Local Volume Change Maps in Nonrigid Registration: When Are Computed Changes Real?

Igor Yanovsky[1], Paul M. Thompson[2], Andrea D. Klunder[2], Arthur W. Toga[2], and Alex D. Leow[2]

[1] Department of Mathematics, University of California, Los Angeles, CA 90095, USA
[2] Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA 90095, USA
{yanovsky@math., thompson@loni., aklunder@loni., toga@loni., feuillet@ }ucla.edu

Abstract. Measures of brain change can be computed from sequential MRI scans, providing valuable information on disease progression. Tensor-based morphometry (TBM) creates maps of these brain changes, visualizing the 3D profile and rates of tissue growth or atrophy. In this paper, we examine the reproducibility and stability of different techniques in TBM. In particular, we compare matching functionals (sum of squared differences and mutual information), and registration schemes (unbiased large-deformation registration and viscous fluid registration) using serial MRI scans of nine normal elderly subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our results show that the unbiased large-deformation method has higher reproducibility. When coupled with unbiased registration, sum of squared differences outperforms mutual information. In contrast, when coupled with fluid registration, mutual information outperforms sum of squared difference. Moreover, the regions with least stability, due to both spatial distortion and intensity inhomogeneity, are the brain stem, thalamus, and ventricles.

Keywords: Mutual information, 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 [1-3]. 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 implementation details. In [6], we systematically examined the statistical properties of Jacobian maps (the determinant of local Jacobian operator applied to deformations), and proposed the unbiased large-deformation image registration (ULDIR). We applied this methodology to 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 when interpreting local structural changes.

This is mainly due to the fact that there is little information regarding the baseline stability, reproducibility, and variability of various nonrigid registration techniques. Thus, the overall goal of this paper is to provide quality calibrations for different non-rigid registration techniques in TBM. In particular, we compare two common cost functionals: L^2, or sum of squared differences, versus mutual information, and two regularization techniques (fluid registration versus the ULDIR technique) to decide which registration method is more reproducible, more reliable with less variability. Following analyses in the preparatory phase of ADNI [4], the foundation of calibrations is based on the assumption that, by scanning a normal control human subject within a 2-week period using the same protocol, any serial MRI scan pair should detect no structural change. Therefore, any regional structural differences detected using TBM can be assumed to be errors, upon which statistical analysis will be applied, providing baseline information on the reliability, reproducibility and variability of different registration techniques.

At this point, we would like to motivate the ULDIR approach, a nonrigid registration technique first proposed by our group in [6], that generates unbiased Jacobian statistics. This method couples the computation of deformations with statistical analyses on the resulting Jacobian maps. As a result, the ULDIR method realizes "correct" deformations by penalizing any bias in the corresponding statistical maps. In the following section, we briefly review the mathematical foundations of this approach.

2 Method

We first define both T and S, on an image domain Ω, as the two images to be registered. Let us also assume, without loss of generality, that the volume of this domain is 1, i.e., $|\Omega| = 1$. We seek to estimate a transformation h such that S matches T when deformed by h. In this paper, we will restrict this mapping to be differentiable, one-to-one, and onto from the image domain onto itself.

The ULDIR method solves for the deformation h minimizing the following combined cost, consisting of the image matching functional C (dependent on T and S, as well as the deformations h and its inverse), and the regularizing symmetric Kullback-Leibler distance (KL) term (scaled by λ):

$$\arg\min_{h \in H} C(T, T \circ h^{-1}, S, S \circ h) + \lambda \left(KL(pdf_h, pdf_{id}) + KL(pdf_{id}, pdf_h) \right). \quad \textbf{(1)}$$

Here the probability density functions (PDFs) are defined as follows

$$pdf_h(x) = |Dh(x)|, \quad pdf_{h^{-1}}(x) = |Dh^{-1}(x)|, \quad pdf_{id}(x) = 1. \quad \textbf{(2)}$$

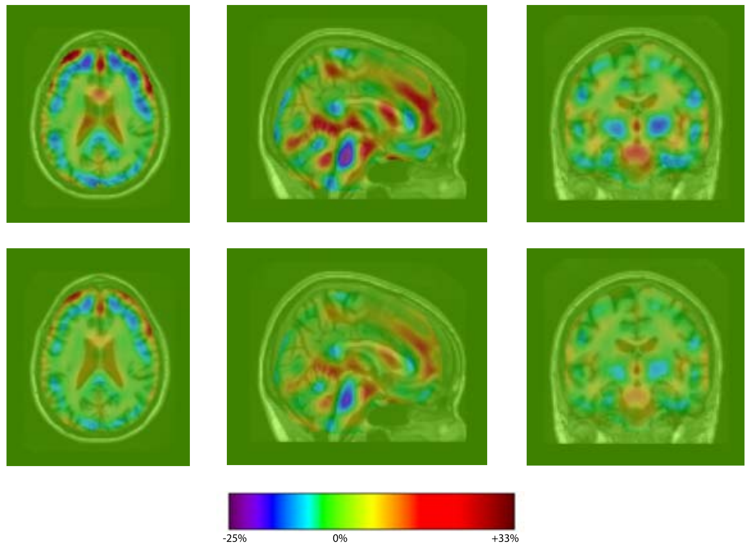


Fig. 1. Nonrigid registration was performed on the serial MRI images (acquired two weeks apart) of nine normal elderly subjects using fluid registration (row 1) and ULDIR (row 2), both driven by mutual information. The mean of the resulting nine Jacobian maps is superimposed on one of the brain volumes. Columns depict slices in axial, sagittal, and coronal planes. ULDIR generates a less noisy mean Jacobian map with values closer to 1. For both deformation models, regions with least stability, due to both spatial distortion and intensity inhomogeneity, are brain stem, thalamus, and ventricles.

The regularization term is linked to statistics on Jacobian maps as follows

$$\begin{aligned} & KL(pdf_h, pdf_{id}) + KL(pdf_{h^{-1}}, pdf_{id}) = KL(pdf_h, pdf_{id}) + KL(pdf_{id}, pdf_h) \\ & = KL(pdf_{id}, pdf_{h^{-1}}) + KL(pdf_{id}, pdf_h) = KL(pdf_{id}, pdf_{h^{-1}}) + KL(pdf_{h^{-1}}, pdf_{id}) \\ & = \int \left(|Dh(x)| - 1 \right) \log |Dh(x)| \, dx = \int \left(|Dh^{-1}(x)| - 1 \right) \log |Dh^{-1}(x)| \, dx. \end{aligned} \quad \textbf{(3)}$$

In this study, the matching functional C takes two forms: L^2 (the sum of squared differences) and MI (mutual information). We refer readers to [6] for more implementation details.

Based on our own approach in [6], we observe that, given no detectable 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 values closer to 1 (i.e., smaller Jacobian deviation translates to better methodology). Mathematically speaking, one way to test the performance is to consider the deviation map *dev* of the Jacobian away from one, defined at each voxel as $dev(x) = |J(x) - 1|$.

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

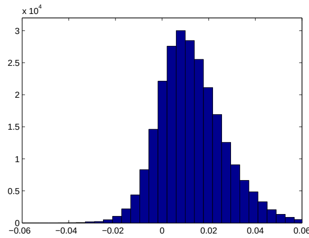


Fig. 2. Histogram for voxel-wise deviation gains (fluid registration over ULDIR), with mutual information as matching functional, for subject 3. The histogram is skewed to the right, indicating the superiority of ULDIR over fluid registration. Paired *t* test shows significance ($p < 0.0001$).

In this study, two distinct types of t-tests are used. Firstly, for each subject, we pooled all voxels inside the region of interest, as defined by the ICBM whole brain mask, and a paired *t* test was conducted to determine whether two methods differ significantly inside the whole brain (for each subject). Secondly, across subjects, we also computed 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: $S^{A,B}(x) > 0$.

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

$$T_{S^{A,B}}(x) = \frac{\sqrt{n} \cdot \overline{S^{A,B}(x)}}{\sigma_{S^{A,B}}(x)}; \qquad \left(\sigma_{S^{A,B}}(x)\right)^2 = \frac{\sum_i \left(S_i^{A,B}(x) - \overline{S^{A,B}(x)}\right)^2}{n-1}, \tag{4}$$

Table 1. Global T statistics for all nine subjects testing whether ULDIR (method B) outperforms fluid registration (method A) when coupled with mutual information.

subject #	1	2	3	4	5	6	7	8	9
$\overline{S^{A,B}}$	0.0287	0.00858	0.0129	0.0121	0.0135	0.0230	0.0386	0.0308	0.0315
$\sigma^2_{S^{A,B}}$	0.00110	0.000154	0.000218	0.000310	0.000304	0.00164	0.00495	0.00130	0.00121
$T_{S^{A,B}}$	430	343	435	341	385	282	272	425	449

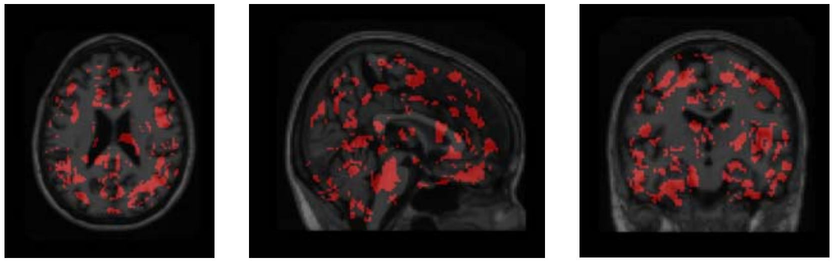


Fig. 3. Voxel-wise paired *t* test empirically thresholded at 3.35 (p = 0.005 on the voxel level with 8 degrees of freedom), showing where ULDIR outperforms fluid registration (regions in red) with statistical significance on a voxel level. In contrast, there are no voxels with T value smaller than -3.35, indicating that fluid registration does not outperform ULDIR at any voxel. Hence, the visualization of voxel-wise paired *t* test with a threshold of -3.35 is omitted.

thus follows the student's t-distribution under null and can be used to determine the p-value of the null hypothesis. If the alternative hypothesis is accepted, we confirm that sequence B outperforms A at point *x*. Otherwise, we would rank A and B equally if the null hypothesis is accepted.

3 Results

In this section, we tested the ULDIR model and compared the results to those obtained with the fluid registration model [5]. In order to obtain a fair comparison, re-gridding was not employed. Of note, re-gridding 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.

In our first study, we maximized mutual information for both models (denoted as MI-fluid and MI-ULDIR). Nine pairs of MRI volumes from the ADNI dataset were used. Each of the nine pairs of scans is 128 by 160 by 128 and subjects were re-scanned identically within a 2-week period. A uniform value of $\lambda = 0.5$ was used for all nine deformations using the ULDIR algorithm.

An example of the histogram of deviation gains (fluid registration over ULDIR) for subject 3 is shown in **Figure 2.** Here, the histogram is skewed to the right, indicating the superiority of ULDIR over fluid registration when coupled with mutual information.

In **Figure 3**, we plot 3D voxel-wise paired T statistics, with an empirical threshold of 3.35 (p = 0.005 on the voxel level with 8 degrees of freedom) to show statistical significance.

In our second experiment, we deformed one of the volumes (subject 3) using the fluid and ULDIR models, coupled with L^2 as the intensity matching term. When

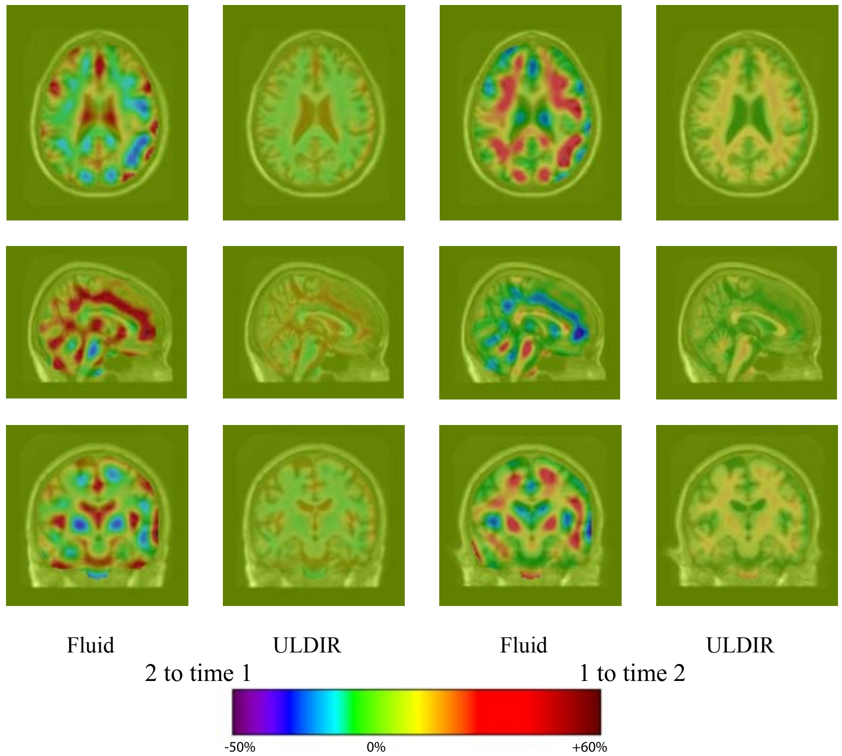
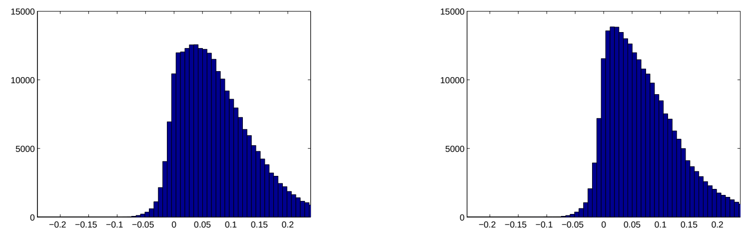


Fig. 4. Deformations from time 2 to time 1 (columns 1 and 2) and time 1 to time 2 (columns 3 and 4) are performed for one of the subjects (subject 3) using the fluid registration (columns 1 and 3) and ULDIR (columns 2 and 4). Both the fluid and ULDIR registrations are coupled with the L^2 matching functional. Jacobian maps of deformations are superimposed on the deformed volumes. Rows depict slices in axial, sagittal, and coronal planes. The ULDIR method generates less noisy Jacobian maps with values closer to 1, which shows the increased stability of the approach when no volumetric change is present.

appropriate, we will refer to these models as L^2-fluid and L^2-ULDIR, respectively. The deformation was performed in both directions (time 1 to time 2, and time 2 to time 1). The parameter λ = 500 was used for both deformations using ULDIR algorithm. The visual difference in results between ULDIR and the fluid registration, when coupled with the L^2 cost functional, is especially noticeable with confirmed statistical significance (**Figures 4 and 5**). We also visually assessed the inverse consistency by concatenating forward and backward Jacobian maps (in ideal situation, this operation should yield identity) in **Figure 6**. Again, we observe noticeable visual difference between the results obtained with ULDIR and fluid registration.

Lastly, we compared the four different registration methods (L^2-ULDIR, MI-ULDIR, L^2-fluid, and MI-fluid) using subject 3. We again conducted *t* tests on the voxel-wise deviation gains for all voxels inside the ICBM brain mask (246,191 voxels are inside the mask). Contrary to our original hypothesis that mutual information



(a) (b)

Fig. 5. Histograms of deviation gains (fluid registration over ULDIR) using L^2 as the matching functional (for subject 3) for (a) forward direction (time 1 to time 2), and (b) backward direction (time 2 to time 1). The histogram is skewed to the right, indicating the superiority of ULDIR over fluid registration. Paired *t* test shows significance in both cases ($p < 0.0001$).

outperforms L^2 regardless of regularization, we showed, surprisingly, that L^2-ULDIR outperforms MI-ULDIR ($t = 170$, $p < 0.0001$), and MI-fluid outperforms L^2-fluid ($t = 367$, $p < 0.0001$). In other words, mutual information performs better only when coupled with fluid registration, while it actually performs worse than L^2 when ULDIR is used.

To explain this seemingly paradoxical result, we postulate that by less constraining the deformations (i.e., fluid registration), assuming intensity 1-to-1 correspondence (i.e., matching using L^2) may lead to local oscillations of deformation maps, as minimizing L^2 forces a local search for smallest intensity differences. The result of this is a Jacobian map with locally extreme values, translating to spurious signals and, in our case, less reproducibility. On the other hand, the ULDIR method eliminates local oscillations, thus allowing for a globally better matching when intensity 1-to-1 correspondence can be assumed (i.e., when L^2 is applicable).

4 Conclusion

This paper systematically investigates the reproducibility and variability of different methodologies in TBM. We showed that the unbiased registration technique performs significantly differently than the fluid registration technique. The results also indicated that, when applied to serial scans obtained using the same protocol, the sum of squared differences yields maps with less spurious signals when coupled with unbiased registration. However, when coupled with fluid registration, the sum of squared differences performs less favorably.

Although various techniques in TBM have been extensively applied to detect disease effects and monitor brain changes, this paper is the first baseline calibration study that compares registration models in tensor-based morphometry. We believe our results are important, as they provide greater insight into the interpretation of TBM results in the future.

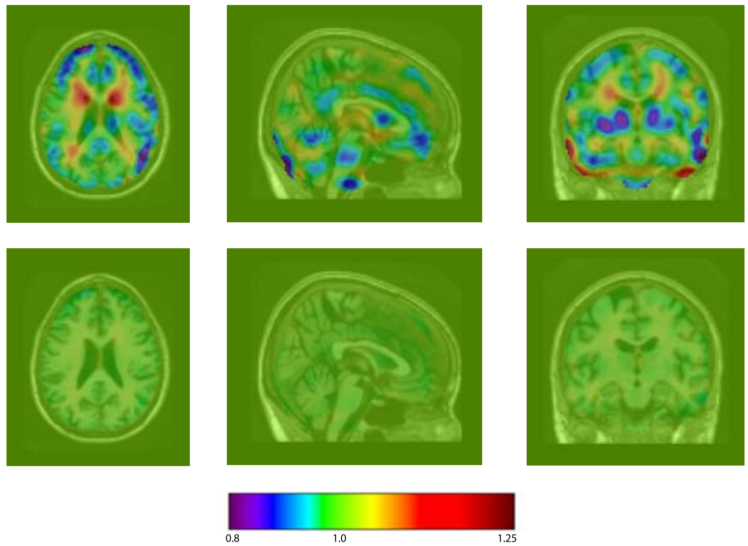


Fig. 6. This figure examines the inverse consistency of two different deformation models in tensor-based morphometry. Products of Jacobian maps generated using fluid registration (row 1) and ULDIR (row 2) for forward direction (time 1 to time 2) and backward direction (time 2 to time 1) are shown. Both the fluid and ULDIR registrations are driven by L^2 intensity matching functional. For the ULDIR method, the visualization of the product of Jacobian maps is less noisy with values closer to 1, showing better inverse consistency.

References

1. P. M. Thompson and A. W. Toga, "A framework for computational anatomy," Computing and Visualization in Science, vol. 5, pp. 13-34, 2002.
2. U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," Quarterly of Applied Mathematics, vol. 56, pp. 617-694, 1998.
3. G. E. Christensen, R. D. Rabbitt, and M. I. Miller, “Deformable templates using large deformation kinematics,” IEEE Transactions on Image Processing, vol. 5, no. 10, pp. 1435-1447, 1996.
4. A. D. Leow, A. D. Klunder, C.R. Jack, A.W. Toga, A.M. Dale, M.A. Bernstein, P.J. Britson, J.L. Gunter, C.P. Ward, J.L. Whitwell, B. Borowski, A. Fleisher, N.C. Fox, D. Harvey, J. Kornak, N. Schuff, C. Studholme, G.E. Alexander, M.W. Weiner, P.M. Thompson, For the ADNI Preparatory Phase Study (2005). “Longitudinal Stability of MRI for Mapping Brain Change using Tensor-Based Morphometry,” NeuroImage, 31(2):627-40, 2006.
5. 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, vol. 7, pp. 565-575, 2003.
6. A. D. Leow, I. Yanovsky, M.-C. Chiang, A. D. Lee, A. D. Klunder, A. Lu, J. T. Becker, S. W. Davis, A. W. Toga, P. M. Thompson, “Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration,” IEEE Transactions on Medical Imaging, 26(6): 822-832, 2007.