Multimodal unbiased image matching via mutual information

Igor Yanovsky[a], Paul M. Thompson[b], Stanley J. Osher[a], Alex D. Leow[b]

[a]Department of Mathematics, University of California, Los Angeles, CA, USA
[b]Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA, USA

ABSTRACT

In the past decade, information theory has been studied extensively in computational imaging. In particular, image matching by maximizing mutual information has been shown to yield good results in multimodal image registration. However, there have been few rigorous studies to date that investigate the statistical aspect of the resulting deformation fields. Different regularization techniques have been proposed, sometimes generating deformations very different from one another. In this paper, we present a novel model for multimodal image registration. The proposed method minimizes a purely information-theoretic functional consisting of mutual information matching and unbiased regularization. The unbiased regularization term measures the magnitude of deformations using either asymmetric Kullback-Leibler divergence or its symmetric version. The new multimodal unbiased matching method, which allows for large topology preserving deformations, was tested using pairs of two and three dimensional serial MRI images. We compared the results obtained using the proposed model to those computed with a well-known mutual information based viscous fluid registration. A thorough statistical analysis demonstrated the advantages of the proposed model over the multimodal fluid registration method when recovering deformation fields and corresponding Jacobian maps.

Keywords: Multimodal matching, mutual information, nonlinear image registration, computational anatomy.

1. INTRODUCTION

In recent years, computational neuroimaging has become an exciting interdisciplinary field with many applications in functional and anatomic brain mapping, image-guided surgery, and multimodality image fusion.[1–6] The goal of image registration is to align, or spatially normalize, one image to another. In multi-subject studies, this reduces subject-specific anatomic differences by deforming individual images onto a population average brain template. When applied to serial scans of human brain, image registration offers tremendous power in detecting the earliest signs of illness, understanding normal brain development or aging, and monitoring disease progression. Recently, there has been an expanding literature on various nonrigid registration techniques, with different image matching functionals, regularization schemes, and numerical implementations. In[7,8] we systematically examined the statistical properties of Jacobian maps (the determinant of the local Jacobian operator applied to the deformations), and proposed an *unbiased* large-deformation image registration approach. In this context, unbiased means that the Jacobian determinants of the deformations recovered between a pair of images follow a log-normal distribution, with zero mean after log-transformation. We argued that this distribution is beneficial when recovering change in regions of homogeneous intensity, and in ensuring symmetrical results when the order of two images being registered is switched. We applied this method to a longitudinal MRI dataset from a single subject, and showed promising results in eliminating spurious signals. We also noticed that different registration techniques, when applied to the same longitudinal dataset, may sometimes yield visually very different Jacobian maps, causing problems in interpreting local structural changes. Given this ambiguity and the increasing use of registration methods to measure brain change, more information is required concerning the baseline stability, reproducibility, and statistical properties of signals generated by different nonrigid registration techniques.

Further author information:
I.Y.: E-mail: yanovsky@math.ucla.edu
P.T.: E-mail: thompson@loni.ucla.edu
S.O.: E-mail: sjo@math.ucla.edu
A.L.: E-mail: feuillet@ucla.edu

In this paper, we present a novel model for multimodal image registration. The proposed method minimizes a purely information-theoretic functional consisting of mutual information matching and unbiased regularization (prior work has focused on the case where unbiased regularization was coupled with the summed squared intensity difference matching). The unbiased regularization term measures the magnitude of deformations using either asymmetric Kullback-Leibler (KL) divergence or symmetric Kullback-Leibler (SKL) distance yielding Asymmetric Unbiased and Symmetric Unbiased models, respectively. We also aim to provide quality calibrations for different non-rigid registration techniques in TBM. In particular, we compare three regularization techniques (fluid registration versus the Asymmetric Unbiased and Symmetric Unbiased techniques). Our experiments are designed to decide which registration method is more reproducible, more reliable, and offers less artifactual variability in regions of homogeneous image intensity. The foundation of our calibrations is based on the assumption that, by scanning healthy normal human subjects from Alzheimer's Disease Neuroimaging Initiative (ADNI) baseline dataset twice over a 2-week period using the same protocol, serial MRI scan pairs should not show any systematic biological change. Therefore, any regional structural differences detected using TBM over such a short interval may be assumed to be errors. We apply statistical analysis to the profile of these errors, providing information on the reliability, reproducibility and variability of different registration techniques. Moreover, serial images of 10 subjects from the ADNI follow-up phase (images acquired one year apart) were analyzed in a similar fashion. In images collected one year apart, real anatomical changes are present; neurobiological changes due to aging and dementia include widespread cell shrinkage, regional gray and white matter atrophy and expansion of fluid-filled spaces in the brain. Thus, a good computational technique should be able to detect physiological changes for the ADNI follow-up (1-year) phase.

At this point, we would like to motivate the unbiased approach, which couples the computation of deformations with statistical analyses on the resulting Jacobian maps. As a result, the unbiased approach ensures that deformations have intuitive axiomatic properties by penalizing any bias in the corresponding statistical maps. In the following sections, we describe the mathematical foundations of this approach, define energy functionals for minimization, and perform thorough statistical analyses to demonstrate the advantages of the multimodal unbiased registration models.

2. UNBIASED LARGE-DEFORMATION IMAGE REGISTRATION

Let Ω be an open and bounded domain in $\mathbb{R}^n$, for arbitrary n. Without loss of generality, assume that the volume of Ω is 1, i.e. $|\Omega| = 1$. Let $I_1 : \Omega \to \mathbb{R}$ and $I_2 : \Omega \to \mathbb{R}$ be the two images to be registered. We seek to find the transformation $\mathbf{g} : \Omega \to \Omega$ that maps the source image I_2 into correspondence with the target image I_1. In this paper, we will restrict this mapping to be differentiable, one-to-one, and onto. We denote the Jacobian matrix of a deformation $\mathbf{g}$ to be $D\mathbf{g}$, with Jacobian denoted by $|D\mathbf{g}(\mathbf{x})|$. The displacement field $\mathbf{u}(\mathbf{x})$ from the position $\mathbf{x}$ in the deformed image $I_2 \circ \mathbf{g}(\mathbf{x})$ back to $I_2(\mathbf{x})$ is defined in terms of the deformation $\mathbf{g}(\mathbf{x})$ by the expression $\mathbf{g}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x})$ at every point $\mathbf{x} \in \Omega$. Thus, we consider the problems of finding $\mathbf{g}$ and $\mathbf{u}$ to be equivalent.

We now describe the construction of the Unbiased Large-Deformation Image Registration. We associate three probability density functions to $\mathbf{g}$, $\mathbf{g}^{-1}$, and the identity mapping $\mathbf{id}$:

$$p_{\mathbf{g}}(\mathbf{x}) = |D\mathbf{g}(\mathbf{x})|, \quad p_{\mathbf{g}^{-1}}(\mathbf{x}) = |D\mathbf{g}^{-1}(\mathbf{x})|, \quad p_{\mathbf{id}}(\mathbf{x}) = 1. \tag{1}$$

By associating deformations with their corresponding global density maps, we can now apply information theory to quantify the magnitude of deformations. In our approach, we choose the Kullback-Leibler (KL) divergence and symmetric Kullback-Leibler (SKL) distance. The KL divergence between two probability density functions, $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, is defined as $KL(p_1(\mathbf{x}), p_2(\mathbf{x})) = \int_\Omega p_1(\mathbf{x}) \log[p_1(\mathbf{x})/p_2(\mathbf{x})]d\mathbf{x} \geq 0$. We define the SKL distance as $SKL(p_1(\mathbf{x}), p_2(\mathbf{x})) = KL(p_1(\mathbf{x}), p_2(\mathbf{x})) + KL(p_2(\mathbf{x}), p_1(\mathbf{x}))$.

The Unbiased method solves for the deformation $\mathbf{g}$ (or, equivalently, for the displacement $\mathbf{u}$) minimizing the energy functional E, consisting of the image matching term F and the regularizing term R which is based on KL divergence or SKL distance. The fidelity term F dependents on I_2 and I_1, as well as the displacement $\mathbf{u}$. The general minimization problem can be written as

$$E(I_1, I_2, \mathbf{u}) = F(I_1, I_2, \mathbf{u}) + \lambda R(\mathbf{u}), \qquad \inf_{\mathbf{u}} E(I_1, I_2, \mathbf{u}). \tag{2}$$

Here, $\lambda > 0$ is a weighting parameter.

2.1 Asymmetric Unbiased Registration

To quantify the magnitude of deformation $\mathbf{g}$, in this paper we introduce a new regularization term R_{KL}, which is an asymmetric measure between $p_{\mathbf{id}}$ and $p_{\mathbf{g}}$: $R_{KL}(\mathbf{g}) = KL(p_{\mathbf{id}}, p_{\mathbf{g}})$. This regularization term can be shown to be

$$R_{KL}(\mathbf{g}) = \int_\Omega p_{\mathbf{id}} \log \frac{p_{\mathbf{id}}}{p_{\mathbf{g}}} \, d\mathbf{x} = \int_\Omega -\log|D\mathbf{g}(\mathbf{x})| d\mathbf{x} = \int_\Omega |D\mathbf{g}^{-1}(\mathbf{y})| \log|D\mathbf{g}^{-1}(\mathbf{y})| d\mathbf{y}. \tag{3}$$

Thus, the energy functional in (2) implementing Asymmetric Unbiased registration can be written as

$$E(I_1, I_2, \mathbf{u}) = F(I_1, I_2, \mathbf{u}) - \lambda \int_\Omega \log|D(\mathbf{x} - \mathbf{u}(\mathbf{x}))| d\mathbf{x}, \tag{4}$$

for some distance measure F between $I_2(\mathbf{x} - \mathbf{u})$ and $I_1(\mathbf{x})$.

2.2 Symmetric Unbiased Registration

In this section, we describe the regularization functional based on the symmetric KL distance between $p_{\mathbf{id}}$ and $p_{\mathbf{g}}$: $R_{SKL}(\mathbf{g}) = SKL(p_{\mathbf{id}}, p_{\mathbf{g}})$. As shown in,[7] the regularization term is linked to statistics on Jacobian maps as follows

$$
\begin{aligned}
R_{SKL}(\mathbf{g}) &= KL(p_{\mathbf{g}}, p_{\mathbf{id}}) + KL(p_{\mathbf{g}^{-1}}, p_{\mathbf{id}}) = KL(p_{\mathbf{g}}, p_{\mathbf{id}}) + KL(p_{\mathbf{id}}, p_{\mathbf{g}}) \\
&= KL(p_{\mathbf{id}}, p_{\mathbf{g}^{-1}}) + KL(p_{\mathbf{id}}, p_{\mathbf{g}}) = KL(p_{\mathbf{id}}, p_{\mathbf{g}^{-1}}) + KL(p_{\mathbf{g}^{-1}}, p_{\mathbf{id}}) \\
&= \int_\Omega \left(|D\mathbf{g}(\mathbf{x})| - 1\right) \log|D\mathbf{g}(\mathbf{x})| d\mathbf{x} = \int_\Omega \left(|D\mathbf{g}^{-1}(\mathbf{y})| - 1\right) \log|D\mathbf{g}^{-1}(\mathbf{y})| d\mathbf{y}.
\end{aligned}
$$

The energy functional employing Symmetric Unbiased registration can be rewritten as

$$E(I_1, I_2, \mathbf{u}) = F(I_1, I_2, \mathbf{u}) + \lambda \int_\Omega \left(|D(\mathbf{x} - \mathbf{u}(\mathbf{x}))| - 1\right) \log|D(\mathbf{x} - \mathbf{u}(\mathbf{x}))| d\mathbf{x}, \tag{5}$$

for some distance measure F. Notice that the symmetric unbiased regularizing functional is pointwise nonnegative, while the asymmetric unbiased regularizer in (3) can take either positive or negative values locally.

3. MUTUAL INFORMATION MATCHING

In,[7, 8] we used the sum of squared intensity differences as a measure of the distance between the deformed image $I_2(\mathbf{x} - \mathbf{u})$ and target image $I_1(\mathbf{x})$, defined as $F_{SSD}(I_1, I_2, \mathbf{u}) = \frac{1}{2} \int_\Omega \left(I_2(\mathbf{x} - \mathbf{u}(\mathbf{x})) - I_1(\mathbf{x})\right)^2 d\mathbf{x}$. The sum of the squared differences matching functional is suitable when the images have been acquired through similar sensors and thus are expected to present the same intensity range and distribution. In this paper, we employ mutual information as the distance metric. Mutual information is a measure of how much information one random variable has about another. One of the main advantages of using mutual information is that it can be used to align images of different modalities, without requiring knowledge of the relationship (joint intensity distribution) of the two registered images. In this section, we will briefly review information theory basics and define the mutual information between the deformed image and the target image.

3.1 Information Theory

Let X be a continuous random variable with continuous cumulative distribution function $P(x) = \Pr\{X \le x\}$. Let $p(x) = P'(x)$ when the derivative is defined. The differential entropy $H(X)$ of a continuous random variable X with the probability density function $p(x)$ is defined as $H(X) = -\int_D p(x) \log p(x) dx$, where D is the support set of the random variable. The joint differential entropy $H(X, Y)$ of a pair of random variables X and Y with joint density function $p(x, y)$ is defined as $H(X, Y) = -\int p(x, y) \log p(x, y) dx dy$, and the conditional differential entropy $H(X|Y)$ is provided by $H(X|Y) = -\int p(x, y) \log p(x|y) dx dy$. Here, $p(x|y)$ is the conditional probability density function. It can be shown that the joint entropy of two random variables is the entropy of one plus the conditional entropy of the other: $H(X, Y) = H(X) + H(Y|X)$.

We now introduce mutual information, which is a measure of the amount of information that one random variable contains about another random variable. Mutual information $MI(X,Y)$ is provided by

$$MI(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y).$$

Hence, the mutual information is the reduction in the uncertainty of X due to the knowledge of Y. It is often more convenient to write the mutual information in terms of the Kullback-Leibler divergence between the joint distribution $p(x,y)$ and the product of marginal distributions $p(x)$ and $p(y)$:

$$MI(X,Y) = KL(p(x,y), p(x)p(y)) = \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy.$$

3.2 Image Matching

To define the mutual information between the deformed image $I_2(\mathbf{x} - \mathbf{u})$ and the target image $I_1(\mathbf{x})$, we follow the notations in,[9] where p^{I_1} and $p_{\mathbf{u}}^{I_2}$ are used to denote the intensity distributions estimated from $I_1(\mathbf{x})$ and $I_2(\mathbf{x} - \mathbf{u})$, respectively. An estimate of their joint intensity distribution is denoted as $p_{\mathbf{u}}^{I_1, I_2}$. In this probabilistic framework, the link between two modalities is fully characterized by a joint density. We let $i_1 = I_1(\mathbf{x})$, $i_2 = I_2(\mathbf{x} - \mathbf{u}(\mathbf{x}))$ denote intensity values at point $\mathbf{x} \in \Omega$. Given the displacement field $\mathbf{u}$, the mutual information computed from I_1 and I_2 is provided by

$$MI_{\mathbf{u}}^{I_1, I_2} = MI(I_1(\mathbf{x}), I_2(\mathbf{x} - \mathbf{u}(\mathbf{x}))) = \int_{\mathbb{R}^2} p_{\mathbf{u}}^{I_1, I_2}(i_1, i_2) \log \frac{p_{\mathbf{u}}^{I_1, I_2}(i_1, i_2)}{p^{I_1}(i_1) p_{\mathbf{u}}^{I_2}(i_2)} di_1 di_2.$$

We seek to maximize the mutual information between $I_2(\mathbf{x} - \mathbf{u})$ and $I_1(\mathbf{x})$, or equivalently, minimize the negative of $MI_{\mathbf{u}}^{I_1, I_2}$:

$$F_{MI}(I_1, I_2, \mathbf{u}) = -MI_{\mathbf{u}}^{I_1, I_2}. \tag{6}$$

The gradient of (6) is given by $\partial_{\mathbf{u}} F_{MI}(\mathbf{u}) = \frac{1}{|\Omega|} \left[Q_{\mathbf{u}} * \frac{\partial \psi}{\partial \xi_2} \right] (I_1(\mathbf{x}), I_2(\mathbf{x} - \mathbf{u})) \nabla I_2(\mathbf{x} - \mathbf{u})$, where $Q_{\mathbf{u}}(i_1, i_2) = 1 + \log \frac{p_{\mathbf{u}}^{I_1, I_2}(i_1, i_2)}{p^{I_1}(i_1) p_{\mathbf{u}}^{I_2}(i_2)}$, and $\psi(\xi_1, \xi_2)$ is a two-dimensional Parzen windowing kernel, which is used to estimate the joint intensity distribution from $I_2(\mathbf{x} - \mathbf{u})$ and $I_1(\mathbf{x})$. Here, we use the Gaussian kernel with variance σ^2: $\psi(\xi_1, \xi_2) = G_\sigma(\xi_1, \xi_2) = \frac{1}{2\pi\sigma^2} e^{\frac{-(\xi_1^2 + \xi_2^2)}{2\sigma^2}}$.

4. MINIMIZATION OF ENERGY FUNCTIONALS

In general, we expect minimizers of the energy functional $E(\mathbf{u})$ to exist. Computing the first variation of the functional in (2), we obtain the gradient of $E(I_1, I_2, \mathbf{u})$, namely $\partial_{\mathbf{u}} E(I_1, I_2, \mathbf{u})$. We define the force field $\mathbf{f}$, which drives I_2 into registration with I_1, as

$$\mathbf{f}(\mathbf{x}, \mathbf{u}) = \partial_{\mathbf{u}} E(I_1, I_2, \mathbf{u}) = \partial_{\mathbf{u}} F_{MI}(I_1, I_2, \mathbf{u}) + \lambda \partial_{\mathbf{u}} R(\mathbf{u}). \tag{7}$$

Here, $R(\mathbf{u})$ is either $R_{KL}(\mathbf{u})$ or $R_{SKL}(\mathbf{u})$. Explicit expressions for components of $\partial_{\mathbf{u}} R(\mathbf{u})$, in both cases, are derived in Appendix A for three dimensional case. We minimize (2) using the fluid flow proposed in.[5] Given the velocity field $\mathbf{v}$, the following partial differential equation can be solved to obtain the displacement field $\mathbf{u}$:

$$\frac{\partial \mathbf{u}(\mathbf{x}, \tau)}{\partial \tau} = \mathbf{v}(\mathbf{x}, \tau) - \mathbf{v}(\mathbf{x}, \tau) \cdot \nabla \mathbf{u}(\mathbf{x}, \tau). \tag{8}$$

Here, τ is an artificial time variable. The instantaneous velocity is obtained[10] by convolving $\mathbf{f}$ with Gaussian kernel G_σ of variance σ^2, $\mathbf{v} = G_\sigma * (-\mathbf{f}(\mathbf{x}, \mathbf{u}))$.

To avoid possible confusion, we summarize the methods we will be referring to in our subsequent analyses. In later discussions, minimization of the following energies

$$E(I_1, I_2, \mathbf{u}) = F_{MI}(I_1, I_2, \mathbf{u}) + \lambda R_{KL}(\mathbf{u}), \tag{9}$$
$$E(I_1, I_2, \mathbf{u}) = F_{MI}(I_1, I_2, \mathbf{u}) + \lambda R_{SKL}(\mathbf{u}) \tag{10}$$

using equations (7), (8) will be referred to as MI-Asymmetric Unbiased and MI-Symmetric Unbiased models, respectively. The model above, provided $\lambda = 0$, will be referred to as the MI-Fluid model.

5. STATISTICAL ANALYSIS

5.1 Statistical testing on the deviation of log Jacobian maps

Based on the authors' approach in,[7] we observe that, given that there is no systematic structural change within two weeks, any deviation of the Jacobian map from one should be considered error. Thus, we expect that a better registration technique would yield $\log |D\mathbf{g}|$ values closer to 0 (i.e., smaller log Jacobian deviation translates into better methodology). Mathematically speaking, one way to test the performance is to consider the deviation map dev of the logged (i.e., logarithmically transformed) Jacobian away from zero, defined at each voxel as

$$dev(\mathbf{x}) = \big| \log |D\mathbf{g}(\mathbf{x})| \big|. \tag{11}$$

For two different registration methods A and B, we define the voxel-wise deviation gain of A over B (denoted by $S^{A,B}$) as

$$S^{A,B}(\mathbf{x}) = dev^A(\mathbf{x}) - dev^B(\mathbf{x}).$$

For the ADNI baseline dataset (in which patients are scanned twice with MRI, two weeks apart), two distinct types of t tests are used, a within-subject paired t test and a group paired t test. A within-subject paired t test is conducted for each subject by pooling all voxels inside a region of interest, as defined by the ICBM whole brain mask. This determines whether two methods differ significantly inside the whole brain (for each subject). A group paired t test, on the other hand, is performed across subjects, by computing a voxel-wise t-map of deviation gains. In this case, to statistically compare the performance of two registration methods, we rely on the standard t test on the voxel mean of S. To construct a suitable null hypothesis, we notice that the following relation would hold, assuming B outperforms A: $S^{A,B} > 0$. Thus, the null hypothesis in this case would be testing if the mean deviation gain is zero: $H_0 : \mu_{S^{A,B}} = 0$. To determine the ranking of A and B, we have to consider one-sided alternative hypotheses. For example, when testing if B outperforms A, we use the following alternative hypothesis $H_1 : \mu_{S^{A,B}} > 0$. The voxel-wise T statistic, defined as

$$T_{S^{A,B}}(\mathbf{x}) = \frac{\sqrt{n} \cdot \overline{S^{A,B}(\mathbf{x})}}{\sigma_{S^{A,B}}(\mathbf{x})},$$

where $\overline{S^{A,B}(\mathbf{x})} = \sum_i S_i^{A,B}(\mathbf{x})/n$, and $(\sigma_{S^{A,B}}(\mathbf{x}))^2 = \sum_i (S_i^{A,B}(\mathbf{x}) - \overline{S^{A,B}(\mathbf{x})})^2/(n-1)$, thus follows the Student's t distribution under the null hypothesis and may be used to determine the p-value that the null hypothesis is true. If the alternative hypothesis is accepted, we confirm that sequence B outperforms A at point $\mathbf{x}$. Otherwise, we would rank A and B equally if the null hypothesis is not rejected.

5.2 Permutation Testing to Correct Multiple Comparisons

To determine the overall global effects of different registration methods on the deviation of log Jacobian maps throughout the brain, we performed permutation tests to adjust for multiple comparisons.[11,12] Following the analyses in,[13] we resampled the observations by randomly flipping the sign of $S_i^{A,B}$ ($i = 1, 2, ..., n$) under the null hypothesis. For each permutation, voxelwise t tests are computed. We then compute the percentage of voxels inside the chosen ROI (in this case the ICBM mask) with T statistics exceeding a certain threshold. The multiple comparisons corrected p value may be determined by counting the number of permutations whose above-defined percentage exceeds that of the un-permuted observed data. For example, we say that sequence B outperforms A on the whole brain if this corrected p value is smaller than 0.05 (that is, less than 5% of all permutations have the above-defined percentage greater than that of the original data). All possible ($2^{10} = 1024$) permutations were considered in determining the final corrected p value.

5.3 Cumulative Distribution Function (CDF)

To visually assess the global significance level of the voxel-wise t tests on deviation gains, we also employed the cumulative distribution function (CDF) plot. In brief, we plot observed cumulative probabilities against the theoretical distribution under the null hypothesis. These CDF plots are commonly created as an intermediate step, when using the false discovery rate (FDR) method to assign overall significance values to statistical maps.[14] As they show the proportion of supra-threshold voxels in a statistical map, for a range of thresholds, these CDF

(a) I_1	(c) $I_2 \circ \mathbf{g}$, MI-Fluid	(e) $|D\mathbf{g}|$, MI-Fluid
(b) I_2	(d) $I_2 \circ \mathbf{g}$, MI-Sym.Unbiased	(f) $|D\mathbf{g}|$, MI-Sym.Unbiased
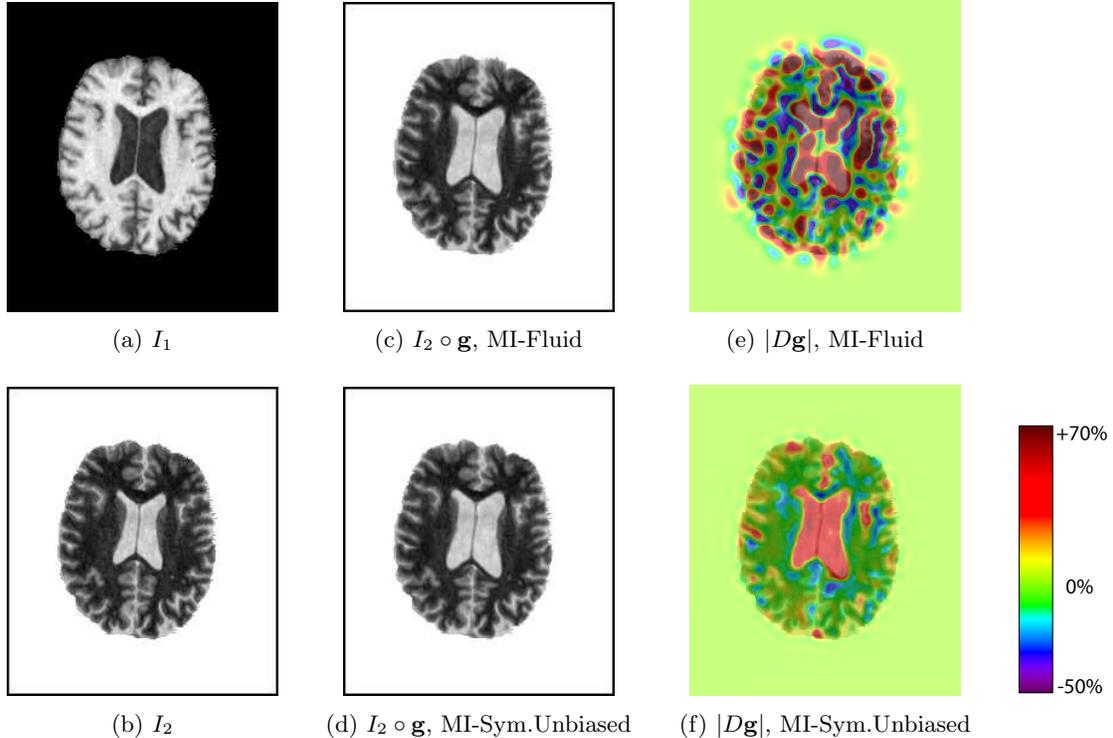
Figure 1. Serial MRI example. (a) image I_1; (b) image I_2; (c) image I_2 is deformed to image I_1 using MI-Fluid model; (d) image I_2 is deformed to image I_1 using MI-Symmetric Unbiased model. Jacobian map of the deformation is superimposed with the deformed image for (e) MI-Fluid model and (f) MI-Symmetric Unbiased model.

plots offer a measure of the effect size in a statistical map. They also may be used to demonstrate which methodological choices influence the effect size in a method that creates statistical maps. In the case of deviation gains S of a worse technique A over a better technique B in the ADNI baseline data, we expect a CDF curve to lie above the Null line, in the sense that a better technique exhibits less systematic changes.

6. RESULTS

In this section, we tested the MI-Asymmetric Unbiased and MI-Symmetric Unbiased models and compared the results to those obtained using the MI-Fluid registration model (see equations (9)-(10)). Of note, even though Asymmetric Unbiased and Symmetric Unbiased methods minimize different energy functionals, our experiments showed that they generate very similar maps.

To obtain a fair comparison, re-gridding was not employed. Re-gridding is a method to relax the energy computed from the linear elasticity prior after a certain number of iterations, which allows large-deformation mappings to be recovered without any absolute penalty on the displacement field (other than via the smoothness constraint on the velocity field which is integrated to give the displacement).[5] It is essentially a memory-less procedure, as how images are matched after each re-gridding is independent of the final deformation before the re-gridding, rendering the comparison of final Jacobian fields and cost functionals problematic. Moreover, we consider the strategy of re-gridding, through the relaxation of deformation fields over time, to be less rigorous from a theoretical standpoint, as the imposition of a regularizer can be used to secure distributional properties in the resulting statistics (e.g., symmetric log-Jacobian).

In order to observe the robustness of mutual information matching and also to gain more insight into the effect of the unbiased regularization, in Figures 1 and 2 we first consider matching pairs of 2D binary slices from a set of serial MRI images (each of size 226 by 256), where visually significant ventricle enlargement
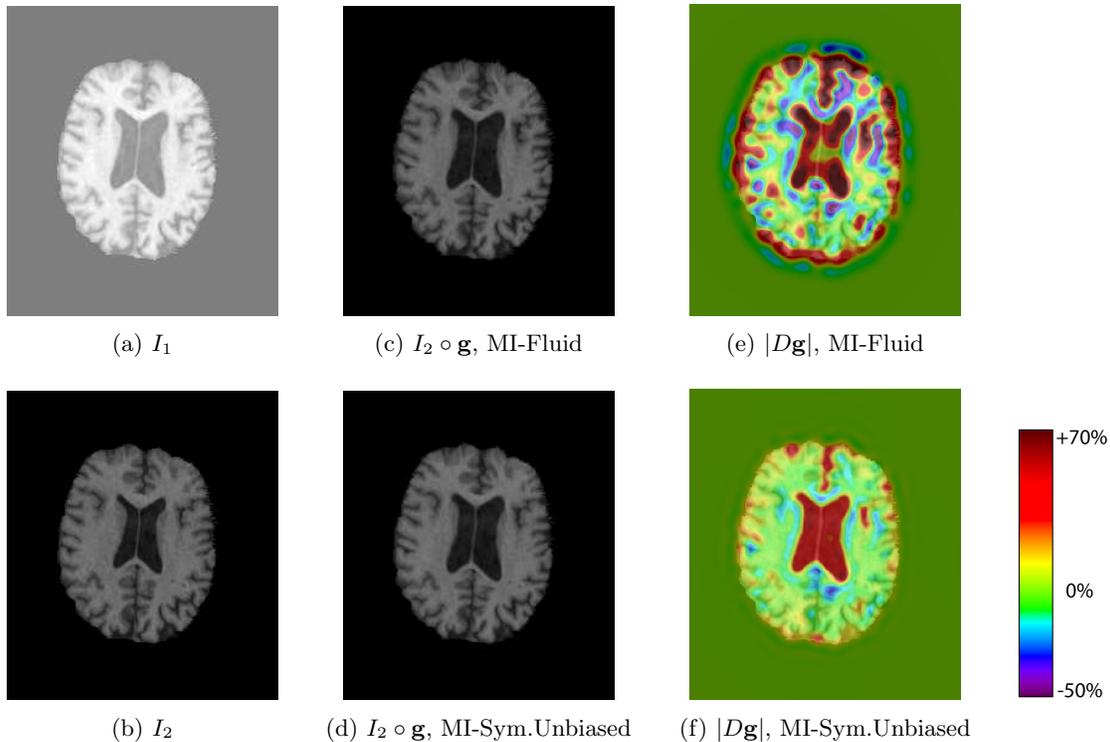
Figure 2. Serial MRI example. (a) image I_1; (b) image I_2; (c) image I_2 is deformed to image I_1 using MI-Fluid model; (d) image I_2 is deformed to image I_1 using MI-Symmetric Unbiased model. Jacobian map of the deformation is superimposed with the deformed image for (e) MI-Fluid model and (f) MI-Symmetric Unbiased model.

is present. In both figures, source and target images are of different contrast, making the sum of squared intensity differences inapplicable as a choice of a data fidelity term. Both MI-Fluid and MI-Symmetric Unbiased methods generated a close match between the deformed image and the target image (Figures 1(c,d) and 2(c,d)). Figures 1(e,f) and 2(e,f) show Jacobian maps of deformations. Here, there is no reason not to evenly distribute the Jacobian field inside the ventricles, as realized using the Unbiased method. In contrast, Fluid method generated noisy mean Jacobian maps with extreme values along the ventricular boundary. Indeed, given the overall longitudinal ventricular dilatation, we argue that the corresponding density change map should be constant inside the ventricle.

We also performed nonlinear registration on a dataset that we shall refer to as the "ADNI Baseline" dataset, collected during the preparatory phase of the ADNI project, which includes serial MRI images of ten normal elderly subjects acquired two weeks apart. Each of the ten pairs of scans is represented on a $128 \times 160 \times 128$ grid. Here, the foundation of calibrations is based on the assumption that, by scanning normal control human subjects serially within a two-week period using the same MRI protocol, no systematic structural changes should be recovered.

Here, we compared Fluid, Asymmetric Unbiased, and Symmetric Unbiased methods coupled with mutual information matching. Uniform values of $\lambda = 5$ and $\lambda = 10$ were used for all deformations using MI-Symmetric Unbiased and MI-Asymmetric Unbiased algorithms, respectively. Since the Asymmetric Unbiased model quantifies only the forward deformation, the weight of the corresponding regularization functional is half the magnitude of that of the Symmetric Unbiased model, and hence, a weighting parameter twice as large should be used.

Figures 3-5 show the results of registering a pair of serial MRI images for one of the subjects. The deformation was computed in both directions (time 2 to time 1, and time 1 to time 2) using all three regularization methods based on mutual information matching. Results indicate the Asymmetric Unbiased and Symmetric Unbiased methods outperform Fluid method, generating more stable inverse consistent maps[15] with less variability.

MI-Fluid

MI-Asymmetric Unbiased

MI-Symmetric Unbiased

| time 2 to time 1 | time 1 to time 2 | products of Jacobians |

-40%　　0%　　+40%　　　　-40%　　0%　　+40%　　　　0.7　　1.0　　1.3
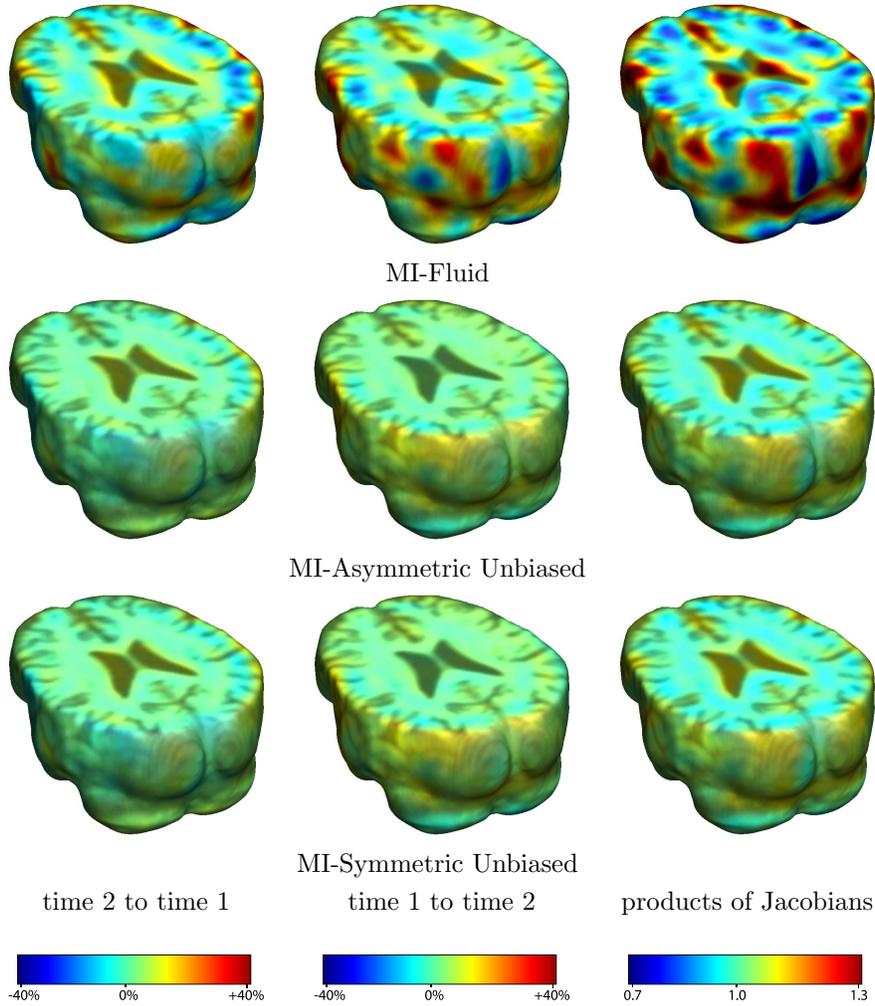
Figure 3. Nonrigid registration was performed on an image pair from one of the subjects from the ADNI Baseline study (serial MRI images acquired two weeks apart) using MI-Fluid (row 1), MI-Asymmetric Unbiased (row 2), and MI-Symmetric Unbiased (row 3) registration methods. Jacobian maps of deformations from time 2 to time 1 (column 1) and time 1 to time 2 (column 2) are superimposed on the target volumes. The unbiased methods generate less noisy Jacobian maps with values closer to 1; this shows the greater stability of the approach when no volumetric change is present. Column 3 examines the inverse consistency of deformation models. Products of Jacobian maps generated using all three models are shown, for the forward direction (time 1 to time 2) and backward direction (time 2 to time 1). For the mutual information-based unbiased methods, the products of the Jacobian maps are less noisy, with values closer to 1, showing better inverse consistency.

In Table 1, we compared MI-Fluid and MI-Symmetric Unbiased methods, conducting a within-subject paired t test inside the ICBM mask for each of the ten subjects. In this case, $p < 0.0001$ for all subjects, indicating that the Symmetric Unbiased registration, when coupled with mutual information matching cost functional, produces more reproducible maps with less variability.

Figure 6(a,b,c) shows the mean Jacobian maps of ten subjects obtained using Fluid, Asymmetric Unbiased, and Symmetric Unbiased registration algorithms coupled with mutual information matching. Jacobian maps generated using unbiased models have values closer to 1, whereas Fluid model generated noisy mean maps. Figures 7(a,b) demonstrate the Unbiased regularization technique outperforming Fluid registration with statistical significance.

To emphasize the differences between the distributions of log Jacobian values for Fluid and unbiased (both

(a) KL divergence   (b) SKL distance
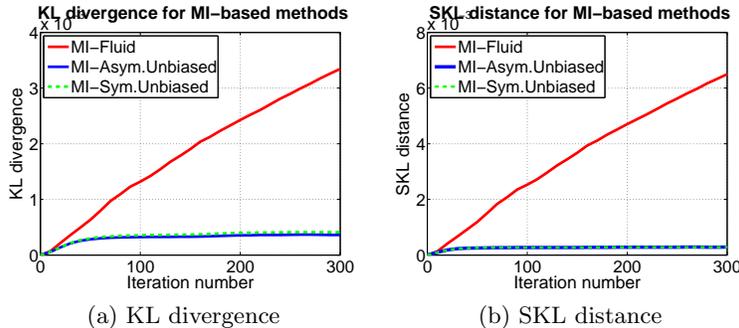
Figure 4. (a) KL divergence and (b) SKL distance per iteration are shown for the MI-Fluid (solid red), MI-Asymmetric Unbiased (solid blue), and MI-Symmetric Unbiased (dashed green) methods. For MI-Fluid, both KL and SKL measures increase. Even though the Asymmetric Unbiased method explicitly minimizes the KL distance, and the Symmetric Unbiased model minimizes the SKL distance, both the KL and SKL measures stabilize for both unbiased methods.

(a) MI-Fluid vs. MI-Asym.Unbiased   (b) MI-Fluid vs. MI-Sym.Unbiased
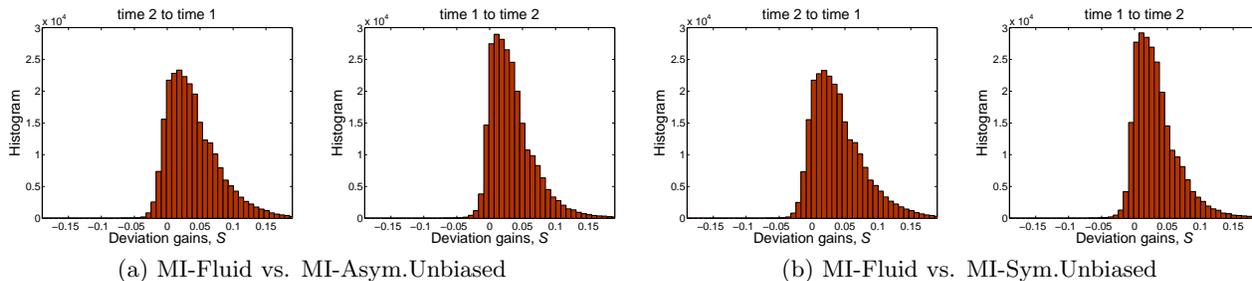
Figure 5. Histograms of voxel-wise deviation gains (a) MI-Fluid over MI-Asymmetric Unbiased and (b) MI-Fluid over MI-Symmetric Unbiased for one of the subjects, for the forward direction (time 2 to time 1) and backward direction (time 1 to time 2). The histograms are skewed to the right, indicating the superiority of Asymmetric Unbiased and Symmetric Unbiased registration methods over Fluid registration. Paired t test shows significance ($p < 0.0001$).

asymmetric and symmetric) methods, in Figure 7(c), we plotted the cumulative distribution function of the p-values in deviation gains ($S^{A,B}$). For a null distribution, this cumulative plot falls along the line $y = x$. Larger upward inflections of the CDF curve near the origin are associated with significant deviation gains, indicating that both Asymmetric Unbiased and Symmetric Unbiased methods outperform Fluid method in being less likely to exhibit structural changes in the absence of systematic biological changes.

We also analyzed a dataset we shall call the "ADNI Follow-up" phase dataset, which includes serial MRI images ($220 \times 220 \times 220$) of ten subjects acquired one year apart. As the images are now one year apart, real anatomical changes are present, which allows methods to be compared in the presence of true biological changes. In Figure 6(d,e,f), nonlinear registration was performed using Fluid, Asymmetric Unbiased, and Symmetric Unbiased methods coupled with mutual information matching. Visually, the Fluid method generates noisy mean Jacobian maps, while maps generated using unbiased methods suggest a volume reduction in gray matter as well as ventricular enlargement. Here, both Asymmetric Unbiased and Symmetric Unbiased methods perform equally well.

7. CONCLUSION

This paper introduced a novel model for multimodal image registration. The proposed method minimizes a purely information-theoretic functional consisting of mutual information matching and unbiased regularization. This work also systematically investigated the reproducibility and variability of different registration methods in TBM. We showed that Asymmetric Unbiased and Symmetric Unbiased models perform significantly better than the fluid registration technique. Although various techniques have been extensively applied to detect disease effects and monitor brain changes with TBM, this paper is the first calibration study to compare registration

Table 1. Global T statistics for all ten subjects testing whether Symmetric Unbiased registration (method B) outperforms Fluid registration (method A) when coupled with mutual information.

Subject #	1	2	3	4	5	6	7	8	9	10
$\overline{S^{A,B}}$	0.0697	0.0262	0.0399	0.0342	0.0379	0.0820	0.0853	0.0774	0.0489	0.0773
$\sigma^2_{S^{A,B}}$	0.0058	0.00077	0.0016	0.0014	0.0014	0.0071	0.0084	0.0070	0.0023	0.0053
$T_{S^{A,B}}$	455	468	501	456	505	484	460	460	504	527

models for tensor-based morphometry. We believe our results are important, as they provide greater insight into the interpretation of TBM results in the future.

APPENDIX A. DERIVATIONS OF GRADIENT OF $R(\mathbf{u})$ IN THREE SPATIAL DIMENSIONS

In this Appendix, we derive an explicit expression for $\partial_{\mathbf{u}} R(\mathbf{u})$ in (7) when $\Omega \subset \mathbb{R}^3$. Let us denote the components of vector $\mathbf{x}$ to be (x_1, x_2, x_3) and the components of vector $\mathbf{u}$ be (u_1, u_2, u_3). We also denote $\partial_j u_i = \partial u_i / \partial x_j$.

To simplify the notation, we let $J = |D\mathbf{g}| = |D(\mathbf{x} - \mathbf{u})|$. Also, denote $L(J) = L_{KL}(J) = -\log J$, when $R = R_{KL}$ and $L(J) = L_{SKL}(J) = (J - 1)\log J$, when $R = R_{SKL}$. Note that $J : \mathcal{M}_{3\times 3}(\mathbb{R}) \to \mathbb{R}$, where $\mathcal{M}_{3\times 3}(\mathbb{R})$ is the set of 3×3 matrices with real elements, and $L : \mathbb{R} \to \mathbb{R}$. Jacobian J is a function of $\partial_j u_i$, for $i, j = 1, 2, 3$, and is given by

$$J\big(\partial_1 u_1, \partial_2 u_1, \partial_3 u_1, \partial_1 u_2, \partial_2 u_2, \partial_3 u_2, \partial_1 u_3, \partial_2 u_3, \partial_3 u_3\big) = (1 - \partial_1 u_1)(1 - \partial_2 u_2)(1 - \partial_3 u_3) - \partial_1 u_2\, \partial_2 u_3\, \partial_3 u_1$$
$$- \partial_2 u_1\, \partial_3 u_2\, \partial_1 u_3 - \partial_3 u_1(1 - \partial_2 u_2)\partial_1 u_3 - \partial_2 u_1\, \partial_1 u_2(1 - \partial_3 u_3) - \partial_3 u_2\, \partial_2 u_3(1 - \partial_1 u_1).$$

We would like to minimize the functional

$$R(\mathbf{u}) = \int_\Omega L\big(J(\partial_j u_i)\big) d\mathbf{x}, \quad 1 \le i, j \le 3.$$

We find the first Euler-Lagrange equation. For some $\eta \in C^1_c(\Omega)$:

$$\frac{dR}{d\varepsilon}(u_1 + \varepsilon\eta, u_2, u_3)\big|_{\varepsilon=0} = \int_\Omega \left[\frac{dL}{dJ} \frac{\partial J}{\partial(\partial_1 u_1)} \partial_{x_1}\eta + \frac{dL}{dJ} \frac{\partial J}{\partial(\partial_2 u_1)} \partial_{x_2}\eta + \frac{dL}{dJ} \frac{\partial J}{\partial(\partial_3 u_1)} \partial_{x_3}\eta \right] d\mathbf{x}$$

$$= -\int_\Omega \left[\frac{\partial}{\partial x_1}\left(\frac{dL}{dJ} \frac{\partial J}{\partial(\partial_1 u_1)} \right) + \frac{\partial}{\partial x_2}\left(\frac{dL}{dJ} \frac{\partial J}{\partial(\partial_2 u_1)} \right) + \frac{\partial}{\partial x_3}\left(\frac{dL}{dJ} \frac{\partial J}{\partial(\partial_3 u_1)} \right) \right] \eta\, d\mathbf{x}.$$

With notation $L' = dL/dJ$, the first Euler-Lagrange equation becomes:

$$-\frac{\partial}{\partial x_1}\left(L' \frac{\partial J}{\partial(\partial_1 u_1)} \right) - \frac{\partial}{\partial x_2}\left(L' \frac{\partial J}{\partial(\partial_2 u_1)} \right) - \frac{\partial}{\partial x_3}\left(L' \frac{\partial J}{\partial(\partial_3 u_1)} \right) = 0.$$

Thus, minimizing the energy $R(\mathbf{u})$ with respect to u_1, for fixed u_2 and u_3, yields the first component of $\partial_{\mathbf{u}} R(\mathbf{u})$:

$$\partial_{u_1} R(\mathbf{u}) = \frac{\partial}{\partial x_1}\Big(\big((1 - \partial_2 u_2)(1 - \partial_3 u_3) - \partial_3 u_2\, \partial_2 u_3\big) L' \Big) + \frac{\partial}{\partial x_2}\Big(\big(\partial_3 u_2\, \partial_1 u_3 + \partial_1 u_2(1 - \partial_3 u_3)\big) L' \Big)$$

$$+ \frac{\partial}{\partial x_3}\Big(\big(\partial_1 u_2\, \partial_2 u_3 + (1 - \partial_2 u_2)\partial_1 u_3\big) L' \Big).$$

Note that $L'_{KL}(J) = -1/J$ and $L'_{SKL}(J) = 1 + \log J - 1/J$.

| Baseline Study | | | Follow-up Study | | |

(a) MI-Fluid (b) MI-A.Unbiased (c) MI-S.Unbiased (d) MI-Fluid (e) MI-A.Unbiased (f) MI-S.Unbiased

-5%   0%   +5%
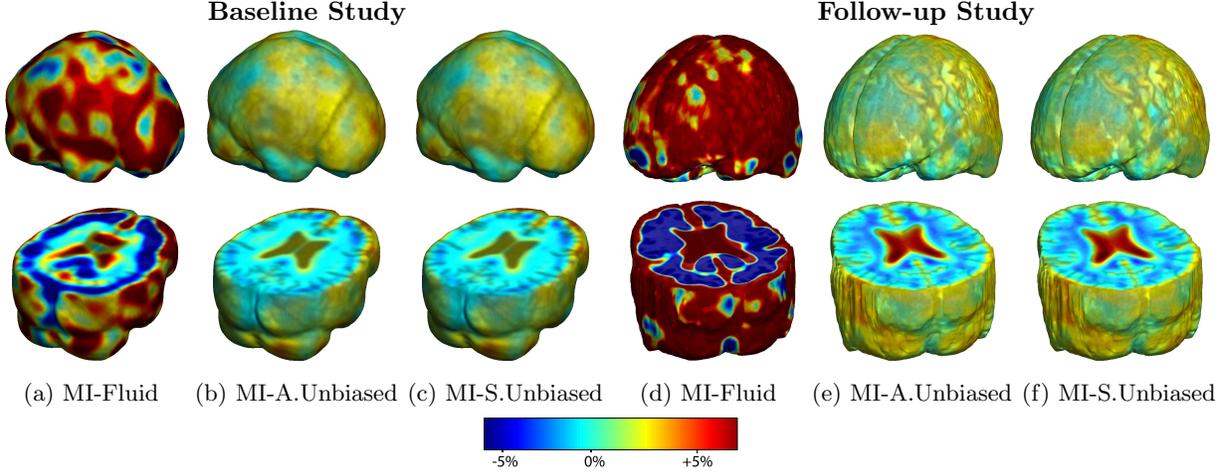
Figure 6. Nonrigid registration was performed on the ADNI Baseline study (columns 1-3) and ADNI Follow-up study (columns 4-6) using MI-Fluid, MI-Asymmetric Unbiased, and MI-Symmetric Unbiased registration methods. For each method, the mean of the resulting 10 Jacobian maps is superimposed on one of the brain volumes. Visually, MI-Fluid generates noisy mean maps. For the ADNI Baseline dataset (serial MRI images acquired two weeks apart), unbiased methods generate less noisy maps with values closer to 1. For the ADNI Follow-up dataset (serial MRI images of patients with Alzheimer's disease acquired 12 months apart), maps generated using unbiased methods suggest a volume reduction in gray matter as well as ventricular enlargement.

Similarly, the other two Euler-Lagrange equations can be found to be:

$$\partial_{u_2} R(\mathbf{u}) = \frac{\partial}{\partial x_1}\Big((\partial_2 u_3\,\partial_3 u_1 + \partial_2 u_1(1-\partial_3 u_3))L'\Big) + \frac{\partial}{\partial x_2}\Big(((1-\partial_1 u_1)(1-\partial_3 u_3) - \partial_3 u_1\,\partial_1 u_3)L'\Big)$$
$$+ \frac{\partial}{\partial x_3}\Big((\partial_2 u_1\,\partial_1 u_3 + \partial_2 u_3(1-\partial_1 u_1))L'\Big),$$
$$\partial_{u_3} R(\mathbf{u}) = \frac{\partial}{\partial x_1}\Big((\partial_2 u_1\,\partial_3 u_2 + \partial_3 u_1(1-\partial_2 u_2))L'\Big) + \frac{\partial}{\partial x_2}\Big((\partial_1 u_2\,\partial_3 u_1 + \partial_3 u_2(1-\partial_1 u_1))L'\Big)$$
$$+ \frac{\partial}{\partial x_3}\Big(((1-\partial_1 u_1)(1-\partial_2 u_2) - \partial_2 u_1\,\partial_1 u_2)L'\Big).$$

ACKNOWLEDGMENTS

REFERENCES

1. P. M. Thompson and A. W. Toga, "A framework for computational anatomy," *Computing and Visualization in Science* **5**, pp. 13–34, 2002.
2. U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quarterly of Applied Mathematics* **56**, pp. 617–694, 1998.
3. B. Avants and J. C. Gee, "Geodesic estimation for large deformation anatomical shape averaging and interpolation," *NeuroImage* **23, suppl. 1**, pp. S139–50, 2004.
4. D. Shen and C. Davatzikos, "Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration," *NeuroImage* **18**(1), pp. 28–41, 2003.
5. G. Christensen, R. Rabbitt, and M. Miller, "Deformable templates using large deformation kinematics," *IEEE Transactions on Image Processing* **5**(10), pp. 1435–1447, 1996.

(a) Voxel-wise paired t test   (b) Multiple Comparison Analysis   (c) CDF Plot
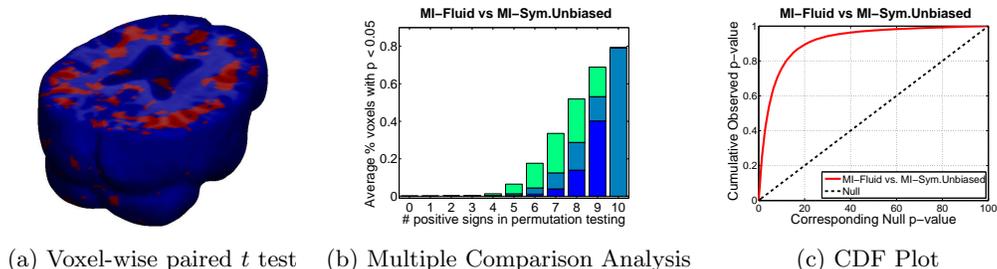
Figure 7. Statistical Assessment of Performance of MI-Symmetric Unbiased registration method in comparison with MI-Fluid method using ADNI baseline dataset. Since the results obtained using Asymmetric Unbiased method are similar to those obtained using Symmetric Unbiased method, they are not shown here. (a) Voxel-wise paired t test for the deviation gain S empirically thresholded at 2.82 ($p = 0.005$ on the voxel level with 9 degrees of freedom), showing where MI-Symmetric Unbiased registration outperform MI-Fluid registration (regions in red) with statistical significance on a voxel level. (b) Multiple Comparison Analysis using permutation testing on the deviation gain S of MI-Fluid over MI-Symmetric Unbiased. Each permutation randomly assigns a positive or negative sign to each of the 10 log-Jacobian maps. Here, results are plotted with respect to the number of positive signs (from 0 to 10) with 10 positive signs indicating the observed data. Dark blue, light blue, and green colors indicate the minimum, average, and maximum percentage of voxels with $p < 0.05$ of all possible permutations with a given number of positive signs. There is only one observation for the observed data, and thus, minimum, maximum, and average values are equal for the rightmost bar. The result indicates that out of 1024 permutations, no permutation gives a greater percentage of voxels with $p < 0.05$ than the observed data does. This indicates that unbiased regularization technique outperforms Fluid method with $p < 0.001$. (c) Cumulative distribution of p-values for the deviation gain S of MI-Fluid over MI-Symmetric Unbiased. Here, the CDF line is well above the Null line ($y = x$), indicating that Symmetric Unbiased method outperforms Fluid method (i.e. less deviation) in being less likely to exhibit structural change in the absence of biological change. Note that the interval $p \in [0, 0.05]$ is of most importance for observation.

6. M. I. Miller, "Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms," *NeuroImage* **23, suppl. 1**, pp. S19–S33, 2004.

7. A. Leow, I. Yanovsky, M.-C. Chiang, A. Lee, A. Klunder, A. Lu, J. Becker, S. Davis, A. Toga, and P. Thompson, "Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration," *IEEE Transactions on Medical Imaging* **26**(6), pp. 822–832, 2007.

8. I. Yanovsky, P. Thompson, S. Osher, and A. Leow, "Topology preserving log-unbiased nonlinear image registration: Theory and implementation," *IEEE Conference on Computer Vision and Pattern Recognition* , pp. 1–8, June 2007.

9. G. Hermosillo, C. Chefd'Hotel, and O. Faugeras, "A variational approach to multi-modal image matching," *INRIA Research report 4117* , 2001.

10. E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," *Medical Image Analysis* **7**, pp. 565–575, 2003.

11. E. Bullmore, J. Suckling, S. Overmeyer, S. Rabe-Hesketh, E. Taylor, and M. Brammer, "Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain," *IEEE Transactions on Medical Imaging* **18**, pp. 32–42, 1999.

12. T. Nichols and A. Holmes, "Nonparametric analysis of PET functional neuroimaging experiments: A primer," *Human Brain Mapping* **15**, pp. 1–25, 2001.

13. A. Leow, A. Klunder, C. Jack, A. Toga, A. Dale, M. Bernstein, P. Britson, J. Gunter, C. Ward, J. Whitwell, B. Borowski, A. Fleisher, N. Fox, D. Harvey, J. Kornak, N. Schuff, C. Studholme, G. Alexander, M. Weiner, and P. Thompson, "Longitudinal stability of MRI for mapping brain change using tensor-based morphometry," *NeuroImage* **31**(2), pp. 627–40, 2006.

14. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society* **B 57**, pp. 289–300, 1995.

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