A Mixture Model for Automatic Diffeomorphic Multi-Atlas Building

Miaomiao Zhang, Hang Shao, P. Thomas Fletcher

School of Computing, University of Utah, Salt Lake City, UT

Abstract. Computing image atlases that are representative of a dataset is an important first step for statistical analysis of images. Most current approaches estimate a single atlas to represent the average of a large population of images, however, a single atlas is not sufficiently expressive to capture distributions of images with multiple modes. In this paper, we present a mixture model for building diffeomorphic multi-atlases that can represent sub-populations without knowing the category of each observed data point. In our probabilistic model, we treat diffeomorphic image transformations as latent variables, and integrate them out using a Monte Carlo Expectation Maximization (MCEM) algorithm via Hamiltonian Monte Carlo (HMC) sampling. A key benefit of our model is that the mixture modeling inference procedure results in an automatic clustering of the dataset. Using 2D synthetic data generated from known parameters, we demonstrate the ability of our model to successfully recover the multi-atlas and automatically cluster the dataset. We also show the effectiveness of the proposed method in a multi-atlas estimation problem for 3D brain images.

1 Introduction

Atlas building through diffeomorphic image registration is an effective way to compute a common coordinate system for comparison across images, as well as to encode shape via the diffeomorphic transformations. As such, atlas estimation becomes a fundamental step in population-based studies, shape quantification, etc. The diffeomorphic framework guarantees a smooth, and smoothly invertible, mapping between the atlas and each individual subject. Early work in this area chose either a standard template or a randomly selected subject from the dataset as the atlas. This, as a result, may bias the statistical analysis made with an atlas that is not representative of the data at hand. Several works [8, 16, 13, 21] have constructed an unbiased single atlas in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [3]. They register a set of input images to a template, which is simultaneously estimated by minimizing the energy function of the sum-of-squared difference between the deformed atlas and each individual image. All these approaches are based on the assumption that the population comes from a distribution of transformations from the atlas centered about a single mode.

However, a single atlas does not provide enough information for robust statistical analysis if significant differences exist between sub-populations. Blezek et al. [5] were the first to investigate the multi-atlas building problem and infer each atlas from the mode of a population through the mean-shift algorithm. They iteratively optimized between a small deformation image registration framework and atlas construction. Later, two major classes of multi-atlas building methods were developed in medical imaging. One of the classes is a two-step strategy, in which the algorithm does clustering, like K-means or affinity propagation, after registration. Another class of multi-atlas building is motivated by probabilistic modeling of multi-atlases. Allassonnière et al. [2] discussed a mixture model of template estimation with small deformations. Sabuncu et al. [12] introduced a joint framework of image registration and clustering using a mixture of Gaussians, although their work was not in a diffeomorphic setting. Tang et al. [14] proposed a random diffeomorphic orbit model to treat the multiple atlases as Gaussian random fields, and then estimated them from the model using maximum a posteriori estimation. These multi-atlas methods are of high importance for related research areas, for example, image segmentation [1, 9, 11, 15, 18], where a priori knowledge about the shapes and structures from the pre-segmented multiple atlases is used to guide the segmentation. Aljabar et al. [1] discussed the issue of multi-atlas selection and showed that multi-atlas based segmentation results in higher accuracy than a single atlas.

In this paper, we propose a generative formulation of the diffeomorphic multi-atlas building problem. We develop an algorithm that can for the first time cluster population-based images into different sub-groups automatically while co-registering them in a diffeomorphic setting with marginalized deformations. To achieve this, we use a MCEM algorithm, where the diffeomorphic image transformations are treated as latent variables and integrated out from the joint distribution of the complete data likelihood. Much like Zhang et al. in [21], we sample on the space of diffeomorphisms by HMC sampling. The reason why we treat diffeomorphisms as latent random variables, rather than parameters to estimate by maximum a posterior (MAP), is that such a mode approximation scheme performs poorly under image noise and is more likely to get stuck in a local minima as shown in [2, 21]. Since we have a generative model, we can forward generate 2D synthetic data with known parameters from our model. We demonstrate that our model can successfully recover multiple atlas as well as automatically cluster the dataset. We also show the effectiveness of the proposed method in a multi-atlas estimation problem for real 3D brain MRI images.

2 Background

We define a mixture model for multi-atlas building in the diffeomorphic setting. Before introducing our model, we first briefly review single diffeomorphic atlas building [13, 17, 19, 21].

Diffeomorphic Atlas Building Given input images $J^1, \ldots, J^N \in L^2(\Omega, \mathbb{R})$, the atlas image I and the diffeomorphic deformations between the atlas and each

individual image are estimated alternatively by minimizing the energy,

$$E(v^n, I) = \sum_{n=1}^{N} \frac{1}{2\sigma^2} \left\| I \circ (\phi^n)^{-1} - J^n \right\|^2 + (Lv^n, v^n), \tag{1}$$

where σ^2 represents noise variance, and the velocities $\{v^n \in L^2([0,1], V)\}_{n=1,\ldots,N}$ are initial velocity fields in a reproducing kernel Hilbert space, V, equipped with a metric, $\mathcal{L} : V \to V^*$, a positive-definite, self-adjoint, differential operator, mapping to the dual space, V^*. The dual to the vector v^n is a momentum, $m^n \in V^*$, such that $m^n = Lv^n$ and $v^n = Km^n$, where K is the inverse operator of L. The notation (m^n, v^n) denotes the pairing of a momentum vector $m^n \in V^*$ with a tangent vector $v^n \in V$. The deformation ϕ^n is generated by the integral flow of time-varying velocity fields v_t^n, that is, $(d/dt)\phi^n(t,x) = v^n(t, \phi^n(t,x))$ with $v_0^n = v^n$. To achieve the optimal solution of the energy above, the geodesic path ϕ^n is constructed via integration of the following EPDiff equation [10]:

$$\frac{\partial v^n}{\partial t} = -K \mathrm{ad}^*_{v^n} m^n = -K \left[(Dv^n)^T m^n + Dm^n \, v^n + m^n \, \mathrm{div}(v^n) \right], \tag{2}$$

where D denotes the Jacobian matrix, and the operator ad^* is the dual of the negative Lie bracket of vector fields, $\mathrm{ad}_v w = -[v, w] = Dvw - Dwv$.

3 Mixture Model of Diffeomorphic Multi-Atlas Building

We assume that the input images $\{J^n\}_{n=1,\ldots,N}$ are generated from multiple atlases I^k, where $k = 1, \ldots, K$ represents the number of clusters and π^k denotes the prior probability of the kth cluster. Each individual image J^n is associated with a k-dimensional binary random variable $\boldsymbol{z}^n$, in which a particular kth element z^{nk} is equal to 1 and all other elements are equal to 0. As in a general mixture model, the prior distribution of $\boldsymbol{z}^n$ is specified by the mixing coefficients π^k as $p(z^{nk} = 1) = \pi^k$, where $\pi^k \in [0,1]$ with $\sum_{k=1}^{K} \pi^k = 1$. We then can write the distribution of $p(\boldsymbol{z}^n)$ as

$$p(\boldsymbol{z}^n) = \prod_{k=1}^{K} (\pi^k)^{z^{nk}}.$$

Let $\boldsymbol{v}^n$ denote a set of initial velocities from each cluster k for the nth image, which is $\{v^{nk}\}$. Similarly, the atlases $\{I^k\}$ and noise variances $\{\sigma^k\}$ will be represented as $\boldsymbol{I}$ and $\boldsymbol{\sigma}$. Consider that our input images and atlases are measured on a discrete grid, we formulate our noise model as i.i.d. Gaussian at each image voxel, with the data likelihood $p(J^n \,|\, \boldsymbol{z}^n, \boldsymbol{v}^n, \boldsymbol{I}, \boldsymbol{\sigma})$ given by

$$\begin{aligned} p(J^n \,|\, \boldsymbol{z}^n, \boldsymbol{v}^n, \boldsymbol{I}, \boldsymbol{\sigma}) &= \prod_{k=1}^{K} N(J^n \,|\, v^{nk}, I^k, \sigma^k)^{z^{nk}} \\ &= \prod_{k=1}^{K} \left[\frac{1}{(2\pi)^{M/2} (\sigma^k)^M} \exp\left(-\frac{\| I^k \circ (\phi^{nk})^{-1} - J^n \|^2}{2(\sigma^k)^2} \right) \right]^{z^{nk}}, \end{aligned} \tag{3}$$

where M is the number of voxels.

We then define a multivariate Gaussian distribution on the initial velocity $\boldsymbol{v}^n$ that guarantees smoothness of the geodesic shooting path. The formulation is given by

$$p(\boldsymbol{v}^n) = \prod_{k=1}^{K} N(v^{nk} \,|\, 0,\, L^{-1})^{z^{nk}} \propto \prod_{k=1}^{K} \left[\exp\left(-\frac{1}{2}(Lv^{nk}, v^{nk}) \right) \right]^{z^{nk}}, \tag{4}$$

where we use a metric of the form $L = -\alpha\Delta + I$, in which Δ is the discrete Laplacian operator, α is a positive regularity parameter, and I denotes an identity matrix. In this paper, we set α with the same value across all clusters.

Putting together equations (3) and (4), we arrive at the log joint posterior distribution with a set of parameters $\theta = \{I^k, \sigma^k, \pi^k\}$ as

$$\log \prod_{n=1}^{N} p(\boldsymbol{z}^n, \boldsymbol{v}^n \,|\, J^n, \theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} z^{nk}\{\log \pi^k - \frac{1}{2(\sigma^k)^2} \left\| J^n - I^k \circ (\phi^{nk})^{-1} \right\|_{L^2}^2 - \frac{M}{2}\log \sigma^k - \frac{1}{2}(Lv^{nk}, v^{nk})\} + const. \tag{5}$$

Figure 1 shows the graphical representation of our model.

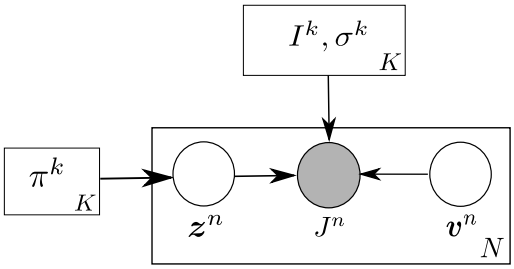

Fig. 1: Graphical representation of our model for a set of i.i.d. images $\{J^n\}$, with corresponding latent variables $\{\boldsymbol{z}^n, \boldsymbol{v}^n\}_{n=1,\ldots,N}$ and parameters $\{I^k, \sigma^k, \pi^k\}_{k=1,\ldots,K}$.

4 Inference

We now present an algorithm for estimating the parameters, θ, of our model described in the previous section. In order to treat the $\boldsymbol{z}^n, \boldsymbol{v}^n$ as latent random

variables, we need to integrate them out over the log posterior given by (5). Marginalizing $\boldsymbol{z}^n$ is straightforward as the Gaussian mixture model. However, marginalizing $\boldsymbol{v}^n$ is intractable in the closed form. We develop a Hamiltonian Monte Carlo procedure to generate samples $\boldsymbol{v}^n$ from the posterior distribution (5), and use these samples in a Monte Carlo Expectation Maximization algorithm to estimate θ. The inference consists of two main steps:

1. E-step To compute the expectation function Q, we integrate out the hidden variables $\boldsymbol{z}^n$ and $\boldsymbol{v}^n$ with the current estimate of the parameters $\theta^{(i)}$ as

$$Q(\theta \,|\, \theta^{(i)}) = E_{\boldsymbol{z}^n, \boldsymbol{v}^n} \left[\sum_{n=1}^{N} \log p(\boldsymbol{z}^n, \boldsymbol{v}^n \,|\, J^n, \theta^{(i)}) \right]. \tag{6}$$

A standard way to approximate (6) is sampling $\boldsymbol{z}^n, \boldsymbol{v}^n$ through Gibbs sampling on the joint posterior distribution (5). We draw S samples, ${\boldsymbol{v}^{nj}}_{\{j=1,\ldots,S\}}$, from the log conditional distribution $\log p(\boldsymbol{v}^n \,|\, \boldsymbol{z}^n, J^n, \theta^{(i)})$ by HMC. Note that $\boldsymbol{v}^{nj}$ denotes a set of jth samples for the nth initial velocity across k clusters. We then use the sample mean to approximate the expectation function Q. To simplify the computation, we develop a closed-form solution for marginalizing $\boldsymbol{z}^n$ from the log conditional distribution $\log p(\boldsymbol{z}^n \,|\, \boldsymbol{v}^n, J^n, \theta^{(i)})$ directly.

Closed-form solution for z^{nk}: Much like [4], the closed-form solution for computing the expectation of z^{nk} is

$$\gamma(z^{nk}) = \frac{\pi^k \prod_{j=1}^{S} N(J^n, v^{njk} \,|\, \theta^{(i)})}{\sum_{k=1}^{K} \pi^k \prod_{j=1}^{S} N(J^n, v^{njk} \,|\, \theta^{(i)})},$$

which is the responsibility that cluster k takes for representing the observed image data J^n. Here v^{njk} is the jth sample for the nth velocity field that belongs to the cluster k.

Hamiltonian Monte Carlo Sampling for v^{nk}: Hamiltonian Monte Carlo [7] is a powerful sampling method that efficiently explores the target distribution with a high acceptance rate. We draw random samples from the joint distribution of latent variables, v^{nk}, which generates diffeomorphic deformations between the atlas and each individual image. The critical step of HMC sampling is computing the gradient with respect to the initial velocity v^{nk} of the distribution (5). We use a fast version of the geodesic shooting algorithm [16] proposed by Zhang and Fletcher [20] (`https://bitbucket.org/zhubomm/flashc`). The steps for computing the gradient w.r.t. the initial velocity v^{nk} are:

- Forward integrate the geodesic evolution equations (2) to generate diffeomorphic deformations at time points $t_1 = 0, t_2, \ldots, t_T = 1$.
- Compute the gradient $\nabla_{v^{nk}} Q$ at $t_T = 1$ as

$$f = -K \left[\frac{1}{\sigma^2} (I^k \circ (\phi^k)^{-1} - J^n) \cdot \nabla I^k \circ (\phi^k)^{-1} \right],$$

– Integrate the gradient backward to $t_1 = 0$ using reduced adjoint Jacobi fields from Bullo [6] to update the gradient. By introducing time-dependent adjoint variables $\hat{v}, \hat{f}$, we write the reduced adjoint Jacobi equations as

$$\frac{d\hat{v}}{dt} = -\mathrm{ad}_v^\dagger \hat{v}, \;\; \frac{d\hat{f}}{dt} = -\hat{v} + \mathrm{sym}_v^\dagger \hat{f}\,,$$

where $\mathrm{sym}_v^\dagger \hat{f} = -\mathrm{ad}_v \hat{f} + \mathrm{ad}_{\hat{f}}^\dagger v$. For more details on the derivation of the reduced adjoint Jacobi field equations, see [6]. The final gradient term w.r.t. v^{nk} is

$$\nabla_{v^{nk}} Q = v^{nk} - f.$$

Finally, the expectation function (6) is ultimately approximated as

$$Q(\theta \,|\, \theta^{(i)}) \approx \frac{1}{S} \sum_{j=1}^{S} \sum_{n=1}^{N} \log p(\gamma(\boldsymbol{z}^n), \boldsymbol{v}^{nj} \,|\, J^n, \theta^{(i)}). \tag{7}$$

2. M-step We then maximize the approximated function $Q(\theta \,|\, \theta^{(i)})$ (7) to update the parameters $\theta = \{I^k, \sigma^k, \pi^k\}$, which turns out to be a closed-form update for all parameters. We set the derivative of the expectation Q w.r.t. each parameter of θ to zero, and the closed-form update is

$$\begin{aligned}
\tilde{\pi}^k &= \frac{N^k}{N}, \text{ where } N^k = \sum_{n=1}^{N} \gamma(z^{nk}), \\
\tilde{I}^k &= \frac{\sum_{n=1}^{N} \sum_{j=1}^{S} \gamma(z^{nk}) \cdot J^k \circ \phi^{njk} |D\phi^{njk}|}{\sum_{n=1}^{N} \sum_{j=1}^{S} \gamma(z^{nk}) \cdot |D\phi^{njk}|}, \\
(\tilde{\sigma}^k)^2 &= \frac{1}{M \cdot S \cdot N^k} \sum_{i=1}^{N} \sum_{j=1}^{S} \gamma(z^{nk}) \cdot \|I^k \circ (\phi^{njk})^{-1} - J^n\|^2.
\end{aligned}$$

5 Results

We demonstrate the effectiveness of our model using both 2D synthetic data and real 3D MRI brain data.

Synthetic Data Because we have a generative model, we can forward simulate random images from the known parameters $\theta = \{I^k, \sigma^k, \pi^k\}$, where we choose $k \in \{1, 2, 3\}$. We use three atlases, which are 2D binary images of a square, triangle, and ellipse with a resolution of 100×100. We then generate 30 initial velocity fields (10 per cluster) from the prior $p(v^{nk})$ given in (4), setting $\alpha = 3.0$. We shoot the initial velocities by the EPDiff equations (2) to generate diffeomorphic deformations, and then use them to transform the atlases. Finally, we add random Gaussian noise with $\sigma = (0.01, 0.025, 0.03)$ to each transformed cluster atlas.

In our testing procedure, we initialize $\sigma = 0.3$ for all K clusters (K is the true number of clusters in this synthetic example). For the HMC sampling procedure, we use a step size of 0.05 for leap-frog integration with 40 samples after a burn-in of 50 samples. Each atlas from the kth cluster is initialized to the linear average of the image intensities over the samples we generated for each cluster, and the $\{\pi^k\}$ are set as the averaged weight, 0.3. Figure 2 compares the ground truth atlases and our estimated atlases, showing that our model is able to accurately recover the true atlases, as well as automatically cluster data into sub-groups. As for other parameters, we get the estimated $\sigma = (0.011, 0.026, 0.031)$ and $\pi = 0.33333$ for each cluster. We compared our multi-atlas approach with a single atlas estimated over all data points using the method of Zhang et al. [21]. The single atlas was completely incapable of representing the synthetic dataset.

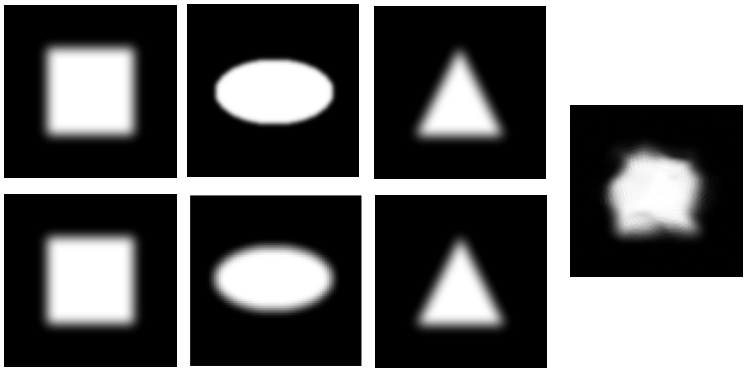

Fig. 2: Estimation of atlases. Top: ground truth atlases of three clusters: square, ellipse, and triangle; Bottom: our estimation; Right: single atlas estimated from the whole dataset.

OASIS Brain Data To show the effectiveness of our model on the real 3D brain data, we applied our algorithm to an OASIS brain MRI dataset with 26 healthy subjects from ages 60 to 90. All the MRI images have resolution $128 \times 128 \times 128$ with the image spacing $1.0 \times 1.0 \times 1.0mm^3$, and are skull-stripped, intensity normalized, and co-registered with affine transforms. We set $\alpha = 0.3$, which was estimated by Zhang et al. [21] with 10 time-steps in geodesic shooting. We ran K-means algorithm with two clusters using image intensity as features, and then used the output as our initialization for I^k, the initial atlas at each cluster. Note that here we use cross-validation to determine the number of clusters. Other alternative ways could also be used, such as the Elbow method, which evaluates the percentage of variance with respect to the number of clusters and information criterion approaches (for instance, Akaike information criterion and Bayesian information criterion).

The first two columns in Figure 3 show sagittal, axial, and coronal views of slices from the output of K-means algorithm, which are the greyscale averages of the clustered images. The middle two columns are atlases estimated from our model. It demonstrates that the final atlases produces sharper averaged images with more details. Meanwhile, the big shape difference between the two estimated atlases shows that multiple atlases gives a better representation of multi-model population by an atlas per mode than a single atlas that mixes up the features across different groups. For a purpose of better visualization, we also add difference maps that represent the absolute value of the intensity differences between our two estimated atlases.

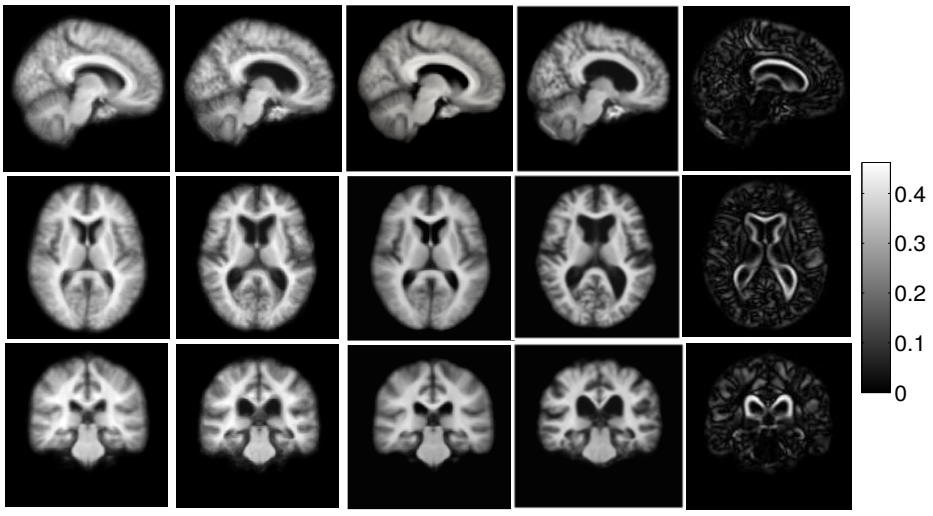

Fig. 3: Initialization and our estimation of atlases. Top to bottom: sagittal, axial, and coronal views of the K-means initialization and our estimated atlases. Left to right: initialization for each cluster (column 1-2), our estimated atlases from two different clusters (column 3-4) and difference maps over image intensity between two atlases.

6 Conclusion

In this paper, we presented a generative Gaussian mixture model of diffeomorphic multi-atlas building. Our method is the first probabilistic model for constructing multiple atlases navigated by unsupervised clustering in a diffeomorphic setting. To treat the diffeomorphic transformations as latent random variables, we developed a MCEM algorithm to integrate the diffeomorphisms out from the

joint distribution of data likelihood via HMC sampling on the space of diffeomorphisms. Our algorithm aggregates data that belongs to the same category automatically and constructs multiple representations of a large image database. This framework can be very useful for further statistical analysis in many areas, such as shape variation quantification and guidance of image segmentation.

Acknowledgments This work was supported by NSF CAREER Grant 1054057.

References

1. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46(3), 726–738 (2009)
2. Allassonnière, S., Amit, Y., Trouvé, A.: Toward a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society, Series B 69, 3–29 (2007)
3. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)
4. Bishop, C.M.: Pattern recognition and machine learning. Springer (2006)
5. Blezek, D.J., Miller, J.V.: Atlas stratification. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, pp. 712–719. Springer (2006)
6. Bullo, F.: Invariant affine connections and controllability of Lie groups (1995)
7. Duane, S., Kennedy, A., Pendleton, B., Roweth, D.: Hybrid Monte Carlo. Physics Letters B pp. 216–222 (1987)
8. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, Supplement1, 151–160 (2004)
9. Koikkalainen, J., Lötjönen, J., Thurfjell, L., Rueckert, D., Waldemar, G., Soininen, H., ADNI: Multi-template tensor-based morphometry: application to analysis of Alzheimer's disease. NeuroImage 56(3), 1134–1144 (2011)
10. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (2006)
11. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21(4), 1428–1442 (2004)
12. Sabuncu, M.R., Balci, S.K., Shenton, M.E., Golland, P.: Image-driven population analysis through mixture modeling. Medical Imaging, IEEE Transactions on 28(9), 1473–1487 (2009)
13. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In: International Symposium on Biomedial Imaging (ISBI) (April 2013)
14. Tang, X., Oishi, K., Faria, A.V., Hillis, A.E., Albert, M.S., Mori, S., Miller, M.I.: Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. PloS one 8(6), e65591 (2013)
15. Van Rikxoort, E.M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M.A., Pluim, J.P., van Ginneken, B.: Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. Medical Image Analysis 14(1), 39–49 (2010)

16. Vialard, F.X., Risser, L., Holm, D., Rueckert, D.: Diffeomorphic atlas estimation using Kärcher mean and geodesic shooting on volumetric images. In: MIUA (2011)
17. Vialard, F.X., Risser, L., Rueckert, D., J.Cotter, C.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision 97(2), 229–241 (2012)
18. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D., ADNI: LEAP: learning embeddings for atlas propagation. NeuroImage 49(2), 1316–1325 (2010)
19. Younes, L., Arrate, F., Miller, M.: Evolutions equations in computational anatomy. NeuroImage 45(1S1), 40–50 (2009)
20. Zhang, M., Fletcher, P.T.: Finite-dimensional Lie algebras for fast diffeomorphic image registration. In: Information Processing in Medical Imaging (2015)
21. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Information Processing in Medical Imaging. pp. 37–48. Springer (2013)