Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2017 ; 2017: 873–877. doi:10.1109/ISBI.2017.7950655.

THE IMPACT OF MATCHING FUNCTIONAL ON ATROPHY MEASUREMENT FROM GEODESIC SHOOTING IN Diffeomorphisms

Greg M. Fleishman and

UCLA Bioengineering, 420 Westwood Plaza, 5121 Engineering V, UCLA, CA 90095-1600

Paul M. Thompson

USC, Imaging Genetics Center, 4676 Admiralty Way, 2nd floor, Marina del Rey, CA 90292

Abstract

Longitudinal registration has been used to map brain atrophy and tissue loss patterns over time, in both healthy and demented subjects. However, we have not seen a thorough application of the geodesic shooting in diffeomorphisms framework for this task. The registration model is complex and several choices must be made that may significantly impact the quality of results. One of these decisions is which image matching functional should drive the registration. We investigate four matching functionals for atrophy quantification using geodesic shooting in diffeomorphisms. We check if the choice of matching functional has an impact on the correlation of atrophy scores with clinical variables. We also check the impact of matching functional choice on estimates of the N80 sample size for hypothetical clinical trials that test for slowing of brain atrophy. We find that the mutual information function, which has primarily been used for linear and multi-modal registration, achieves comparable correlation with clinical variables to other matching functionals while yielding better sample size estimates.

Index Terms

LDDMM; geodesic shooting; correlation coefficient; mutual information; longitudinal registration

1. INTRODUCTION

Nonlinear image registration represents the shape change evident in a pair of anatomical images as a displacement vector field. The LDDMM (large deformation diffeomorphic metric mapping) framework for diffeomorphic registration [1] restricts the displacement field to be a diffeomorphism, which is constructed by integrating the flow of a time-dependent velocity field. Consequently, LDDMM constructs a path within a space of diffeomorphisms; Beg et al. show that at optimality this path is a geodesic. In the geodesic shooting in diffeomorphisms (GSiD) formulation, rather than optimizing the entire velocity flow, only the initial momentum of the flow is considered a free variable, and the path is constructed by integrating the appropriate Euler-Poincaré differential equation (EPdiff) [2, 3, 4, 5, 6]; this guarantees the path will be a geodesic throughout optimization. Further, as shown by Miller et al. [2], at optimality the initial momentum is proportional to the moving

image gradient. Hence, the objective of GSID is to find a scalar momentum field that parameterizes an optimal matching between the given images.

Like most registration methods, GSID requires an image matching functional that takes two images as input and returns a scalar value quantifying how well aligned the images are. Many matching functionals have been proposed in the literature [7]. Generally speaking, the primary difference between them is the extent to which they attempt to normalize unwanted features from the image intensities. For example, the sum of squared differences functional makes no attempt to account for meaningless image intensity differences, and hence would be a poor choice for noisy or cross-modality registrations, whereas the mutual information functional is invariant to some transformations between the image intensity histograms.

A great deal of the theoretical work on LDDMM and GSID is agnostic about the choice of matching functional, and we have not seen an evaluation of the impact of matching functional choice on the quality of measurements in the GSID context. In particular we are interested in the application of GSID to longitudinal MRI time-series of the brain to quantify atrophy. In that case, deformations are often very low amplitude even relative to the spatial resolution of the images. However, atrophy of as little as 5% in critical brain areas can be associated with significant clinical effects [8]. Hence, it is crucial to measure atrophy, and therefore longitudinal deformations, with the highest degrees of accuracy and precision possible. In such a case, the selection of image matching functional may significantly impact our ability to capture atrophy evident in the image time-series.

2. METHODS

2.1. GSID

A complete discussion of the GSID model is beyond the scope of this paper; for a thorough discussion of the following equations, please see [3]. For a moving image I and fixed image J , the GSID objective function is:

$$E(P_0) = \frac{1}{\sigma^2} \langle P_0 \nabla I, K(P_0 \nabla I) \rangle_{L_2} + \mathcal{D}(I \circ \phi_1^{-1}, J) \quad (1)$$

where $\mathcal{D}(\cdot, \cdot)$ is a functional taking two images as input and returning a scalar value that quantifies how well the images are matched. Equation (1) must be minimized with respect to the initial scalar momentum field P_0 . The initial momentum P_0 provides the remaining initial condition for the EPdiff equation(s), which govern the time evolution of the momentum and moving image:

$$\begin{cases} \partial_t I + \nabla I \cdot v = 0 \\ \partial_t P + \nabla \cdot (Pv) = 0 \\ v + K(P \nabla I) = 0 \end{cases} \quad (2)$$

The third equation relates momentum and velocity, where K plays the role of an inertia; K is a smoothing kernel and $K(w)$ is taken to mean the convolution of vector field w with K . The path of diffeomorphisms ϕ_t is constructed from the velocity flow $v(x, t)$ according to the ODE:

$$\begin{cases} \partial_t \phi_t = v_t(\phi) \\ \phi_0 = \text{Id} \end{cases} \quad (3)$$

this yields the geodesic path of diffeomorphisms ϕ_t , where the end point ϕ_1 is used to match I and J . The matching residual $\nabla_{\phi_1} \mathcal{D}(\cdot, \cdot) / \nabla I$, which resides in the coordinate system of J , must be brought back to the coordinate system of I while respecting the geodesicity of the whole path ϕ_t . This is done by integrating the adjoint system backwards in time with initial conditions $\hat{I}_1 = \nabla_{\phi_1} \mathcal{D}(\cdot, \cdot) / \nabla I$ and $\hat{P}_0 = 0$:

$$\begin{cases} \partial_t \hat{I} + \nabla \cdot (v \hat{I}) + \nabla \cdot (P \hat{v}) = 0 \\ \partial_t \hat{P} + v \cdot \nabla \hat{P} - \nabla I \cdot \hat{v} = 0 \\ \hat{v} + K(\hat{I} \nabla I - P \nabla \hat{P}) = 0 \end{cases} \quad (4)$$

The solution of system (4) provides the gradient of equation (1) with respect to the initial momentum P_0 enabling optimization by some form of gradient descent.

2.2. Matching Functionals

Let $I(x)$ and $J(x)$ be two images such that $x \in \Omega \subset \mathbb{R}^3$ with $\partial\Omega$ the boundary of Ω and $I(\partial\Omega) = J(\partial\Omega) = 0$. We review the functional form, interpretation, and some implementation details (including any free parameters) for four different matching functionals. We do not derive the gradients of the functionals; for those details see [9].

Sum of squared differences—This is the simplest functional:

$$SSD(I, J) = \|I - J\|_{L_2}^2 = \int_{\Omega} (I(x) - J(x))^2 dx \quad (5)$$

The $SSD(\cdot)$ functional considers the input images elements of a Euclidean vector space; it deals directly with the input image intensities, and has no free parameters.

Global Correlation Coefficient—Let $\hat{I}(x) = I(x) - \frac{1}{|\Omega_I^*|} \int_{\Omega_I^*} I(x) dx$ (where Ω_I^* is the support of $I(x)$ and $|\cdot|$ denotes volume); i.e. $\hat{I}(x)$ is $I(x)$ adjusted such that the mean intensity value over its support is 0. Let $\hat{J}(x)$ be defined similarly. The global correlation coefficient is then:

$$GCC(I, J) = \frac{COV(I, J)^2}{VAR(I) \times VAR(J)} = \frac{\left(\int_{\Omega_I^* \cap \Omega_J^*} \hat{I}(x) \hat{J}(x) dx \right)^2}{\int_{\Omega_I^*} \hat{I}(x)^2 dx \int_{\Omega_J^*} \hat{J}(x)^2 dx} \quad (6)$$

The GCC ranges from 0 for distinct images of random noise to 1 for images that differ only by a linear mapping of the image intensities. Due to this invariance, GCC is more robust to global confounds of the image intensities that might occur due to scanner drift (for images taken at different times) or scanner differences (for images taken at different sites). GCC also has no free parameters.

Local Correlation Coefficient—This functional is the application of the GCC formula to all patches of a fixed window size in the image support. That is, if w_x is a window centered at x and x' is a coordinate local to w_x then the LCC is:

$$LCC(I, J) = \int_{\Omega_I^* \cap \Omega_J^*} GCC[I(w_x), J(w_x)] dx = \int \frac{\left(\int_{w_x} (I(x') - \hat{I}_{w_x}(x)) (J(x') - \hat{J}_{w_x}(x)) dx' \right)^2}{\int_{w_x} (I(x') - \hat{I}_{w_x}(x))^2 dx' \int_{w_x} (J(x') - \hat{J}_{w_x}(x))^2 dx'} dx \quad (7)$$

where $\hat{I}_{w_x}$ and $\hat{J}_{w_x}$ are mean filtered images with window size w . As opposed to GCC , LCC accounts for local rather than global image intensity statistics. This makes LCC more robust to nonlinear transformations of the image intensity histogram, which might occur under various circumstances including if one or more intensity gradients or confounds due to a large nonlinear field inhomogeneity are present.

Computing the LCC requires mean filtering both images, which can be efficiently implemented using summed area tables (faster than FFT methods). The LCC has one free parameter, the window size w , which should be selected based on the size scale of features the registration is attempting to match.

Mutual Information—Mutual information has several equivalent definitions; we will present only one. First, let $p_I(i)$ and $p_J(j)$ be the normalized intensity histograms for images I and J and let $p_{IJ}(i, j)$ be the normalized joint intensity histogram for both images. Here, i and j are image intensity values. Then, mutual information is defined as the Kullback-Leibler divergence of the joint intensity distribution from the joint distribution under the assumption of independence:

$$MI(I, J) = \int \int p_{IJ}(i, j) \ln \left(\frac{p_{IJ}(i, j)}{p_I(i)p_J(j)} \right) di dj \quad (8)$$

$MI(I, J)$ is minimal when $I(x)$ contains no information about $J(x)$; that is, when knowing the intensity at a particular location in I tells you nothing about what intensity might be at the same location in image J . In that case, I and J are independent and $p_{IJ}(i, j) = p_I(i)p_J(j)$ and $MI(I, J) = 0$. $MI(I, J)$ is maximal when $I(x)$ fully determines $J(x)$ (and vice versa); in that

case, $p_{IJ}(i, j) = p_I(i)p_J(j) = p_J(j) = p_I(i)$ and $MI(I, J)$ reduces to $\int \mathcal{R} p_I(i) \ln \left(\frac{1}{p_I(i)} \right) di$ which is the Shannon entropy of the image.

MI requires estimation of the joint intensity distribution (the individual image distributions are then obtained by marginalizing the joint distribution). First, a number of bins must be selected in which to count the image intensities. Second, the joint distribution is constructed by Parzen-window density estimation. This can be efficiently implemented by first constructing the joint intensity histogram and then Gaussian smoothing. Hence, with this implementation, MI requires two user parameters: the number of bins and the width of the smoothing kernel.

2.3. Histogram matching

We will consider the previous four matching functionals both with and without histogram matching (HM) of the input data. HM is a nonlinear transformation of the image intensities of one image such that its histogram matches that of another. HM may be particularly appropriate for longitudinal image pairs considering anatomical structures are expected to be comparable in size and intensity, modulo any atrophy. Hence, HM of a follow-up image to a baseline image, for example, is expected to compensate for some noise and intensity differences due to scanner drift or other acquisition confounds.

3. EXPERIMENTAL RESULTS

We took 100 randomly chosen subjects from the ADNI-2 longitudinal MRI dataset - publicly available at adni.loni.usc.edu- and registered their baseline scans to their 24-month follow-up scans using GSiD. We repeated these registrations with five different choices for matching functional: *SSD*, *GCC*, *LCC* with a window size of $w = 11 \times 11 \times 11 \text{ mm}$, *LCC* with a window size of $w = 21 \times 21 \times 21 \text{ mm}$, and MI with 256 bins and an isotropic Gaussian smoothing kernel standard deviation of 2.5 bins. Further, we repeated these five experiments on the same data set but first histogram matched the follow-up image to the baseline image. Hence, we present results for a total of 10 experimental conditions. Before GSiD, the images were preprocessed according to the protocol detailed in [10, 11].

After GSiD, the Jacobian determinants of the deformations mapping the baseline to the 24 month follow-up images were moved to a common coordinate system. The Jacobian determinants were averaged in a region where the rate of atrophy is significantly associated with Alzheimer's Disease (AD; Fig. 1) to produce a scalar value atrophy score that represents the percent volume loss within the region for each subject [8]. The region was constructed from registrations of baseline to 24 month follow-up images from a completely separate data set of healthy controls and AD subjects from ADNI-1. Using the Jacobian determinant maps from those registrations, a statistical test was performed at every voxel to

test for significant differences between healthy controls and AD subjects. The region used here is made up of the voxels whose p-value was below 10^{-14} .

We first ask if the choice of matching functional significantly affects atrophy measurements. We performed a pairwise t-test between the atrophy scores for all pairs of experimental conditions. Figure 2 shows the p-values of those tests. Using a significance threshold of $0.001 \approx 0.05/45$. We find that nearly all matching functionals produce atrophy scores significantly different from atrophy scores produced by the other matching functionals with a few exceptions. In particular, it seems that the correlation based matching functionals are less likely to produce distinct measurements from each other than from the non-correlation based functionals. This may indicate that for this dataset, the local intensity statistics are sufficiently similar to the global intensity statistics, in which case *GCC* may be preferred for speed. Another important observation is that the mutual information functional produced atrophy scores that were different from all other methods, with the exception of itself when HM was included.

We also ask how the choice of matching functional impacts the correlation of atrophy scores with known clinical measures of dementia. Atrophy scores obtained by longitudinal registration have been shown to correlate with the diagnostic group of the subject and with their performance on clinical assessments of dementia. The left side of Figure 3 shows the Pearson's correlation coefficient with diagnostic group (DX corr) and mini mental state exam scores (MMSE corr). In both cases, the choice of matching functional does not seem to have a large affect on the correlation between the atrophy scores and clinical variable with one exception. The *LCC* with the larger window size correlates less with the clinical variables than the other results, something which histogram matching seems to correct.

The correlation is a measurement of the strength of a linear relationship between two variables, however it says nothing about the slope of that relationship. We also would like to assess the impact of matching functional on the distribution of atrophy scores within diagnostic groups. One measurement that captures information about these distributions is the N80 sample size estimate; in words the N80 is the estimated number of individuals required to detect a 25% reduction in the mean rate of atrophy, with 80% power, and with 95% confidence in the result. As a formula, the N80 is:

$$N80 = \frac{2\sigma^2(z_{1-0.05/2} + z_{0.8})^2}{(0.25\mu)^2} \quad (9)$$

where μ is the average atrophy score for a population, σ^2 is its standard deviation, and z_α is the value at which the cumulative standard normal distribution equals α . Substituting in the values for z reduces the N80 formula simply to $N80 = 250.88 \times (\sigma/\mu)^2$.

The N80 is a function of the breadth of the distribution of atrophy scores normalized by the average amplitude of the atrophy signal. A lower N80 indicates a stronger signal, less variance in the signal, or both. The right side of Figure 3 shows N80s for all 10 experimental conditions for each of four diagnostic groups: healthy controls (HC), early mild cognitive

impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). In general, the N80s tend to decrease for matching functionals that are increasingly invariant to confounds in the intensity matching. In particular, *MI* seems to offer some of the lowest N80 sample size estimates while retaining correlations to the clinical variables comparable to the other functionals. Finally, we acknowledge a few matching functionals which we did not include in this study, but may evaluate in future studies [12, 13].

Acknowledgments

The authors thank Dmitry Isaev for help troubleshooting our MI implementation. This work was supported, in part, by NIH grant U54 EB020 403 to the ENIGMA Center for Worldwide Medicine, Imaging & Genomics.

References

1. Faisal Beg M, Miller Michael I, Trouvé Alain, Younes Laurent. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vision*. Feb; 2005 61(2):139–157.
2. Miller, Michael I., Trouvé, Alain, Younes, Laurent. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*. 2006; 24(2):209–228. [PubMed: 20613972]
3. Vialard, François-Xavier, Risser, Laurent, Rueckert, Daniel, Cotter, Colin J. Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *Int J Comput Vision*. Apr; 2012 97(2):229–241.
4. Niethammer, Marc, Huang, Yang, Vialard, François-Xavier. Geodesic regression for image time-series. In: MICCAI. , editor. *Lecture Notes in Computer Science*. Vol. 6892. Springer; 2011. p. 655–662.
5. Ashburner, John, Friston, Karl J. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage*. 2011; 55(3):954–967. [PubMed: 21216294]
6. Singh NP, Hinkle J, Joshi S, Fletcher PT. A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. *Proceedings of the 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*. 2013:1219–1222.
7. Sarrut, David, Miguët, Serge. Similarity measures for image registration. In *First European Workshop on Content-Based Multimedia Indexing*. 1999:263–270.
8. Hua, Xue, Ching, Christopher R K., Mezher, Adam, Gutman, Boris, Hibar, Derrek P., Bhatt, Priya, Leow, Alex D., Jack, Clifford R., Jr, Bernstein, Matt, Weiner, Michael W., Thompson, Paul M. MRI-based brain atrophy rates in ADNI phase 2: acceleration and enrichment considerations for clinical trials. *Neurobiology of Aging*. 2016; 37:26–37. [PubMed: 26545631]
9. Hermosillo, Gerardo, Chef'd'hotel, Christophe, Faugeras, Olivier. Variational methods for multimodal image matching. *Int J Comput Vision*. Dec; 2002 50(3):329–343.
10. Fleishman, Greg M., Gutman, Boris A., Thomas Fletcher, P., Thompson, Paul M. *International Conference on Information Processing in Medical Imaging*. Springer; International Publishing: 2015. Simultaneous longitudinal registration with group-wise similarity prior; p. 746–757.
11. Fleishman, Greg M., Thomas Fletcher, P., Gutman, Boris A., Prasad, Gautam, Wu, Yingnian, Thompson, Paul M. Geodesic refinement using James-Stein estimators. *Mathematical Foundations of Computational Anatomy*. :60.
12. Chiang, Ming-Chang, Dutton, Rebecca A., Hayashi, Kiralee M., Lopez, Oscar L., Aizenstein, Howard J., Toga, Arthur W., Becker, James T., Thompson, Paul M. 3d pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. *NeuroImage*. 2007; 34(1):44–60. [PubMed: 17035049]
13. Leow, Alex D., Yanovsky, Igor, Chiang, Ming-Chang, Lee, Agatha D., Klunder, Andrea D., Lu, Allen, Becker, James T., Davis, Simon W., Toga, Arthur W., Thompson, Paul M. Statistical properties of jacobian maps and the realization of unbiased large-deformation nonlinear image registration. *IEEE transactions on medical imaging*. 2007; 26(6):822–832. [PubMed: 17679333]

Fig. 1.
ROI significantly associated with atrophy in AD used to compute atrophy scores

	SSD	SSD_HM	GCC	GCC_HM	LCC_W11	LCC_W11_HM	LCC_W21	LCC_W21_HM	MI	MI_HM
SSD		4.03E-06	< 5E-17	2.22E-15	< 5E-17	1.11E-16	2.21E-01	< 5E-17	< 5E-17	< 5E-17
SSD_HM			< 5E-17	8.88E-16	< 5E-17	< 5E-17	4.93E-01	< 5E-17	< 5E-17	< 5E-17
GCC				1.58E-09	2.33E-02	4.14E-06	1.04E-14	7.42E-02	< 5E-17	< 5E-17
GCC_HM					1.12E-02	1.63E-02	3.80E-12	4.34E-02	< 5E-17	< 5E-17
LCC_W11						1.04E-09	< 5E-17	3.37E-08	7.51E-11	1.76E-10
LCC_W11_HM							< 5E-17	7.00E-04	5.86E-14	1.98E-14
LCC_W21								< 5E-17	< 5E-17	< 5E-17
LCC_W21_HM									9.57E-14	4.14E-14
MI										4.31E-02

Fig. 2.

p-values for pairwise Student's t-test on atrophy scores for every pair of experimental conditions, bold indicates not significant at a threshold of $0.05/45 \approx 0.001$; all other entries are significant. w11: $11 \times 11 \times 11 \text{ mm}$ window; w21: $21 \times 21 \times 21 \text{ mm}$ window.

	DX corr	MMSE corr		HC n80	EMCI n80	LMCI n80	AD n80
SSD	0.521	-0.763		282	192	179	31
SSD_HM	0.519	-0.762		279	204	193	31
GCC	0.52	-0.729		167	135	120	32
GCC_HM	0.524	-0.74		179	173	138	30
LCC_W11	0.517	-0.709		149	127	97	20
LCC_W11_HM	0.534	-0.705		135	173	99	16
LCC_W21	0.42	-0.539		193	352	122	22
LCC_W21_HM	0.536	-0.727		125	137	101	18
MI	0.525	-0.729		124	93	116	16
MI_HM	0.528	-0.724		143	99	108	14

Fig. 3.

Left: correlation of atrophy scores with diagnostic group (DX corr) and mini mental state exam scores (MMSE corr). Right: N80 sample size estimates for healthy controls (HC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD)