



Published in final edited form as:

Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2012 ; : 956–963. doi:10.1109/CVPR.2012.6247771.

From label fusion to correspondence fusion: a new approach to unbiased groupwise registration

Paul A. Yushkevich, Hongzhi Wang, John Pluta, and Brian B. Avants

Department of Radiology, University of Pennsylvania, Philadelphia, USA

Abstract

Label fusion strategies are used in multi-atlas image segmentation approaches to compute a consensus segmentation of an image, given a set of candidate segmentations produced by registering the image to a set of atlases [19, 11, 8]. Effective label fusion strategies, such as local similarity-weighted voting [1, 13] substantially reduce segmentation errors compared to single-atlas segmentation. This paper extends the label fusion idea to the problem of finding correspondences across a set of images. Instead of computing a consensus segmentation, weighted voting is used to estimate a consensus coordinate map between a target image and a reference space. Two variants of the problem are considered: (1) where correspondences between a set of atlases are known and are propagated to the target image; (2) where correspondences are estimated across a set of images without prior knowledge. Evaluation in synthetic data shows that correspondences recovered by fusion methods are more accurate than those based on registration to a population template. In a 2D example in real MRI data, fusion methods result in more consistent mappings between manual segmentations of the hippocampus.

1. Introduction

Multi-atlas segmentation has emerged as a top performer in a number of difficult segmentation problems. This approach warps expert segmentations from multiple atlas images into the space of a target image using deformable registration, and combines the warped segmentations into a consensus segmentation using a *label fusion* strategy. This multi-atlas label fusion (MALF) approach has been shown to perform consistently better than single-atlas segmentation. Among MALF methods, those that incorporate image similarity in assigning relative weights to atlases during label fusion (*e.g.*, [7]) tend to outperform methods that do not use similarity (majority voting [11, 8], STAPLE [19]), and methods that allow these weights to vary spatially across the target image (*e.g.*, [1, 13, 17]) tend to outperform methods with global weighting. MALF is effective because it reduces both random and systematic errors in segmentation, the former by averaging data from multiple sources, and the latter because it reduces biases due to atlas-target dissimilarity. MALF is particularly useful for heterogeneous populations, as long as the atlas set is representative of the underlying heterogeneity. However, a limitation of MALF is its inability to estimate a pointwise correspondence map between the structures segmented in the target image and the atlas images, something that is explicitly provided by approaches based on registration to a single atlas.

Such correspondences, are, of course, essential in many applications. For example, a correspondence between segmented structures is needed to analyze functional MRI data in a structure of interest across a population. Correspondence maps can be obtained by applying shape analysis techniques to the segmented structures, but such shape-based correspondences ignore the underlying intensity information. On the other hand, using registration to a single labeled atlas or population-derived template [9, 3] provides correspondences at the cost of reduced accuracy relative to MALF. Thus, there is an acute need for a technique that shares the segmentation performance characteristics of MALF, while also providing pointwise correspondences between subjects in a population.

This paper aims to develop such a technique. It builds on the MALF framework, adapting it from the problem of label fusion to the problem of *correspondence fusion*: given a set of candidate mappings from the target image to a common coordinate space, weighted voting-based fusion is used to derive a consensus coordinate mapping. Two algorithms are proposed. In *fusion-based correspondence propagation (FCP)*, one is given a set of atlases with known correspondences between all pairs of atlases, and seeks to extend this correspondence onto a new target image. Correspondence is represented by a coordinate vector at every voxel of every image, such that corresponding points on different images share the same coordinates, and such that the mapping between each image and the coordinate domain is diffeomorphic. FCP is a straightforward extension of MALF, with existing label fusion strategies applied to coordinate maps instead of label maps. To ensure that corresponding maps obtained by fusion are diffeomorphic, the fusion problem is formulated as an optimization problem and solved using the greedy approach [9]. The second approach, *fusion-based groupwise correspondence (FGC)*, performs correspondence propagation simultaneously among a set of images, without any prior known correspondences. It serves the same purpose as population template methods [9, 3], i.e., to find pointwise correspondences across a set of images. Solving FGC for n images involves simultaneously solving n FCP problems.

We hypothesize that FCP and FGC generate better correspondences than approaches based on registering images to a population template. This hypothesis is tested on synthetic data with known correspondences, as well as in slices from real MRI data, with pairwise overlap between manual segmentations used to test correspondence quality.

2. Methods

2.1. Overview of Multi-Atlas Label Fusion

We review the MALF problem briefly. Given a set of k “atlases,” each consisting of an intensity image I_i and a manually generated label image L_i , we seek to obtain a segmentation of the target image I_T . For simplicity, we assume that the segmentation problem involves only two labels, and $L_i(x)$ describes the probability of voxel x having the foreground label. Deformable registration is used to warp each atlas to the target image, producing a warped intensity image \hat{I}_i and a warped label image \hat{L}_i . The task of label fusion is to derive a consensus label image L_T by combining information from the set of images $\{I_T, \hat{I}_1, \dots, \hat{I}_k, L_1, \dots, L_k\}$.

Label fusion with local similarity-weighted voting has proven particularly successful in practice. In this framework, the consensus label image is computed as

$$L_T(x) = \sum_{i=1}^k w_i(x) \hat{L}_i(x) \quad \text{subj. to} \quad \sum_{i=1}^k w_i(x) = 1, \quad (1)$$

where $w_i(x)$ is a spatially varying weight function, typically derived from the similarity between images \hat{I}_i and I_T in the neighborhood of x . Thus, segmentations from atlases that closely resemble I_T near x are assigned a larger weight. Several weighting functions have been proposed in the literature [1, 13, 7]. For example, Artaechevarria *et al.* [1] proposed the following inverse polynomial weighting scheme:

$$w_i(x) = \frac{1}{Z(x)} \left[\sum_{y \in \mathcal{N}(x)} [I_T(y) - \hat{I}_i(y)]^2 \right]^{-\beta}, \quad (2)$$

where Z is a normalizing constant, $\mathcal{N}(x)$ is a neighborhood of a voxel x , and $\beta > 0$ is a parameter controlling the weight distribution.

2.2. Fusion-Based Correspondence Propagation

Let $\{I_1 \dots I_n\}$ be a set of atlas images with known correspondences, and let I_T be the target image. For simplicity, assume that all images have been rigidly aligned and are defined on a common domain $\Omega \in \mathbb{R}^d$. For each atlas i , let the known correspondence be specified by the diffeomorphic map $\phi_i: \Omega \rightarrow \Omega$, such that a point x in image I_i corresponds to the point y in image I_j if and only if $\phi_i^{-1}(x) = \phi_j^{-1}(y)$. Conversely, for a point $z \in \Omega$, $\phi_i(z)$ and $\phi_j(z)$ are corresponding points in images I_i and I_j for all $i, j \in [1, n]$. Suppose that non-linear diffeomorphic registration is performed between I (as a fixed image) and each atlas. Let ψ_i represent the resulting deformation from I_T to I_i . Then the i -th *candidate correspondence map* for image I_T is given by $\psi^{-1} \circ \phi_i$. FCP fuses these candidate maps into a single consensus correspondence map ϕ_T .

Let $w_i: \Omega \rightarrow \mathbb{R}$ be a *weight map* for the atlas I_i , subject to $\sum_{i=1}^n w_i(x) = 1$. The weight map represents our confidence in atlas I_i being well matched to I_T in the neighborhood of a point x in I_T . These weight maps are computed in the same way as in similarity-weighted MALF methods above, i.e., $w_i(x)$ is a function of intensity similarity between $I_i \circ \psi_i$ and I_T , computed in the neighborhood of x . Our formulation is agnostic to exactly how $w_i(x)$ is computed, and a specific choice is discussed later in Sec. 2.5.2.

Following the logic of label fusion, we want the inverse consensus correspondence map ϕ_T^{-1} to be as close as possible to the weighted sum of the inverse candidate correspondence maps $\{\phi_i^{-1} \circ \psi_i\}$, with weighting provided by $\{w_i\}$. To ensure that ϕ_T is diffeomorphic, we solve

$$\phi_T^* = \arg \min_{\phi_T \in \text{Diff}(\Omega)} \left\| \phi_T^{-1} - \sum_{i=1}^n w_i \cdot \phi_i^{-1} \circ \psi_i \right\|_{L^2}^2 + \rho(\phi_T), \quad (3)$$

where $\text{Diff}(\Omega)$ is the group of diffeomorphic maps on Ω , $\|\cdot\|_{L^2}$ is the standard L^2 norm, and ρ is some spatial regularization term, *e.g.*, favoring smooth deformations.

Minimization (3) can be solved using existing tools for diffeomorphic image registration. Simply denote J_p as the image formed by the p -th component of the map

$\sum_{i=1}^n w_i \cdot \phi_i^{-1} \circ \psi_i$, and let \mathcal{J}_p denote the p -th component of the identity map, i.e., $\mathcal{J}_p(x_1, \dots, x_d) = x_p$. We must then solve the simple image registration problem

$$\phi_T^* = \arg \min_{\phi_T \in \text{Diff}(\Omega)} \sum_d \|\mathcal{I}_d \circ \phi_T^{-1} - J_d\|_{L^2}^2 + \rho(\phi_T). \quad (4)$$

This registration problem can be solved by any image registration tool with diffeomorphic constraints that supports the mean square difference metric and can accept multiple pairs of fixed/moving images as input. Specific implementation issues are discussed in Sec. 2.5.1.

2.3. Fusion-Based Groupwise Correspondence

Atlas sets with predefined correspondences are seldom available, and can be expensive to generate. Thus, the main focus of this paper is on leveraging MALF ideas for the computation of groupwise correspondences between a set of unannotated images. Finding such correspondences is at the heart of deformation-based morphometry. We hypothesize that by making locally similar images have locally similar correspondences, we can improve the overall correspondence between a group of images.

The FGC problem involves k images I_1, \dots, I_k defined over a common domain Ω . For each image, only the intensity information is available. We formulate the FGC problem as a set of k concurrent FCP problems (3) in which each image is treated as the target image and the remaining images are treated as atlases. As a preliminary step, image registration is performed between all pairs of images, yielding a set of transformations $\{\psi_{ij}\}$. Specifically, ψ_{ij} is the transformation that is used to optimally warp image I_j into the space of image I_i , according to the chosen registration method. Next, for each image i , we compute a set of

weight maps $\{w_{ij}\}$, satisfying $\sum_{j=1}^k w_{ij}(x) = 1$. For each i and j , $w_{ij}(x)$ reflects the local similarity between image I_i and the warped image $I_j \circ \psi_{ij}$ near x , according to some weighting strategy (the specific choice for this paper is given in Sec. 2.5.2). To simplify the subsequent expressions, let $w_{ii}(x) = 0$ and let $\psi_{ii}(x) = x$ for all i and x .

FGC seeks to find a set of diffeomorphic transformations $\{\phi_i : \Omega \rightarrow \Omega\}$ between a reference space and each of the k images, which minimizes the overall objective function:

$$\{\phi_i^*\} = \arg \min_{\{\phi_i\} \in \text{Diff}(\Omega)} \sum_{i=1}^k \left\| \phi_i^{-1} - \sum_{j=1}^k w_{ij} \cdot \phi_j^{-1} \circ \psi_{ij} \right\|_{L^2}^2 + \sum_{i=1}^k \rho(\phi_i). \quad (5)$$

Despite the similarity between (5) and (3), the FGC objective function in (5) can no longer be optimized using existing registration tools because it cannot be reformulated as a set of pairwise image registrations. Instead, we minimize the objective using a variational approach. Let $\psi_{ij}(x) = x + \mathbf{v}_{ij}(x)$ and $\phi_i^{-1} = x + \mathbf{u}_i(x)$. Let

$$\Delta_i = \phi_i^{-1} - \sum_{j=1}^k w_{ij} \cdot \phi_j^{-1} \circ \psi_{ij} = \mathbf{u}_i - \sum_{j=1}^k w_{ij} \cdot [\mathbf{v}_{ij} + \mathbf{u}_j \circ \psi_{ij}] \quad (6)$$

Then we may write the component of the objective function in (5) not related to regularization as

$$E(\mathbf{u}_1, \dots, \mathbf{u}_k) = \sum_{i=1}^k \|\Delta_i\|_{L^2}^2. \quad (7)$$

Let h_m be a variation in the displacement field \mathbf{u}_m . The Gateaux derivative of E with respect to h_m is

$$\begin{aligned}
& \frac{1}{2} \delta_{h_m} E \\
&= \sum_{i=1}^k \langle \Delta_i, \delta_{h_m} \Delta_i \rangle \\
&= \sum_{i=1}^k \langle \Delta_i, \delta_{i,m} \cdot h_m \\
&\quad - w_{im} \cdot h_m \circ \psi_{im} \rangle \tag{8} \\
&= \langle \Delta_m, h_m \rangle \\
&\quad - 2 \sum_{i=1}^k \langle \Delta_i, w_{im} \cdot h_m \circ \psi_{im} \rangle \\
&= \langle \Delta_m - \sum_{i=1}^k (\Delta_i \circ \psi_{im}^{-1}) (w_{im} \circ \psi_{im}^{-1}) |D\phi_{im}^{-1}|, h_m \rangle,
\end{aligned}$$

where $\delta_{i,m}$ is the Kronecker delta and D is the Jacobian operator. From this, the gradient of E is given by

$$\frac{1}{2} D_{\mathbf{u}_m} E = \Delta_m - \sum_{i=1}^k (w_{im} \cdot \Delta_i) \circ \psi_{im}^{-1} \cdot |D\phi_{ij}^{-1}|. \tag{9}$$

To restrict the space of solutions to diffeomorphic transformations and to provide an implicit regularization prior, we adopt a greedy iterative strategy [9]. At the beginning, the

displacement fields $\{\mathbf{u}_i^0\}$ are initialized at zero. At each iteration, the displacement fields are updated as

$$\Delta \mathbf{u}_m^t = -\epsilon \cdot G_\gamma * D_{\mathbf{u}_m^t} E, \tag{10}$$

$$\mathbf{u}_m^{t+1} = \mathbf{u}_m^t \circ (\mathbf{x} + \Delta \mathbf{u}_m^t) + \Delta \mathbf{u}_m^t, \tag{11}$$

where ϵ is the step size, and G_γ is a Gaussian convolution kernel with standard deviation γ . Smoothing of the gradient with a Gaussian kernel in (10) provides implicit regularization for the deformation field. Successive composition of small deformations in (11) ensures that

$\mathbf{x} + \mathbf{u}_m^t$ is diffeomorphic. The above iteration produces the set of transformations $\{\phi_i^{-1}\}$ that minimize the energy E (plus an implicit regularization term). To obtain transformations ϕ_i , the inverse of these transformations is computed directly, using the fixed point approach in [6].

The minimization (5) is highly parallelizable and, in theory, can be implemented efficiently in a parallel environment with k compute nodes. During each iteration, Δ_i would first be computed at each node i . Then the fields $\{\Delta_i\}$ would be distributed across all the nodes. Next, each node would perform the update step (10,11) in parallel. Lastly, the updated fields $\{\mathbf{u}_i^{t+1}\}$ would be distributed across all the nodes. Overall, the greatest computational

expense associated with this method comes from performing the initial set of $O(k^2)$ pairwise registrations.

2.4. Comparison Method: Population Templates

In the evaluation presented below, we compare FGC to the leading alternative approach for computing groupwise correspondence between sets of images, which, for brevity, we call the *unbiased population template (UPT)* approach. UPT, which was originally developed by Joshi *et al.* [9], constructs an unbiased, population-specific template from a set of images by iterative application of diffeomorphic registration and intensity averaging. By construction, this method provides a set of correspondences between the input images and the template. Our implementation of UPT is based on subsequent work by Avants *et al.* [3], which includes a shape averaging step and allows the use of the cross-correlation image match metric for similarity computation during diffeomorphic registration.

2.5. Implementation Details

2.5.1 Registration—All pairwise registrations between images are performed using the greedy diffeomorphic Symmetric Normalization (SyN) algorithm [2], implemented as part of the open-source ANTS software package. SyN registrations used the cross-correlation metric with a 9×9 window; 3 resolution levels with maximum 50 iterations per level; step size 0.25; Gaussian regularization with standard deviation of 3 pixels. Our UPT implementation uses SyN and the accompanying ANTS scripts for unbiased atlas construction.

2.5.2 Weight Maps—Weight maps, i.e., the term w_i in (3) and the term w_{ij} in (5), were generated using a rank-based scheme proposed by [20]. As before, let I_T be the target image, and let $\hat{I}_1 \dots \hat{I}_k$ be a set of k images registered to I_T . Let $\nu_i(x)$ denote the similarity map between \hat{I}_i and I_T . We use normalized cross-correlation over a 9×9 pixel neighborhood as the local similarity measure. Let $\rho_i(x)$ be the local rank of the image \hat{I}_i in terms of similarity:

$$\rho_i(x) = |\{j \in [1, k] : \nu_j(x) > \nu_i(x)\}|. \quad (12)$$

Then the weight function is given as

$$w_i(x) = \left[\frac{1}{Z} e^{-\alpha \rho_i(x)} \right] * G_\sigma(x), \quad (13)$$

where $Z = \sum_{i=1}^k e^{-\alpha k}$ is a normalization constant, G_σ is the Gaussian convolution kernel with standard deviation σ . Parameter α controls bias towards best-matching atlases in the weighting, and σ controls the spatial regularization of the weight map. This paper uses $\alpha = 0.5$ and $\sigma = 1.2$ (in pixel units). The appeal of this rank-based weighting scheme is that the weights are not affected by absolute values of the similarity metric, presumably making the scheme application-independent. However, note that any of the many spatially-varying, similarity-weighted label fusion schemes proposed recently (e.g., [1, 13]) could also be used instead.

3. Experiments

3.1. Simulated Data

Simulated 2D datasets are constructed to evaluate the FCP (Sec. 2.2) and FGC (Sec. 2.3) algorithms using known true correspondences. A high-resolution (512×512 pixels) image

of a histological slice of the hippocampus (Figure 1) is randomly deformed to generate simulated images. The i -th random deformation is constructed by generating a random time-varying velocity field $V_i(x, t)$, with $t \in [0, 1]$, and solving the flow ODE for the diffeomorphism ξ_i :

$$\partial \xi_i(x, t) / \partial t = V_i(\xi(x, t), t); \quad \xi_i(x, 0) = 0. \quad (14)$$

For brevity, let $\xi_i(x) = \xi_i(x, 1)$ denote the end-point diffeomorphism. Random velocity fields are constructed as follows. A set of N_p control points is chosen randomly in the unit cube in $x_1 \times x_2 \times t$ space. A random displacement is applied to the x_1 and x_2 coordinates of each point, by sampling from a zero mean isotropic normal distribution with standard deviation σ_p . Additional control points are sampled regularly from the faces of the unit cube and assigned a zero displacement. Field $V_i(x, t)$ is computed by applying radial basis function interpolation to the control points. Parameters N_p and σ_p influence the complexity and the magnitude of the random displacement, respectively. The flow ODE (14) is solved using the semi-Lagrangian approach [4] with time step 0.04. The high-resolution image is warped by the transformation ξ_i^{-1} , resampled to 96×96 resolution, and corrupted by adding uncorrelated Gaussian noise with standard deviation σ_n . Examples of images obtained by this random deformation process are shown in Figure 1.

We generate 51 random simulation datasets. In order to search the parameter space uniformly and without bias, the simulation parameters, N_p , σ_p , σ_n , as well as the number of simulated images, k , are sampled randomly from uniform distributions: $N_p \sim U(100, 1000)$, $\sigma_p \sim U(10, 100)$, $\sigma_n \sim U(0, 50)$, and $k \sim U(2, 48)$.

3.2. FCP Evaluation in Simulated Data

FCP (3) is evaluated using a leave-one-out strategy. The q -th simulation image is treated as the target image and the other $k - 1$ images are treated as atlases with known correspondences to a common reference space (provided by the mappings $\{\xi_i\}$). To evaluate the accuracy of FCP for image q , the consensus correspondence mapping ϕ_q computed using (3) is compared to the ground truth correspondence mapping ξ_q . Specifically, for each point x in the q -th simulated image, ξ_q^{-1} gives the true corresponding location in the undeformed image, and $\phi_q^{-1}(x)$ gives the FCP estimate of this location. The Euclidean distance between $\xi_q^{-1}(x)$ and $\phi_q^{-1}(x)$ is a measure of FCP error. This error is integrated over the foreground voxels in the q -th simulation image (as we are not interested in background correspondence), and averaged over all k leave-one-out experiments. This leads to the leave-one-out FCP error metric

$$\mathcal{E}_{\text{FCP}} = \frac{1}{k} \sum_{q=1}^k \frac{1}{|M_q|} \int_{M_q} \|\phi_q^{-1}(x) - \xi_q^{-1}(x)\| dx, \quad (15)$$

where M_q is the foreground region in the q -th image, and $|M_q|$ is its area.

For comparison, we compute the average correspondence error from performing direct registration between each simulated image and the undeformed source image (Fig. 1a) resampled to 96×96 resolution. The undeformed image can be viewed as an idealized population-specific template, and correspondence error between the undeformed image and the simulated images provides the lower bound for correspondence errors that would result from registering simulated images to a population-derived template. The error from direct registration is denoted \mathcal{E}_{DR} .

Results—For brevity, we only report the results in terms of the relative improvement of FCP over direct registration:

$$\Delta_{\text{FCP}} = \frac{\mathcal{E}_{\text{DR}} - \mathcal{E}_{\text{FCP}}}{\mathcal{E}_{\text{DR}}} \cdot 100\% \quad (16)$$

Figure 2 plots Δ_{FCP} against the number of atlases, k , in all 51 simulation experiments. FCP clearly underperforms DR when fewer than 10 atlases are available, but as the number of atlases increases, FCP reaches a consistent improvement. When $k = 16$ atlases are used, $\Delta_{\text{FCP}} > 10\%$ in 54% of experiments and $\Delta_{\text{FCP}} > 5\%$ in 93% of experiments. Among the experiments with $k = 16$, Δ_{FCP} is negatively correlated with the additive noise parameter σ_n ($R = -0.64$, $p < 0.001$), but there is no significant correlation with parameters N_p and σ_p . We conclude that, given a relatively small number of atlases, FCP consistently recovers the ground truth correspondence better than direct registration to the template.

3.3. FGC Evaluation in Simulated Data

To evaluate the group-wise FGC approach, we perform the optimization (5) among all k simulated images, yielding correspondence maps ϕ_1, \dots, ϕ_k from a common reference space to the simulated images. These recovered correspondences are correct if for each voxel x and each pair of images (i, j) , we have

$$\xi_i^{-1}(\phi_i(x)) = \xi_j^{-1}(\phi_j(x)) \quad (17)$$

Note that this does not necessarily imply $\phi_i = \xi_i$. Thus to define the error associated with the correspondence maps ϕ_1, \dots, ϕ_k , we can measure the degree to which (17) is violated, i.e., how far apart, on average, the points $\{\xi_1^{-1}(\phi_1(x)), \dots, \xi_k^{-1}(\phi_k(x))\}$ are from each other. This leads to the *groupwise correspondence* error metric

$$\mathcal{E}_{\text{GRP}} = \frac{1}{k^2|M|} \cdot \int_M \left[\sum_{i=1}^k \sum_{j=1}^k \|\xi_i^{-1}(\phi_i(x)) - \xi_j^{-1}(\phi_j(x))\|^2 \right]^{\frac{1}{2}} dx. \quad (18)$$

To evaluate FGC, we compare its groupwise correspondence error to that of UPT (Sec. 2.4), and our analysis focuses on their relative difference:

$$\Delta_{\text{FGC}} = \frac{\mathcal{E}_{\text{GRP}}^{\text{UPT}} - \mathcal{E}_{\text{GRP}}^{\text{FGC}}}{\mathcal{E}_{\text{GRP}}^{\text{UPT}}} \cdot 100\%. \quad (19)$$

Results—Figure 3 visualizes the correspondences and correspondence errors produced by FGC and UPT in one simulation experiment. The top row shows the mean correspondence difference maps, i.e., the integrand in (18), for the two methods, and the bottom row shows the mean intensity image computed by warping the simulated images into the common reference space and averaging their intensity. FGC produces lower correspondence error in the hippocampus, and its mean image is more crisp, and matches the undeformed image in Fig. 1a more closely.

Figure 4 plots Δ_{FGC} against the size of the simulated image set, k , across all simulation experiments. Unlike FCP, there is little effect of k on Δ_{FGC} . Among all experiments, $\Delta_{\text{FGC}} > 5\%$ in 92% of experiments, $\Delta_{\text{FGC}} > 10\%$ in 78% of experiments, and $\Delta_{\text{FGC}} > 20\%$ in 31% of

experiments. Δ_{FGC} is significantly negatively correlated with σ_p ($R = -0.73$, $p < 0.001$) and positively correlated with σ_n ($R = 0.35$, $p = 0.01$). Overall, we conclude that FGC achieves consistently lower errors than UPT across all parameter values.

3.4. FGC Evaluation in MRI Slices

As a preliminary evaluation of FGC in real imaging data, we use it to compute groupwise correspondences between a set of 2D MRI slices from 32 subjects from a study of aging and cognitive impairment [10]. The data consists of T2-weighted MR images of the hippocampal region with highly anisotropic resolution ($0.4\text{mm} \times 0.4\text{mm} \times 2.0\text{mm}$), acquired in an oblique coronal plane angled orthogonally to the long axis of the hippocampus. Manual segmentation has been used to divide each hippocampus formation (HF) into head, tail and body regions, with the body region typically extending over 3-5 slices and partitioned into 7 *subfields*. For this evaluation, we extract for each subject a single slice near the center of the body of the left HF, and crop a 96×96 pixel region around the center of the manual segmentation. An example of such an image and its segmentation are shown in Fig. 5. We perform rigid registration to remove differences in pose between the input images. We then proceed to use the FGC and UPT methods to compute groupwise correspondences between the images.

Lacking the availability of a “ground truth” correspondence between the images, we evaluate the groupwise correspondence by measuring how well it matches labels assigned by manual segmentation to pairs of corresponding voxels. Specifically, given correspondence maps $\phi_1 \dots \phi_k$ from a common reference space to each of the input images, and given segmentation label maps $L_1 \dots L_k : \Omega \rightarrow \{0, \dots, 7\}$, we measure the total *label agreement error*

$$\mathcal{E}_{\text{LAB}} = \frac{1}{k^2 |\Omega|} \int_{\Omega} \sum_i^k \sum_j^k \delta [L_i(\phi_i(x)), L_j(\phi_j(x))] dx, \quad (20)$$

where $\delta[\bullet, \bullet]$ is the Kronecker delta function. Underlying the use of this metric is the assumption that manual segmentation correctly identifies corresponding anatomical regions in different images.

Results—Figure 6 shows the average images constructed by averaging the intensity of the input images warped into a common reference space using FCP and UPT correspondence maps, as well as the map of average pointwise label agreement error, *i.e.*, the integrand in (20). The total label agreement error is 0.0330 for FGC and 0.0358 for UPT, *i.e.* the FGC error is 7.8% lower. Thus, we can conclude that the results of the real data experiment were generally consistent with the simulated data experiments.

4. Discussion and Conclusions

We have presented a new approach to groupwise image correspondence derivation that uses multi-atlas registration and similarity-weighted voting-based fusion techniques, which, to our knowledge, have only been used for image segmentation until now. We have shown this fusion-based approach (FGC) to perform better than UPT in simulated and real data experiments, although the improvements in simulated data (10-20%) were greater than in the real data experiment (8%). These improvements do come at a considerable additional computational cost, since k^2 registrations are required as input to FGC. By contrast, UPT only performs $O(mk)$ registrations, where m is the number of iterations, typically < 10 . However, this added cost, which is similarly a drawback of MALF segmentation, has not

prevented MALF from being widely adopted in the biomedical imaging field. Hence, we expect that if the error reduction demonstrated in this paper can be shown to extend to 3D imaging datasets, the proposed FGC framework would also find numerous applications in this field. Thus, our future work will focus on implementing FGC in the parallel computing environment and evaluating it in 3D image datasets.

Although our evaluation focused on a comparison with the UPT method, there are a number of other groupwise correspondence finding methods that are also available. Among the closest to our work are methods that optimize an information theoretic groupwise objective function, *e.g.*, the congealing work of Zollei *et al.* [21], the minimum description length work of Twining *et al.* [16], or the groupwise approaches of Studholme and Cardenas [15] and Bhatia *et al.* [5]. Although we cannot speak to the relative performance of FGC and these methods, we can note that the objective function minimized in FGC is substantially simpler (involving essentially quadratic forms) and that FGC has an advantage of allowing any image registration algorithm and metric to be used for pairwise image registration. A number of methods have also derived groupwise correspondences from the results of k^2 registrations. For instance, Seghers *et al.* [14] compose deformations from pairwise registrations to construct a population mean. However, this approach does not assign preferences to the pairwise registrations with greater similarity as FGC does, nor does it guarantee diffeomorphic registrations. To simplify the complexity of the $O(k^2)$ groupwise registration problem, some authors proposed hierarchical clustering approaches [18, 12]. Similar strategies could prove useful to reduce the complexity of FGC. We also note the actual label fusion strategy (13) that we use has not been extensively optimized, and more advanced label fusion strategies (*e.g.* [13, 1, 17]) may lead to further improvements in FGC performance.

In conclusion, this paper presented a new groupwise image registration approach that uses the spatially-varying similarity-weighted voting strategy that has proven very successful in multi-atlas segmentation, to establish group-wise correspondences. In simulated and read 2D image sets, this new approach consistently improved upon unbiased population template generation, one of the leading approaches to groupwise image registration.

References

1. Artaechevarria, Xabier; Munoz-Barrutia, Arrate; Ortiz-de Solorzano, Carlos. Combination strategies in multi-atlas image segmentation: application to brain mr data. *IEEE Trans Med Imaging*. Aug; 2009 28(8):1266–77. 1, 2, 4, 7. [PubMed: 19228554]
2. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. Feb 4; 2008 12(1):26–41. [PubMed: 17659998]
3. Avants, Brian; Gee, James C. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage*. 2004; 23(Suppl 1):S139–S150. 1, 2, 4. [PubMed: 15501083]
4. Beg, M. Faisal; Miller, Michael I.; Trouvé, Alain; Younes, Laurent. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision*. 2005; 61(2):139–157. 4.
5. Bhatia, KK.; Hajnal, JV.; Puri, BK.; Edwards, AD.; Rueckert, D. Biomedical Imaging: Nano to Macro, 2004. *IEEE International Symposium on*. IEEE; 2004. Consistent groupwise non-rigid registration for atlas construction.; p. 908-911.7
6. Chen M, Lu W, Chen Q, Ruchala KJ, Olivera GH. A simple fixed-point approach to invert a deformation field. *Medical physics*. 2008; 35:81. 3. [PubMed: 18293565]

7. Collins, D. Louis; Pruessner, Jens C. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI. *Med Image Comput Comput Assist Interv.* 2009; 5762 1, 2.
8. Heckemann, Rolf A.; Hajnal, Joseph V.; Aljabar, Paul; Rueckert, Daniel; Hammers, Alexander. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage.* Oct; 2006 33(1):115–126. 1. [PubMed: 16860573]
9. Joshi S, Davis Brad, Jomier Matthieu, Gerig Guido. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage.* 2004; 23(Suppl 1):S151–S160. 1, 2, 3, 4. [PubMed: 15501084]
10. Mueller, Susanne G.; Weiner, Michael W. Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. *Hippocampus.* Jun; 2009 19(6):558–564. 6. [PubMed: 19405132]
11. Rohlfing, Torsten; Brandt, Robert; Menzel, Randolph; Maurer, Calvin R. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage.* Apr; 2004 21(4):1428–1442. 1. [PubMed: 15050568]
12. Sabuncu M, Balci S, Golland P. Discovering modes of an image population through mixture modeling. *Med Image Comput Comput Assist Interv.* 2008:381–389. 7. [PubMed: 18982628]
13. Sabuncu, Mert R.; Thomas Yeo, BT.; Van Leemput, Koen; Fischl, Bruce; Golland, Polina. A generative model for image segmentation based on label fusion. *IEEE Trans Med Imaging.* Oct; 2010 29(10):1714–29. 1, 2, 4, 7. [PubMed: 20562040]
14. Seghers D, D'Agostino E, Maes F, Vandermeulen D, Suetens P. Construction of a brain template from mr images using state-of-the-art registration and segmentation techniques. *Med Image Comput Comput Assist Interv.* 2004:696–703. 7.
15. Studholme C, Cardenas V. A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recognition Letters.* 2004; 25(10):1191–1202. 7.
16. Twining C, Cootes T, Marsland S, Petrovic V, Schestowitz R, Taylor C. A unified information-theoretic approach to groupwise non-rigid registration and model building. *Information Processing in Medical Imaging.* 2005:167–198. Springer. 7.
17. Wang H, Suh J, Pluta J, Altinay M, Yushkevich P. Optimal weights for multi-atlas label fusion. *Information Processing in Medical Imaging.* 2011:73–84. Springer. 1, 7. [PubMed: 21761647]
18. Wang Q, Chen L, Yap PT, Wu G, Shen D. Groupwise registration based on hierarchical image clustering and atlas synthesis. *Human brain mapping.* 2010; 31(8):1128–1140. 7. [PubMed: 20063349]
19. Warfield, Simon K.; Zou, Kelly H.; Wells, William M. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging.* Jul; 2004 23(7):903–921. 1. [PubMed: 15250643]
20. Yushkevich, Paul A.; Wang, Hongzhi; Pluta, John; Das, Sandhitsu R.; Craige, Caryne; Avants, Brian B.; Weiner, Michael W.; Mueller, Susanne. Nearly automatic segmentation of hippocampal subfields in in vivo focal t2-weighted mri. *Neuroimage.* Dec; 2010 53(4):1208–1224. 4. [PubMed: 20600984]
21. Zöllei, Lilla; Learned-Miller, Erik G.; Grimson, W. Eric L.; Wells, William M., III Efficient population registration of 3d data. In *Proc. Computer Vision for Biomedical Image Applications.* 2005:291–301. 7.

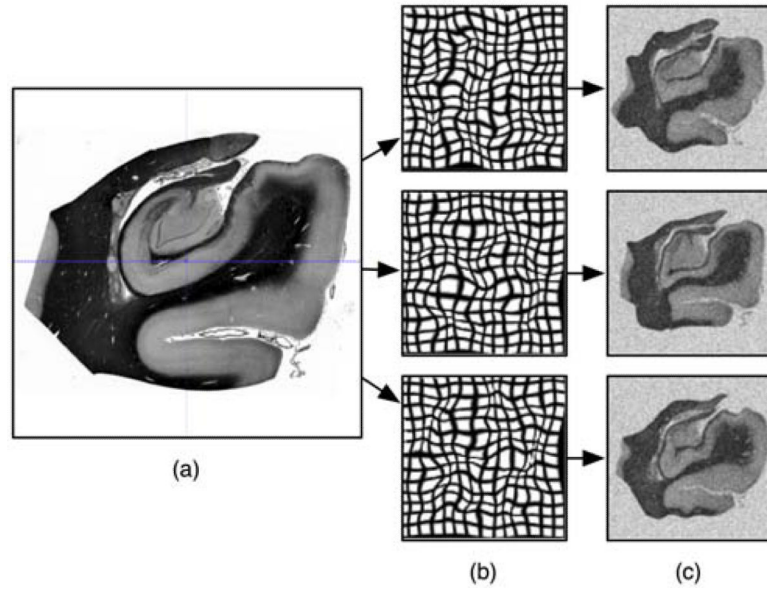


Figure 1. Generation of synthetic images for evaluation. A high-resolution image (a) undergoes random diffeomorphic deformations (b), subsampling, and additive noise (c).

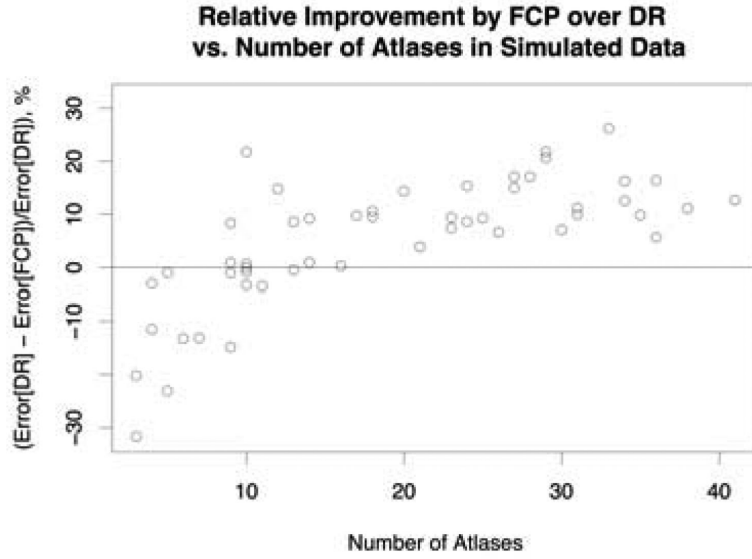


Figure 2. Relative improvement Δ_{FCP} of fusion-based correspondence propagation (FCP) over direct registration (DT) in simulated data experiments, plotted against the number of simulated atlases in each experiment, k . Positive values mean lower error with FCP.

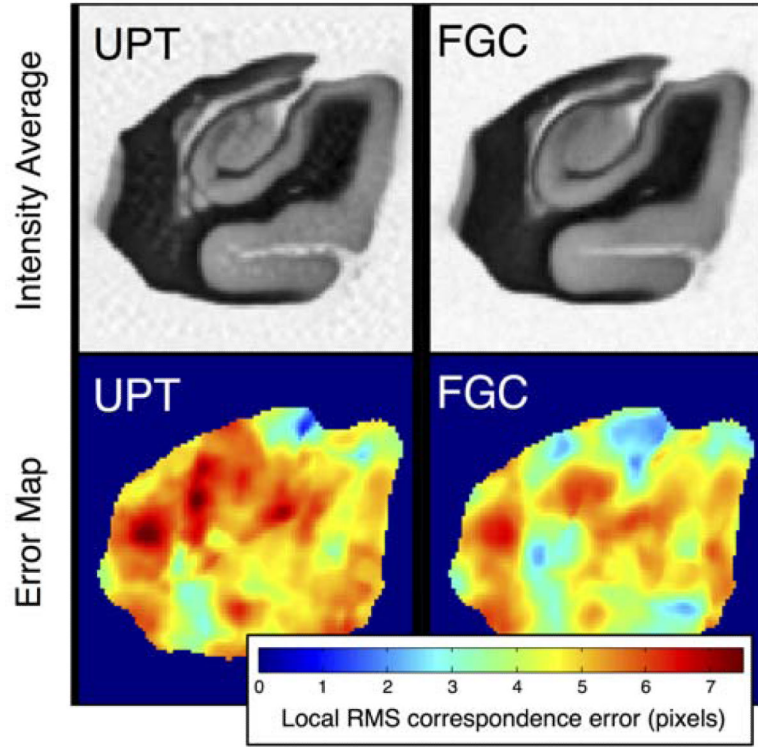


Figure 3. Average intensity images and pointwise local correspondence error maps computed in the common reference space after obtaining groupwise correspondences using the unbiased population template (UPT) and fusion-based (FGC) methods in one of the synthetic data experiments. The simulation experiment chosen for this figure has a very large deformation parameter value ($\sigma_p = 90$) in order to emphasize the difference between the methods.

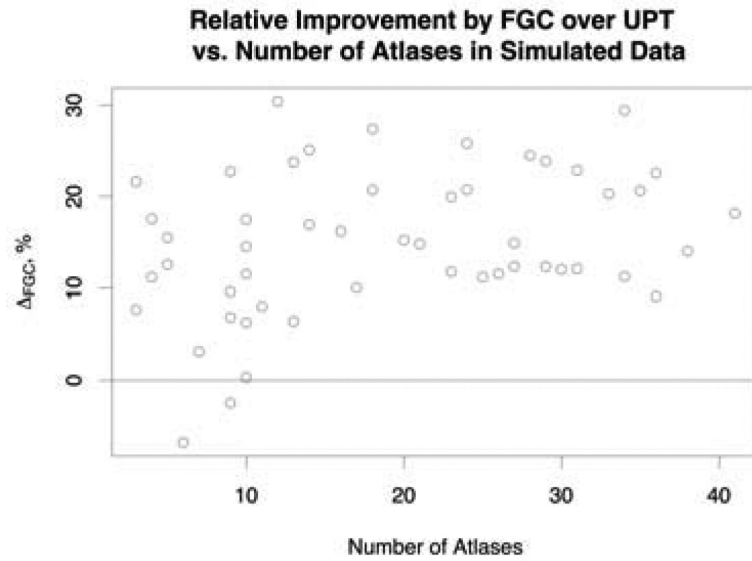


Figure 4. Relative improvement of the FGC over UPT across all simulated data experiments.

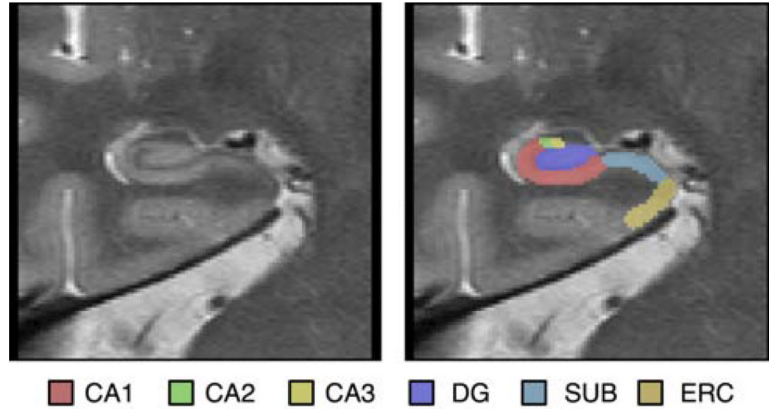


Figure 5. An example T2-weighted MR image of the hippocampus from the evaluation dataset, and its manual segmentation. Color labels correspond to different subfields of the hippocampus.

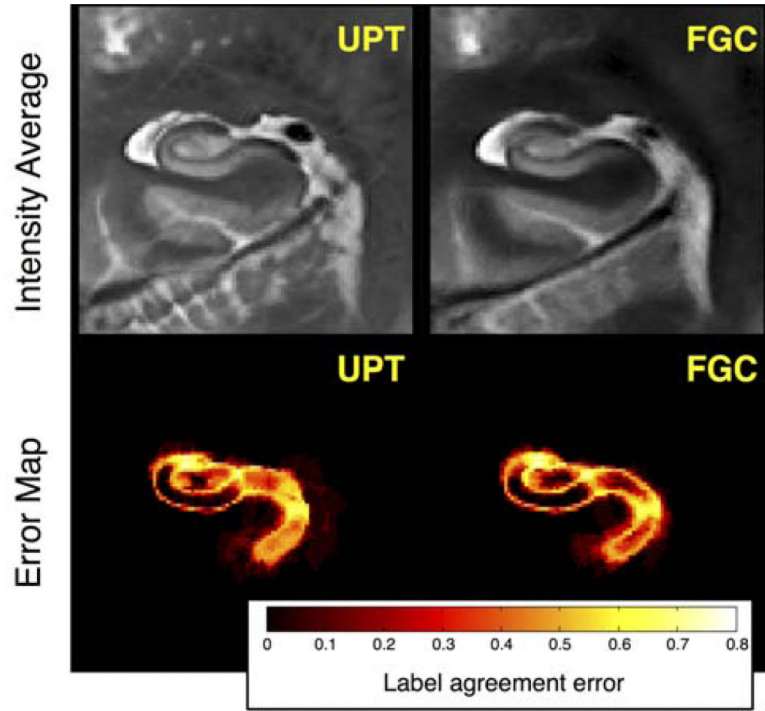


Figure 6. Average intensity images (top) and label agreement maps (bottom) computed from the UPT and FGC groupwise correspondences in the T2-weighted MRI dataset.