image at the end of Branch I, e.g., by comparing appearances within the two respective patches circled in green (c.f. Figure 2.4(a)). The deformation associated with the subject point under consideration is then predictable, in that the subject point shares a common correspondence in the template space with the identified point in the intermediate image. Meanwhile, for the point in the right half of the subject image (e.g., encircled by the yellow patch), its correspondence in the template, as well as the associated deformation, is also revealed with respect to the guidance from another intermediate image (i.e., at the end of Branch III). Note that the patch-based correspondence detection is conducted within the neighborhood of each point-of-interest, since only local correspondence is meaningful in non-rigid brain MR image registration.

2.3.2 The Prediction-Reconstruction Protocol

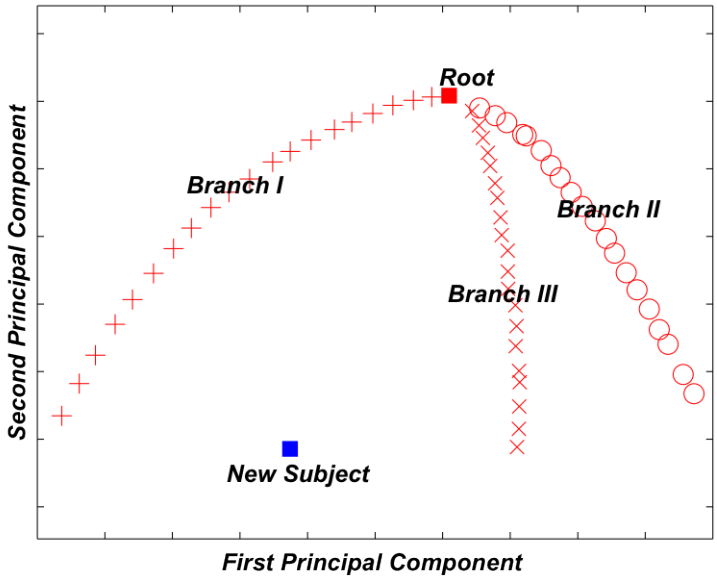
To effectively utilize the patch-scale guidance from the collection of the intermediate images and apply it towards brain MR image registration, a novel prediction-reconstruction strategy, namely *the P-R protocol*, is proposed in Wang *et al.* (2013b). The P-R protocol consists of two coupled steps:

1 *Predict* the deformations associated with a subset of key points, which are scattered in the image space to cover the entire brain volume;

2 *Reconstruct* the dense deformation field based on all key points and their predicted



(a)



(b)

Figure 2.4: The population of simulated images in Section 2.3 incorporated with an additional subject as highlighted by the blue rectangle in (a). The new subject is registered with the root by utilizing patch-scale guidance. The distribution of all images revealed by PCA is shown in (b).

deformations to register the subject image with the template.

In the *prediction* step, it is critical to establish point-to-point correspondences between the subject and the template, by utilizing the highly reliable correspondences identified between the subject and the intermediate images. To this end, the patch-based sparsity learning technique, which is widely applied in computer vision (Wright *et al.*, 2010), is applied here. For a specific subject point, it aims to estimate the linear and sparse representation of its surrounding patch given all possible candidate patches from the intermediate images. The optimal linear representation determined by the sparsity learning locates a sparse set of intermediate patches, the appearances of which are highly similar to the subject patch under consideration. Therefore, all center points of the intermediate patches qualified by the sparsity learning can be regarded as the correspondence candidates of the subject point, and then help identify the correspondence of the subject point in the template image. The subject-template point correspondence predicts the (local) deformation, following which the specific subject point is expected to deform.

The prediction procedure is applied to a subset of selected key points, each of which typically identifies several possible correspondences between the subject and the template via sparsity learning. In other words, multiple predictions of the deformation associated with a certain key point are often collected. All key points and the multiple predictions of their individual deformations are further integrated for the *reconstruction* of the dense deformation field across the entire image space. That is, varying confidences of individual key points and all of their predicted deformations are computed. Then an adaptive interpolation approach, which is based on a special family of compact-support radial basis functions (RBFs), is applied to reconstruct the dense deformation field. The reconstruction considers the computed confidences of predictions, as the predicted deformation of higher confidence plays a more important role. Meanwhile,

the reconstructed deformation field complies with the smoothness constraint, in order to suppress unrealistic warping (i.e., folding) of brain tissues.

2.3.3 Prediction of Correspondences

It is critically important to predict the correspondence, as well as the associated deformation, of each key point in the image space. For convenience, mathematical notations with respect to the P-R protocol are introduced first. In particular, the subject S needs to be registered with the fixed template image T following the deformation $\phi(\cdot) : \Omega_T \rightarrow \Omega_S$, while the point $x \in \Omega_T$ in the template space locates its correspondence at $\phi(x) \in \Omega_S$ in the subject image space. Conversely, $\phi^{-1}(\cdot)$ is capable of deforming the template towards the subject image space. To help estimate $\phi(\cdot)$, the patch-scale guidance is contributed by the collection of the intermediate images $\{M_i | i = 0, \cdots, m\}$. Given each M_i, the deformation $\psi_i(\cdot)$ that registers it with the template is already known. That is, $\psi_i(x) \in \Omega_{M_i}$ indicates the correspondence of the template point $x \in \Omega_T$. Note that the template T is also referred as M_0, as $\psi_0(\cdot)$ is simply an identity transform that registers the template to itself. For easy reference, a list of important variables here is summarized in Table 2.1. Moreover, since only the prediction of the non-rigid deformation is investigated here, all images are necessarily pre-processed, including being aligned to a common space by affine registration (i.e., FLIRT (Jenkinson and Smith, 2001; Jenkinson *et al.*, 2002)).

The deformation is predictable in that point-to-point correspondences can be identified from similar patch-scale appearances between images. Figure 2.5 helps illustrate the rationale of the patch-scale guidance. In Fig. 2.5(a) particularly, three individual patches (in the top-bottom order) from the template T, a certain intermediate image M_i, and the subject S are enumerated. All three patches are represented by circles, while their center points are noted by $x \in \Omega_T$, $\tilde{y} \in \Omega_{M_i}$, $\tilde{z} \in \Omega_S$, respectively. Without

Variable	Note
$\mathcal{T}$	Template image
$\mathcal{S}$	Subject image
$\mathcal{M}_i$	The i-th intermediate image (i is the index)
$\Omega_\mathcal{T}, \Omega_\mathcal{S}, \Omega_{\mathcal{M}_i}$	Individual image spaces
x	Template point
$y, \tilde{y}$	Points in the intermediate images
$z, \tilde{z}$	Points in the subject image
t	Resolution
$\phi(\cdot)$	Transformation field to register $\mathcal{S}$ with $\mathcal{T}$
$\psi_i(\cdot)$	Transformation field to register $\mathcal{M}_i$ with $\mathcal{T}$
$\vec{\phi}, \vec{\psi_i}$	Signature vectors of $\phi^t(x)$ and $\psi_i(x)$
$\vec{u}, u_i$	Confidence of $\psi_i(x)$ in prediction
$\vec{\xi}, \vec{\theta}_{ij}$	Signatures of patches at $\phi^{t-1}(x)$ and y_{ij}
$\vec{v}, v_{ij}$	Confidence of y_{ij} in prediction
r_c	Maximal radius allowed in correspondence detection
$\vec{w}, w_{ij}, \mathbf{W}$	Confidences of combined predictions
$k(\cdot), \mathbf{K}$	RBF kernel function and kernel matrix
σ, c	Control the size of the support of $k(\cdot)$
$\vec{\gamma}_x, \mathbf{\Gamma}$	RBF kernel coefficients
$\mathbf{\Phi}$	Predicted transformations

Table 2.1: Summary of important variables with respect to the Prediction-Reconstruction protocol.

losing generality, x and $\tilde{y}$ are defined to be correspondences to each other such that $\tilde{y} = \psi_i(x)$ or $x = \psi_i^{-1}(\tilde{y})$. Immediately this leads to

Proposition 1 *If $\tilde{y} \in \Omega_{M_i}$ and $\tilde{z} \in S$ are correspondences to each other, then*

$$\phi^{-1}(\tilde{z}) = \psi_i^{-1}(\tilde{y}). \tag{2.1}$$

Proof 1 *Given the facts that (1) $\tilde{y}$ and $\tilde{z}$ are correspondences to each other, (2) x and $\tilde{z}$ are correspondences to each other, the correspondence relationship could be established between x and $\tilde{y}$. Here, the correspondence is defined such that any two points under*

consideration share similar appearances in their surrounding patches. Therefore, the proposition is derived following (1) $x = \psi_i^{-1}(\tilde{y})$ and (2) $x = \phi^{-1}(\tilde{z})$ accordingly.

The proposition above allows the prediction of $\phi^{-1}(\cdot)$ from the inverse of the collection $\{\psi_i(\cdot)\}$. The model is further improved to predict $\phi(\cdot)$ directly from $\psi_i(\cdot)$, instead of its inverse. As in Figure 2.5(b), there is

Proposition 2 *If (1) $y \in \Omega_{M_i}$ and $z \in S$ are correspondences to each other AND (2) y is spatially close to $\psi_i(x)$, then*

$$\phi(x) - z = \nabla\phi(x) \left(\nabla\psi_i(x)\right)^{-1} \left(\psi_i(x) - y\right), \tag{2.2}$$

where ∇ indicates the Jacobian operator.

Proof 2 *As y is spatially close to $\psi_i(x)$, it is natural to assume that $y = \psi_i(x + \delta x)$ where δx is a infinitesimal perturbation of x. Moreover, it is implied by $\psi_i(\cdot)$ that the point $(x + \delta x)$ locates its correspondence as y. Thus the points $(x + \delta x)$ and z are also correspondences to each other, via the bridge of $y \in \Omega_{M_i}$:*

$$y = \psi_i(x + \delta x) = \psi_i(x) + \nabla\psi_i(x)\delta x + \mathcal{O}(\delta x^T \delta x), \tag{2.3}$$

$$z = \phi(x + \delta x) = \phi(x) + \nabla\phi(x)\delta x + \mathcal{O}(\delta x^T \delta x). \tag{2.4}$$

After subtracting the two equations and eliminating the perturbation variable δx, it leads to the conclusion

$$\left(\nabla\psi_i(x)\right)^{-1} \left(y - \psi_i(x)\right) = \left(\nabla\phi(x)\right)^{-1} \left(z - \phi(x)\right), \tag{2.5}$$

from which (2.2) is derived.

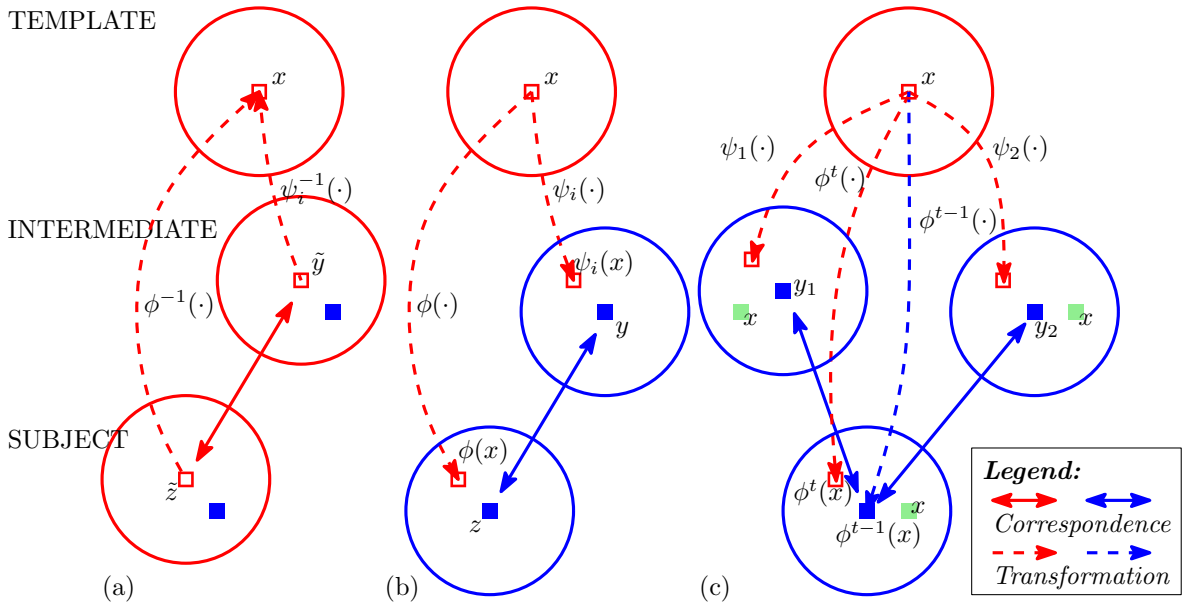


Figure 2.5: Illustration of the predictability of the deformation: (a) The correspondence between the template point x and the subject point $\tilde{z}$ is established as both points identify $\tilde{y}$ as their correspondence in the intermediate image; (b) The subject deformation $\phi(x)$ is predictable from the intermediate deformation $\psi_i(x)$, if y and z are correspondences to each other; (c) Multiple correspondence candidates of y might be detected, thus resulting in multiple predictions upon the subject deformation $\phi(x)$.

The rule in Proposition 2 allows the prediction of $\phi(x)$ from the intermediate collection $\{\psi_i(x)\}$ of the intermediate images. All variables (i.e., i, y, x, and z in (2.2)) need to be handled carefully for brain MR image registration. In particular, a set of key points is selected from the template image space, with more details regarding the selection of key points provided in Section 2.3.5. Each key point is then fed as an instance of the variable x to (2.2). Furthermore, z is related to the tentative estimation of $\phi(x)$; thus (2.2) is converted to the incremental optimization style as

$$\phi^t(x) := \phi^{t-1}(x) + \nabla\phi^{t-1}(x)\left(\nabla\psi_i(x)\right)^{-1}\left(\psi_i(x) - y\right). \tag{2.6}$$

Here, t records the timing in optimizing $\phi(\cdot)$. The term $\nabla\phi^{t-1}(x)$ is used to approximate the Jacobian of $\phi^t(x)$ by assuming that $\phi^t(x)$ can only be generated by changing

$\nabla\phi^{t-1}(x)$ mildly. The variables i (as well as $\psi_i(x)$) and y will be determined later, such that $y \in \Omega_{M_i}$ is the correspondence to the previously estimated $\phi^{t-1}(x) \in \Omega_S$ given the patch-scale appearances of the two center points.

Given the key point x, several intermediate images with their individual contributions as $\psi_i(x)$ might be available. Moreover, multiple instances of y can potentially be identified as correspondences to $\phi^{t-1}(x)$ as well. Though the number of correspondences can be arbitrarily reduced to 1 for each instance of x, allowing multiple correspondences can greatly improve the robustness and the accuracy for correspondence detection (Chui and Rangarajan, 2003). To this end, the sparsity learning technique (Wright *et al.*, 2010) is applied for the determination of i and y, respectively. In general, the sparsity learning technique allows multiple, yet only a limited number of, instances of $\psi_i(x)$ and y to contribute to the prediction in (2.6). Meanwhile, varying confidences are attained for the instances of $\psi_i(x)$ and y that are active in the prediction. The product of the confidences of $\psi_i(x)$ and y is further regarded to measure the reliability of the resulted predictions. All key points and their multiple predicted deformations, along with the varying confidences, are passed to the reconstruction of the dense deformation field in Section 2.3.4.

Determination of i and $\psi_i(x)$

The correspondence between $\phi^{t-1}(x)$ and y implies that the locations of the two points should be close to each other, especially in brain MR images after affine registration. Therefore, $\psi_i(x)$ better predicts $\phi^t(x)$ if the two deformations are more similar. In Fig. 2.5(c), for example, the point $\phi^{t-1}(x)$ is assumed to identify its correspondence y_1 from M_1 and another correspondence y_2 from M_2. However, the challenges in determining y_1 and y_2 are different concerning $\psi_1(x)$ and $\psi_2(x)$. As $\psi_2(x)$ is closer to $\phi^t(x)$ than $\psi_1(x)$ (in reference to the marked location of x), the correspondence detection for y_1

thus should be conducted in a much larger area in that $\|y_1 - \phi^{t-1}(x)\| > \|y_2 - \phi^{t-1}(x)\|$. In this case, $\psi_2(x)$ is obviously a better selection for the sake of predicting $\phi(x)$.

In order to determine a $\psi_i(x)$ that is similar to $\phi^t(x)$ and compute the accompanying confidence, the sparse representation of $\phi^t(x)$ over the dictionary spanned by $\psi_i(x)$ is investigated. Assuming that $\phi^t(x)$ and $\psi_i(x)$ are signified by the vectors $\vec{\phi}$ and $\vec{\psi}_i$, respectively, the sparsity learning aims to solve

$$\begin{aligned} \vec{u} = \quad & \arg\min_{\vec{u}} \|\vec{\phi} - \mathbf{\Psi}\vec{u}\|^2 + \alpha\|\vec{u}\|_1, \\ \text{s.t.} \quad & \mathbf{\Psi} = [\vec{\psi}_1, \cdots, \vec{\psi}_i, \cdots, \vec{\psi}_M], \\ & \vec{u} = [u_1, \cdots, u_i, \cdots, u_M]^T, \\ & u_i \geq 0, \forall i. \end{aligned} \tag{2.7}$$

Here, $\vec{u}$ indicates the vector of the coefficients for the linear representation of $\vec{\phi}$ given the dictionary $\mathbf{\Psi}$, which consists of potential contributions from all intermediate images. The l_1 constraint $\|\vec{u}\|_1$, weighted by the non-negative scalar α, favors a sparse subset of column items from $\mathbf{\Psi}$ to represent $\vec{\phi}$. The coefficient u_i, yielded by the sparsity learning, also acts as a similarity indicator between $\vec{\phi}$ and $\vec{\psi}_i$ (Wright *et al.*, 2010).

To attribute signatures to both $\phi^t(x)$ and $\psi_i(x)$, the corresponding deformations are vectorized into $\vec{\phi}$ and $\vec{\psi}_i$, respectively. On the other hand, $\vec{\phi}$ cannot be acquired directly, in that $\phi^t(x)$ is still pending for estimation. As an alternative, the signature $\vec{\phi}$ is generated from $\phi^{t-1}(x)$, based on the assumption of mild changing between $\phi^{t-1}(x)$ and $\phi^t(x)$. In general, via the optimization in (2.7), several intermediate images are identified (or *activated*), with their contributions $\{\psi_i(x)\}$ and the non-negative coefficients $\{u_i\}$, for the sake of the prediction of $\phi^t(x)$. The term u_i is further regarded as the measure of the confidence in predicting $\phi^t(x)$ from $\psi_i(x)$.

Determination of y

Candidates of y can be identified via the correspondence detection, centered at the location of $\phi^{t-1}(x)$, within each active intermediate image after the determination of i. For convenience, all possible candidates of y are annotated by $\{y_{ij}\}$, as y_{ij} indicates the j-th candidate from the i-th intermediate image. The candidate collection $\{y_{ij}\}$ often consists of each grid point $y_{ij} \in \Omega_{M_i}$ if $\|\phi^{t-1}(x) - y_{ij}\| \leq r_c$ and r_c is the maximally allowed radius for correspondence detection. The signatures of the points $\phi^{t-1}(x)$ and y_{ij} are defined as $\vec{\xi}$ and $\vec{\theta}_{ij}$, respectively, as the similarity between the two points can thus be acquired by comparing their signature vectors. In particular, the same sparsity learning technique is used for the purpose

$$\begin{aligned} \vec{v} = \quad & \arg\min_{\vec{v}} \|\vec{\xi} - \mathbf{\Theta}\vec{v}\|^2 + \beta\|\vec{v}\|_1, \\ \text{s.t.} \quad & \mathbf{\Theta} = [\cdots, \vec{\theta}_{1j}, \cdots, \vec{\theta}_{ij}, \cdots, \vec{\theta}_{Mj}, \cdots], \\ & \vec{v} = [\cdots, v_{1j}, \cdots, v_{ij}, \cdots, v_{Mj}, \cdots]^T, \\ & v_{ij} \geq 0, \forall i, j. \end{aligned} \tag{2.8}$$

In (2.8), the matrix $\mathbf{\Theta}$ indicates the dictionary of contributions from the candidate collection $\{y_{ij}\}$, the vector $\vec{v}$ records the coefficients for the linear sparse representation of $\vec{\xi}$ given $\mathbf{\Theta}$, and the non-negative scalar β controls the sparsity of $\vec{v}$. The vectorized patch is defined as the signature (i.e., $\vec{\theta}_{ij}$) for the center point (i.e., y_{ij}). Then, v_{ij} captures the similarity between the two patches centered at $\phi^{t-1}(x)$ and y_{ij}. Higher v_{ij} obviously implies that the correspondence between $\phi^{t-1}(x)$ and y_{ij} is more reliable given their individual patch-scale appearances. As the result, v_{ij} is regarded as the confidence for predicting $\phi^t(x)$ from y_{ij}.

The correspondence detection via (2.8) is independent of x, and thus may cause inconsistent outputs for points that are even neighboring each other. To this end, the

consistency in correspondence detection is enforced via the $l_{2,1}$-norm constraint (Liu *et al.*, 2009). In particular, the optimization problem in (2.8) is modified as

$$\begin{aligned} \mathbf{V} = \quad & \arg\min_{\mathbf{V}} \|\mathbf{\Xi} - \mathbf{\Theta}\mathbf{V}\|^2 + \beta\|\mathbf{V}\|_{2,1}, \\ \text{s.t.} \quad & \mathbf{\Xi} = [\cdots, \vec{\xi}_{x+\Delta}, \cdots], \|\Delta\| \leq \epsilon, \\ & \mathbf{\Theta} = [\cdots, \vec{\theta}_{1j}, \cdots, \vec{\theta}_{ij}, \cdots, \vec{\theta}_{Mj}, \cdots], \\ & \mathbf{V} = [\cdots, \vec{v}_{x+\Delta}, \cdots], \|\Delta\| \leq \epsilon, \\ & v_{x+\Delta,ij} \geq 0, \forall\Delta, \forall i, \forall j. \end{aligned} \tag{2.9}$$

(2.9) aims to detect correspondence candidates for the centering point x, as Δ in the subscript is associated with the point $(x + \Delta)$ that is neighboring x. Similarly, $\vec{\xi}_{x+\Delta}$ represents the signature vector for $(x + \Delta)$. The matrix $\mathbf{\Xi}$ captures signatures for all points that are located within the radius of ϵ to the point x, while their representation coefficients on $\mathbf{\Theta}$ are stored in individual columns of the matrix $\mathbf{V}$. Identical to (2.8), the coefficient vector for $(x + \Delta)$, namely $v_{x+\Delta,ij}$ is encouraged to be sparse. Meanwhile, neighboring points are inclined to share similar coefficients, as their patch-scale appearances could not change drastically. Therefore, besides the l_1 constraint, the $l_{2,1}$ constraint to the matrix $\mathbf{V}$ (Liu *et al.*, 2009) is enforced. That is, each column of $\mathbf{V}$ satisfies the sparsity requirement, while the sparsity patterns of individual columns are expected to be highly similar. Finally, the column for $\Delta = 0$ in $\mathbf{V}$ tells all possible correspondence candidates of the point $\phi^{t-1}(x)$.

Any arbitrary combination of $\psi_i(x)$ and y_{ij} yields an attempt in predicting $\phi^t(x)$. In particular, the confidence w_{ij} is defined for the attempt as the product of the confidences of $\psi_i(x)$ and y_{ij}, or $w_{ij} = u_i v_{ij}$. The sparsity enforced in selecting $\psi_i(x)$ and y_{ij} results in multiple, but a limited number of, predictions with non-zero confidences. In this

way, the method can (1) avoid local minima if only acquiring a single but incorrect prediction for the key point, and (2) suppress a majority of predictions of low reliability. The confidence of each key point is further normalized by $w_{ij} \leftarrow w_{ij} / \sum w_{ij}$, to impose the same priors on all key points.

The order of determining i first and then y provides good scalability in handling a large-scale population of (intermediate) images. Specifically, the determination of i can be much more efficient than y, in that the column size of the dictionary $\mathbf{\Psi}$ is identical with the number of intermediate images, or $\mathcal{O}(m)$. Meanwhile, multiple correspondences may exist in even a single intermediate image. The dictionary $\mathbf{\Theta}$ has to enumerate all possible instances of y, and thus increases the column size to $\mathcal{O}(r_c^3 m)$, as r_c represents the radius in searching for correspondences. By determining i first, the number of *activated* intermediate images, only from which the contributions to the determination of y should be counted, is well controlled. Thus, the complexity in determining y is well scaled regardless of the size of the collection of the intermediate images, as most intermediate images are *deactivated* already in the determination of y.

2.3.4 Reconstruction of Deformation Field

The dense deformation field is reconstructed to fit the multiple predictions of all key points. To this end, the radial basis function (RBF) provides a feasible means for the representation of the deformation field. Suppose that the RBF kernel function is $k(\cdot)$ and $\vec{\gamma}_x$ is the radial basis function (RBF) coefficient vector for the key point x, the dense field associated with the arbitrary location $x' \in \Omega_T$ is then computed by

$$\phi(x') = \sum_x k\left(\|x' - x\|\right) \vec{\gamma}_x. \tag{2.10}$$

The kernel matrix $\mathbf{K}$ is further defined, in which the entry at the junction of the m-

th row and the n-th column is calculated by feeding the Euclidean distance between the m-th and the n-th key points to the kernel function $k(\cdot)$. If only a single prediction was ever attempted for each key point, the residuals for the dense field to fit the predicted deformations of all key points could then be easily computed in the matrix form as $\|\mathbf{\Phi} - \mathbf{K\Gamma}\|^2$. Here, the predicted deformation (in the transposed row vector form) of the m-th key point is recorded in the m-th row of $\mathbf{\Phi}$ and its transposed RBF coefficient vector in the m-th row of $\mathbf{\Gamma}$.

In order to accommodate multiple predictions of each key point, the expanded matrix $\mathbf{\Phi}$ and the confidence matrix $\mathbf{W}$ are introduced for fitting. All predictions, as well as the confidences, are enumerated in $\mathbf{\Phi}$ and $\mathbf{W}$. Supposing that the p-th row of $\mathbf{\Phi}$ records a certain prediction for the m-th key point weighted with the confidence w_{ij}, the entry of $\mathbf{W}$ at the junction of the p-th row, and the m-th column is set to w_{ij} while all other entries in the p-th row are zero. The overall residuals in fitting predictions, weighted by varying confidences, then become $\|\mathbf{\Phi} - \mathbf{WK\Gamma}\|^2$.

Smoothness regularization is critically important to the reconstruction of the dense field, in order to suppress any unrealistic warping that might be applied to brain tissues (Rueckert and Schnabel, 2011). To this end, the kernel function $k(\cdot)$ is usually designed in the style of low-pass filters (Myronenko and Song, 2010). Further, if $\mathbf{K}$ is positive-definite, the regularization can be attained by solving (Girosi *et al.*, 1995)

$$\mathbf{\Gamma} = \arg\min_{\mathbf{\Gamma}} \|\mathbf{\Phi} - \mathbf{WK\Gamma}\|^2 + \lambda \operatorname{Tr}\left(\mathbf{\Gamma}^T \mathbf{K\Gamma}\right), \tag{2.11}$$

where λ controls the strength of the smoothness constraint. The RBF coefficients $\mathbf{\Gamma}$, which are needed to generate the dense deformation field according to (2.10), are thus solvable in the following

$$\left(\mathbf{K} + \lambda \left(\mathbf{W}^T \mathbf{W}\right)^{-1}\right) \mathbf{\Gamma} = \left(\mathbf{W}^T \mathbf{W}\right) \mathbf{W}^T \mathbf{\Phi}. \tag{2.12}$$

In (2.12), $\mathbf{W}^T\mathbf{W}$ is a positive-definite diagonal matrix, where the m-th diagonal entry equals the sum of squares of the confidences for all predictions for the m-th key point.

The kernel $k(\cdot)$ is designed such that $\mathbf{K}$ is positive definite and $k(\cdot)$ has low-pass response. Abundant choices of RBF kernels are available, e.g., the thin plate splines (TPS) with polynomial decay in frequency domain (Bookstein, 1989; Chui and Rangarajan, 2003). Most RBF kernels, however, are globally supported, leading to a very dense matrix $\mathbf{K}$ and thus suffering from scalability and numerical instability. As an alternative, the compactly supported kernel (Genton *et al.*, 2001) is used for the reconstruction of the deformation field

$$k(\|x' - x\|) = \begin{cases} \left(1 - \dfrac{\|x' - x\|^2}{c}\right) \cdot \exp\left(-\dfrac{\|x' - x\|^2}{2\sigma^2}\right), & \text{if } \|x' - x\| \leq c; \\ 0, & \text{if } \|x' - x\| > c. \end{cases} \tag{2.13}$$

The kernel $k(\cdot)$ is obviously a truncated Gaussian, as it is set to 0 if beyond the compact support ($\|x' - x\| > c$). The resulted kernel matrix $\mathbf{K}$ is sparse and thus benefits solving (2.12).

To alleviate the concern over the optimal parameters of the kernel, the multi-kernel strategy (Floater and Iske, 1996) is further applied to recursively reconstruct the deformation field. To derive a set of RBF kernels $k_h(\cdot)$, σ is fixed in (2.13) and c is adjusted. The size of the compact support for $k_h(\cdot)$, denoted by c_h, is defined following $c_h = c_{h-1}/2$. The reconstruction starts with the kernel c_1. Then, the residuals after using the kernel $k_{h-1}(\cdot)$ are further fitted by the kernel $k_h(\cdot)$, which shows better capability in modeling deformations at higher frequencies. The iterative procedure terminates when the stopping criterion is met, i.e. the residual $\|\mathbf{\Phi} - \mathbf{WK\Gamma}\|^2$ is tiny enough, or the number of allowed kernels is exhausted. In the end, the dense deformation field is represented by integrating contributions from all kernels involved.

2.3.5 The Prediction-Reconstruction Hierarchy

The P-R protocol can be naturally embedded into a hierarchical framework in order to better tackle the high complexity in brain MR image registration. The hierarchy gives a schematic solution that supports multi-resolution optimization of the deformation field. That is, the deformation field predicted in an earlier level initializes the next level of the higher resolution. In particular, by relating the variable t in (2.6) to the low-middle-high resolutions, the P-R hierarchy is summarized as follows:

Load $\mathcal{T}$, $\mathcal{S}$, $\{\mathcal{M}_i\}$, and $\{\psi_i(\cdot)\}$;
Initialize $\phi(\cdot)$ to the identity transform;
Select a set of template key points $\mathbb{X} \subseteq \Omega_{\mathcal{T}}$;
for $level \in \{1, 2, 3\}$ **do**
    Select a subset of key points $\mathbb{X}_{level} \subseteq \mathbb{X}$;
    for $x \in \mathbb{X}_{level}$ **do**
        Determine i to activate $\psi_i(x)$ ((2.7));
        Determine candidates of y that are correspondences to $\phi^{level-1}(x)$ ((2.9));
        Acquire multiple predictions of $\phi^{level}(x)$ ((2.6));
    end for
    Reconstruct the dense deformation field $\phi^{level}(\cdot)$ (Section 2.3.4);
end for
Save $\phi^3(\cdot)$ as the final output of $\phi(\cdot)$.

The P-R hierarchy above functions similarly to state-of-the-art multi-resolution image registration methods, which are needed for registering all intermediate images with the template and for refining the predicted deformation field. In particular, HAMMER (Shen and Davatzikos, 2002) is used to register all intermediate images to the template, and explicitly match correspondence points for estimating the deformation field in reg-

istration. The resulting deformation fields of the intermediate images are then used for the prediction of the deformation that registers a new subject with the template. The key points are abundant in context information and thus crucial to accurate alignment of neuroanatomical structures. Meanwhile, the set of key points $\mathbb{X}$ can be pre-computed once the template image is fixed. As in HAMMER, the key points are mostly located at the transitions of individual brain tissues (i.e., WM, GM, and CSF). Then, $\mathbb{X}_{level}$ that corresponds to a certain resolution by sampling $\mathbb{X}$ randomly is acquired. The subset of key points $\mathbb{X}_{level}$ enlarges its size gradually when the level increases (i.e., 1.0×10^4 for the size of $\mathbb{X}_1$, 4.0×10^4 for $\mathbb{X}_2$, and 1.6×10^5 for $\mathbb{X}_3$ in the end). Finally, after the P-R hierarchy predicts the dense deformation field that registers the subject with the template, the estimated deformation field is refined, e.g., by feeding the field as initialization and running diffeomorphic Demons (Vercauteren *et al.*, 2009) and HAMMER (Shen and Davatzikos, 2002) at the high resolution only, respectively.

2.3.6 Experimental Results

The P-R hierarchy is applied to both simulated and real populations for the evaluation of its performance. For the sake of refining the deformation field predicted by the P-R hierarchy, two state-of-the-art registration methods, i.e., diffeomorphic Demons (Vercauteren *et al.*, 2009) and HAMMER (Shen and Davatzikos, 2002), are used. The refinements are conducted within the original image resolution (or the *high* resolution) only and follow recommended configurations of the two methods. Details related to the experiments on the individual populations are reported in the following.

Simulated Data.

With the representation in B-Splines, 100 deformation fields are simulated. The simulated image population is then generated by deforming a pre-selected *template* in

accordance to the simulated deformation fields. The template is arbitrarily selected as the fourth image in the LPBA40 dataset (Shattuck *et al.*, 2008). In the pre-processing steps, the template is isotropically resampled to the size of $220 \times 220 \times 184$ and the spacing of $1 \times 1 \times 1\text{mm}^3$. The control points of B-Splines in simulating deformation fields are placed 32mm apart, while the coefficients for control points along all axes are uniformly sampled from -20mm to $+20$mm. Exemplar slices from the template and simulated images are shown in Figure 2.6. From all the simulated images, one is designated as the *subject* while the rest are used to form up the *intermediate* collection. The deformation field that registers each intermediate image to the template is acquired by inverting the field for simulation directly via ITK (`http://www.itk.org`). Moreover, the inverse of the simulated field of the subject is regarded as the *ground truth*, against which the deformation produced in image registration can be quantitatively compared. Further, 10 different subjects are selected from the collection in order to repeat all tests. The results are examined and compared from three perspectives, including the errors of the predicted deformations of the key points, the errors incurred in the reconstruction process, and the errors after the final refinement.

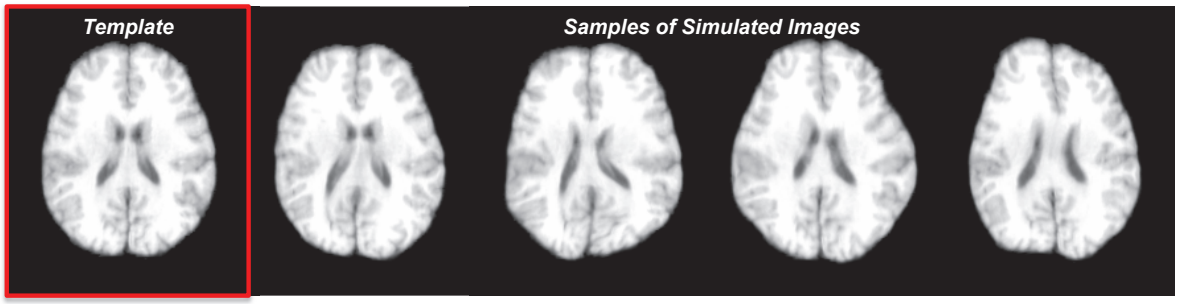


Figure 2.6: The template and samples of simulated images for the evaluation in Section 2.3.6.

Errors of Predicted Deformations of Key Points. The predicted deformation for each key point is examined and the error with respect to the ground truth is

computed. For each subject, the mean error for all key points is first calculated. The errors are further averaged across all 10 tests. The overall prediction error of the P-R hierarchy is 2.116 ± 0.975mm compared to the ground truth. By loosening the requirements for consistency, the model of correspondence detection can be downgraded from (2.9) to the simpler (2.8). Then, the overall error for all key points and all tests increases to 2.304 ± 0.984mm. Therefore, it is concluded that the correspondences for key points can be predicted more accurately by examining the correspondence consistency of neighboring points in the image space.

Errors of Reconstructed Deformation Fields. A set of compactly supported RBF kernels is used for the reconstruction of the deformation field, based on the key points and their previously predicted deformations. Each reconstructed deformation field is compared with the ground truth, in order to evaluate the residual errors between the two fields. First, only a single RBF kernel is used for the reconstruction of the field ($c = 9$mm, $\lambda = 0.05$). The parameters are determined such that (1) no folding occurs in the predicted deformation field due to λ, and (2) the residual error is made minimal (3.115 ± 1.185mm) by manually inspecting outputs of different parameters (i.e., the configuration of c). Then, the complete reconstruction method in Section 2.3.4 is adopted. In particular, three compactly supported kernels ($\lambda = 0.05$) are cascaded, with $c_1 = 10$mm for the first kernel. The overall error is 3.020 ± 1.134mm, implying that the reconstruction accuracy is improved without the need to tune the parameter c. Note that the reconstruction errors are calculated by excluding non-brain areas in the images already. The errors, as well as the standard deviations, for 54 regions of interest (ROIs) are plotted in Fig. 2.7. The index, as well as the specific region of interest (ROI), is referred to in Table 2.2. In Fig. 2.7, the red color represents the errors caused by using a single RBF kernel, while cyan is for the errors caused by using cascaded kernels for the reconstruction.

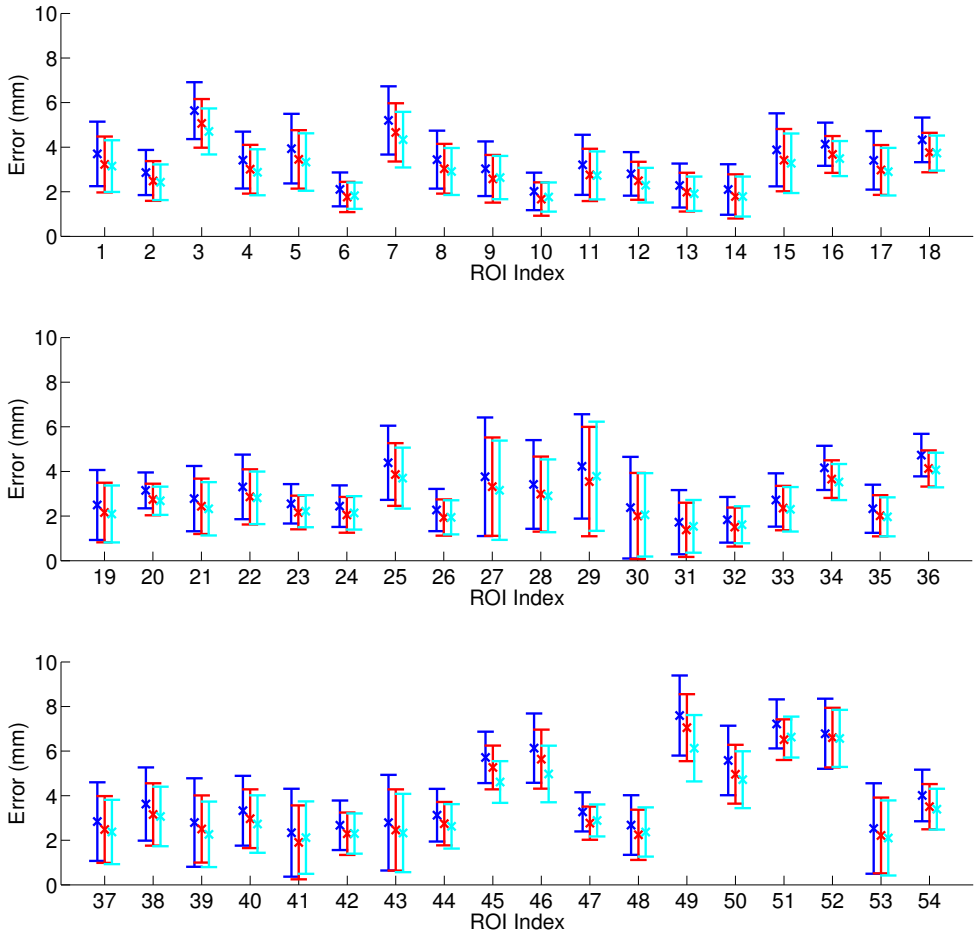


Figure 2.7: The errors and the standard deviations of the deformation fields reconstructed by individual methods: blue - reconstructed by TPS; red - reconstructed by a single compactly supported RBF kernel; cyan - reconstructed by multiple compactly supported RBF kernels. The ROI names corresponding to their indices are listed in Table 2.2.

The comparison of the reconstruction error also involves TPS, which is often applied for interpolating the deformation field in the literature (Chui and Rangarajan, 2003; Wu *et al.*, 2010; Yap *et al.*, 2010). The error associated with the TPS-based interpolation is 3.540 ± 1.384mm, as detailed errors in individual ROIs are colored in blue in Fig. 2.7. Note that the compact-support RBFs provide better numerical scalability than TPS-based interpolation here. In fact, TPS yields a very dense kernel matrix that requires partitioning the image space into several blocks. Then, the deformation

fields in individual blocks are independently interpolated for further integration. Distortions and errors are thus inevitably introduced through adjacency of neighboring blocks. However, the compact-support RBFs result in a highly sparse kernel matrix, thus solving the problem much more conveniently. Moreover, though TPS enables the estimation of the affine transformation, in addition to the non-rigid deformation, between two images, it is not necessary here since all images are affinely registered already and the focus of this work is on deformable image registration.

Errors of Refined Deformation Fields. The predicted deformation fields are further refined via Demons (Vercauteren *et al.*, 2009) and HAMMER (Shen and Davatzikos, 2002), respectively. Each refined deformation field is compared with the ground truth, so the residual errors between them can be calculated. The errors in 54 anatomical regions averaged over 10 tests, as well as the standard deviations, are plotted in Fig. 2.8. With the P-R hierarchy for initialization, the average errors of deformation fields are 0.730 ± 0.211mm (refined by Demons) and 0.462 ± 0.062mm (refined by HAMMER). Alternatively, each subject can be registered with the template directly by using registration methods in the traditional way. The error for the direct registration via Demons is 0.837 ± 0.238mm. Similarly, the error for HAMMER is 0.532 ± 0.072mm. Therefore, the P-R hierarchy provides good initializations for image registration, as the *initialization-refinement* strategy can effectively reduce the errors of deformation fields compared to using conventional registration methods alone.

Real Data

Two real datasets of brain MR images are used in the following section for the evaluation and comparison of the P-R hierarchy. In each dataset, a *template* and a *subject* are randomly selected and designated; other images are then tagged as the *intermediate*. All intermediate images are segmented and then registered with the template via

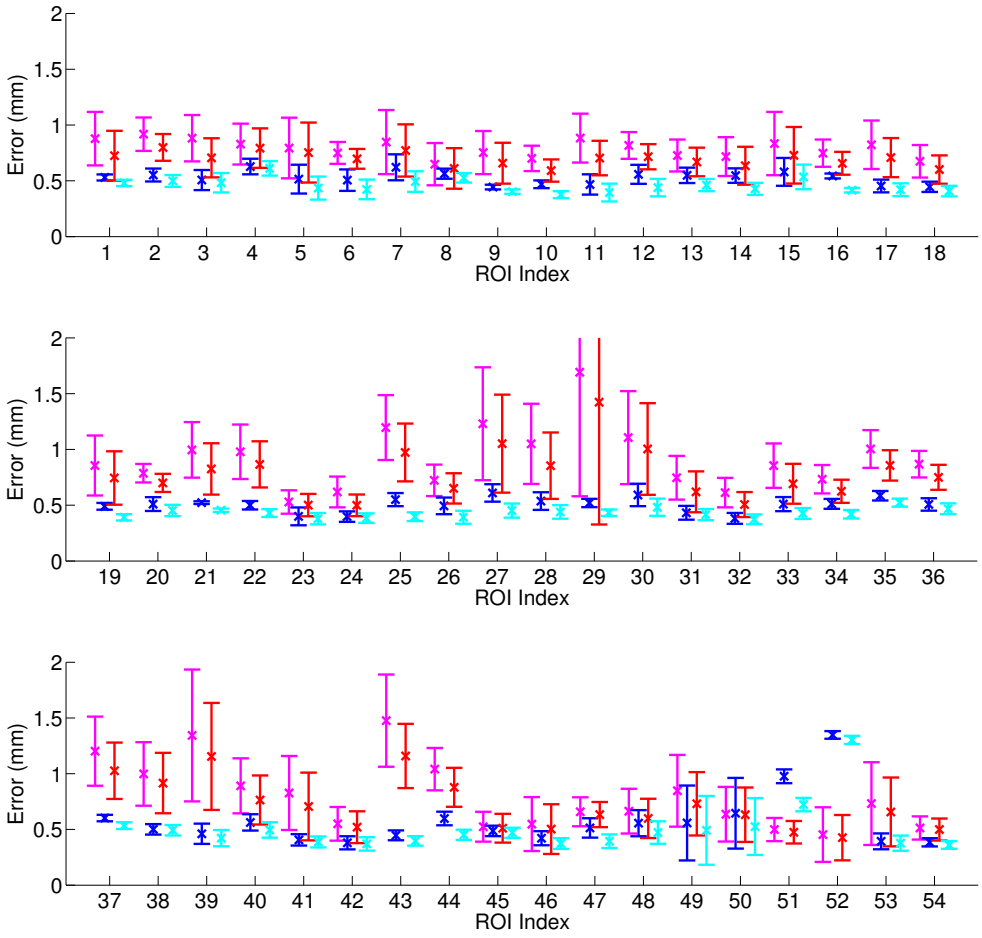


Figure 2.8: The errors and the standard deviations of the deformation fields estimated/refined by individual methods: magenta - estimated by Demons; blue - refined by Demons; red - estimated by HAMMER; cyan - refined by HAMMER. The ROI names corresponding to their indices are listed in Table 2.2.

HAMMER (Shen and Davatzikos, 2002), while the key points in the template image space are determined at the same time. The predicted deformation field for the subject can be further refined via existing registration methods. Moreover, after registering the subject with the template, the Dice ratio of anatomical ROIs can be computed as an indicator of the accuracy of the registration. The Dice ratio measures the overlap of corresponding ROIs in the deformed subject and the template, as the higher measure typically implies that the two images are registered more accurately (Klein *et al.*, 2009;

Index	ROI	Index	ROI
1	L Superior Frontal Gyrus	2	R Superior Frontal Gyrus
3	L Middle Frontal Gyrus	4	R Middle Frontal Gyrus
5	L Inferior Frontal Gyrus	6	R Inferior Frontal Gyrus
7	L Precentral Gyrus	8	R Precentral Gyrus
9	L Middle Orbitofrontal Gyrus	10	R Middle Orbitofrontal Gyrus
11	L Lateral Orbitofrontal Gyrus	12	R Lateral Orbitofrontal Gyrus
13	L Gyrus Rectus	14	R Gyrus Rectus
15	L Postcentral Gyrus	16	R Postcentral Gyrus
17	L Superior Parietal Gyrus	18	R Superior Parietal Gyrus
19	L Supramarginal Gyrus	20	R Supramarginal Gyrus
21	L Angular Gyrus	22	R Angular Gyrus
23	L Precuneus	24	R Precuneus
25	L Superior Occipital Gyrus	26	R Superior Occipital Gyrus
27	L Middle Occipital Gyrus	28	R Middle Occipital Gyrus
29	L Inferior Occipital Gyrus	30	R Inferior Occipital Gyrus
31	L Cuneus	32	R Cuneus
33	L Superior Temporal Gyrus	34	R Superior Temporal Gyrus
35	L Middle Temporal Gyrus	36	R Middle Temporal Gyrus
37	L Inferior Temporal Gyrus	38	R Inferior Temporal Gyrus
39	L Parahippocampal Gyrus	40	R Parahippocampal Gyrus
41	L Lingual Gyrus	42	R Lingual Gyrus
43	L Fusiform Gyrus	44	R Fusiform Gyrus
45	L Insular Cortex	46	R Insular Cortex
47	L Cingulate Gyrus	48	R Cindulate Gyrus
49	L Caudate	50	R Caudate
51	L Putamen	52	R Putamen
53	L Hippocampus	54	R Hippocampus

Table 2.2: The List of ROIs in the LONI LPBA40 Dataset

Rohlfing, 2012). All images in each dataset are tested as the template and the subject exhaustively, as the detailed performances report in the following.

NIREP NA0 Dataset There are 16 images in the NIREP dataset, each of which is labeled by 32 ROIs. The ROI indices and names are provided in Table 2.3. First, the refinement upon the predicted deformation fields is conducted via Demons (Vercauteren *et al.*, 2009). Compared to the direct registration via Demons, the overall

Dice ratio after refinement increases by 1.49%. Then, the deformation fields are refined by HAMMER (Shen and Davatzikos, 2002), and the overall Dice ratio after refinement is 2.57% higher than directly registering two images via HAMMER. The average Dice ratios, as well as standard deviations, with respect to individual anatomical ROIs are detailed in Fig. 2.9.

Index	**ROI**	**Index**	**ROI**
1	L Occipital Lobe	2	R Occipital Lobe
3	L Cingulate Gyrus	4	R Cingulate Gyrus
5	L Insula Gyrus	6	R Insula Gyrus
7	L Temporal Gyrus	8	R Temporal Gyrus
9	L Superior Temporal Gyrus	10	R Superior Temporal Gyrus
11	L Infero Temporal Region	12	R Infero Temporal Region
13	L Parahippocampal Gyrus	14	R Parahippocampal Gyrus
15	L Frontal Lobe	16	R Frontal Lobe
17	L Superior Frontal Gyrus	18	R Superior Frontal Gyrus
19	L Middle Frontal Gyrus	20	R Middle Frontal Gyrus
21	L Inferior Gyrus	22	R Inferior Gyrus
23	L Orbital Frontal Gyrus	24	R Orbital Frontal Gyrus
25	L Precentral Gyrus	26	R Precentral Gyrus
27	L Superior Parietal Lobule	28	R Superior Parietal Lobule
29	L Inferior Parietal Lobule	30	R Inferior Parietal Lobule
31	L Postcentral Gyrus	32	R Postcentral Gyrus

Table 2.3: The List of ROIs in the NIREP NA0 Dataset

LONI LPBA40 Dataset The LPBA40 dataset contains 40 images, each of which is labeled by 54 ROIs. The ROI indices and names are provided in Table 2.2. Similar to the experiment on the NIREP dataset, the predicted deformation fields are refined via Demons (Vercauteren *et al.*, 2009) and HAMMER (Shen and Davatzikos, 2002), respectively. Compared to the direct registration via Demons, the overall Dice ratio after refinement increases by 1.47%. Compared to the direct registration via HAMMER, the outcomes of the P-R hierarchy and the refinement improve the overall Dice ratio by 1.88%. The average Dice ratios, as well as standard deviations, with respect to

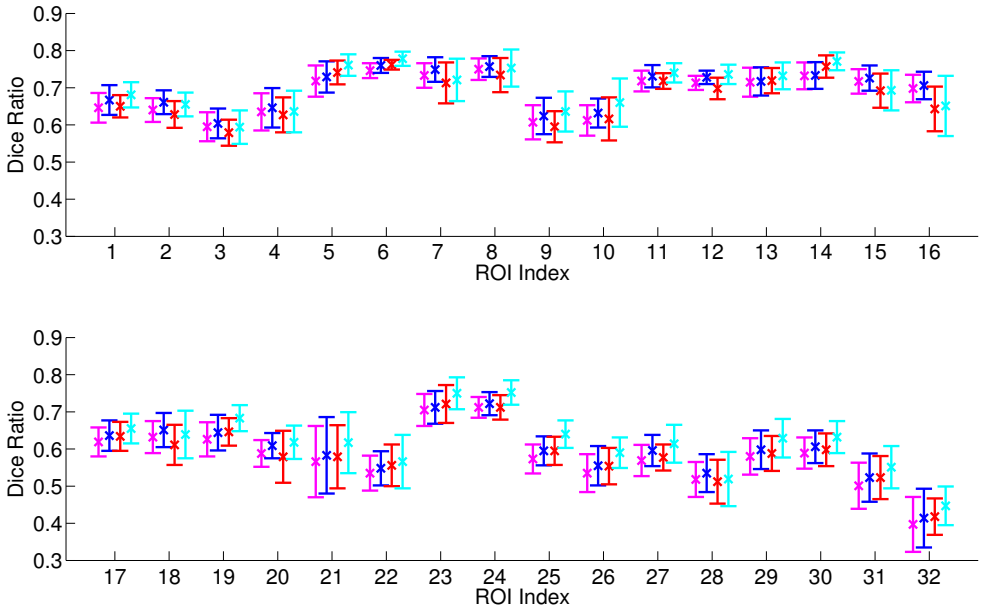


Figure 2.9: The overall Dice ratios and the standard deviations yielded after individual methods in the registration of the NIREP dataset: magenta - directly registered by Demons; blue - refined by Demons; red - directly registered by HAMMER; cyan - refined by HAMMER. The ROI names corresponding to their indices are listed in Table 2.3.

individual anatomical ROIs are detailed in Fig. 2.10.

In general, on both real datasets, the Dice ratios increase by refining the outputs from the P-R hierarchy, compared to applying state-of-the-art registration methods directly. The improvement upon the registration accuracy can be attributed to the introduction of the predicted deformation field, which acts as initialization to the existing registration methods. In particular, the prediction provides an input deformation for registration to optimize at the *high* resolution. Since the prediction is accurate and reliable, the refinement via state-of-the-art registration methods essentially leads to more accurate alignment of the subject and the template images.

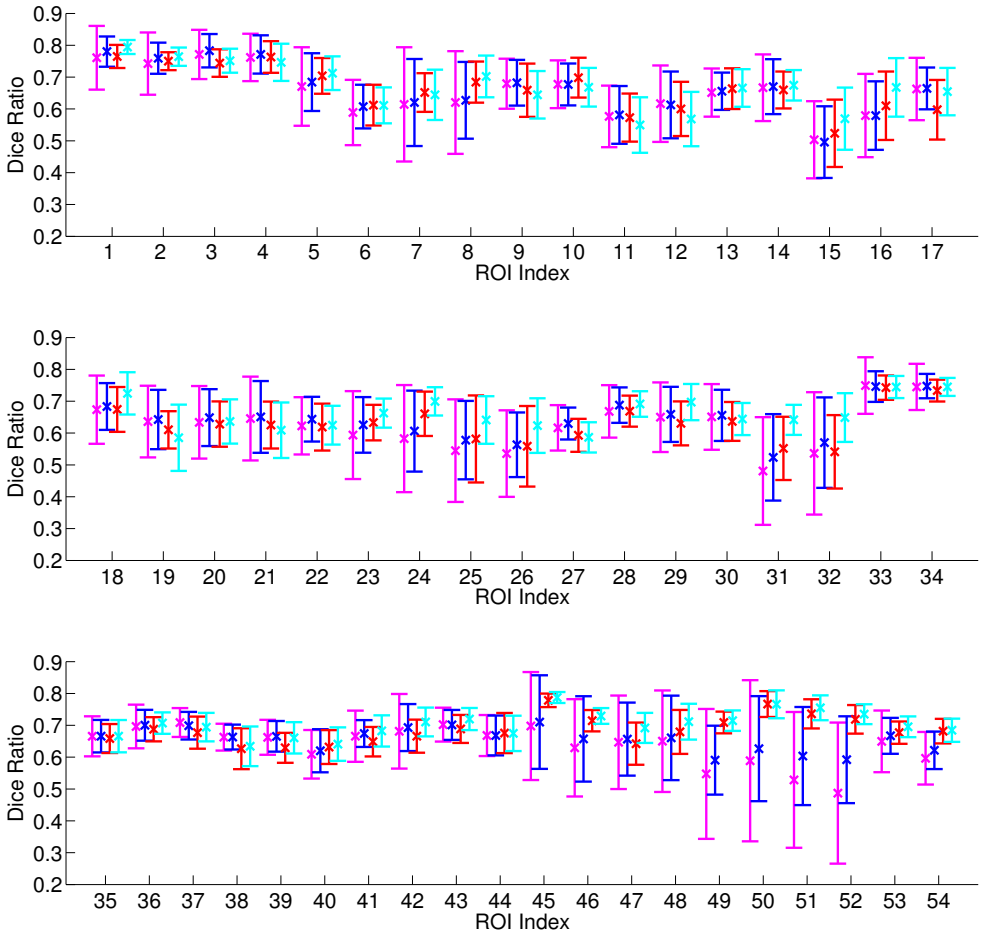


Figure 2.10: The overall Dice ratios and the standard deviations yielded after individual methods in the registration of the LPBA40 dataset: magenta - directly registered by Demons; blue - refined by Demons; red - directly registered by HAMMER; cyan - refined by HAMMER. The ROI names corresponding to their indices are listed in Table 2.2.

2.4 Summary

In the conventional style, only two images are involved for a specific registration task - the moving subject is deformed to the space of the fixed template following the estimated deformation field. The pairwise scheme, though simple, faces enormous challenges, especially given high anatomical variation between the subject and the template. In this case, the guidance from the intermediate images is able to predict the deformation for the subject to complete its registration with the template.

The intuitive motivation here is that similar images should share similar deformation when they are registered with the same template. This leads to image-scale guidance, in which an intermediate image similar to the subject is always identified. The granularity is further reduced to the patch scale, such that individual parts within the subject can flexibly seek for guidance from different intermediate images. Similar to image-scale guidance, patch-scale guidance relies on the observation that similar patches, though from different images, are associated with similar deformations when all images are registered with the template. Moreover, patch-scale guidance can be interpreted from the perspective of correspondence detection.

Guidance from the intermediate images in the population is capable of better solving the pairwise registration between the subject and the template. It is true that the introduction of the intermediate images has incurred additional complexity, as the question has evolved beyond the scope of only two images. However, the advantages associated with the intermediate images are undeniable. The intermediate images provide reliable deformation fields, which are usually manually inspected, to predict the subject's registration. The prediction, combined with the refinement, leads to higher accuracy compared with the direct registration scheme. The guidance can also speed up the estimation of the deformation in cases when time cost is more sensitive than the potentially slight loss of accuracy (Chou *et al.*, 2013; Kim *et al.*, 2012).

Chapter 3

Groupwise Registration of Image Population

3.1 Overview

The high anatomic variability within individual human brains and the corresponding MR scans leads to the dilemma, in which any single template can hardly yield a full success in terms of atlas-based analysis or clinical diagnostic system (Toga and Thompson, 2001). Although it is applicable to register each subject in a large-scale population to a template via pairwise registration, the registration scheme introduces systematic bias towards the template. For instance, when examining impacts of certain diseases upon the brain morphology (e.g., brain tissue atrophy caused by AD), registration is often necessary for analyzing the large-scale population that consists of both patient images and normal controls. Although a normal control is commonly selected from the population to serve as the template, the selection may increase the difficulty in registering certain patient images with the template. Thus, additional noise for the statistical analysis is introduced by registration, since the overall quality used to register patient images with the template is less reliable compared to the registration of

normal controls. To increase the signal-to-noise ratio (SNR) for statistical analysis, the registration of a large-scale population of individual images needs to be considered within a unified registration framework, so that the overall registration accuracy for all images in a population is maximized and the anatomical differences within or across populations is better delineated. To this end, the groupwise registration technique has gained increasing research interest recently due to its capability of avoiding the determination of the template as well as the bias (Joshi *et al.*, 2004; Wu *et al.*, 2012a).

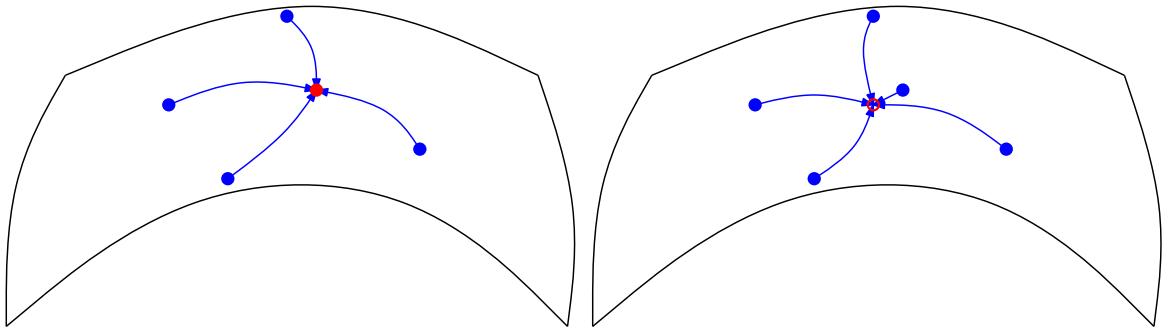


Figure 3.1: Comparison of the schemes of pairwise registration and groupwise registration. In (a), a certain image (indicated by the red dot) in the population is selected as the template, while other images (in blue) are independently registered with the template in the pairwise style. In (b), all images (in blue) are registered in the groupwise style and thus deformed towards the common space of the population (represented by the red circle).

To alleviate the limitations of pairwise registration, groupwise registration is capable of registering all images towards the common space of the population simultaneously. Figure 3.1 compares the schemes of the conventional pairwise registration with its groupwise counterpart. In the pairwise style, the template is often determined to be a certain image in the population (e.g., the red node in Figure 3.1). All other images in the population (indicated by the nodes in blue) are then registered with the template independently. The selection of the template is obviously critical - the registration could be extremely difficult if an *outlier* (e.g., S_2 in Figure 2.1(c)) is selected as the

template. By contrast, in groupwise registration, all images are deformed towards the common space of the population (represented by the red circle in Figure 3.1(b)), while the common space itself is revealed spontaneously during the registration process, instead of being manually designated in the beginning.

Different from the objective function in pairwise registration (c.f. (1.1)), all images in the population are involved simultaneously for groupwise registration. Assuming that the large-scale population consists of subject images $\mathbf{S} = \{S_i | i = 1, \cdots, n\}$, the task of groupwise registration is to estimate a collection of deformation fields $\mathbf{\Phi} = \{\phi_i(\cdot) | i = 1, \cdots, n\}$, such that

$$\mathbf{\Phi} = \arg\max_{\mathbf{\Phi}} \{\text{Similarity}(\mathbf{S}, \mathbf{\Phi}) + \text{Regularization}(\mathbf{\Phi})\}. \tag{3.1}$$

All images in $\mathbf{S}$ are warped to the common space following the deformation fields $\mathbf{\Phi}$. The (smoothness) regularization upon $\mathbf{\Phi}$ is typically enforced on each individual deformation field, or in the form $\sum_i \text{Regularization}(\phi_i)$. In the common space, the similarities amongst all warped images (measured by $\text{Similarity}(\mathbf{S}, \mathbf{\Phi})$) are maximized. The common space is also referred to as the *atlas* of the population. The atlas helps reveal the image variation within the population, while comparing the atlases of individual populations provides an intuitive glimpse of the inter-population analysis (Wang *et al.*, 2010b). Moreover, each pair of input images is connected via the common space (c.f. Figure 3.2). That is, by warping S_i in accordance with its deformation $\phi(\cdot)_i$ and then applying $\phi_j^{-1}(\cdot)$ to it, the newly generated image $S_i \circ \phi_i \circ \phi_j^{-1}$ shares the same space with the subject S_j. In other words, S_i is registered with S_j following the concatenated deformation field $\phi_i \circ \phi_j^{-1}$.

In accordance with the various implementations related to $\text{Similarity}(\mathbf{S}, \mathbf{\Phi})$ in (3.1), most contemporary groupwise registration methods fall into three categories.

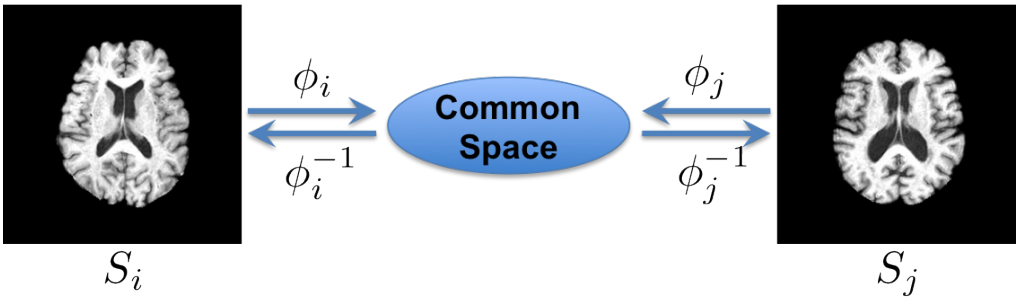


Figure 3.2: In groupwise registration, each pair of subject images, S_i and S_j, is connected via the common space. S_i is deformed to the common space following its deformation ϕ_i, and further deformed to the space of S_j following the inverse of the deformation field of S_j, or ϕ_j.

1. Pairwise registration is capable of evolving into the form of groupwise registration, e.g., by constructing the atlas of the population and then registering each subject image with the atlas. A characteristic of groupwise registration is the lack of a pre-determined template to serve as the destination towards which the subject is deformed, as in pairwise registration. However, with an estimated atlas that indicates the common space of the population, groupwise registration is reduced to functioning as a set of pairwise registration tasks between each subject in the population and the atlas. Mathematically, the similarities amongst all warped images are measured by $\sum_i \text{Similarity}(S_i \circ \phi_i, \bar{S})$, where $\bar{S}$ indicates the group mean image. The group-mean-based similarity measure allows each subject S_i to identify the deformation field $\phi_i(\cdot)$ and to complete the pairwise registration between S_i and $\bar{S}$.

2. Without the need to construct the atlas, the population similarity is observable as $\sum_{i,j,i\neq j} \text{Similarity}(S_i \circ \phi_i, S_j \circ \phi_j)$ other than the previous group-mean-based measure. Moreover, the evaluation of the similarity is often combined with a graph-based structure that approximates the manifold of the image population. Thus, only for pairs of images that are connected in the graph and similar in ap-

pearances, the similarities between them need to be considered. With knowledge of the image manifold modeled by the graph, a certain subject can estimate it respective deformation pathways to other similar images in the population, i.e., via pairwise registration. The averaging upon all acquired deformations results in the pathway that corresponds to the geodesic on the manifold to deform the subject under consideration towards the common space, even though the common space is never explicitly constructed (Jia *et al.*, 2010; Ying *et al.*, 2014).

3. Conventional pairwise techniques are used in both above categories, for registering either a certain subject with the estimated group-mean image or a pair of similar subjects. Nevertheless, it is possible to directly capture and maximize the coherence within all images of the population in a complete groupwise sense, without turning to pairwise registration for help. In this case, the similarity term in (3.1) can typically be represented by $\int_{x\in\Omega} \text{Coherence}(\mathbf{S}, \mathbf{\Phi}, x)dx$. The coherence with respect to x in the common space Ω, for example, is observable as the inverse of the variance of intensities belonging to the set of points $\{S_i \circ \phi_i(x) | i = 1, \cdots, n\}$. Thus, the task of groupwise registration is to find $\{\mathbf{\Phi}(x), \forall x \in \Omega\}$ such that the coherence of all deformed images at $x \in \Omega$ is maximized. Meanwhile, it is also implied that all points in the set $\{\phi_i(x), \forall i\}$ should be correspondences to each other and share the same location x in the common space Ω.

All three categories of groupwise registration methods and their implementations will be detailed in Sections 3.2-3.4, respectively. Then, the issue of preserving sub-group consistency in groupwise registration will be discussed in Section 3.5. A brief summary of this chapter is provided later in Section 3.6.

3.2 The "Mean" of the Population

A straightforward and well-known solution of groupwise registration is to reduce the problem into a series of pairwise registration tasks, with the participation of the iteratively estimated group-mean image. In this way, the complex groupwise registration is successfully implemented via pairwise registration, for which a lot of sophisticated methods are available in the literature. The group-mean-based groupwise registration aims to solve the following problem

$$\mathbf{\Phi} = \arg\max_{\mathbf{\Phi}} \left(-\sum_i \|S_i \circ \phi_i - \bar{S}\|^2 + \sum_i \text{Regularization}(\phi_i) \right). \tag{3.2}$$

In particular, two iterative steps are involved in this optimization, which is also illustrated by Figure 3.3:

1. Estimate the group mean image $\bar{S}^t$ (corresponding to the t-th iteration) to reveal the common space of the population, based on all subject images $\{S_i\}$ and their previously estimated deformation fields $\{\phi_i^{t-1}\}$;

2. Update ϕ_i^t based on the tentative group mean image $\bar{S}^t$, all subjects $\{S_i\}$, and their previously estimated deformations $\{\phi_i^{t-1}\}$.

As in Figure 3.3(a), for example, the current group mean image, which is represented by the red circle, is derived from all to-be-registered subjects (in the blue nodes). Then, in (b), each subject starts to deform and tries to shorten their (geodesic) distances to the tentative group mean image. Next, in (c), the group mean is further updated according to the latest subjects, while each subject can further refine its deformation with respect to the new group mean image.

A high-quality atlas is needed to reflect the common space of the population, while all subjects regard the atlas as their destination in pairwise registration. Based on the

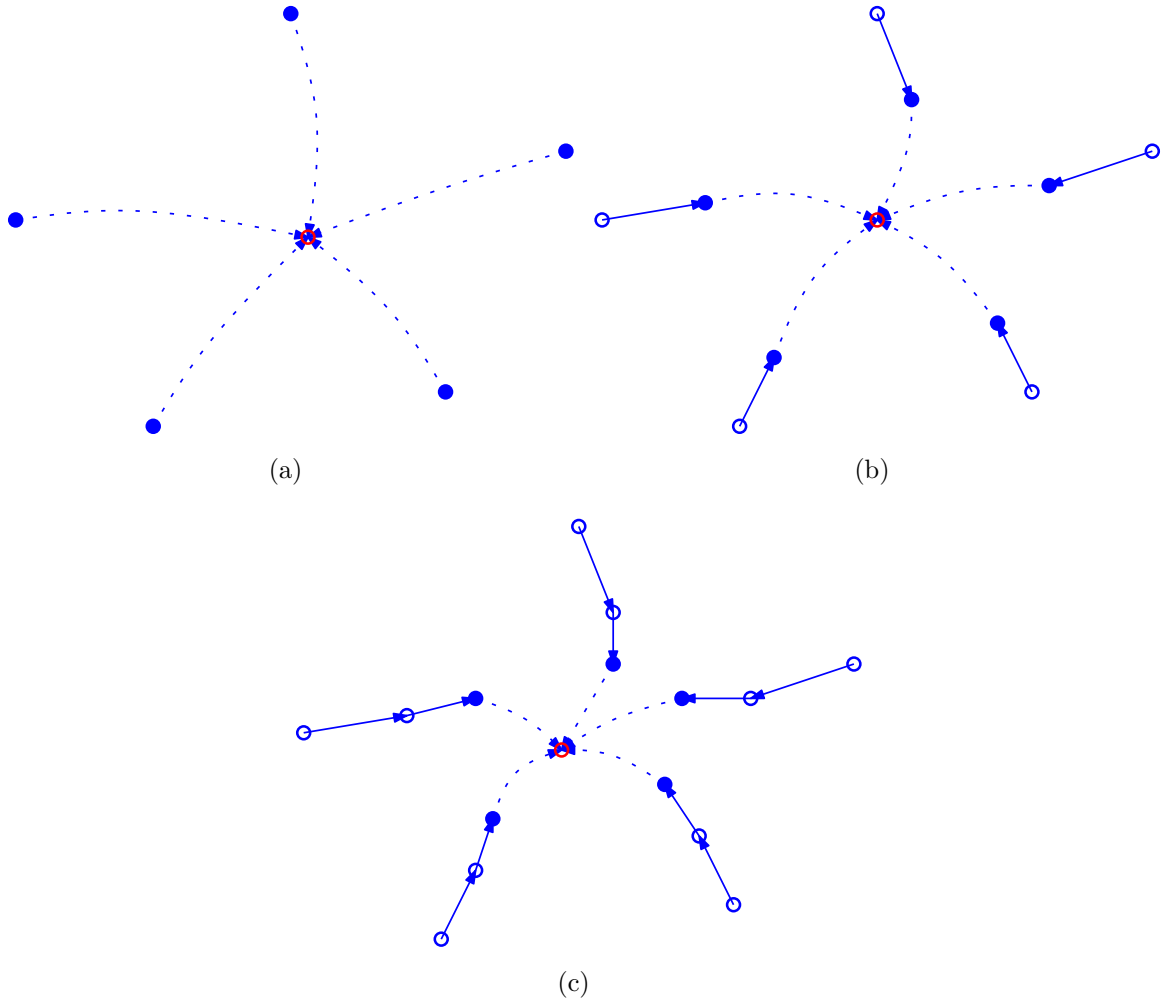


Figure 3.3: Illustration of the iterative group-mean-based groupwise registration method. In (a), the group mean image (in the red circle) is estimated from all images in the population (represented by the blue nodes). In (b), each subject starts to deform towards the tentatively estimated group mean via pairwise registration. In (c), the group mean is updated according to the newly deformed subjects, while all subjects continue deforming to the latest group mean image.

theory of the manifold of diffeomorphism, Joshi *et al.* (2004) derived an elegant solution for the atlas in the "group-mean" form

$$\bar{S} = \frac{1}{n} \sum_i S_i \circ \phi_i. \tag{3.3}$$

The group-mean formulation has inspired several follow-up works (Fletcher *et al.*, 2009; Ma *et al.*, 2008; Zhang *et al.*, 2013), in which the computation of the mean image is shown to be non-trivial.

A major drawback of the conventional group-mean method (Joshi *et al.*, 2004) is that all subjects are treated equally when constructing the atlas, especially in the very beginning of the groupwise registration process. Since the subject images are not well aligned with each other prior to groupwise registration, averaging them in the intensity domain typically results in a very fuzzy mean image. The blurry group-mean image fails to provide clear guidance to the subsequent pairwise registration tasks, and leads to (1) the loss of anatomical details which can hardly be recovered from the initial fuzzy group-mean image; (2) the degradation of the alignment in each iteration of groupwise registration due to the difficulty of establishing reliable correspondences between sharp subject images and the fuzzy group-mean image; 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.

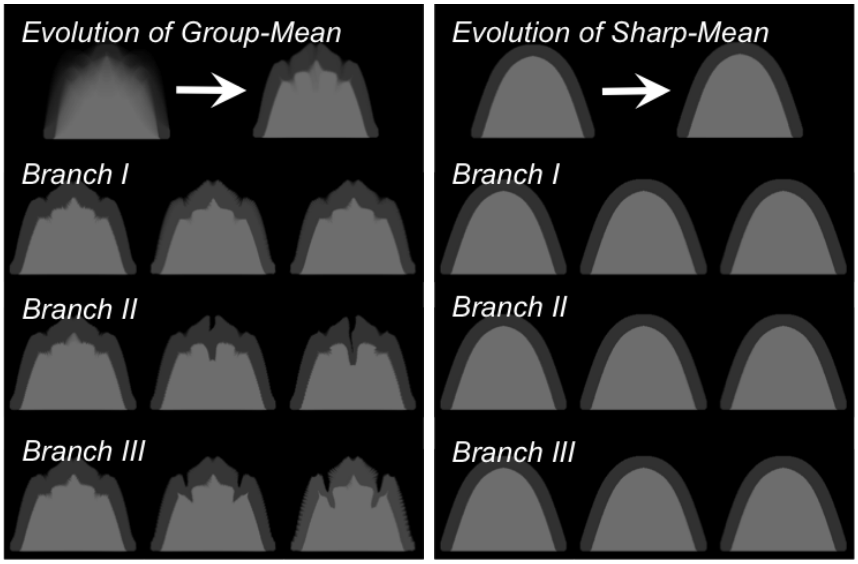
3.2.1 Sharp Mean: Motivation

The quality of the group-mean image, or its "sharpness" in particular, is important for the performance of the group-mean-based groupwise registration. Indeed, the fuzzy group-mean image is detrimental to groupwise registration due to the difficulty in registering an individual subject of clear anatomical structures with the group mean image of fuzzy structures. Also, a fuzzy group-mean image challenges the convergence of optimization since it provides insufficient anatomical information to guide registration. For example, in Figure 3.4, the group-mean method is applied to the simulated dataset (c.f. Figure 2.2) for the sake of groupwise registration. A group-mean image is constructed in the conventional way by following (3.3) and averaging all input images. The

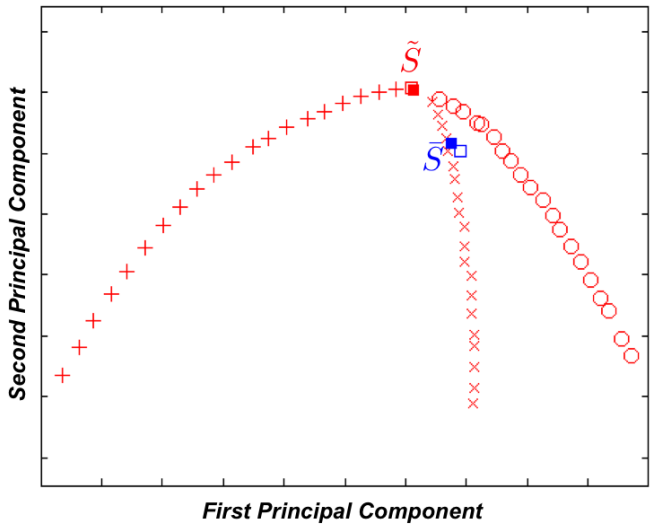
iterative evolution of the group mean image starts from a very fuzzy instance, while the finally estimated mean image (corresponding to the end of groupwise registration) contains a lot of artifacts (c.f. Figure 3.4(a)). Moreover, further examinations confirm that similar artifacts are observable on all deformed subjects, the samples of which are also available in the figure. The reason is that the low-quality group-mean image lacks clear structural information to guide the pairwise registration between each subject and the mean. Thus, the errors in estimating the subject deformation fields occur from the very beginning of groupwise registration, and can never be corrected in the following optimization. Although all subjects become similar to each other and yield a less fuzzy mean image in the end, the artifacts basically reflect the fact that huge errors are accumulated when the registration converges to an incorrect solution.

The drawback of the conventional group-mean method is outlined in Wu *et al.* (2011), where a new way to compute the "sharp" group-mean image is proposed for modeling the atlas in groupwise registration. The sharp-mean work consists of two contributions:

1. It generalizes the conventional group-mean method by adaptively treating all subject images throughout the registration process. Only the registered subjects that are close enough to the tentatively estimated group-mean image are involved in updating the sharp-mean, since treating the subjects equally in the early registration stage leads to the irreversible loss of structural/anatomical details. As registration progresses, deformed subjects are more likely to agglomerate to the common space of the population. Then, more subjects will be allowed to participate in the construction of the group-mean image, and their contributions gradually become similar to each other. In this scenario, the group-mean image is capable of approaching the common space iteratively, while its high "sharpness" is well preserved.



(a) (b)



(c)

Figure 3.4: Comparison of the conventional group-mean method and novel sharp-mean method. In (a), the group-mean image is fuzzy in the beginning and contains artifacts in the end. The low-quality mean image leads to failed cases in registering images in all three branches. In (b), the sharp-mean retains consistent high quality throughout the entire registration process, and leads to satisfactory results for all images in the population. In (c), the evolutions of the group-mean $\bar{S}$ and the sharp-mean $\tilde{S}$ are projected onto the 2D PCA plane. The evolution starts from the hollow square and moves to the solid square for both mean images.

2. The performance of groupwise registration, measured by the accuracy in aligning each subject with the group-mean image, is enhanced by introducing a tree-structure for describing the distribution of the entire image population. It is non-trivial for pairwise registration methods to handle the prevalent high anatomical differences between individual brain images. As a remedy, a tree-based hierarchical registration method (Hamm *et al.*, 2010; Jia *et al.*, 2011a) is incorporated to register each subject with the latest group-mean image. Specifically, the sharp-mean is set as the root of the MST for the population, while each subject is represented by a certain node. The tree only allows similar images to be directly connected, while every subject is able to identify its path to the root. The benefit of the tree-based registration is that (1) each subject is registered with a connected counterpart that owns similar anatomies only; (2) the complete registration of the subject to the group-mean image is estimated by concatenating all deformation fields along its pathway to the root.

The advantages of the sharp-mean method, as compared to the conventional group-mean method, are illustrated by Figure 3.4(b). The sharpness of the mean image is preserved consistently during its iterative evolution. Meanwhile, with the help of MST to identify the optimal deformation pathway between each subject and the sharp group-mean image, all deformed subject images are similar to each other and to the sharp group-mean image at the end of groupwise registration. In Figure 3.4(c), the evolution of the sharp group-mean image, as well as the conventional group-mean image, is projected onto the 2D PCA plane. Both the sharp group-mean (in red, noted by $\tilde{S}$) and the conventional group-mean images (in blue, noted by $\bar{S}$) evolve from the hollow square (corresponding to the start of registration) to the solid square (corresponding to the end of registration). Note that the sharp-mean image coincides well with the root that is used to simulate the entire dataset, while the conventional group-mean image

deviates from the distribution of the population (i.e., in three branches of the tree).

3.2.2 Objective Function and Optimization

Instead of using arbitrarily equal weights, the contributions from individual subjects to the construction of the sharp-mean image should be integrated dynamically during the groupwise registration process. Moreover, each anatomical region in a certain subject may have its own unique characteristic other than the respective region in the group-mean. Therefore, applying the same weight to all anatomical regions of a certain subject may lead to uncontrolled fuzziness across different anatomical areas of the group-mean, even though varying weights are assigned to individual subjects. To mathematically formulate these issues and solve the construction of the sharp-mean, two strategies are utilized:

1. A distance measure for each point $x \in \Omega$, with respect to the tentative mean image $\tilde{S}^t$ and the previously deformed subject S_i^{t-1}, is defined as

$$D(S_i^{t-1}, \tilde{S}^t, x, b) = \sum_{y, \|y-x\| \leq b} \|S_i^{t-1}(y) - \tilde{S}^t(y)\|^2. \tag{3.4}$$

   Here, b is related to the size of the patch centered at the point x. The term $D(S_i^{t-1}, \tilde{S}^t, x, b)$ captures the overall intensity difference between the corresponding local image patches (centered at x) in the images S_i^{t-1} and $\tilde{S}^t$. Although other advanced distances are applicable, the simple intensity difference is used here for easy computation. Meanwhile, recall that in $S_i^{t-1} = S_i \circ \phi_i^{t-1}$, the distance measure is also dependent on the deformation field ϕ_i^{t-1}, albeit implicit. It is worth noting that $D(\cdot)$ approaches to the distance between the two entire images if b is large enough, while it becomes a voxelwise difference when b shrinks to 1. Since registration usually follows global to local refinement, the value of b is high in the

initial registration stage and then decreases gradually.

2. In order to treat each subject adaptively, the variable $\omega_i^t(x)$ (satisfying $\sum_i \omega_i^t(x) = 1$) weights the contribution of the warped subject S_i^{t-1} to constructing the group-mean image at a particular location x. In the initial stage of groupwise registration, all subjects are not well aligned and their averaging typically results in a fuzzy mean image. Thus, to keep the group-mean sharp throughout registration, only those deformed subjects $\{S_i^{t-1}\}$ that are similar enough to the tentative group-mean image $\tilde{S}^t$ are qualified with high weights $\omega_i^t(x)$. With the progress of registration, all subjects are likely to agglomerate towards the population center. At that moment, all subjects $\{S_i^{t-1}\}$ are allowed to contribute equally for the unbiased estimation of the group-mean, as long as they are similar enough to the group-mean image. In other words, given deformed subjects that are well aligned, little fuzziness is introduced and the group-mean image keeps high sharpness. The dynamic changes of the weights are controlled by requiring the entropy of the set $\{\omega_i^t(x)\}$, i.e., $-\omega_i^t(x) \cdot \log(\omega_i^t(x))$, to increase while registration progresses.

With the two strategies in the above, the sharp-mean based groupwise registration method needs to maximize the following objective function

$$\sum_i \left(- \int_\Omega \left(\omega_i(x) \cdot D(S_i \circ \phi_i, \tilde{S}, x, b) + r \cdot \omega_i(x) \cdot \log(\omega_i(x)) \right) dx + \text{Regularization}(\phi_i) \right), \tag{3.5}$$

where $\sum_i \omega_i^t(x) = 1$ is satisfied for any $x \in \Omega$ and the scalar r controls the penalty of high distance from S_i to $\tilde{S}$. Compared with the conventional group-mean method, the new formulation is generalized by introducing the adaptive weights $\omega_i(x)$ for not only each subject S_i but also every spatial location x in the common space.

The optimization related to (3.5) is accomplished by decoupling it into two alternating sub-problems, i.e., estimating the sharp-mean image and then the deformation

fields for all subjects. First, given the previously deformed subjects S_i^{t-1} (with respect to ϕ_i^{t-1}), the optimal ω^t is computed. The mean image $\tilde{S}^t$ is tentatively estimated according to ω^t. Second, using ϕ_i^{t-1} as the initialization, the pairwise registration algorithm, e.g., diffeomorphic Demons (Vercauteren *et al.*, 2009), is used to calculate the deformation field ϕ_i^t for each subject S_i towards $\tilde{S}$. The iterative solution allows all subjects to deform toward the mean image, while the deformed subjects become more similar to each other gradually.

Adaptive Estimation of Sharp Mean.

Given the subject image S_i and its previously estimated deformation ϕ_i^{t-1} (with respect to the mean $\tilde{S}^{t-1}$), the latest deformed subject is $S_i^{t-1} = S_i \circ \phi_i^{t-1}$. After isolating all variables other than ω and $\tilde{S}$ in (3.5), the optimal solution to $\omega_i^t(x)$ is computed by letting the derivative of the objective function vanish:

$$\omega_i^t(x) = \exp\left(-D(S_i^{t-1}, \tilde{S}^{t-1}, x, b)/r\right). \tag{3.6}$$

The weight $\omega_i^t(x)$ is further normalized following $\omega_i^t(x) \leftarrow \omega_i^t(x)/\sum_i \omega_i^t(x)$. Two parameters are involved here in determining $\omega_i^t(x)$, i.e., the temperature r and the patch size b. In particular, r controls the fuzziness of the mean image, acting as the inverse temperature in the annealing system. Initially, the value of r is low, i.e., the contribution of a certain subject to constructing the group-mean image decays in an exponential way unless it is very similar to the tentative mean image. With the progress of groupwise registration, all subjects become more similar to the common space. Then, the temperature r increases to encourage equal weights for all deformed subjects. In general, the temperature $r(t)$, corresponding to the t-th iteration, is determined as $r(t) = r_0 + \Delta_r \cdot t$. The term r_0 is the initial temperature and Δ_r denotes the iterative temperature increase. To cater for possible misalignment, especially in the initial stage of groupwise

registration, b is able to adaptively control the scale, i.e., from global to local fashion, in measuring the image similarity. That is, the size of the patch, controlled by b, drops from the entire image space to a single voxel iteratively and linearly during the registration process. In this way, the weight $\omega_i^t(x)$ is different across not only the subjects but also spatial locations.

After determining the weight $\omega_i^t(x)$, the sharp-mean image is computed as

$$\tilde{S}^t(x) = \sum_i \tilde{\omega}_i^t(x) \cdot S_i^{t-1}(x), \tag{3.7}$$

where $\tilde{\omega}_i^t(x) \propto \sum_{y,\|y-x\|\leq b} \omega_i^t(y)$ and is further normalized to satisfy $\sum_i \tilde{\omega}_i^t(x) = 1$. It is clear that (1) the group mean image is the weighted average of all previously registered subjects with respect to the common space; (2) the weight $\omega_i^t(x)$ is adaptive to each subject; (3) the weight $\omega_i^t(x)$ is also locally adaptive as it is computed from the local patch discrepancy.

It is worth noting that $\tilde{S}^1$ still follows the conventional simple average in the first iteration. Alternatively, the weights $\omega_i^t(x)$ are calculated by regarding the median image in the population as $\tilde{S}^0$ for reference. The sharpness of $\tilde{S}^1$ is then guaranteed. To identify the median image, the distance between two images, i.e., S_i and S_j, is defined by (3.4) as $\int_x D(S_i, S_j, x, b)dx$. Then, the median image is determined to be the subject, with minimized overall distances to all other images in the population. Moreover, no bias in estimating the mean image is introduced by the median image in that: (1) the median image is only used as the reference to calculate the contribution of each subject, instead of being directly used as a template for registration; (2) the temperature r is gradually increased to ensure that all warped subjects have an equal opportunity to contribute to constructing the group-mean image, as long as the subject images are well registered in the common space.

Incremental Optimization of Deformation.

With the latest sharp-mean image $\tilde{S}^t$, the task for the subject S_i is to incrementally update its deformation $\phi_i^t(\cdot)$ from the existing field $\phi_i^{t-1}(\cdot)$ via pairwise registration. However, the high anatomical variations among the images may hinder the pairwise registration problem from being solved correctly and efficiently. Assuming all subjects reside on the manifold, it is relatively easy for the pairwise registration algorithm to align two nearby subjects, as opposed to two faraway subjects. In light of this, a tree-based registration method is applied, where each subject only needs to be registered with other similar images.

For the sake of tree-based registration, the image distance between each pair of images in the augmented population $\{\{S_i^{t-1}\}, \tilde{S}^t\}$ is computed first. Then, a fully connected graph is built by regarding each subject as a node and assigning image distances to respective edges. Next, MST is extracted from the graph, using Kruskal's algorithm, where $\tilde{S}^t$ is always set as the root node since the goal here is to estimate the deformation from each subject to the mean image. In this way, all images are 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 registered more robustly and accurately to the tentative group-mean, especially for the subjects that are far away from the common space. Note that, in estimating $\phi_i^t(\cdot)$, both $\phi_i^{t-1}(\cdot)$ and the deformations related to individual paths of the tree are only used as initializations. The deformation $\phi_i^t(\cdot)$ is always optimized in the direct pairwise registration between S_i and $\tilde{S}^t$.

Compared to the conventional group-mean method, the advantages of sharp-mean-based groupwise registration include: (1) the sharp-mean image is adaptively constructed from all aligned subjects, instead of a simple arithmetic average; (2) the con-

tribution of each subject is dynamically adjusted throughout groupwise registration in the annealing scenario; (3) the sharpness of the group-mean image is always preserved; and (4) the accuracy of groupwise registration in each iteration is improved by using MST to guide pairwise registration between the subject and the mean image. In conclusion, the method of the sharp-mean in groupwise registration is summarized as follows

Initialize parameters and variables, including t, r_0, Δ_r, b_0, $\{S_i\}$, $\{\phi_i^0(\cdot)\}$;
Compute the median image of the population and set it as $\tilde{S}^0$;
loop
    Compute the adaptive weights $\{\omega_i^t(x)\}$;
    Compute the sharp-mean $\tilde{S}^t$ in accordance with $\{\omega_i^t(x)\}$;
    Compute image distances and build MST;
    Register S_i to $\tilde{S}^t$ in pairwise registration;
    Deform S_i in accordance with $\{\phi_i^t(\cdot)\}$ to generate S_i^t;
    if converged **then**
        Abort loop;
    else
        $t \leftarrow t + 1$
    end if
end loop
Save $\{\phi_i^t(\cdot)\}$ as the outputs.

3.2.3 Experimental Results

Evolution of Group-Mean Images.

It is claimed that the sharp-mean method is capable of producing an atlas of higher "sharpness" than the conventional group-mean method. To verify the quality of the

mean image after groupwise registration, a population of 18 elderly brain MR images is used here. The size of each image in the population is $256 \times 256 \times 124$ and the spatial resolution is $0.9375 \times 0.9375 \times 1.5\text{mm}^3$ per voxel. Sample slices of all 18 images are arranged in the MST and shown in Figure 3.5. Note that the MST is also used by the sharp-mean method for estimating the deformation field between each individual subject and the group-mean image.

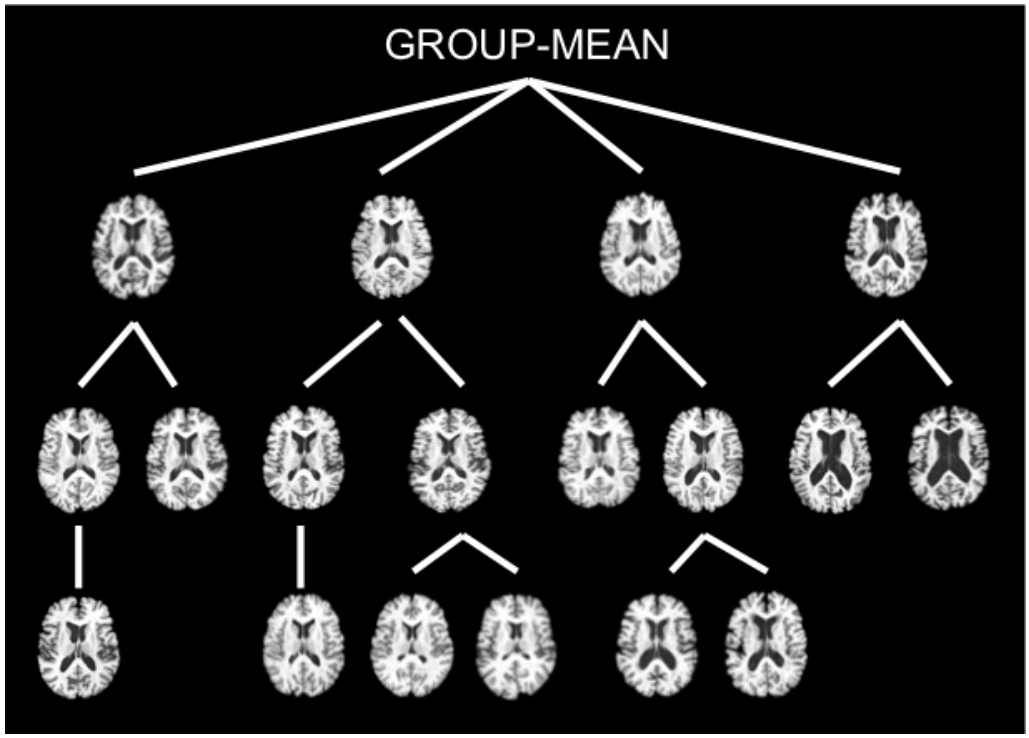


Figure 3.5: The MST consists of sample slices from 18 elderly brain MR images, which need to be registered towards the common space of the population.

The conventional group-mean method starts from a very fuzzy group mean, the 3D rendering of which is shown at the left end of the top row in Figure 3.6. By contrast, the sharp-mean with detailed anatomical structures can be constructed even from the very beginning of groupwise registration as in the bottom row of Figure 3.6. Though the quality of the conventional group-mean improves quickly during the iterative optimization, the final atlas still suffers from loss of detail compared to the outcome of the

sharp-mean method. The fuzzy atlas in the conventional group-mean method implies that the quality of the groupwise registration of all images in the population is lower. Meanwhile, it is worth noting that the initial sharp-mean is highly similar to the final atlas, indicating the capability of the method to approximate the unbiased atlas precisely and reliably even though all images are not well registered yet.

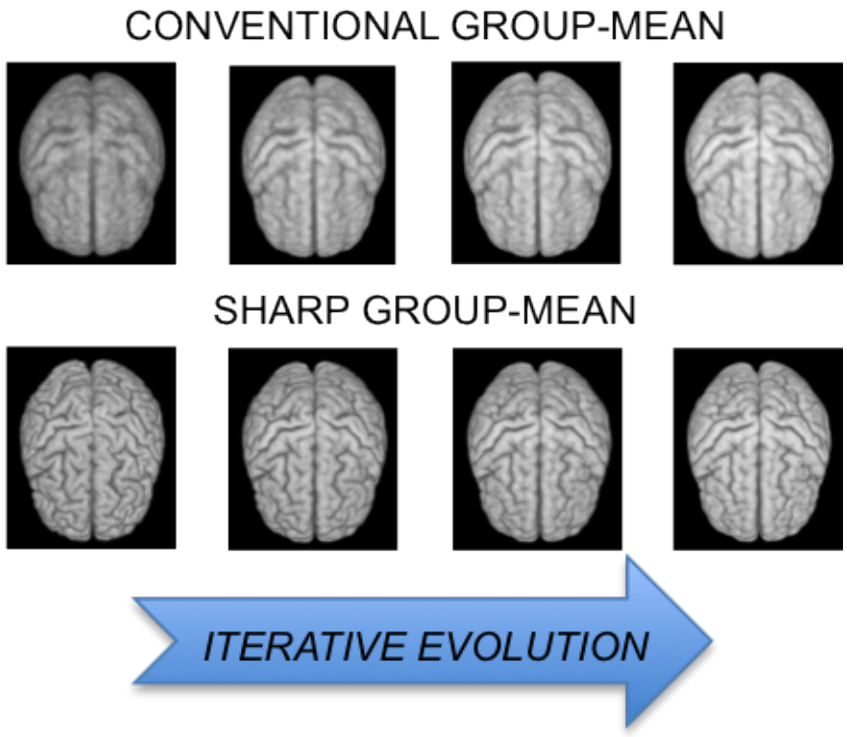


Figure 3.6: The conventional group-mean method starts with a fuzzy atlas, while the sharp-mean method is able to construct a high-quality atlas in the beginning of groupwise registration. Though the mean images are iteratively improved, the final atlas of the conventional method is still worse than the sharp mean image.

Dice Ratios.

To quantitatively evaluate the accuracy of groupwise registration, the Dice ratios are evaluated using two individual populations of real images, i.e., the NIREP NA0 dataset and the LONI LPBA40 dataset. The NIREP dataset consists of 16 images, each with 32

anatomical ROIs, while 40 images with 54 ROIs are available in the LONI dataset. After groupwise registration, all images as well as their ROIs are deformed to the common space, where the Dice ratios can be computed. Different from pairwise registration, the ROIs belonging to the common space need to be *voted* first. That is, the label for a certain point in the common space is determined as the most frequent ROI label for all points at the same location in deformed subjects. After all ROIs are *voted* in the common space, each registered image can compute the Dice ratio with respect to the common space for every ROI.

The overall Dice ratio for the sharp-mean method, averaged over all subjects and all ROIs in the NIREP dataset, is 79.31%, compared to 75.75% yielded by the conventional group-mean method. Moreover, the results are also compared with pairwise registration, in particular, 70.60% for HAMMER. Similarly, the scores for the LONI dataset are 79.26% (sharp-mean), 78.07% (conventional group-mean), 65.22% (HAMMER), respectively. Note that in pairwise registration via HAMMER here, the template is determined to be the median of the population, while the sum of the image distances (in SSD) from the median to all other images in the population are minimal. The Dice ratio again is evaluated in accordance to the ROIs voted in the common space, instead of the template-associated ROIs, in order to be consistent in comparing with other groupwise registrations. Detailed Dice ratios with respect to individual ROIs are plotted in Figure 3.12 (for the NIREP dataset) and Figure 3.13 (for the LONI dataset), respectively.

The scores reported in the above lead to the conclusion that both group-mean-based groupwise registrations achieve higher registration accuracy than the pairwise method. Although pairwise registration is designed to handle each subject and the template well, no connection between individual subject images is enforced during their independent registration towards the template. In groupwise registration, however, all

subjects are registered with the group-mean image, while the mean image is also iteratively refreshed according to the deformed subjects. In this way, the relationship across different subjects is incorporated into groupwise registration, which contributes to better registration accuracy in the end. Moreover, the sharp-mean method outperforms the conventional group-mean method on both real datasets, partly due to two reasons: (1) the sharp group-mean and its clear anatomical structures provide more accurate context information to guide image registration, and (2) the MST structure in the sharp-mean method alleviates the challenge of difficult image registration with high anatomical variation.

3.3 Graph Theory in Groupwise Registration

In the group-mean method (c.f. Section 3.2), the common space is explicitly instantiated first and all subject images are then registered with the mean image independently. The interaction between a certain subject and the group-mean image is *global*, in the sense that pairwise registration has to be conducted between the subject and the group-mean, even though the two images might be very different anatomically and faraway on the image manifold. As an alternative, Jia *et al.* (2010) proposed a method named ABSORB, which is short for Atlas Building by Self-Organized Registration and Bundling, for the sake of groupwise registration. ABSORB avoids the explicit estimation of the common space, and thus differs from the group-mean method.

In ABSORB, a graph is built to approximate the manifold of the image population. In order to register each subject image to the common space in an iterative manner, two strategies, namely *self-organized registration* and *image bundling*, are utilized.

1. The self-organized registration estimates the deformations from each subject to a subset of its neighbors in the graph, which are (1) similar to the subject under

consideration in appearance and (2) closer to the common space than the subject. By warping the subject in accordance with the averaged deformation, the self-organized registration is capable of condensing the distribution of the population iteratively, though only *local* interactions between similar subjects are considered.

2. With the progress of ABSORB, certain nearby subjects become close enough to each other and are bundled into a sub-group spontaneously. To this end, the image bundling leads to the hierarchical registration structure in ABSORB. That is, at the high level, the registration considers the representative images of all sub-groups bundled from the previous low level. A pyramid of images is thus built automatically, allowing each subject to access neighbors from nearby sub-groups and to quickly deform towards the common space. The atlas at the common space of the population is revealed once the registration arrives at the utmost level in the hierarchy.

3.3.1 Self-Organized Registration

To deform a subject S_i towards the common space at iteration t, the self-organized registration is first performed. Specifically, a subset of neighbors of S_i^{t-1}, resulting from S_i and its previous deformation ϕ_i^{t-1}, is selected according to the graph for $\mathbf{S}^{t-1}$. Then the incremental deformation field that warps S_i^{t-1} to S_i^t is calculated by combining all deformation fields between S_i^{t-1} and its selected neighbors. Note that the deformation fields between S_i^{t-1} and its selected neighbors can be obtained via any pairwise registration methods, while diffeomorphic Demons (Vercauteren *et al.*, 2009) is applied for demonstration here. The iterative solution of the self-organized registration is illustrated in Figure 3.7. Suppose each node, corresponding to a subject, is connected with its two qualified neighbors in the graph. The deformations from each subject

to its neighbors, indicated by dashed arrows, are estimated via pairwise registration. Then, in Figure 3.7(b), every node is allowed to move following the averaged deformations that are previously estimated. The tentative warps of all subjects thus result in a more *compact* distribution of the population. The above procedures are further applied iteratively, until all images are well aligned in the common space.

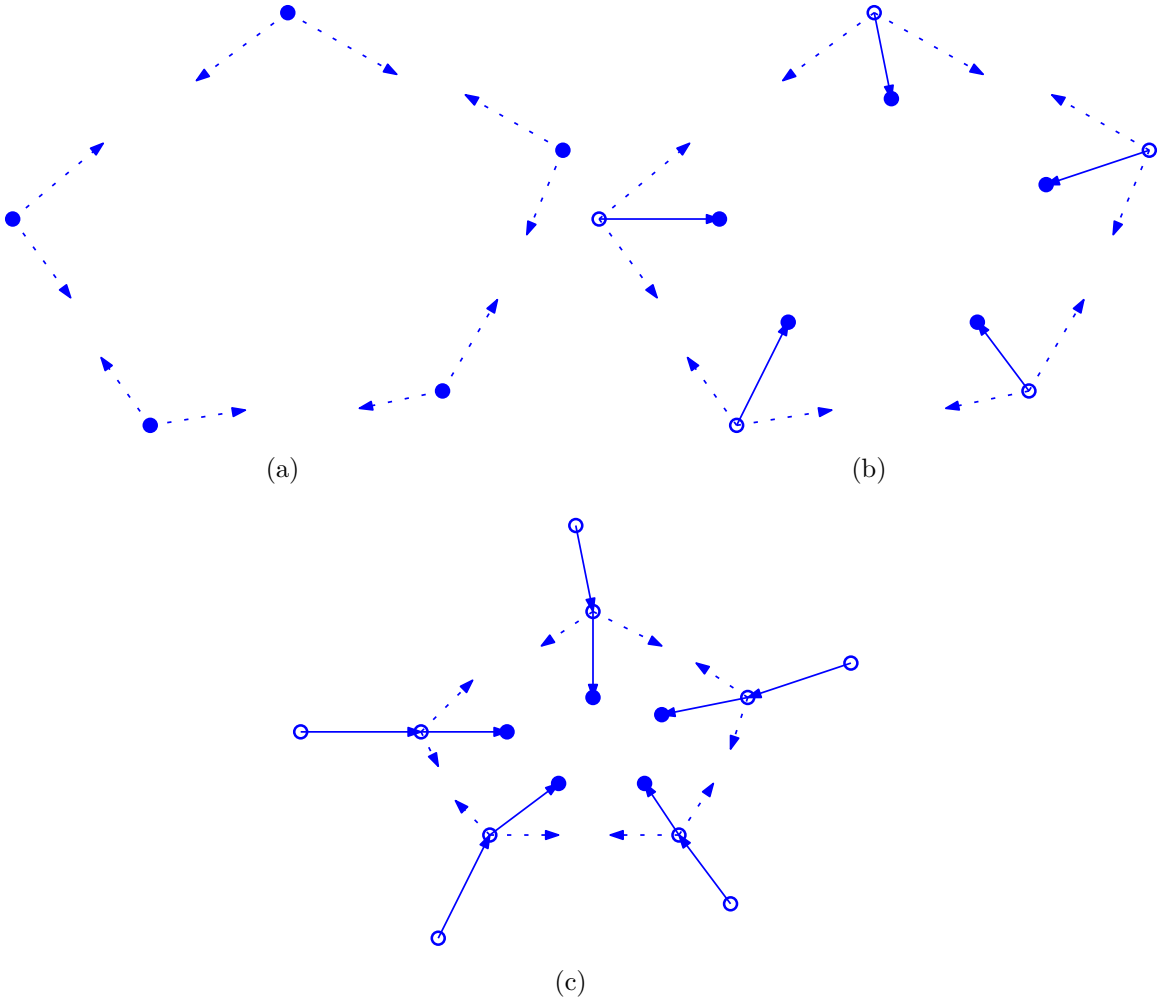


Figure 3.7: Illustration of self-organized registration. In (a), each subject estimates its deformations towards all qualified neighbors. In (b), the deformations are averaged and applied to warp individual subjects. In (c), the above procedures are iteratively invoked until all subjects are registered to the common space in the groupwise manner.

Selection of Neighbors.

The selection of neighboring subjects is critical to self-organized registration. A procedure is thus designed to adaptively qualify the subset of neighbors for each subject by considering both local and global information with respect to the distribution of the population. Specifically, a metric is first defined to measure the distance between any pair of subjects in the population. Then a graph is built and updated iteratively to help locate the global center, which approximates the common space in groupwise registration. Finally, the selection of neighboring subjects is applied to each subject and its deformation field is later computed from the qualified neighbors.

In ABSORB, the intensity difference, or SSD, is used to capture the distance between two images, though other metrics are also applicable (Hamm *et al.*, 2010; Munsell *et al.*, 2012; Seghers *et al.*, 2004). To better instantiate the manifold that captures the intrinsic distribution of the population, a k-NN isomap (Tenenbaum *et al.*, 2000) is constructed from the pairwise image distances. Then, the distance of S_i and S_j, denoted by $d(S_i, S_j)$, is updated as the shortest geodesic distance between the two images on the k-NN isomap. Note that the superscript $t-1$ or t related to the current iteration is temporarily ignored for convenience.

To ensure that the groupwise registration of the population is conducted on the learned manifold, the interactions between individual subjects in a single iteration are constrained within local neighborhoods only. That is, for the subject image S_i at the iteration t, the incremental deformation of the subject is locally determined with respect to its neighbors, while the subject approaches the common space of the population in the iterative manner. The local restriction is important, especially in the early stage of registration, because it is always much easier to register two nearby subjects with similar structures than to register images that are faraway from each other. As the result, both global and local information for the distribution of the population is

combined by embedding the global information into the selection of local neighbors. For example, when qualifying the neighbors of a certain subject, only images that are closer to the global center than the subject under consideration are chosen from its neighborhood. It is worth noting that the global center is not serving as the common space or the template, to which each subject is registered directly. The iteratively updated global center approximates the common space, and provides guidance for the selection of qualified neighbors only.

The global distribution of the population is embedded into the selection of local qualified neighbors, as only those closer to the global center are chosen. It implies that the determination of the global center is also critical to the performance of the algorithm. In particular, the median subject of the population is selected from the learned manifold as the global center due to its robustness to the outliers (Hamm *et al.*, 2010). The median of the population is defined as the subject that minimizes the overall distances from this subject to all other subjects in the population. To this end, an undirected graph, the edges of which are assigned the distances between connected subjects, helps the identification of the median subject. Instead of using a fixed graph throughout the registration process, a dynamic graph is used, which is updated from all tentatively deformed subjects after each iteration. The global center is then acquired from the graph, or the manifold, accordingly.

The determination of the global center helps subjects adjust their incremental deformations adaptively. However, if the graphs in different iterations are generated independently, the global center may change dramatically, especially in the early stage of registration. Thus, the estimated deformation for each subject would lack smoothness from the perspective of the entire registration process. A stable global center can provide a consistent guidance for subjects to avoid unnecessarily zigzagging paths. To obtain such a consistent but not fixed global center, an iterative neighborhood graph

(ING) is derived from the k-NN isomap. The weight assigned to the edge connecting two images S_i^t and S_j^t in ING is defined as

$$w(S_i^t, S_j^t) = \begin{cases} d(S_i^t, S_j^t) + \alpha \cdot w(S_i^{t-1}, S_j^{t-1}), & t > 0; \\ d(S_i^t, S_j^t), & t = 0. \end{cases} \tag{3.8}$$

The scalar $\alpha \in (0, 1)$ provides relaxation in terms of the previously calculated distances to include the distance information in previous iterations. Based on the weight assigned to each edge in ING, the median image, or the global center at the current iteration, can thus be selected.

With ING and the global center, the neighbors for each subject are examined and qualified accordingly. The desired neighbors should satisfy two criteria: (1) the neighbors are similar to the subject (i.e., within its neighborhood), and (2) the neighbors are more similar to the global center compared with the subject under consideration. Note that, for a certain subject, the number of its qualified neighbors is allowed to be zero only if the subject is the global center. In general, the neighbor selection strategy successfully embeds the global information of the population into the selection of the local neighbors for each subject, which is a desirable property of the self-organized registration.

Average of Deformations.

To complete self-organized registration, each subject needs to estimate its incremental deformation by averaging the deformation fields with respect to its selected neighbors. The averaging here is conducted within the domain of deformations, which significantly differs from the averaging of intensities as in Section 3.2. In particular, the subject S_i^{t-1} follows an unknown incremental deformation $\Delta\phi_i^t$ to deform towards the common space. With the subset of selected neighbors $\{\dot{S}_j^{t-1} | j = 1, \cdots, m_i\}$, pairwise

registration is performed between S_i^{t-1} and each $\dot{S}_j^{t-1}$ via diffeomorphic Demons (Vercauteren *et al.*, 2009), respectively. The resultant deformation field, $\psi_{i\to j}$, is further inverted and denoted by $\psi_{i\to j}^{-1}$ to describe the deformation that registers $\dot{S}_j^{t-1}$ to S_i^{t-1} (Christensen and Johnson, 2001). The overall incremental deformation for S_i^{t-1} is thus computed as

$$\Delta\phi_i^t = \left(\frac{1}{m}\cdot\sum_{j=1}^{m}\psi_{i\to j}^{-1}\right)^{-1}. \tag{3.9}$$

To emphasize the effect of the neighbors which are closer and more similar to the subject, image distances can also be incorporated as

$$\Delta\phi_i^t = \left(\frac{\sum_{j=1}^{m}\exp\left(w(S_i^{t-1},\dot{S}_j^{t-1})/2\sigma^2\right)\cdot\psi_{i\to j}^{-1}}{\sum_{j=1}^{m}\exp\left(w(S_i^{t-1},\dot{S}_j^{t-1})/2\sigma^2\right)}\right)^{-1}. \tag{3.10}$$

For each S_i^{t-1} in iteration t, the above procedure yields the incremental deformation field $\Delta\phi_i^t$. Then, the subject is deformed following $S_i^t = S_i^{t-1}\circ\Delta\phi_i^t$. The newly deformed subject S_i^t moves closer to the current global center. Note, it is possible that the global center cannot identify any qualified neighbors from the population. In this case, an identity transformation is temporarily used. Due to the fact of the iteratively dynamic global center, the specific image is capable of locating its deformation towards the real common space in the following iterations via self-organized registration.

3.3.2 Image Bundling

With the self-organized registration introduced in the above, each subject in the population deforms towards the common space iteratively. However, it is possible that several nearby subjects converge spontaneously, thus partitioning a sub-group from the entire population. Each sub-group may become stable and consists of compactly distributed subjects, while individual sub-groups are far away from each other. In this

case, qualified neighbors for a certain subject are mostly restricted within the sub-group to which the subject belongs. In order to break the barriers of sub-groups and to further optimize the results of groupwise registration, image bundling is a necessary measure here.

Individual bundles, which correspond to sub-groups of various sizes, need to be identified first. Given a set of tentatively registered subject images, a clustering method i.e., affinity propagation (AP) (Frey and Dueck, 2007), is adopted to bundle the images into sub-groups. Then, the representative image of each sub-group, determined automatically by AP, forms a new population of a much smaller size. The same procedure of self-organized registration is applied to the new population (at a higher level), to further register sub-groups of images together.

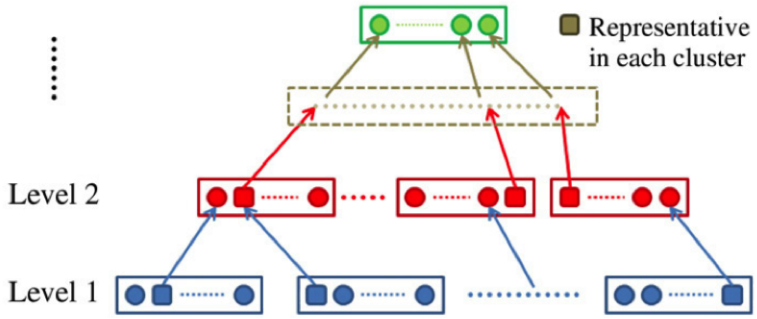


Figure 3.8: Image bundling leads to a hierarchical registration structure. The representative images (in squares) of individual low-level sub-groups (in boxes) are considered at the high level, where self-organized registration is recursively applied. When all images under consideration are within a single bundle, the image pyramid reaches its top level to complete groupwise registration.

The detail of the hierarchical registration structure is illustrated in Figure 3.8. Initially, all the subjects in the population are placed on the bottom level (Level 1), where self-organized registration is performed on each of them, including the selection of neighboring subjects and the combination of multiple deformations to qualified neighbors. The AP clustering method is then applied to detect whether the registered

images have fallen into a stable sub-group. If the clustering results on $\mathbf{S}^{t-1}$ and $\mathbf{S}^{t}$ are not exactly the same, both self-organized registration and image bundling are repeated for iteration t. If the clustering results in two consecutive iterations do not change, i.e., both the sub-groups and their representative images are exactly the same, the group-wise registration framework goes to the higher level and initiates a new population containing all representative images. The same procedures are repeated on this new population, and the iterative registration will terminate once the representative images are clustered into a single sub-group, or when the registration reaches the top of the hierarchical structure.

It is worth noting that the registration accuracy does not decrease as the registration procedure moves upwards in the hierarchical structure. Also, the smoothness of the estimated deformation fields is preserved, as justified by the following reasons. First, the images in each sub-group have been well aligned to each other upon identification of the sub-group. Therefore, it is reasonable to use the representative image to select a common set of qualified neighbors for all non-representative images in the sub-group. Second, each image always has a chance to be registered individually with respect to its new qualified neighbors (selected by its representative image). This indicates that the deformation field estimated for the representative image is directly applied to non-representative images. Therefore, the registration accuracy is guaranteed for the non-representative images, since all of them are separately registered with the qualified neighbors (selected from the representative images of other sub-groups). Third, the previously estimated deformation field for each image is regarded as a good initialization for the current registration of the same image, thereby no discontinuities of the deformation field are accumulated. When each image is registered to its new qualified neighbors in every iteration, the smoothness of the deformation field is always enforced.

In the utmost level, all subjects are registered very close to each other, and the rest of the representative images are compactly distributed within a single bundle. The task of groupwise registration then finishes. The average of all deformed subjects is regarded as the final atlas, which reflects the common space of the population. The sharpness of the atlas is well preserved, with clear anatomical structures and boundaries.

3.3.3 Experimental Results

Different from the group-mean method in Section 3.2, subject images only interact with other images in the population that are similar in appearance. However, with local interactions only, all subjects are still capable of identifying and agglomerating to the common space. After groupwise registration, the mean of all deformed subjects is computed and regarded as the atlas of the population. Figure 3.9 compares the atlas of the NIREP dataset estimated by ABSORB with the outcome of the group-mean method (Joshi *et al.*, 2004). The atlas of ABSORB clearly outperforms the fuzzy mean image by Joshi *et al.* (2004), especially in gyri and sulci highlighted by red arrows. It is worth noting that the atlas is never optimized directly in ABSORB, though the quality of the atlas is better and implies more superior registration quality.

Quantitatively, the overall Dice ratios for ABSORB are 78.86% for the NIREP dataset and 79.50% for the LONI dataset. Both scores are significantly better than the conventional group-mean method Joshi *et al.* (2004) and comparable with the sharp-mean method Wu *et al.* (2011). Detailed Dice ratios with respect to individual ROIs are plotted in Figure 3.12 (for the NIREP dataset) and Figure 3.13 (for the LONI dataset), respectively. The results indicate the capability of ABSORB to achieve top-class groupwise registration accuracy, even though the atlas or the common space is not explicitly estimated.

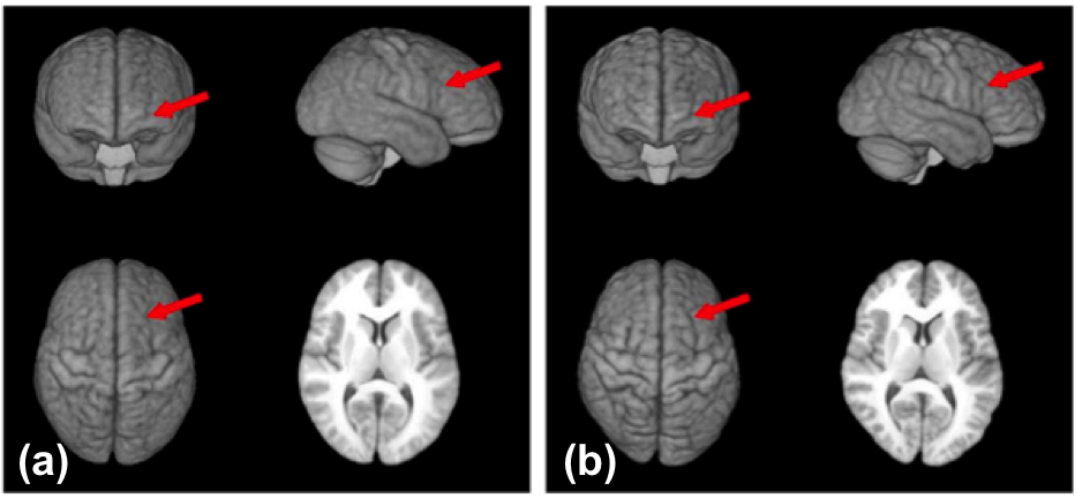


Figure 3.9: The atlas images of the NIREP brain dataset built by the group-mean method (a), ABSORB (b). The atlas generated by ABSORB can keep more anatomical details than that generated by the group-mean method, especially on the cortical regions marked by red arrows.

3.4 Features and Groupwise Correspondence

It is noted that, in pairwise registration, a point in the subject is deformed to the location of its correspondence in the template image space. Similarly, groupwise registration can also be interpreted from the perspective of correspondence detection. That is, a set of correspondence points from individual subject images should be deformed to the same location in the common space, even though the common space (e.g., the atlas) might not be explicitly constructed. To reduce ambiguities in correspondence detection for groupwise registration, the concept of multiple features is introduced by Wang *et al.* (2010a) and further extended in Wu *et al.* (2012b). In both works, an *attribute vector* $\vec{a}_i(x) = \{a_i^j(x)\}$ is assigned to describe the point x of the subject S_i. A metric then measures the coherence of correspondences with respect to the location x in the common space Ω and in accordance with the set of attribute vectors $\{\vec{a}_i(\phi_i(x))\}$ acquired from individual subject images. In general, the objective function of groupwise

registration becomes

$$\mathbf{\Phi} = \arg\max_{\mathbf{\Phi}} \left(\int_{x \in \Omega} \text{Coherence}\left(\{\phi_i(x)\}\right) \cdot dx + \sum_i \text{Regularization}(\phi_i) \right). \quad (3.11)$$

Note that the term Coherence $(\{\phi_i(x)\})$ is defined from the attribute vectors $\{\vec{a}_i(\phi_i(x))\}$. By solving the optimization problem in the above, the deformation fields for all images in the population can be estimated simultaneously in groupwise registration.

3.4.1 Attribute Vector

Multiple features encapsulated by the attribute vector enable the correspondence detection to be conducted more precisely for better image registration. The composition of the attribute vector is quite flexible and can be easily fine-tuned to cater for different applications. Besides complicated feature descriptors that are popular in computer vision, even simple attributes can contribute to yield high registration accuracy. Under the assumption that intensity transitions should be aligned at the same location after two images are well registered, the intensity gradients are taken as effective feature descriptors (Haber and Modersitzki, 2006; Staring *et al.*, 2009). Inspired by these works, the attribute vector in Wang *et al.* (2010a) is defined as consisting of the point intensity, as well as the gradient of intensities.

More sophisticated features, which are anatomically meaningful, are adopted by Wu *et al.* (2012b). After segmenting each image into tissues of WM, GM, and CSF, the attribute vector consists of three parts: (1) the tissue edge information related to the point; (2) the tissue type at the point; and (3) the geometric-moment-invariants (Lo and Don, 1989) of the point with respect to the three tissues. The edge information reflects whether the point is adjacent to tissue boundaries, while the tissue type of a certain point is also recorded in the attribute vector. The geometric-moment-invariants

model the distributions of individual tissues within the neighborhood of a center point. Note that the composition of the attribute vector in Wu *et al.* (2012b) is identical with that in HAMMER (Shen and Davatzikos, 2002), where the attribute vector proves to be successful in pairwise registration due to its capability of telling individual points apart.

3.4.2 Importance Sampling

Even though the attribute vector brings in more discrepancy capability than a single feature (e.g., the intensity only), the task of correspondence detection is still difficult in brain MR images. The ambiguity is high and prevalent, leading to the issue that accurate and exact correspondences are usually hard to acquire for a majority of points in the image space. To this end, it is better to apply correspondence detection to only a limited number of *key points*, where salient features exist and lead to relatively reliable correspondence detection. Denoting the set of key points as $\mathbb{X}$, the coherence term in (3.11) is then discretized from $\int_{x\in\Omega} \text{Coherence}(\cdot)dx$ to $\sum_{x\in\mathbb{X}} \text{Coherence}(\cdot)$, while each $x \in \mathbb{X}$ can be treated independently. The key points and their correspondences essentially contribute to the estimation of the dense deformation fields.

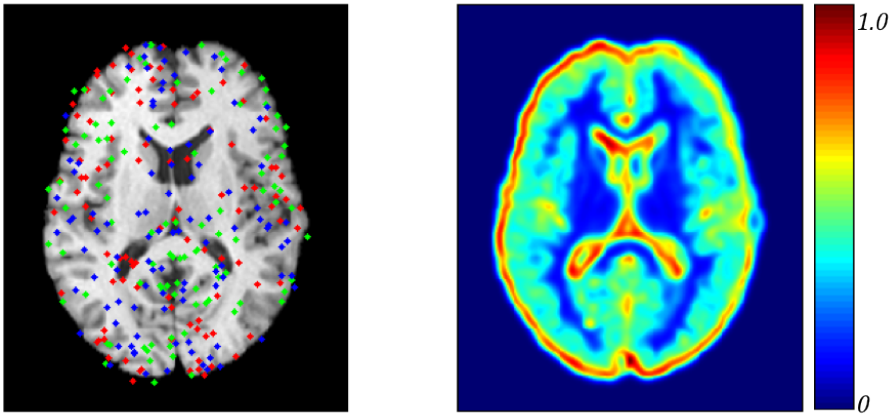
The importance sampling in Wang *et al.* (2010a), which collects the set of key points $\mathbb{X}$ from the implicit common space Ω, is in close relationship with the intensity gradient that is already part of the attribute vector. In particular, the gradient magnitude maps of all subject images are averaged, smoothed, and then normalized in the entire image space. The produced values signify the importance of each point, given the assumption that aligning tissue transitions (of high intensity gradients) is critical in image registration. Based on the importance map, a number of key points (around 1% of the brain volume size) are randomly sampled as the registration is completely driven by the selected key points. Higher importance implies that the point is more likely to

be drawn from the non-uniform sampling, while lower value denotes more possibility of rejection. Based on the importance map in Figure 3.10(b), for instance, a set of key points, colored red in Figure 3.10(a), is selected initially, while more key points are sampled to gradually participate to the progress of registration (red followed by green, and blue). The spatial distribution of sampled key points is more concentrated in boundary abundant areas, where accurate alignment of anatomical structures is fundamentally important in image registration.

Key points are also sampled in Wu *et al.* (2012b) to guide groupwise registration. However, the sampling in Wu *et al.* (2012b) is conducted within the spaces of individual images, as each subject S_i keeps its own set of key points $\mathbb{X}_i$. All points in $\{\mathbb{X}_i\}$ will contribute their correspondence information to groupwise registration. Moreover, due to the availability of the tissue segmentation (e.g., in WM, GM, and CSF), the key points are acquired from boundaries or transitions of individual tissues. The segmentation specifies anatomical information to each key point - only points of the same tissue type should be aligned in image registration. The tissue boundaries also bring to the key points rich context features, which are necessary and helpful to reliable correspondence detection.

3.4.3 Divergence Minimization

To maximize the coherence, or to minimize the divergence, within the set of attribute vectors $\{\vec{a}_i(\phi_i(x))\}$, a metric has to be defined first. Moreover, the divergence is better captured from not only a single point but its surrounding patch as well (Wang *et al.*, 2010a). Given x in a certain image, the neighborhood of the point is regarded as its local pattern. Groupwise registration is then formulated to find the deformations which could minimize the divergence among local patterns from different images, with respect to the same location in the common space. For the subject S_i and the point $\phi_i(x) \in \Omega_i$,



(a) (b)

Figure 3.10: Illustration of the key points acquired via importance sampling. In (a), red points are first selected as key points to drive groupwise registration, while more points in green and yellow participate later as registration progresses. The sampling of key points is based on the importance map shown in (b).

a probability density function (PDF) $Pr(a^j|i, \phi_i(x))$ is further estimated to reflect the distribution of the j-th feature $a_i^j(\phi_i(x))$ within the neighborhood of $\phi_i(x)$. The PDF can be generated by using the Parzen windowing technique.

The probability density functions (PDFs) of all features here act as the signature of the local pattern for the center point. Assuming that all input images are already registered (c.f. Figure 3.11), the set of PDFs $\{Pr(a^j|i, x), \forall i\}$ should be highly coherent. Therefore, a well-behaved groupwise registration algorithm needs to minimize the divergence within $\{Pr(a^j|i, \phi_i(x)), \forall i\}$. In order to measure this divergence, a popular information-theoretic metric, the Jensen-Shannon divergence (Lin, 1991), is employed following

$$\text{Divergence}(\{Pr(a^j|i, \phi_i(x)), \forall i\}) = \frac{1}{n}\sum_{i=1}^{n}\left(Pr(a^j|i, \phi_i(x)) \cdot \log\frac{Pr(a^j|i, \phi_i(x))}{\sum_s Pr(a^j|s, \phi_s(x))}\right). \tag{3.12}$$

Note that (3.12) computes the divergence related to the specific key point $x \in \Omega$ and a single feature a^j. The overall Coherence $(\{\phi_i(x)\})$ in (3.11) is counted by combining

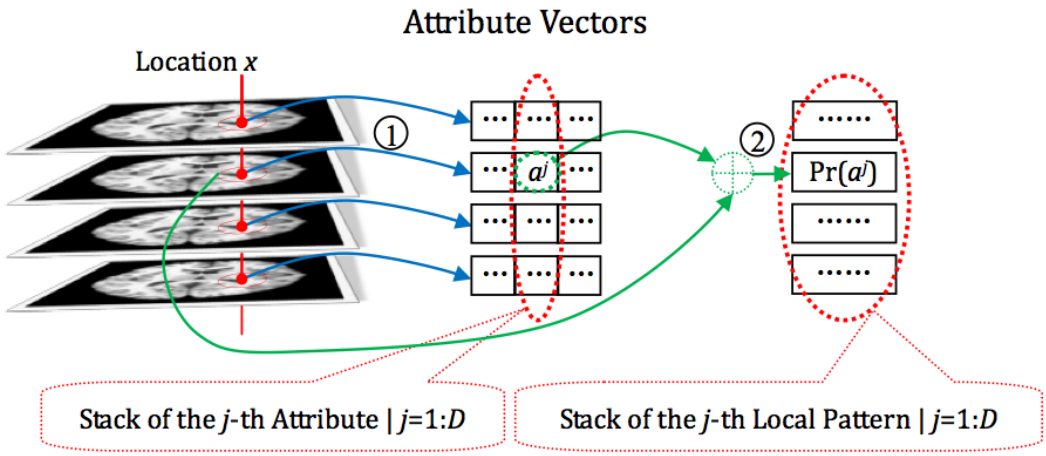


Figure 3.11: Each point in the image is described by its unique attribute vector, while the j-th feature in the attribute vector is underlined by the PDF $Pr(a^j)$ that is calculated from near the point.

contributions from all key points and all features. In particular, the weight of each key point $x \in \mathbb{X}$ is associated with its computed importance in sampling. Individual features are equally treated, though more advanced methods (Wu *et al.*, 2007) are applicable here to learn the optimal weights in the adaptive manner.

The optimization of the overall objective function in (3.11) is achieved by analytically solving for its gradient and then feeding it as the steepest descent direction to a first order solver. To further improve the robustness, a multi-resolution registration strategy is adopted. In each resolution, except for the highest one, images are down-sampled first. The attributes are then calculated on those down-sampled images. This approach ensures that the generated attribute vector is scale-related, and thus can better capture anatomical structures in different resolutions. B-Splines (Rueckert *et al.*, 1999) are employed to describe the deformation for each subject image. To better avoid local maxima in registration, the number of B-Spline control points is associated with the resolution of registration (e.g., 16 along each direction in low resolution, 32 in middle resolution, and 64 in high resolution). Therefore, at a coarser resolution, the optimization problem is effectively reduced to a lower degree-of-freedom problem,

which is achieved by using a smaller number of B-Spline control points. The estimated deformation is progressively refined as the registration gradually moves from the low level to the high level in the hierarchical framework.

3.4.4 Correspondence Detection via Common Space

Bearing the notions of attribute vectors and correspondence detection for groupwise registration, the solution in Wu *et al.* (2012b) is significantly different. Instead of defining the divergence with respect to the key point sampled in the common space, the coherence in Wu *et al.* (2012b) is computed by examining point-to-point correspondences between all possible pairs of images:

$$\text{Coherence}\left(x, x \in \mathbb{X}_i, \forall i\right) = \sum_{j, j \neq i} \text{Similarity}\left(x, \phi_j \circ \phi_i^{-1}(x)\right), \tag{3.13}$$

though the similarity between $x \in \Omega_i$ and $\phi_j \circ \phi_i^{-1}(x) \in \Omega_j$ is derived from respective attribute vectors of the two points.

The model in the above is built upon the fact the common space is able to bridge point-to-point correspondences across individual images. For example, given a certain key point $x \in \mathbb{X}_i$ in the subject image S_i, its correspondence in the atlas (whose existence is assumed here) is defined as $\phi_i^{-1}(x) \in \Omega$. Then, with respect to another subject S_j, the correspondence can be identified following ϕ_j and denoted by $\phi_j \circ \phi_i^{-1}(x) \in \Omega_j$. As a result, in (3.13), to maximize the similarity of the attribute vectors of $x \in \Omega_i$ and $\phi_j \circ \phi_i^{-1}(x) \in \Omega_j$ is equivalent to revealing the correspondence between x and a certain point in S_j. The correspondence essentially contributes to the estimation of both ϕ_i and ϕ_j in terms of groupwise registration.

The optimization of the overall coherence integrated upon all key points from individual subject images involves two steps: (1) to identify correspondences for each

key point $x \in \Omega_i$ in other subjects; (2) to interpolate the dense deformation fields of all images from the key points and their detected correspondences. The two steps are alternatively and iteratively applied. That is, in the first step, the correspondence information is refreshed according to the previously estimated deformation fields, while the updated correspondence leads to an incremental refinement to the deformation in the second step.

A correspondence field ψ_i is created for the subject S_i in the current iteration t. The correspondence field is sparse in that only correspondences of key points need to be identified, while the values of other points can simply be set to zero. In particular, the correspondence field ψ_i aims to maximize

$$\sum_{x \in \mathbb{X}_i} \sum_{j, j \neq i} \text{Similarity}\left(\vec{a}_i(x), \vec{a}_j\left(\phi_j \circ \psi_i^{-1}(x)\right)\right) - \sum_{x \in \mathbb{X}_i} \|\psi_i^{-1}(x) - \phi_i^{-1}(x)\|^2. \tag{3.14}$$

Note that, in the above equation, ϕ_i and ϕ_j indicate the deformation fields estimated in the previous iteration, while their superscript $(t-1)$ is ignored for simplicity. Each key point $x \in \mathbb{X}_i$ is encouraged to search for its correspondence in S_j by deviating $\psi_i(x)$ from the previously estimated $\phi_i(x)$. At the same time, the magnitude of the update of $\psi_i(x)$ is restricted (i.e., by limiting $\|\psi_i(x) - \phi_i(x)\|^2$), in that the incremental adjustment to $\phi_i(x)$ should be tiny in each iteration.

In particular, a greedy search method is applied for correspondence detection. Given the key point $x \in \Omega_i$, its correspondence in S_j is searched for within the window centered at $\phi_j \circ \phi_i^{-1}(x)$. For x and each point in the search window, their attribute vectors are examined such that only points of the same tissue type and tissue edge information could possibly be correspondences to each other. Then, the dissimilarity between the two points is calculated as the Euclidean distance of the geometric-invariant-moments in their attribute vectors. Moreover, the evaluation involves not only the two points

but also their individual local patches - to compare patches yields more reliable correspondence than to compare the centered points only.

The correspondence detection for the key point $x \in \Omega_i$ needs to be conducted with respect to all subject images other than S_i. Due to the prevalent uncertainty in correspondence detection, the concept of soft correspondence is introduced. That is, for a certain subject S_j, the key point x is allowed to identify multiple points as its potential correspondences. Each correspondence candidate is assigned a confidence, which is the point similarity evaluated in the above. All soft correspondences are then averaged, in accordance with their individual confidences, to contribute to $\psi_i^{-1}(x)$ that indicates the updated correspondence location in the common space for x.

The sparse correspondence field $\psi_i^{-1}(\cdot)$ encodes the updated correspondence information on all key points in $\mathbb{X}_i$, which differ from the previously estimated deformation field (i.e., $\phi_i^{-1}(\cdot)$). The incremental deformations $\Delta\phi_i$ for individual key points can thus be acquired by subtracting the two fields. Then, the dense incremental deformation field is interpolated from key points and their incremental moves. In particular, TPS is used here for the interpolation such that the resulted incremental deformation field remains smooth. The deformation field for S_i can then be updated following $\phi_i \leftarrow \phi_i \circ \Delta\phi_i$. Note that the smoothness requirement upon ϕ_i is achieved in the fluid style, as in every iteration the incremental $\Delta\phi_i$ remains smooth.

The whole method is further embedded into a hierarchical and multi-resolution framework. As registration progresses, more key points are involved for accurate alignment of anatomical structures. The number of key points in each image is around 10% of the brain volume size in the beginning of groupwise registration, while the rate increases to about 50% when registration finishes. Also, the soft correspondence is iteratively tightened. The reason is that multiple correspondences provide higher robustness in correspondence detection, especially when all images are not well registered

in the initial stage. However, after the deformed subject images are mostly similar to each other in the common space, the correspondence should be detected within a smaller-sized search window, and the one-to-one correspondence is then able to provide higher accuracy in estimating the (incremental) deformation field.

3.4.5 Experimental Results

The registration accuracy with respect to the two methods described above, namely Div-Min for Wang *et al.* (2010a) and GLIRT for Wu *et al.* (2012b), is evaluated and reported in the following. Again, the Dice ratios are computed from the NIREP dataset and the LONI dataset. The Div-Min method achieves the overall Dice ratio of 71.11% for the NIREP dataset and 75.36% for the LONI dataset. The corresponding scores for GLIRT are 76.52% and 76.64%, respectively. Note that both methods gain higher Dice ratios compared with the conventional pairwise registration via HAMMER. The detailed Dice ratios of all ROIs are plotted in Figure 3.12 (for the NIREP dataset) and Figure 3.13 (for the LONI dataset), respectively. The figures also contain scores yielded by HAMMER (pairwise registration) and other groupwise registration methods including the conventional group-mean method (Joshi *et al.*, 2004), the sharp-mean method (Wu *et al.*, 2011), and ABSORB (Jia *et al.*, 2010).

The scores of Div-Min are lower than GLIRT, due to the following possible reasons. First, the number of key points sampled in Div-Min is much lower than that of GLIRT, and the low sampling rate makes the evaluation of the objective function in registration less accurate. Second, the attribute vector in GLIRT is much more powerful due to the anatomical information conveyed by the features. Specifically, in correspondence detection, only points of the same tissue type could possibly be correspondences to each other. The anatomical interpretation of the attribute vector significantly reduces ambiguities in correspondence detection. Finally, each key point is allowed to search

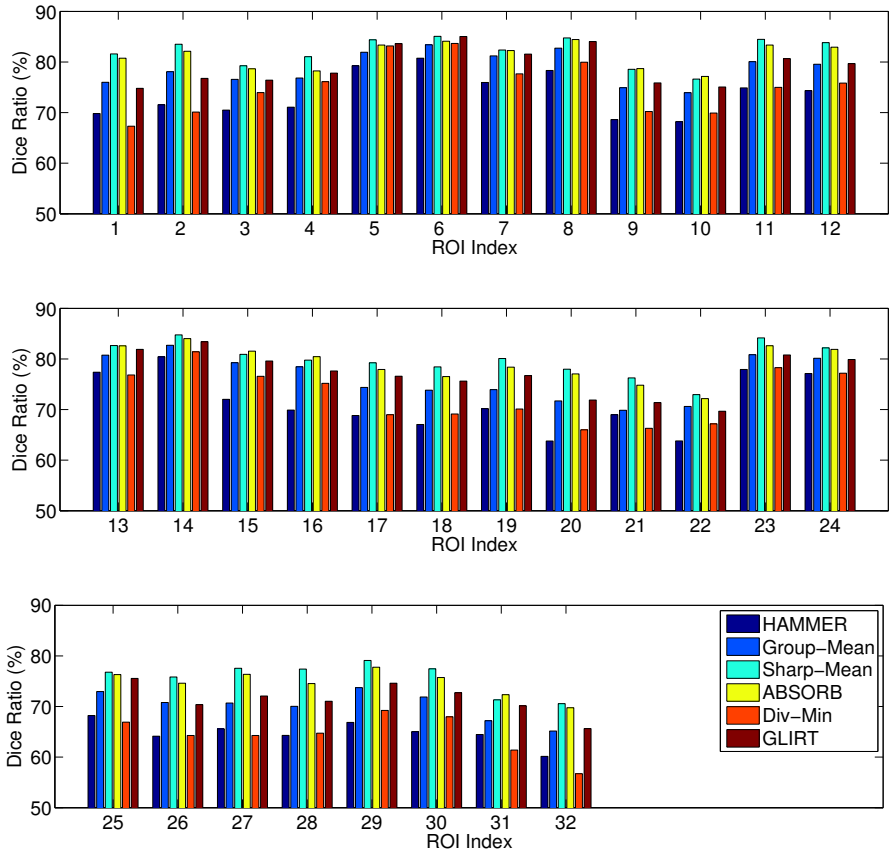


Figure 3.12: The Dice ratios for individual groupwise registration methods, as well as the pairwise registration via HAMMER (Shen and Davatzikos, 2002), on the NIREP dataset. The ROI names corresponding to their indices are listed in Table 2.3.

for its correspondences, while multiple correspondence candidates are allowed. Thus the detected correspondences are more reliable for the estimation of the deformation fields.

3.5 Sub-Group Consistency

It is often necessary to consider sub-group consistency in groupwise registration of a large-scale image population and in subsequent analyses. For example, given a population of both normal controls and patient data, the images can be naturally partitioned into two sub-groups. Though groupwise registration is capable of warping all images

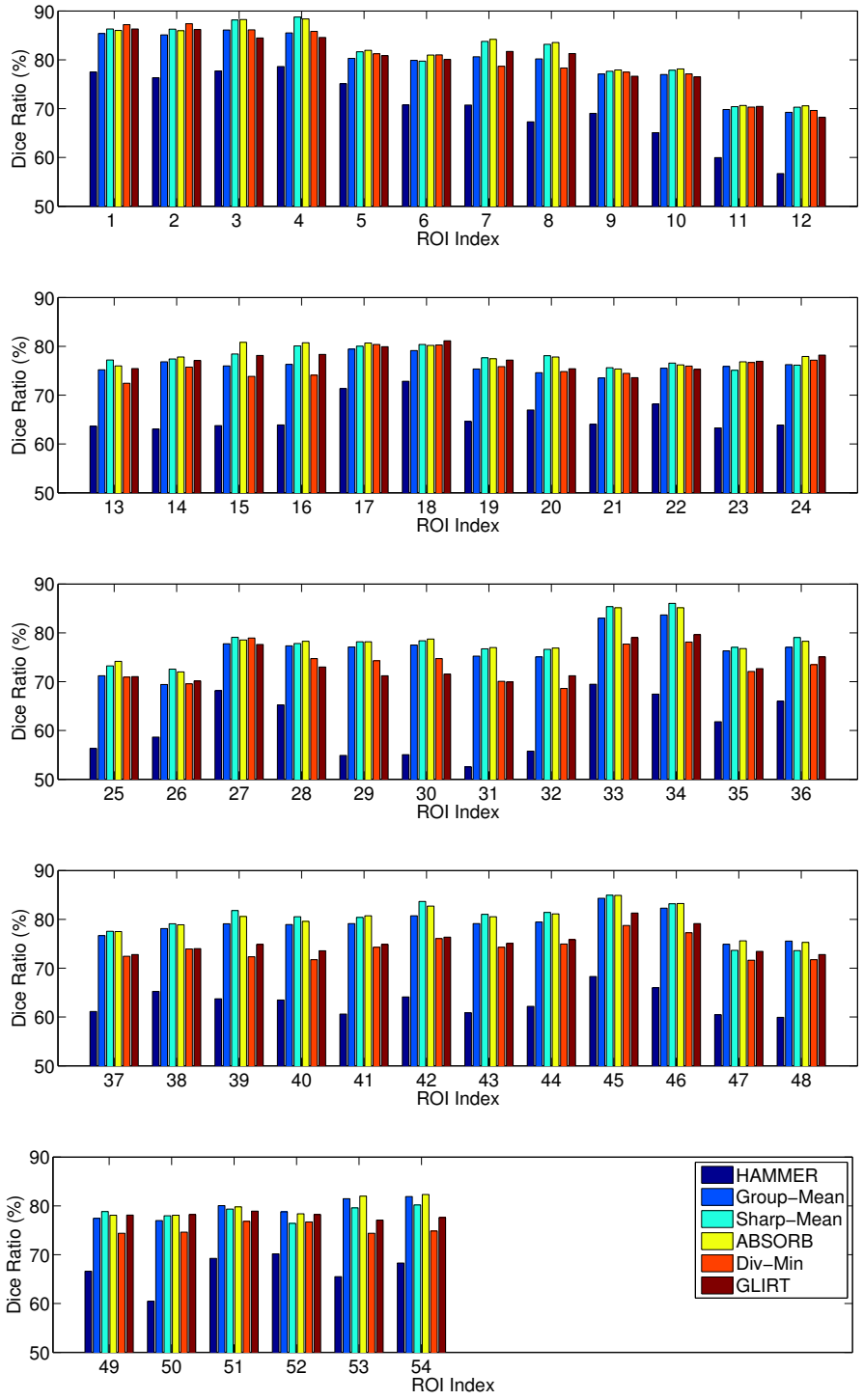


Figure 3.13: The Dice ratios for individual groupwise registration methods, as well as pairwise registration by HAMMER (Shen and Davatzikos, 2002), on the LONI dataset. The ROI names corresponding to their indices are listed in Table 2.3.

to a common space, the relationship amongst images in each sub-group might be interrupted since a certain image needs to interact with all other images regardless of their belongings to individual sub-groups. If the sub-group information (e.g., the distribution of the entire population in two sub-groups) was discarded during image registration, additional difficulties would possibly be encountered for differentiating and comparing the two sub-groups of images in subsequent studies.

3.5.1 Hierarchical Groupwise Registration

Wang *et al.* (2010b) proposed a hierarchical framework to tackle the sub-group consistency issue in groupwise registration. In particular, a clustering method is first invoked to produce a hierarchical clustering of the entire image population. Then, groupwise registration is applied to each clustered sub-group. Note that images in a certain sub-group are usually similar to each other in appearances, thus making it relatively easy to solve (groupwise) registration in the sub-group. Thereafter, an atlas for each sub-group can be constructed and regarded as the representative image of the sub-group. All representative images of individual sub-groups are further registered towards the common space of the entire population, in a new round of groupwise registration at a higher level. The deformation fields of the representative images are also applied to images in respective sub-groups. Thus, all images in the population could complete their groupwise registration with respect to the common space by following (1) the deformation fields towards the representative images in their individual sub-groups and (2) the deformation fields from representative images to the final common space of the population.

Figure 3.14 illustrates the process of hierarchical groupwise registration (Wang *et al.*, 2010b). Specifically, two sub-groups, as well as their exemplar images, are shown in the bottom of Figure 3.14. The atlas of each sub-group can thus be estimated via groupwise

registration. Then, the atlases of all sub-groups are collected and registered towards the common space of the entire population (in the top of the figure). Note that the depth of the hierarchical callbacks of groupwise registration is not arbitrarily restricted, while more levels are apparently allowed if a certain sub-group needs to be partitioned into more clusters of finer granularity to cater for the complicated distribution of images in the population.

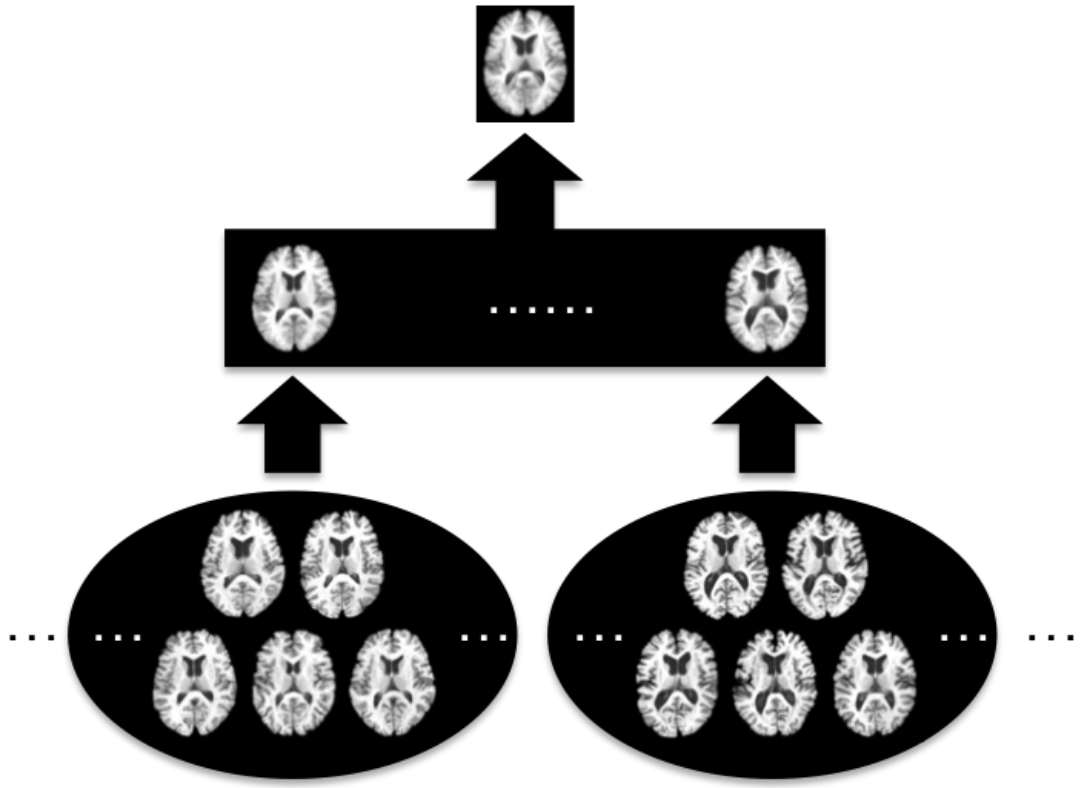


Figure 3.14: Illustration of hierarchical groupwise registration. The large-scale population is first clustered into several sub-groups (bottom). Groupwise registration is then applied to each sub-group, where an atlas is estimated accordingly (middle). The atlases of individual sub-groups are applied to a new round of groupwise registration for the estimation of the common space of the entire population (top). Finally, all images are registered with the common space in the groupwise style.

The hierarchical clustering and then registration of the large-scale population allows more precise description of the "multi-mode" distribution of brain MR images (Sabuncu *et al.*, 2009). That is, for all images in a certain sub-group, their high-level registration

(e.g., to the final common space) follows the same deformation field that is estimated from the representative image of their sub-group. In this way, the entire population of images is registered in the unbiased groupwise manner such that all images can be quantitatively compared. In addition, the interactions across individual sub-groups are eliminated; thus the subtle difference between sub-groups of images can be better revealed.

The power of the hierarchical clustering is demonstrated by the example in Figure 3.15. In particular, the population consists of 24 brain MR images and can be partitioned into two sub-groups. The first sub-group contains 12 normal controls. The second sub-group is simulated by introducing tissue atrophies to every image in the first sub-group. The tissue atrophies, which account for 10% of the original WM volumes, are applied to the regions of pre-central gyrus (PCG) and superior temporal gyrus (STG), respectively. Exemplar slices (of segmented tissues in WM and GM) from a certain image in the first sub-group are shown in Figure 3.15(a) and (c), while the corresponding image slices with atrophies in the second sub-group are shown by (b) and (d). The atrophies are highlighted by red arrows in the figure.

The entire population of the two sub-groups is then processed via hierarchical groupwise registration. After all images are registered with the common space, the Jacobian determinants of their individual deformation fields are computed at all voxels. Note that the Jacobian determinant reflects the morphological changes estimated by image registration: a determinant above 1 indicates the spatial expansion, while a value lower than 1 corresponds to the shrinkage. Further, paired t-test is applied to examine the Jacobian determinants of the deformation fields of images in the two respective sub-groups. Statistically significant differences are thus detected in the regions of PCG and STG (c.f. Figure 3.15(e) and (f)), implying the capability of hierarchical groupwise registration in revealing subtle yet systematic morphological changes between the two

sub-groups.

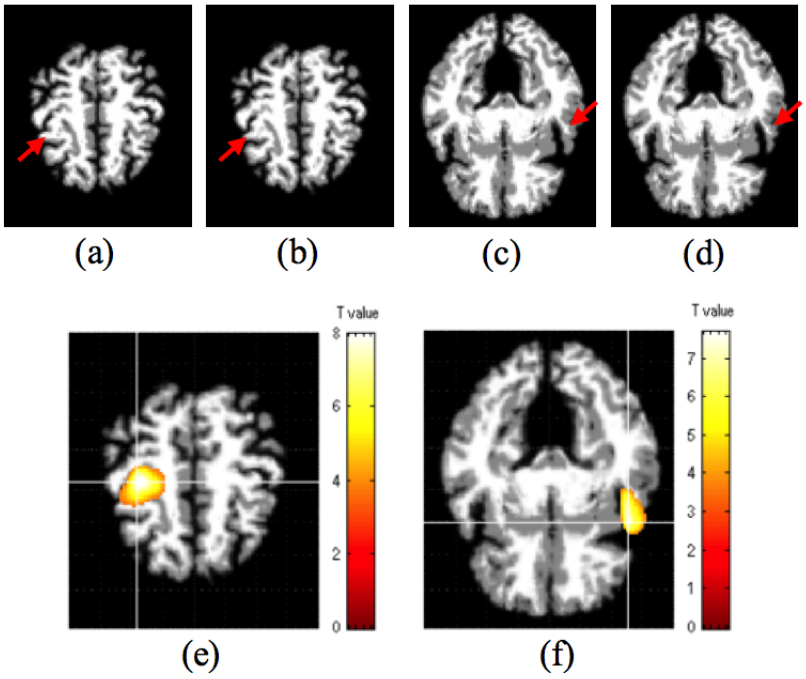


Figure 3.15: Slices of the original image, (a) and (c), and the corresponding slices with simulated atrophies in PCG (b), and STG (d), are provided, respectively. The morphological changes between the two sub-groups, with respect to the simulated atrophies, are detected via paired t-test. The distributions of t-values are shown in (e) and (f), respectively.

3.5.2 Longitudinal Constraint

Longitudinal sequences of brain MR images, which are usually referred to as 4D (three spatial dimensions and one temporal dimension) image data, need to be specially addressed in groupwise registration. Typically, a longitudinal sequence for a certain patient contains acquisitions of different time points (e.g, 3- or 6-month interval between two consecutive time points), and thus is capable of monitoring the temporal impacts of neural diseases. In that the morphological changes of brain anatomies are usually gentle and slow, the similarity amongst images of the same longitudinal sequence can be high even though their acquisition time points are different. By contrast, the inter-patient

image similarity is much lower due to the high anatomical variation across patients. To this end, the longitudinal sequence for a certain patient can be regarded as a sub-group. Then, the hierarchical groupwise registration method (c.f. Section 3.5.1) is applicable to register multiple longitudinal sequences, as well as their individual acquisitions, to a common space. After registration, quantitative analysis and comparison on the entire population of longitudinal sequences of images can be conducted.

In fact, the longitudinal sequence conveys a temporal constraint, which is far beyond regarding a certain sequence as a sub-group only. That is, when registering a certain longitudinal sequence to the common space of the population, the deformation fields associated with respective time points in the sequence should be *temporally smooth*. The reasons are apparent: (1) images corresponding two consecutive time points in the same sequence are similar to each other in appearances, and (2) similar images should share similar deformation fields with respect to the same common space. In the hierarchical groupwise registration framework, however, the temporal smoothness is ignored. Though all time points in a certain sequence are grouped together and then follow the same deformation field towards the common space of the population, the deformation fields of individual time points with respect to the representative image of the sequence are subject to independent smoothness constraint only.

For better registration of longitudinal sequences, Wu *et al.* (2012c) improved the feature-based groupwise registration method in Section 3.4.4 by introducing spatio-temporal heuristics. In the groupwise registration method designed for 3D images, each individual deformation field is required to be smooth. By contrast, in the 4D groupwise registration setting, an additional temporal smoothness constraint is applied. That is, for two consecutive time points of the same patient, the temporal changes are required to be minimal. After denoting the deformation field for the i-th sequence and the t-th time point as $\phi_{i,t}$, the regularization of $\mathbf{\Phi} = \{\phi_{i,t}\}$ then comprises of not only the

smoothness of each individual $\phi_{i,t}$, but also the magnitude of $\partial\phi_{i,t}/\partial t$.

Method	**WM**	**GM**	**CSF**
GLIRT (Wu *et al.*, 2012b)	$89.89 \pm 0.96\%$	$81.92 \pm 2.57\%$	$86.27 \pm 1.37\%$
Hierarchical (Wang *et al.*, 2010b)	$90.51 \pm 0.97\%$	$82.44 \pm 2.73\%$	$86.41 \pm 1.34\%$
Longitudinal (Wu *et al.*, 2012c)	$91.28 \pm 1.26\%$	$83.70 \pm 3.12\%$	$87.21 \pm 1.71\%$

Table 3.1: Tissue Dice rations (mean±standard deviation) concerning the consistency within individual longitudinal sequences.

In order to evaluate the effects of different registration schemes on longitudinal image sequences, a 4D image population is extracted from the Alzheimers Disease Neuroimaging Initiative (ADNI) repository (`http://www.adni-info.org/`). The population consists 5 sequences of normal controls and 5 other sequences of AD patients, while each sequence has 3 time points. All 30 images are necessarily pre-processed, including skull-stripping, tissue segmentation, and affine registration. Then three different methods are applied to register the entire population in the groupwise manner. For fair comparison, all methods use the same features to identify correspondences in registration. For hierarchical registration, each sequence is regarded as a sub-group and the entire population is registered in two levels.

After all images are registered to the common space, the performances of individual methods are evaluated as the tissue overlapping within each sequence. In particular, given all time points in a sequence that are registered to the common space, the Dice ratio of tissues (including WM, GM, and CSF) are computed between each pair of them. The mean scores averaged across individual sequences, as well as the standard deviations, are reported in Table 3.1. In the hierarchical registration scheme, the Dice ratios are higher than GLIRT (Wu *et al.*, 2012b), in that a sequence only adopts a single deformation field in registering all its member images to the common space. An additional temporal smoothness constraint is introduced to the longitudinal registration scheme; thus the Dice ratios are further improved.

3.6 Summary

Due to the capability of handling a large-scale population of images in an unbiased manner, groupwise registration is becoming a more preferred choice in the area of medical image analysis. Though it is more complex to optimize the deformation fields of multiple brain MR images simultaneously, the population provides additional information that guides all images to be registered more accurately with respect to the common space. The performance of groupwise registration thus has drawn a lot of interest, as demonstrated by the utilization of the technique in many applications and studies.

Groupwise registration differs from pairwise registration in that the template is not used. Thus, a possible solution for groupwise registration is to estimate the "template", or the atlas of the population first. Then, all images can be simply registered with the atlas via existing pairwise methods. To this end, the estimation of the atlas in the group-mean form is investigated. The mean image can be adaptively created in accordance with the distribution of all images in the population, while the sharp-mean substantially increases the performance of mean-based groupwise registration.

Moreover, the explicit common space is not always necessary. For example, a certain image can predict its deformation pathway towards the unknown common space by interacting with other similar images only. The interaction happens between closely distributed subjects on the image manifold, and can be regarded as being local. On the other hand, the common space bridges point-to-point correspondences across individual images. Thus, even though the common space is explicit, the task of groupwise registration can be accomplished by revealing groupwise correspondences with respect to all images in the population.

Chapter 4

Population-Based Registration and Multi-Atlas Labeling

4.1 Multi-Atlas Labeling

It is often necessary in many studies that certain medical images should be labeled into different anatomical ROIs, in order to facilitate the following region-based analysis. Manual labeling, though accurate and probably the best if done by well-trained experts, has a high cost, especially for large-scale populations of brain MR images. By contrast, an automatic labeling method has the advantage of reducing the need for human interaction. Meanwhile, the inconsistency among individual experts might be suppressed after replacing manual labeling with the automatic method. Therefore, the automatic labeling method is highly desirable for analysis of brain MR images and has been intensively investigated recently.

To *label* (also known as to *segment* or to *parcellate*) brain MR images can be accomplished in many different ways. Among them, atlas-based segmentation provides an efficient solution and yields comparable accuracy with respect to manual labeling (Klein *et al.*, 2009). Note that the definition of the term *atlas* here is different from

the scenario of image registration discussed in the previous sections. In the context of image registration, the atlas is typically constructed from an image population (e.g., as the mean of all registered images). The atlas thus reflects the common space of the population and is applicable to guide image registration (i.e., the group-mean based groupwise registration method in Section 3.2). In terms of atlas-based labeling, however, the atlas refers to certain individual images whose segmentations are already known. Specifically, the atlases are often manually labeled, as their labeling information (e.g., in anatomical ROIs) is perceived as the gold standard. Then, after registering the atlases to an unlabeled image, the labeling information associated with the atlases can propagate to segment the unlabeled image. If the atlas can be accurately registered with the unlabeled image, the well-established (anatomical) correspondences between the two images guarantee that the propagated labeling is highly accurate.

Though convenient to use, the atlas-based labeling method inevitably suffers from errors in registering the atlas with the unlabeled image. To this end, a multi-atlas strategy has been more preferred in recent studies to alleviate the impact of errors in image registration. After registering multiple atlases instead of a single one with the unlabeled image, the to-be-determined labeling can be estimated by fusing contributions from all atlases after registration. The accuracy of multi-atlas labeling depends on two factors, i.e., the quality of *image registration* and the reliability of *label fusion*. The impact of registration on multi-atlas labeling is the focus of this chapter, as abundant discussions over label fusion are available in the literature.

A typical scenario for multi-atlas labeling is to segment a population of images, in which only a few *atlases* are pre-labeled by experts and the labels of other images are unknown. For the sake of propagating labeling information from the atlases to unlabeled images, all atlases are usually registered with each unlabeled image following the direct pairwise registration style. More advanced registration schemes bring in

additional benefits. For instance, Wolz *et al.* (2010) utilized the concept of *image-scale guidance* and allowed the propagation to happen between similar images only. In particular, the image population, which consists of both atlases and unlabeled images, is first embedded into a high-dimensional manifold where similar images are closely distributed. Then, each atlas starts propagating its labeling to images that are close on the manifold. The registration of the atlas to each unlabeled image is accurate due to the high similarity between them, and thus results in more reliable labeling for the unlabeled image. Once the labeling is determined, the specific unlabeled image is admitted as an atlas and helps propagate the labeling information further to other images in the population. From the perspective of the entire population, the method called LEAP (Wolz *et al.*, 2010) achieves significantly better labeling results, since the propagation always involves similar images with more reliable registration in between.

After registering atlases with a certain unlabeled image, the labels of the unlabeled image can be determined, or fused, in various ways (Coupé *et al.*, 2011; Rousseau *et al.*, 2011; Wang *et al.*, 2013a; Warfield *et al.*, 2004). In the simple yet powerful majority voting scheme, for example, each grid point in the unlabeled image is assigned the most frequent label from the same locations in all registered atlases. The use of multiple atlases thus alleviates the concern that a single atlas might not necessarily be optimal. That is, even if the registration of a certain atlas and the unlabeled image is not sufficiently accurate, the labels contributed by the atlas are generally regarded as *outliers* and are likely be omitted in majority voting. As the result, with multiple atlases, the fused labeling for the unlabeled image is usually more accurate since the potential errors in registering certain atlases with the unlabeled image become less influential.

To further investigate the importance of image registration in multi-atlas labeling, a population-based registration scheme is demonstrated in this chapter. In the traditional

way, all atlases are registered with each unlabeled image by direct and independent pairwise registration. However, as in the next example, the entire image population is embedded into an MST, which helps register all images to the image space defined by the root. Then, in the root space, all unlabeled images are parcellated according to the deformed atlases in the population. The tentative labeling of the unlabeled images provides additional information that is helpful to image registration, i.e., by adjusting the MST for the population and refining the deformation fields with respect to the root space. Then, with more accurate registration, the labeling for the unlabeled images are iteratively optimized until satisfactory results are generated.

4.2 Registration and Label Fusion

To investigate the role of image registration in multi-atlas labeling specifically for a large-scale image population, a novel population-based labeling method is described. Specifically, the entire image population, including atlases and unlabeled images, is registered to a common space first. Then, the labeling of each labeled image can be estimated in the common space. Moreover, the label fusion for a certain unlabeled image incorporates contributions from not only registered atlases but also tentative labeling of other unlabeled images in the population. Further, the (tentative) labeling of all images contributes to registering all images more accurately in the common space. In general, the method consists of two alternating steps: (1) to register all images to the common space; and (2) to determine the labeling for each unlabeled image. The steps are iteratively applied.

4.2.1 Tree Based Image Registration

All atlases and unlabeled images in the population are registered to a common space by taking advantage of the distribution of the entire image population. In particular, all images are first embedded into a fully connected graph, where the nodes indicate individual images and the edge linking each pair of images records their in-between distance, i.e., in SSD. An MST is then extracted from the graph. The root of the tree is determined to represent the geometric median image in the population, from which the sum of distances to other images in the population is minimized. All images are connected with the root of the tree either directly or via other images/nodes.

The node at the root of the tree, or the median image of the population, is designated as the common space to which all images in the population are registered. It is worth noting that the common space of the median image is different from the definition in groupwise registration (c.f. Chapter 3). However, the potential bias of the median image from the real common space in groupwise registration does not affect the efficiency of multi-atlas labeling. Specifically, as shown later, the common space of the median image here only provides an image space for all images to share their labeling information. Also, it is known that the median image can be regarded as an approximation of the unbiased common space in the sense of groupwise registration.

The learned MST helps register all images in the population to the root in a recursive manner. In particular, given each non-root image, the path that it traverses along edges to the root of the tree can be easily identified. If the parent node of the image under consideration is the root, the direct registration via diffeomorphic Demons (Vercauteren *et al.*, 2009) will be computed immediately. Otherwise, the non-root image will utilize the deformation belonging to its parent node as an initialization and further refine to generate its own deformation towards the root. The recursive callbacks can eventually deform all images to the common space. Compared with the direct registration of two

images that might be very different in anatomies, the MST provides robust initialization in estimating the deformation field. Thus, to organize images in the MST helps achieve higher accuracy overall to register all images to the common space.

Due to the essentially high-dimensional image data and the limited size of the image population, the estimation of the MST might be inaccurate. As a remedy, an augmented population that consists of more simulated images is generated for building the MST. The simulated images are derived by perturbing the pre-determined median image in five steps (Jia *et al.*, 2012b):

1. A set of images is directly registered with the median image;
2. All deformations are then inverted for the sake of deforming the median image;
3. PCA is applied to capture the variation within all inverted deformations;
4. By perturbing coefficients in the learned PCA model, a set of deformations can be simulated;
5. All simulated deformations are applied to warp the median image and generate a set of simulated images in the final.

The same setting in Jia *et al.* (2012b) is followed to specify the number of the simulated images to be twice the size of the original image population. The augmented population, including atlases, unlabeled images, and simulated images, leads to an MST that better captures the distribution of the image population.

4.2.2 Consistent Label Fusion

After all images are registered to the common space, the labeling for the unlabeled images can be fused from the atlases. In particular, the local voting strategy is adopted for stochastic label fusion. By denoting the m-th registered atlas as $\dot{S}_m$ and its labeling

as $\dot{L}_m$, the label for the n-th unlabeled image S_n, or L_n, can be assigned with the label l at the location x, following the likelihood

$$p(L_n(x) = l) \propto \sum_m w(\dot{S}_m, S_n, x) \cdot \delta(\dot{L}_m(x), l). \tag{4.1}$$

In the equation above, $\delta(\dot{L}_m(x), l)$ is the Dirac function and returns 1 if and only if $\dot{L}_m(x) = l$; otherwise, it returns 0. The weight $w(\dot{S}_m, S_n, x)$ indicates the contribution of $\dot{S}_m$ to label S_n by $\dot{L}_m$, and obviously relates to the similarity between $\dot{S}_m$ and S_n at x. By using $d(\dot{S}_m, S_n, x)$ to denote the SSD of the two respective intensity patches centered at x of both $\dot{S}_m$ and S_n (with the size $3 \times 3 \times 3$ in voxel), it is defined that $w(\dot{S}_m, S_n, x) = \exp\left(-d^2(\dot{S}_m, S_n, x)/2\sigma^2\right)$ as σ relates to the standard deviation of all patch-to-patch dissimilarity measures. The likelihood $p(L_n(x) = l)$ is further normalized such that $\sum_l p(L_n(x) = l) = 1, \forall x$. The exact label of $L_n(x)$ is determined as the value l of the maximal likelihood in the final result.

Most methods reported in the literature apply the mono-directional label fusion by propagating the labeling from the atlases to the unlabeled image only. However, recent studies show that the segmentation of an unlabeled image can also benefit from other unlabeled images (Jia *et al.*, 2012b; Wang *et al.*, 2013a). That is, the labeling should propagate not only from atlases to unlabeled images, but also among unlabeled images if they are similar. Specifically, a certain unlabeled image can finally be segmented from fusing both the labels of atlases and the tentative segmentation of other unlabeled images. In this way, not only is the consistency across the segmentation of each unlabeled image and the atlases well preserved, but also the intrinsic consistency among all unlabeled images.

For the consistency of the population in label fusion, the likelihood in labeling $S_n(x)$

can be calculated by

$$p(L_n(x) = l) = \sum_m w(\dot{S}_m, S_n, x) \cdot \delta(\dot{L}_m(x), l) + \sum_{n'} w(S_{n'}, S_n, x) \cdot \delta(L_{n'}(x), l). \quad (4.2)$$

A stable solution to the above can be iteratively attained (Jia *et al.*, 2012b). (4.2) implies that the label $L_n(x)$ complies with both the atlases and other unlabeled images. Also note that the simulated images, though participating in the registration, are not included in the label fusion. All simulated images are instantiated by perturbing the median image, which would arbitrarily dominate the segmentation result with the simulated images included in the label fusion.

4.2.3 Interaction of Registration and Label Fusion

Though the unlabeled images can fuse their labeling from the atlases, it is worth noting that the tentative labeling is derived from the yet imperfect registration. By contrast, the tentative labeling is capable of feeding back for more accurate registration of the population. As all images become better registered with the common space, the labeling of the unlabeled images can be determined more easily and accurately. The registration benefits from the tentative labeling in two ways:

1. The initial MST is estimated prior to the non-rigid registration. The high variation among all images, as well as the simple image distance measure (i.e., SSD of intensities), may lead to improperly estimated MST and thus limit the registration accuracy. On the other hand, after all images are roughly registered to the common space, the distribution of the entire image population is relatively compact and can be better learned by considering the consistency of the labeling of all images. That is, the MST can be updated by considering the tentative labeling.

2. The registration should also favor the consistency within the labeling of individual images, which is only pursued in the label fusion part of conventional methods. In particular, the (tentative) labeling of all images is regarded as an additional image descriptor other than image intensities. In addition to maximizing the intensity coherence, image registration also aims to directly eliminate the labeling inconsistency as well. In particular, the registration is required to align the boundaries of corresponding labels of individual images. The estimated deformation fields are then applied to register all images more accurately in the common space.

Update MST.

To learn the MST for representing the distribution of the entire population, the image distance is evaluated according to the inconsistency between their labeling after all images are (roughly) registered to the common space. Hence the distance measure is different from the SSD for the initial MST. For any two images in the population, their overall Dice ratio is computed and averaged over all labels. The distance of the two images is then derived by inverting the Dice ratio in an exponential decay style. Given pairwise distances of all images, a new MST is thus constructed. Note the median image that is previously selected in the initial MST is still preserved as the root of the new tree. In this way, the common space in registration is fixed, though each non-root image will further refine its own deformation field towards the root. Moreover, the updated MST consists of only atlases and unlabeled images, while the simulated images are not incorporated. The reason is that, after the initial registration, the atlases and the unlabeled images distribute compactly in the vicinity of the median image. Thus the simulated images are not necessary for the update of the tree.

Update Registration

The (tentative) labeling of all images is directly applied to the refinement of the deformation fields in the registration, in order to compensate for the inconsistency within the labeling. After deforming all images towards the median image of the population, the boundaries of all anatomical areas labeled on each deformed image are extracted by applying the Canny edge detector to the tentative labels. Points on the label boundaries form a discrete pointset for each image, while to align the boundary pointsets of individual images is critically important in image registration and meaningful for anatomy parcellation. To this end, a Gaussian kernel is applied to smooth the detected boundaries and convert the discrete boundary pointset into a continuous volume of Gaussian mixtures in the image space (Wang *et al.*, 2008). The volumes of Gaussian mixtures for the labeling of a pair of images can thus be easily registered, i.e., via diffeomorphic Demons (Vercauteren *et al.*, 2009). The newly updated MST is also applied in the above, as the registration upon the label boundaries is performed by following edges in the MST recursively.

Note that all images further refine their registration to the common space after being warped following their previously estimated deformation fields. Therefore, the feedback from the (tentative) labeling leads to an iterative solution, as all images refine their deformation fields to the common space and the unlabeled images update their tentative labeling alternatively. The previous deformation field of each image and its new deformation for refinement are concatenated into a single field, which warps the image from its original space to the common space. To compensate for potential errors in the above, the concatenated deformation functions as an initialization, and is further refined by directly minimizing the intensity of the inhomogeneity between the specific image and the median image designating the common space. In this way, both

image intensities and the (tentative) labeling contribute to update the registration. The interaction between registration and label fusion finally results in more accurate parcellation for the unlabeled images.

4.3 Experimental Results

The performance, or the accuracy of multi-atlas labeling, is measured in the Dice ratio upon the NIREP NA0 dataset. All 16 images in the dataset are manually labeled for the ground truth. Then, 8 of them are regarded as atlases and used for the labeling of the remaining images. The Dice ratios are then computed between the estimated labeling of the 8 unlabeled images and their ground truth.

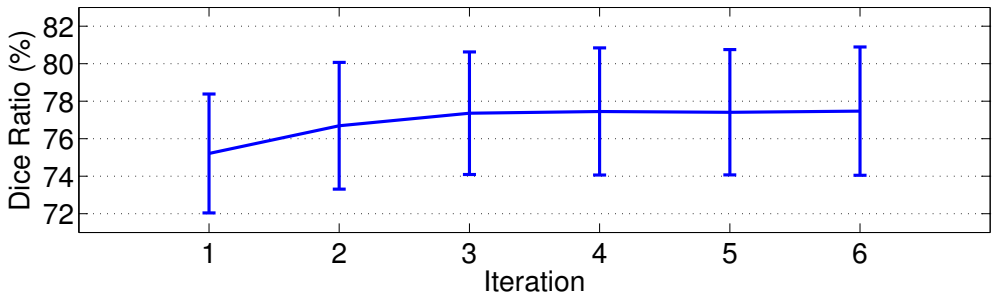


Figure 4.1: The Dice ratio converges quickly during the iterative interaction between registration of the population and label fusion for the unlabeled images.

In the conventional way of multi-atlas labeling, all atlases are registered with each unlabeled image. The label fusion is then conducted in the one-shot format from the registered atlas to the unlabeled image. The overall Dice ratio, averaged upon all unlabeled images and their ROIs, is 73.09%. In comparison with the introduction of the MST and the consistency requirement across the labeling of all unlabeled images, the Dice ratios increases to 75.31%. Further, by allowing the (tentative) labeling to feed back, the registration of the population is improved while the corresponding Dice ratio

reaches 77.45%, implying that more accurate multi-atlas labeling results are yielded. Note that the interaction between registration and label fusion leads to an iterative solution. The reported scores are achieved after 4 iterations. The evolution of the Dice ratio with respect to the increasing number of iterations is plotted in Figure 4.1. Meanwhile, detailed Dice ratios for all 32 ROIs in the NIREP dataset are shown in Figure 4.2.

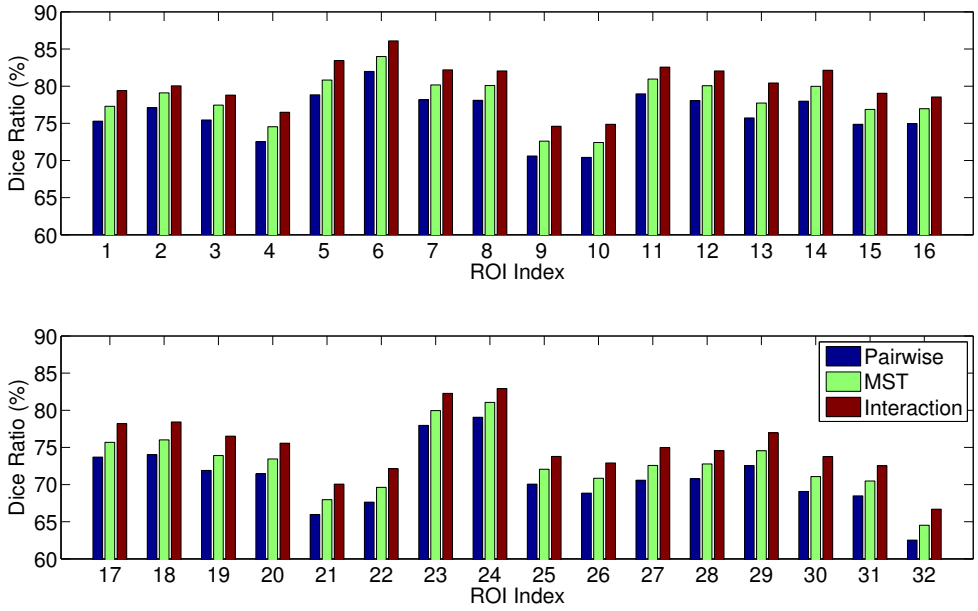


Figure 4.2: Dice ratios of individual multi-atlas labeling methods on the NIREP dataset. The ROI names corresponding to their indices are listed in Table 2.3.

4.4 Summary

Image registration plays an important role in many related applications, including multi-atlas labeling as demonstrated in this chapter. In particular, the impact of population-based registration upon the accuracy of multi-atlas labeling is investigated and examined. Given a population that consists of both atlases and unlabeled images, an MST is created to guide the registration of each image with the root. Then,

within the space of the root image, the unlabeled images can fuse their labels from the registered atlases. Moreover, the interaction between registration and label fusion is considered, such that the (tentative) labeling of all images helps improve image registration, while better registration in turn leads to more accurate label fusion results. The experiments clearly show that the performance of multi-atlas labeling is improved by considering information from the entire image population.

Chapter 5

Summary and Conclusion

As pointed out from the very beginning of this dissertation, image registration is fundamentally important in the area of medical image analysis. In studies of brain MR images specifically, the non-rigid registration technique is often utilized to align individual images and then capture the anatomical variation within them. That is, deformed images become similar to each other after registration, while their intrinsic dissimilarity is encoded by the deformation field that is estimated during image registration. The registration technique enables comparison of individual images after warping them to the same space. Meanwhile, quantitative analysis can be conducted accurately in accordance with the deformation field associated with each image in the population.

Though many efforts have been devoted to the study of brain MR image registration, the problem is still not fully resolved yet. Image registration is typically perceived as an optimization problem with two images involved, by estimating the deformation field for the subject image to maximize the similarity between the deformed subject and the template. A major challenge to image registration is related to its conventional pairwise design - only the subject and the template images are considered, while in practice, most applications of image registration nowadays need to handle large-scale populations of images. Specifically, given an image population, a certain image needs to be designated as the template while the template selection is usually arbitrary. All

other images in the population are registered with the template independently via the conventional pairwise method. Additionally, the concept of "population" is mostly ignored during the registration of images in the population.

In contrast with the traditional method above, this dissertation investigates the role of the "population" in image registration, and concludes that the population information benefits the registration of all images in the population. In particular, the population can be utilized in two ways:

1. The pairwise registration of a certain subject and the template benefits from the guidance provided by other intermediate images in the population;

2. The population allows groupwise registration, in which all images are registered towards the common space of the population simultaneously and accurately.

The guidance from the intermediate images can be utilized due to the observation that two similar images share similar deformation fields when registered to the same template. Thus, if the subject identifies another (intermediate) image from the population that is similar in appearance, the deformation field of the intermediate image provides a well-behaving approximation and then initialization in estimating the subject deformation. The granularity of the guidance can be further reduced to the patch scale, such that different parts of the subject can utilize the guidance from multiple intermediate images flexibly. In general, the introduction of the intermediate images reduces the difficulties faced by the registration of the subject. The robustness and accuracy of pairwise registration are thus improved from the perspective of the entire image population.

To avoid the potential bias caused by the arbitrary template, groupwise registration is proposed, such that all images deform towards the unbiased common space of the population simultaneously. Compared to the traditional pairwise registration, the lack

of the template in groupwise registration incurs additional complexity in estimating the common space, besides the deformation fields for individual subjects. A mean image of the population can thus be explicitly estimated for the description of the common space. Then, groupwise registration is decomposed into two steps, i.e., to refine the group-mean and to compute the deformation field between each subject and the mean image. Alternatively, the explicit construction of the mean image, or the atlas, can even be skipped for groupwise registration. For example, by maximizing the coherence of all deformed images, the entire population can be registered to the common space where point-to-point correspondences across individual subject images are well established.

More powerful image registration techniques facilitate research and clinical studies of medical images substantially. In this dissertation, a simple demonstration is provided to show that improved registration leads to better solutions in terms of multi-atlas labeling, which is commonly used for automatic parcellation and understanding of brain MR images. Furthermore, other studies can also benefit from incorporating population information into image registration, with regard to better registration performance (e.g., robustness, accuracy, time cost, etc.). Also, note that several tools used in this dissertation are publicly available already. More detains can be found from NITRC (`http://www.nitrc.org/projects/absorb/`, `http://www.nitrc.org/projects/hammer_suite/`, `http://www.nitrc.org/projects/glirt/`). To conclude this dissertation, though population information is introduced and applied to image registration, surely more investigation is necessary, especially for related applications of medical image analysis.

BIBLIOGRAPHY

Aggarwal, N., Rana, B., and Agrawal, R. (2012). Classification of alzheimer's from t2 trans-axial brain mr images: A comparative study of feature extraction techniques. *International Journal of Computer Vision and Image Processing (IJCVIP)*, **2**(3), 50–63.

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

Aljabar, P., Wolz, R., and Rueckert, D. (2012). Manifold learning for medical image registration, segmentation, and classification. *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis. IGI Global*.

Amit, Y. (1994). A nonlinear variational problem for image matching. *SIAM Journal on Scientific Computing*, **15**(1), 207–224.

Amit, Y., Grenander, U., and Piccioni, M. (1991). Structural image restoration through deformable templates. *Journal of the American Statistical Association*, **86**(414), 376–387.

Andersson, J., Smith, S., and Jenkinson, M. (2008). Fnirt-fmrib's non-linear image registration tool. In *Fourteenth Annual Meeting of the Organization for Human Brain Mapping*, pages 15–19.

Ardekani, B. A. and Bachman, A. H. (2009). Model-based automatic detection of the anterior and posterior commissures on mri scans. *Neuroimage*, **46**(3), 677–682.

Ashburner, J. and Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human brain mapping*, **7**(4), 254–266.

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, **12**(1), 26–41.

Balafar, M. A., Ramli, A. R., Saripan, M. I., and Mashohor, S. (2010). Review of brain mri image segmentation methods. *Artificial Intelligence Review*, **33**(3), 261–274.

Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, **61**(2), 139–157.

Bookstein, F. L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **11**(6), 567–585.

Camara, O., Schnabel, J. A., Ridgway, G. R., Crum, W. R., Douiri, A., Scahill, R. I., Hill, D. L., and Fox, N. C. (2008). Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal alzheimer's disease images. *NeuroImage*, **42**(2), 696–709.

Casey, B., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, **54**(1–3), 241 – 257.

Chou, C.-R., Frederick, B., Mageras, G., Chang, S., and Pizer, S. (2013). 2d/3d image registration using regression learning. *Computer Vision and Image Understanding*, **117**(9), 1095 – 1106.

Christensen, G. E. and Johnson, H. J. (2001). Consistent image registration. *Medical Imaging, IEEE Transactions on*, **20**(7), 568–582.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996). Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, **5**(10), 1435–1447.

Chui, H. and Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, **89**(2–3), 114 – 141.

Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., and Collins, D. L. (2011). Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, **54**(2), 940–954.

Cuadra, M. B., Pollo, C., Bardera, A., Cuisenaire, O., Villemure, J.-G., and Thiran, J.-P. (2004). Atlas-based segmentation of pathological mr brain images using a model of lesion growth. *Medical Imaging, IEEE Transactions on*, **23**(10), 1301–1314.

Dalal, P., Shi, F., Shen, D., and Wang, S. (2010). Multiple cortical surface correspondence using pairwise shape similarity. In T. Jiang, N. Navab, J. Pluim, and M. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, volume 6361, pages 349–356. Springer Berlin Heidelberg.

Davies, R. H., Twining, C. J., Cootes, T. F., Waterton, J. C., and Taylor, C. J. (2002). A minimum description length approach to statistical shape modeling. *Medical Imaging, IEEE Transactions on*, **21**(5), 525–537.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1997). Support vector regression machines. *Advances in neural information processing systems*, pages 155–161.

Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E. M., Brammer, M. J., Murphy, C., and Murphy, D. G. (2010). Investigating the predictive value of whole-brain structural mr scans in autism: a pattern classification approach. *Neuroimage*, **49**(1), 44–56.

Fletcher, P. T., Venkatasubramanian, S., and Joshi, S. (2009). The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage*, **45**(1), S143–S152.

Floater, M. S. and Iske, A. (1996). Multistep scattered data interpolation using compactly supported radial basis functions. *J. Comp. Appl. Math*, **73**, 65–78.

Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. *science*, **315**(5814), 972–976.

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural mri in alzheimer disease. *Nat Rev Neurol*, **6**(2), 67–77.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

Genton, M. G., Cristianini, N., Shawe-taylor, J., and Williamson, R. (2001). Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, **2**, 299–312.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., and Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal mri study. *Nat Neurosci*, **2**(10), 861–863.

Girosi, F., Jones, M., and Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, **7**(2), 219–269.

Glocker, B., Sotiras, A., Komodakis, N., and Paragios, N. (2011). Deformable medical image registration: Setting the state of the art with discrete methods*. *Annual review of biomedical engineering*, **13**, 219–244.

Haber, E. and Modersitzki, J. (2006). Intensity gradient based registration and fusion of multi-modal images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 726–733. Springer.

Hajnal, J. V. and Hill, D. L. (2010). *Medical image registration*. CRC press.

Hamm, J., Ye, D. H., Verma, R., and Davatzikos, C. (2010). Gram: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis*, **14**(5), 633 – 642.

Heimann, T. and Meinzer, H.-P. (2009). Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, **13**(4), 543 – 563.

Holland, B. A., Haas, D. K., Norman, D., Brant-Zawadzki, M., and Newton, T. H. (1986). Mri of normal brain maturation. *American Journal of Neuroradiology*, **7**(2), 201–8.

Holland, D. and Dale, A. M. (2011). Nonlinear registration of longitudinal images and measurement of change in regions of interest. *Medical image analysis*, **15**(4), 489–497.

Jack, C. R., Shiung, M. M., Gunter, J. L., O'Brien, P. C., Weigand, S. D., Knopman, D. S., Boeve, B. F., Ivnik, R. J., Smith, G. E., Cha, R. H., Tangalos, E. G., and Petersen, R. C. (2004). Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, **62**(4), 591–600.

Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, **5**(2), 143 – 156.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, **17**(2), 825 – 841.

Jia, H., Wu, G., Wang, Q., and Shen, D. (2010). Absorb: Atlas building by self-organized registration and bundling. *NeuroImage*, **51**(3), 1057 – 1070.

Jia, H., Yap, P.-T., Wu, G., Wang, Q., and Shen, D. (2011a). Intermediate templates guided groupwise registration of diffusion tensor images. *NeuroImage*, **54**(2), 928–939.

Jia, H., Wu, G., Wang, Q., Kim, M., and Shen, D. (2011b). itree: Fast and accurate image registration based on the combinative and incremental tree. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1243–1246.

Jia, H., Wu, G., Wang, Q., Wang, Y., Kim, M., and Shen, D. (2012a). Directed graph based image registration. *Computerized Medical Imaging and Graphics*, **36**(2), 139 – 151.

Jia, H., Yap, P.-T., and Shen, D. (2012b). Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage*, **59**(1), 422 – 430.

Jiang, H., Robb, R. A., and Tainter, K. S. H. (1992). New approach to 3-d registration of multimodality medical images by surface matching. In *Visualization in biomedical computing*, pages 196–213. International Society for Optics and Photonics.

Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, **23**, S151–S160.

Kim, M., Wu, G., Yap, P.-T., and Shen, D. (2012). A general fast registration framework by learning deformation-appearance correlation. *Image Processing, IEEE Transactions on*, **21**(4), 1823–1833.

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., *et al.* (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, **46**(3), 786–802.

Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., Fox, N. C., Jack, C. R., Ashburner, J., and Frackowiak, R. S. (2008). Automatic classification of mr scans in alzheimer's disease. *Brain*, **131**(3), 681–689.

Laakso, M. P., Frisoni, G. B., Könönen, M., Mikkonen, M., Beltramello, A., Geroldi, C., Bianchetti, A., Trabucchi, M., Soininen, H., and Aronen, H. J. (2000). Hippocampus and entorhinal cortex in frontotemporal dementia and alzheimer's disease: a morphometric {MRI} study. *Biological Psychiatry*, **47**(12), 1056 – 1063.

Learned-Miller, E. G. (2006). Data driven image models through continuous joint alignment. *Journal IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(2), 236–250.

Lenroot, R. K. and Giedd, J. N. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews*, **30**(6), 718 – 729.

Li, G., Nie, J., Wu, G., Wang, Y., and Shen, D. (2012). Consistent reconstruction of cortical surfaces from longitudinal brain mr images. *NeuroImage*, **59**(4), 3805–3820.

Lin, J. (1991). Divergence measures based on the shannon entropy. *Information Theory, IEEE Transactions on*, **37**(1), 145–151.

Liu, J., Ji, S., and Ye, J. (2009). Multi-task feature learning via efficient l2, 1-norm minimization. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 339–348, Arlington, Virginia, United States. AUAI Press.

Lo, C.-H. and Don, H.-S. (1989). 3-d moment forms: their construction and application to object identification and positioning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **11**(10), 1053–1064.

Lyu, I., Kim, S. H., Seong, J.-K., Yoo, S. W., Evans, A. C., Shi, Y., Sanchez, M., Niethammer, M., and Styner, M. A. (2013). Group-wise cortical correspondence via sulcal curve-constrained entropy minimization. In *Information Processing in Medical Imaging*, pages 364–375. Springer.

Ma, J., Miller, M. I., Trouvé, A., and Younes, L. (2008). Bayesian template estimation in computational anatomy. *NeuroImage*, **42**(1), 252–261.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, **16**(2), 187–198.

Maurer Jr, C. R., Fitzpatrick, J. M., Wang, M. Y., Galloway Jr, R. L., Maciunas, R. J., and Allen, G. S. (1997). Registration of head volume images using implantable fiducial markers. *Medical Imaging, IEEE Transactions on*, **16**(4), 447–462.

Modersitzki, J. (2004). *Numerical Methods for Image Registration (Numerical Mathematics and Scientific Computation)*. Oxford university press USA.

Mulnard, R. A., Cotman, C. W., Kawas, C., van Dyck, C. H., Sano, M., Doody, R., Koss, E., Pfeiffer, E., Jin, S., Gamst, A., Grundman, M., Thomas, R., Thal, L. J., and for the Alzheimer's Disease Cooperative Study (2000). Estrogen replacement therapy for treatment of mild to moderate alzheimer disease: A randomized controlled trial. *JAMA*, **283**(8), 1007–1015.

Munsell, B. C., Temlyakov, A., Styner, M., and Wang, S. (2012). Pre-organizing shape instances for landmark-based shape correspondence. *International Journal of Computer Vision*, **97**(2), 210–228.

Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **32**(12), 2262–2275.

Oliveira, F. P. and Tavares, J. M. R. (2012). Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, (ahead-of-print), 1–21.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., Rapoport, J. L., and Evans, A. C. (1999). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science*, **283**(5409), 1908–1911.

Paus, T., Collins, D., Evans, A., Leonard, G., Pike, B., and Zijdenbos, A. (2001). Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin*, **54**(3), 255 – 266.

Pelizzari, C. A., Chen, G. T., Spelbring, D. R., Weichselbaum, R. R., and Chen, C.-T. (1989). Accurate three-dimensional registration of ct, pet, and/or mr images of the brain. *Journal of computer assisted tomography*, **13**(1), 20–26.

Pham, D. L., Xu, C., and Prince, J. L. (2000). Current methods in medical image segmentation. *Annual review of biomedical engineering*, **2**(1), 315–337.

Pluim, J. P., Maintz, J. A., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, **22**(8), 986–1004.

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B., and Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the mcdonald criteria. *Annals of Neurology*, **69**(2), 292–302.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., and Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, **15**(11), 1676–1689.

Resnick, S. M. and Maki, P. M. (2001). Effects of hormone replacement therapy on cognitive and brain aging. *Annals of the New York Academy of Sciences*, **949**(1), 203–214.

Resnick, S. M., Goldszal, A. F., Davatzikos, C., Golski, S., Kraut, M. A., Metter, E. J., Bryan, R. N., and Zonderman, A. B. (2000). One-year age changes in mri brain volumes in older adults. *Cerebral Cortex*, **10**(5), 464–472.

Rohlfing, T. (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE Transactions on Medical Imaging*, **31**(2), 153–163.

Rousseau, F., Habas, P. A., and Studholme, C. (2011). A supervised patch-based approach for human brain labeling. *Medical Imaging, IEEE Transactions on*, **30**(10), 1852–1862.

Rueckert, D. and Schnabel, J. (2011). Medical image registration. In T. M. Deserno, editor, *Biomedical Image Processing*, pages 131–154. Springer Berlin Heidelberg.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, **18**(8), 712–721.

Sabuncu, M. R., Balci, S. K., Shenton, M. E., and Golland, P. (2009). Image-driven population analysis through mixture modeling. *Medical Imaging, IEEE Transactions on*, **28**(9), 1473–1487.

Schnabel, J. A., Rueckert, D., Quist, M., Blackall, J. M., Castellano-Smith, A. D., Hartkens, T., Penney, G. P., Hall, W. A., Liu, H., Truwit, C. L., *et al.* (2001). A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, pages 573–581. Springer.

Seghers, D., D'Agostino, E., Maes, F., Vandermeulen, D., and 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*, pages 696–703. Springer.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., and Toga, A. W. (2008). Construction of a 3d probabilistic atlas of human cortical structures. *NeuroImage*, **39**(3), 1064 – 1080.

Shen, D. and Davatzikos, C. (2002). Hammer: hierarchical attribute matching mechanism for elastic registration. *Medical Imaging, IEEE Transactions on*, **21**(11), 1421–1439.

Smith, S. M., De Stefano, N., Jenkinson, M., and Matthews, P. M. (2001). Normalized accurate measurement of longitudinal brain change. *Journal of computer assisted tomography*, **25**(3), 466–475.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., and Toga, A. W. (1999a). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nat Neurosci*, **2**(10), 859–861.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Batth, R., Jernigan, T. L., and Toga, A. W. (1999b). Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage*, **9**(6), 587 – 597.

Staring, M., van der Heide, U. A., Klein, S., Viergever, M. A., and Pluim, J. (2009). Registration of cervical mri using multifeature mutual information. *Medical Imaging, IEEE Transactions on*, **28**(9), 1412–1421.

Tang, S., Fan, Y., Wu, G., Kim, M., and Shen, D. (2009). Rabbit: Rapid alignment of brains by building intermediate templates. *NeuroImage*, **47**(4), 1277 – 1287.

Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**(5500), 2319–2323.

Thompson, P. M., Moussai, J., Zohoori, S., Goldkorn, A., Khan, A. A., Mega, M. S., Small, G. W., Cummings, J. L., and Toga, A. W. (1998). Cortical variability and asymmetry in normal aging and alzheimer's disease. *Cerebral Cortex*, **8**(6), 492–509.

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., and Toga, A. W. (2000). Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature*, **404**(6774), 190–193.

Toga, A. W. and Thompson, P. M. (2001). The role of image registration in brain mapping. *Image and Vision Computing*, **19**(1), 3–24.

Vemuri, B. C., Ye, J., Chen, Y., and Leonard, C. M. (2003). Image registration via level-set motion: Applications to atlas-based segmentation. *Medical image analysis*, **7**(1), 1–20.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, **45**(1, Supplement 1), S61 – S72.

Viola, P. and Wells III, W. M. (1997). Alignment by maximization of mutual information. *International journal of computer vision*, **24**(2), 137–154.

Wang, F., Vemuri, B. C., Rangarajan, A., and Eisenschenk, S. J. (2008). Simultaneous nonrigid registration of multiple point sets and atlas construction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **30**(11), 2011–2022.

Wang, H., Suh, J., Das, S., Pluta, J., Craige, C., and Yushkevich, P. (2013a). Multi-atlas segmentation with joint label fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **35**(3), 611.

Wang, Q., Wu, G., Yap, P.-T., and Shen, D. (2010a). Attribute vector guided groupwise registration. *NeuroImage*, **50**(4), 1485–1496.

Wang, Q., Chen, L., Yap, P.-T., Wu, G., and Shen, D. (2010b). Groupwise registration based on hierarchical image clustering and atlas synthesis. *Human brain mapping*, **31**(8), 1128–1140.

Wang, Q., Kim, M., Wu, G., and Shen, D. (2013b). Joint learning of appearance and transformation for predicting brain mr image registration. In *Information Processing in Medical Imaging*, pages 499–510. Springer.

Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, **23**(7), 903–921.

West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer Jr, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., *et al.* (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, **21**(4), 554–568.

Wolz, R., Aljabar, P., Hajnal, J. V., Hammers, A., and Rueckert, D. (2010). Leap: Learning embeddings for atlas propagation. *NeuroImage*, **49**(2), 1316 – 1325.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., and Mazziotta, J. C. (1998a). Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of computer assisted tomography*, **22**(1), 139–152.

Woods, R. P., Grafton, S. T., Watson, J. D., Sicotte, N. L., and Mazziotta, J. C. (1998b). Automated image registration: Ii. intersubject validation of linear and nonlinear models. *Journal of computer assisted tomography*, **22**(1), 153–165.

Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T., and Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, **98**(6), 1031–1044.

Wu, G., Qi, F., and Shen, D. (2007). Learning best features and deformation statistics for hierarchical registration of mr brain images. In *Information Processing in Medical Imaging*, pages 160–171. Springer.

Wu, G., Yap, P.-T., Kim, M., and Shen, D. (2010). Tps-hammer: Improving {HAMMER} registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation. *NeuroImage*, **49**(3), 2225 – 2233.

Wu, G., Jia, H., Wang, Q., and Shen, D. (2011). Sharpmean: Groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage*, **56**(4), 1968–1981.

Wu, G., Jia, H., Wang, Q., Shi, F., Yap, P.-T., and Shen, D. (2012a). *Emergence of Groupwise Registration in MR Brain Study*, pages 1–22. CRC Press.

Wu, G., Wang, Q., Jia, H., and Shen, D. (2012b). Feature-based groupwise registration by hierarchical anatomical correspondence detection. *Human brain mapping*, **33**(2), 253–271.

Wu, G., Wang, Q., and Shen, D. (2012c). Registration of longitudinal brain image sequences with implicit template and spatial–temporal heuristics. *Neuroimage*, **59**(1), 404–421.

Wu, G., Kim, M., Wang, Q., and Shen, D. (2012d). S-hammer: Hierarchical attribute-guided, symmetric diffeomorphic registration for mr brain images. *Human brain mapping*.

Yap, P.-T., Wu, G., Zhu, H., Lin, W., and Shen, D. (2010). F-timer: Fast tensor image morphing for elastic registration. *Medical Imaging, IEEE Transactions on*, **29**(5), 1192–1203.

Ying, S., Wu, G., Wang, Q., and Shen, D. (2014). Hierarchical unbiased graph shrinkage (hugs): A novel groupwise registration for large data set. *NeuroImage*, **84**, 626–638.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage*, **55**(3), 856–867.

Zhang, M., Singh, N., and Fletcher, P. T. (2013). Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In *Information Processing in Medical Imaging*, pages 37–48. Springer.