ANTS: Advanced Open-Source Tools for Normalization And Neuroanatomy

Brian B. Avants, Nicholas J. Tustison, Gang Song, James C. Gee

Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Computational anatomy (CA) seeks to quantify natural variation in biological shape and function with roots that reach back to the seminal works of Charles Darwin and D'arcy Thompson. CA is currently applied to study health, disease and epidemiology and uses deformable mappings between images as a basic technique. However, there is a lack of standards and reproducibility in the field that is due, in part, to the use of proprietary software and private data. To facilitate reproducibility in CA measurements and advancement of imaging sciences, NIH has recently committed significant support to open-source data and software resources. Here, we report a recent product of this commitment: Advanced Normalization Tools (ANTS), an ITK-based toolkit for CA and related areas. The ANTS open-source library consists of a well-evaluated suite of state-of-the-art normalization and template-building tools for quantitative morphometric analysis. We highlight the prominent features of ANTS and demonstrate its utility by performing a detailed analysis on openly-available anatomically labeled brain data from the non-rigid image registration evaluation project (NIREP). The results from this analysis evidences the high level of accuracy achievable with ANTS using intensity-based registration alone. In addition, we show the significant performance gains may be achieved by coupling intensity-based image metrics and point-set-based metrics from specific, sensibly selected cortical structures.

Key words: ANTS, image registration, nueroanatomy, neuroimaging, normalization, open-source, point-set registration

1. Introduction

The rapid advancement of biological and medical imaging technologies has caused a proliferation in the development of quantitative tools for computational anatomy. The principal tools of this emerging field are meaningful deformable mappings between images whether they be driven by similarity metrics which are intensity-based, point-set based, or both. Several categories of mappings have been proposed in the literature. Of particular recent interest are diffeomorphic transformations which, by definition, preserve topology. Topology preservation is fundamental to making comparisons between objects in the natural world as such transformations permit comparisons to be made across time points in an individual's disease process or to study development patterns across a large population.

Despite the number of proposed algorithms, our limited assessment of published research mirrors the experience of many others who prefer a working paradigm of what has been referred to as *reproducible research.* As described by Dr. Kovacevic "[reproducible research] refers to the idea that, in 'computational' sciences, the ultimate product is not a published paper but, rather, the entire environment used to produce the results in the paper (data, software,etc.)." After an informal survey of 15 published papers, she found "none had code available" and "in only about half the cases were the parameters [of the algorithm] specified" (**?**). Recent discussions within the computational sciences research community, particularly among advocates of "open science," have also voiced similar concerns (**??**). In this paper, we discuss our contribution to the open-source medical image analysis research community which we call ANTS (<u>A</u>dvanced <u>N</u>ormalization <u>T</u>ools). Built on an ITK framework, this software package comprises a suite of tools for image normalization and template building based on previously published research.

Perhaps the most persuasive evidence motivating the use of our contributions discussed in this paper is the recent outcome of a large-scale comparative image registration algorithm assessment (**?**). In the largest evaluation study to date involving 14 popular non-linear registration algorithms, our Symmetric Normalization (SyN) transformation model (**?**) discussed below, was consistently one of the top two performers across all tests. Overlap and distance measures used for assessment employed three completely independent analysis methods (permutation tests, one-way ANOVA tests, and indifference zone ranking). Unlike some of the other algorithms used in this brain registration evaluation study, all of our methods (not just SyN) are available as open-source.

We first provide an overview of the different components included in ANTS, such as the available transformation models and similarity metrics offered. This is followed by an extensive experimental analysis that builds upon the

results from the recent image normalization evaluation of **?** which was limited to a single configuration of ANTS. This will provide a useful benchmark for evaluating other possible ANTS configurations.

2. Theoretical Overview of ANTS

A useful classification schema of normalization techniques is based upon the following three principal components (**??**):

- the *transformation model*,
- the *similarity (or correspondence) measures*, and
- the *optimization strategy*.

In general, image normalization is the process of finding the optimal transformation, ϕ, within a specified transformation space which maps each $\mathbf{x}$ of image $\mathcal{I}(\mathbf{x})$ to a location in image $\mathcal{J}(\mathbf{z})$ such that a specified cost function, $\mathcal{C}$, describing the similarity between $\mathcal{I}$ and $\mathcal{J}$, is minimized. A summary of available transformation models and similarity measures are provided in Table **??**. Details are given in subsequent sections.

2.1. ANTS Transformation Models

A variety of transformation models have been proposed in the literature with varying degrees of freedom (illustrated in Figure **??**). For deformable transformations, one approach is to optimize within the space of non-topology preserving, yet physically motivated transformations—an approach pioneered by Bajcsy (**?**). Elastic and related models, such as HAMMER (**?**), statistical parametric mapping (SPM) (**?**), free-form deformations (FFD) (**?**), and Thirion's demons (**?**) operate in the space of vector fields, which does not preserve topology. In other words, barring ad hoc constraints to prevent otherwise, these algorithms allow the topology to change in an uncontrolled way which makes the deformable mappings difficult to interpret in functional or anatomical studies.

These topology concerns have led to current advocation within the computational anatomy research community of diffeomorphic mappings for transformation modeling. Optimizing directly within this space has shown remarkable success in various computational anatomy studies involving longitudinal (**??**), functional (**?**), and population data (**?**). We include three such diffeomorphic algorithms in the ANTS toolbox based on previously reported research.

Regardless of current research trends, however, we recognize that selection of the transformation model is ultimately application-specific, that no single choice is optimal for all scenarios (**?**), and, therefore, the transformation model must be chosen in a principled fashion. In fact, several non-diffeomorphic algorithms performed quite well in Klein's comparative study of nonrigid registration algorithms (**?**). For this reason, we also include elastic-related methods for a panoply of transformation model options.
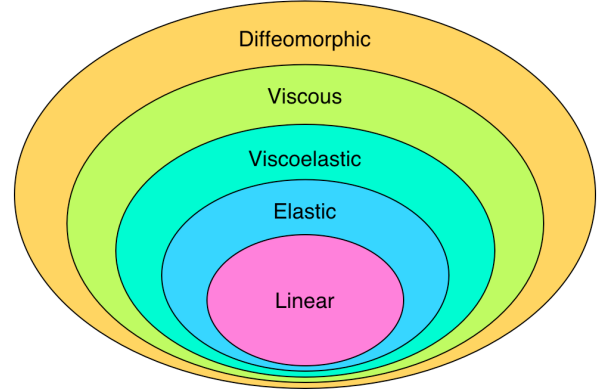

Transformation Model Hierarchy

Figure 1: Diagrammatic illustration of the transformation model hierarchy where the encompassing transformation spaces are characterized by increasing degrees of freedom.

2.1.1. Rigid and Affine Linear Transformations

Image registration strategies often begin with a linear transformation for initial global alignment which precedes a deformable transformation with increased degrees of freedom. The linear transformations available within ANTS optimize either a mean-squared difference (MSQ) or mutual information (MI) similarity metric which are optimized with respect to translation, rotation, and, in the case of affine transformations, scaling and shearing. The successive optimization of each component of the linear transformation allows for careful control over increasing degrees of freedom. ANTS also composes the affine transformation with the deformable transformation field before performing any interpolation or downsampling. In this way, ANTS normalization never requires more than a single image interpolation step and is able to always refer back to the original full-resolution image sources.

2.1.2. Vector Field Operators for Regularization

Antecedent to discussion of available deformable transformation models, we point out that deformable normalization strategies typically invoke a deformation regularization step which smooths the displacement field, $\mathbf{u}$, or velocity field, $\mathbf{v}$, or both by a linear operator such as the Laplacian or Navier-Stokes operator. One may write this regularization as a variational minimization in terms of its linear operator or in terms of a kernel function operating of the field itself, e.g., $\mathbf{u}_{smooth} = K \star \mathbf{u}_{not\ smooth}$, where $K\star$ denotes convolution with the Green's kernel, K, for the linear operator, L.

Regularization models operate on either the whole mapping ϕ or the gradient of the similarity term or both. The same regularization schema is available for both diffeomorphic and the recently proposed directly manipulated free-form deformation (DMFFD) **?**, allowing regularization of both total deformation and deformation update. Viewed from this perspective, hybrid configurations incorporat-

Category	Transformation, ϕ	Similarity Measures	Brief Description
Linear	Rigid†	MI, MSQ	translation and rotation
	Affine†	MI, MSQ	rigid, scaling, and shear
Elastic	Deformable	CC, PR, MI, MSQ, JHCT, PSE	Demons-like algorithm
	DMFFD	CC, PR, MI, MSQ, JHCT, PSE	FFD variant
Diffeomorphic	Exponential	CC,PR, MI, MSQ, JHCT, PSE	minimizes $\boldsymbol{v}(\mathbf{x})$
	Greedy SyN†	CC, PR, MI, MSQ, JHCT, PSE	minimizes $\boldsymbol{v}(\mathbf{x}, \mathbf{t})$ locally in time
	Geodesic SyN†	CC, PR, MI, MSQ, JHCT, PSE	minimizes $\boldsymbol{v}(\mathbf{x}, \mathbf{t})$ over all time

Table 1: Available transformations and similarity metrics available in ANTS. Similarity metric acronyms: CC = cross correlation, PR = probabilistic matching, MSQ = mean squared difference, MI = mutual information, JHCT = Jensen-Havrda-Charvat-Tsallis divergence, PSE = point-set expectation. ANTS also provides the inverse of those transformations denoted by the '$\dagger$' symbol.

ing discretized FFD strategies and diffeomorphisms can be combined for novel image normalization approaches.

ANTS enables a variety of choices for K including the Gaussian with varying σ and a variety of B-spline functions, both of which induce adequate regularity for normalization models used in ANTS. While additional physical operators will be included in future releases, current B-spline options provide many orders of flexibility **?**.

2.1.3. Diffeomorphic Transformations

In contrast to many transformation models which reside in the domain of vector spaces, a diffeomorphism is a differentiable mapping with a differentiable inverse (**??**). Modeling transformations with diffeomorphisms ensures certain desirable topological properties that cannot be guaranteed with other methods.

ANTS assumes the diffeomorphism, ϕ, is defined on the image domain, Ω, an maintains an affine transform at the boundary such that $\phi(\partial\Omega) = A(\mathbf{Id})$ where $A(\mathbf{Id})$ is an affine mapping applied to the identity transformation. ϕ, over time, parameterizes a family of diffeomorphisms, $\phi(\mathbf{x}, t) : \Omega \times t \to \Omega$, which can be generated by integrating a time-dependent, smooth velocity field, $\boldsymbol{v} : \Omega \times t \to \mathbb{R}^d$, through the ordinary differential equation (o.d.e.)

$$\frac{d\phi(\mathbf{x}, t)}{dt} = \boldsymbol{v}(\phi(\mathbf{x}, t), t), \quad \phi(\mathbf{x}, 0) = \mathbf{x}. \quad (1)$$

The existence and uniqueness theorem for o.d.e.'s implies that integrating Equation (**??**) generates a diffeomorphism. The deformation field yielded by ϕ is $\mathbf{u}(\mathbf{x}) = \phi(\mathbf{x}, 1) - \mathbf{x}$.

Dupuis et al. (**?**) motivated the usage of diffeomorphisms for CA by showing that the variational form

$$D(\mathcal{I}, \mathcal{J}) = \int_0^1 ||L\boldsymbol{v}||^2 dt, \ \mathcal{I}(\phi(\mathbf{x}, 1)) = \mathcal{J}(\mathbf{z}) \quad (2)$$

represents a true mathematical metric between anatomical instances $\mathcal{I}$ and $\mathcal{J}$ given an appropriate norm, L, on the velocity field, $\boldsymbol{v}$. An optimal solution, $\boldsymbol{v}^*$, minimizes the metric $D(\mathcal{I}, \mathcal{J})$ with respect to L. Dupuis (**?**) also showed

that such a solution is guaranteed to be well-posed. Intuitively, Equation (**??**) provides a sense of distance between two anatomical shapes. It also illustrates that the optimal diffeomorphic solution is analogous to finding the geodesic curve between two points in a curved space.[1]

In most real-world applications, however, a diffeomorphic path connecting the anatomical instance $\mathcal{J}$ with $\mathcal{I}$ is non-existent (due, for example, to the photometric variation or the presence/absence of a tumor in neuroanatomical images). Therefore, the following minimizing variational form is used for optimization in diffeomorphic normalization to accommodate inexact matching (**??**)

$$\boldsymbol{v}^* = \operatorname*{argmin}_{\boldsymbol{v}} \left\{ \int_0^1 ||L\boldsymbol{v}||^2 dt + \lambda \int_\Omega ||\mathcal{I} \circ \phi(\mathbf{x}, 1) - \mathcal{J}|| d\Omega \right\}. \quad (3)$$

The Euler-Lagrange equations characterizing the optimizing time-varying velocity field, $\boldsymbol{v}^*$, were derived in **?** and later used in formulating the gradient-descent optimization scheme known as *large deformation diffeomorphic metric-matching* (LDDMM) (**?**) with the similarity metric, or data term, for LDDMM being the squared intensity difference with weight λ.

To accommodate a variety of medical image normalization tasks, one typically encounters more complex intensity transfers between one anatomical instance $\mathcal{J}$ and another instance $\mathcal{I}$. Thus, ANTS enables not only a variety of similarity metric possibilities beyond the conventional squared difference metric but it also permits any number of different similarity metrics for a particular image normalization task. This leads to the following generalization of Equation (**??**):

$$\boldsymbol{v}^* = \operatorname*{argmin}_{\boldsymbol{v}} \left\{ \int_0^1 ||L\boldsymbol{v}||^2 dt + \lambda \int_\Omega \Pi_\sim(\mathcal{I}, \phi(\mathbf{x}, 1), \mathcal{J}) d\Omega \right\} \quad (4)$$

---

[1]It is important to note the similarity between the definition of curve length, $\int ||\mathcal{C}'(t)|| dt$, for the parametric curve $\mathcal{C}(t)$ and Equation (**??**). In this sense, the solution for Equation (**??**) is the geodesic diffeomorphism, where $\boldsymbol{v}$ is the tangent vector of the diffeomorphism, such that the shape distance, D, between $\mathcal{I}$ and $\mathcal{J}$ is minimized.

where $\Pi_\sim$ is a similarity metric depending on the images and the mapping and λ controls the degree of exactness in the matching. If $\Pi_\sim$ is selected as cross-correlation, then one is estimating the diffeomorphism under more robust illumination constraints, as described in **?**.

Exploiting the fact that the diffeomorphism, ϕ, can be decomposed into two components ϕ_1 and ϕ_2, **?** construct a *symmetric* alternative to Equation (**??**). This leads to the symmetric variant of Equation (**??**)

$$
\{\boldsymbol{v}_1^*, \boldsymbol{v}_2^*\} = \operatorname*{argmin}_{\boldsymbol{v_1}, \boldsymbol{v_2}} \Bigg\{
$$
$$
\int_0^{0.5} \left(||L\boldsymbol{v}_1(x,t)||^2 + ||L\boldsymbol{v}_2(x,t)||^2 \right) dt
$$
$$
+ \lambda \int_\Omega \Pi_\sim \left(\mathcal{I} \circ \phi_1(\mathbf{x}, 0.5), \mathcal{J} \circ \phi_2(\mathbf{x}, 0.5) \right) d\Omega \Bigg\}. \tag{5}
$$

The corresponding symmetric Euler-Lagrange equations are similar to (**?**). Finding $\boldsymbol{v_1}^*$ minimizes the variational energy from $t = 0$ whereas $\boldsymbol{v_2}^*$ minimizes from $t = 1$. Thus, gradient-based iterative convergence deforms $\mathcal{I}$ and $\mathcal{J}$ along the geodesic diffeomorphism, ϕ, to a fixed point midway (intuited by the notion of shape distance) between $\mathcal{I}$ and $\mathcal{J}$ thus motivating the denotation of the solution strategy as Symmetric Normalization (SyN).

Other diffeomorphic flavors have since been reported in the research literature (e.g. DARTEL (**?**) and Diffeomorphic Demons (**??**). We include three diffeomorphic transformation models for parameterizing $\phi(\cdot)$. These include Geodesic SyN, Greedy SyN, and exponential mapping. As summarized in Table **??**, each of these transformation models can utilize a host of similarity measures both individually and in mutual combination.

Geodesic SyN. Using a gradient-based optimization strategy, minimization implied by Equation (**??**), which we denote as Π, first requires a specified discretization of $t \in [0, 1]$. Calculation of the gradient of Π with respect to the diffeomorphisms ϕ_1 and ϕ_2 is performed at each of the t values which is given by

$$
\nabla\Pi(\boldsymbol{x}, t) = \frac{\partial}{\partial \phi_i} \Pi_\sim (\mathcal{I}(\phi_1^{-1}(\boldsymbol{x}, t)), \mathcal{J}(\phi_2^{-1}(\boldsymbol{x}, 1-t)) \tag{6}
$$

for $i \in \{1, 2\}$. The corresponding velocity fields at each t are then updated from the previous iteration according to

$$
\boldsymbol{v}(\mathbf{x}, t) = \boldsymbol{v}(\mathbf{x}, t) + K \star \nabla\Pi(\mathbf{x}, t). \tag{7}
$$

We then integrate Equation (**??**) using Runge-Kutta methods to yield $\phi_i(\mathbf{x}, t)$ for each $t \in [0, 1]$ and $i \in \{1, 2\}$. We cycle through these steps until convergence or iterative exhaustion.

Greedy SyN. Although the Geodesic SyN algorithm conforms most closely to the theoretical diffeomorphic foundations culminating with Equation (**??**), the computational

and memory cost is significant due to the dense-in-time gradient calculations and requisite reintegration of the diffeomorphisms after each iterative update. As a lower-cost alternative, we offer a greedy variant which performs quite well for most medical image normalization problems we have encountered. Additionally, this was the strategy used in the large-scale comparative image registration algorithm assessment of **?**.

Greedy optimization of Equation (**??**) requires calculation of the gradient of the gradient taken at the mid-point of the full diffeomorphism, i.e. at $t = 0.5$

$$
\nabla\Pi = \frac{\partial}{\partial \phi_i} \Pi_\sim (\mathcal{I}(\phi_1^{-1}(\mathbf{x}, 0.5)), \mathcal{J}(\phi_2^{-1}(\mathbf{x}, 0.5))) \tag{8}
$$

for $i \in \{1, 2\}$. $\phi_1(\mathbf{x}, 0.5)$ and $\phi_2(\mathbf{x}, 0.5)$ are then updated from the previous iteration according to

$$
\phi_i(\mathbf{x}, 0.5) = \phi_i(\mathbf{x}, 0.5) + \delta(K \star \nabla\Pi(\phi_i(\mathbf{x}, 0.5))). \tag{9}
$$

where δ is a user-specifed step parameter. The gradient is then mapped back to the origin of each diffeomorphism to update the full mapping by enforcing $\phi^{-1}(\phi(\mathbf{x}, 1)) = \mathbf{x}$.

Exponential Mapping. Ashburner introduced DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra) as a rapidly computed alternative to time parameterized diffeomorphic schemes (**?**). The key difference between a time-varying diffeomorphism and a diffeomorphism generated by an exponential mapping (**?**) is that the exponential mapping maintains only a single vector field that is constant in time.

By exponentiation of a constant velocity field, a diffeomorphism can be generated from the following o.d.e (cf Equation (**??**))

$$
\frac{d\phi(\mathbf{x}, t)}{dt} = \boldsymbol{v}(\phi(\mathbf{x}, t)), \quad \phi(\mathbf{x}, 0) = \mathbf{x}. \tag{10}
$$

Note that there is no explicit time parameter in the velocity field. Theoretically, restricting the velocity field to be constant in time reduces the size of the space that may be generated (**?**) in a way that is similar to the difference between real and rational numbers, the latter of which are sparsely distributed through the reals.

2.1.4. Vector Space Transformations

Potential mapping solutions to the image matching problem operating in vector spaces are constructed in a similar variational form as that for the diffeomorphic formulation. We write this general variational energy, Π, as

$$
\Pi(\mathcal{I}, \mathcal{J}, \phi) = \int_\Omega \left(\Pi_\sim(\mathcal{I}, \mathcal{J}, \phi)(\mathbf{x}) + \Pi_R(\phi)(\mathbf{x}) \right) d\Omega, \tag{11}
$$

where $\mathcal{I}$ and $\mathcal{J}$ are, again, the moving and fixed images, respectively, and ϕ is the transformation which maps between $\mathcal{I}$ and $\mathcal{J}$. $\Pi_\sim$ is the similarity metric and Π_R is the explicit regularization term.

Gaussian-Regularized Elastic Deformation. A simple and efficient, yet powerful image normalization algorithm is the approach known as Thirion's demons (**?**). Using an optical flow based similarity, the solution is obtained by iterating between the calculation of image forces and subsequent Gaussian regularization. In ANTS, we extend this basic approach to include the similarity metrics available for deformable registration (see Table **??**).

Directly Manipulated Free-Form Deformation. Another top performer in Klein's study (**?**) was the Image Registration Toolkit (IRTK) based on the research originally reported in **?** in which mutual information and a free-form deformation (FFD) transformation model were used to analyze breast deformation. In ANTS we provide an implementation of a variant of the well-known FFD transformation model for image registration known as *directly manipulated free-form deformation* (**?**). The DMFFD model replaces the standard FFD gradient used in **?** with an intuitive preconditioned gradient to overcome problematic energy topographies intrinsic with the traditional approach.

For n-D images, the FFD (and DMFFD) transformational model, ϕ_{FFD}, is defined as

$$\phi_{FFD} = \sum_{i_1=1}^{M_1} \cdots \sum_{i_n=1}^{M_n} \mathbf{P}_{i_1,\ldots,i_n} \prod_{j=1}^{n} B_{i_j,d_j}(u_j) \qquad (12)$$

where $\mathbf{P}_{i_1,\ldots,i_n}$ is an n-D grid of control points and $B_{i_j,d_j}(u_j)$ is the B-spline in the i_j^{th} direction of order d_j.

The gradient of the image normalization energy, Π, with respect to the control points used during gradient-based optimization is easily calculated to be

$$\frac{\partial \Pi}{\partial \mathbf{P}_{i_1,\ldots,i_n}} = \sum_{c=1}^{N_\Omega} \left(\frac{\partial \Pi_\sim}{\partial \phi} + \frac{\partial \Pi_R}{\partial \phi} \right)_c \prod_{j=1}^{n} B_{i_j,d_j}(u_j^c) \qquad (13)$$

which is the gradient used in (**?**). In contrast, the DMFFD approach uses a preconditioned gradient given by

$$\frac{\partial \Pi}{\partial \mathbf{P}_{i_1,\ldots,i_n}} = \left(\sum_{c=1}^{N_\Omega} \left(\frac{\partial \Pi_\sim}{\partial \phi} + \frac{\partial \Pi_R}{\partial \phi} \right)_c \prod_{j=1}^{n} B_{i_j,d_j}(u_j^c) \right.$$
$$\left. \cdot \frac{\prod_{j=1}^{n} B_{i_j,d_j}^2(u_j^c)}{\sum_{k_1=1}^{d_1+1} \cdots \sum_{k_n=1}^{d_n+1} \prod_{j=1}^{n} B_{k_j,d_j}^2(u_j^c)} \right)$$
$$\cdot \left(\frac{1}{\sum_{c=1}^{N_\Omega} \prod_{j=1}^{n} B_{i_j,d_j}^2(u_j^c)} \right). \qquad (14)$$

The difference between the two gradients is seen to reside strictly in terms of the B-spline shape functions which serve to normalize the DMFFD gradient in a unique fashion so as to minimize its susceptibility to hemstitching during the course of optimization.

2.2. ANTS Intensity-Based Similarity Metrics

Several intensity-based image metrics have been proposed in the literature with varying levels of performance dependent upon specific applications. We have included several of the most popular similarity metrics within ANTS. In addition, our software framework facilitates the development of other image metrics.

Both mutual information (**?**) and mean-squared difference similarity metrics are available for the linear transformations. Also included are the cross-correlation (**?**), probabilistic matching (a more aggressive version of cross-correlation), local mutual information (**??**), and mean squared difference similarity metrics for the non-linear transformation models. The parameters for the different metrics are discussed in the ANTS documentation (**?**).

2.3. ANTS Point-Set-Based Similarity Metrics

In addition to intensity-based metrics, ANTS also contains similarity metrics for registering point-sets. These include a landmark matching metric and two point-set metrics which can accommodate point-sets of different cardinality. These point-set metrics can be used alone for strict point-set registration or in conjunction with intensity-based metrics for dual intensity/point-set registration.

2.3.1. Exact Landmark Matching

Generalizing the B-spline fitting algorithm of **?**, we developed a scattered data approximation algorithm (**?**) and contributed the code to the ITK library (**?**). This code is included in ANTS and forms the basis of our exact landmark matching where this metric seeks to minimize the weighted sum of distances between corresponding landmarks using a hierarchical approach.

2.3.2. Point-Set Expectation

In (**?**), the point-set matching problem was formulated in the context of incomplete label matching but is equally applicable to the general scenario of registering point-sets not necessarily of equivalent cardinality. Given two point-sets, X and Y, the essential idea underlying the point-set expectation matching algorithm is that the optimal solution minimizes the distance between each point $y \in Y$ with its corresponding *expected* point in X.

We calculate the expected point using a Bayesian formulation and a non-parametric Parzen windowing scheme. This allows one to define the probability of the point $x \in X$ given a point $y \in Y$ as

$$\mathbf{P}(X = x | Y = y) = G(y; x, \sigma_X) \qquad (15)$$

where $G(y; x, \sigma_X)$ is a normalized Gaussian with mean x and standard deviation σ_X. The expected point $\mathrm{E}(X|y)$ is then calculated to be

$$\mathrm{E}(X|Y = y) = \sum_{j=1}^{|X|} \mathbf{P}(X = x_j | Y = y) x_j$$
$$= \frac{1}{|X|} \sum_{j=1}^{|X|} G(y; x_j, \sigma_X) x_j \qquad (16)$$

5

where $|\cdot|$ denotes cardinality. The weighted sum of distances between the points in Y and their corresponding expected points in X is calculated from Equation (**??**), i.e.

$$\text{PSE}(X,Y) = \frac{1}{|Y|} \sum_{i=1}^{|Y|} \left\| y_i - \frac{1}{|X|} \sum_{j=1}^{|X|} G(y_i; x_j, \sigma_X) x_j \right\|^2 . \tag{17}$$

2.3.3. Jensen-Havrda-Charvat-Tsallis Divergence

Recent information theoretic approaches have been proposed for point-set registration. A previous open-source contribution (**?**) generalizes the Jensen-Shannon divergence to the Jensen-Havrda-Charvat-Tsallis (JCHT) divergence which permits a fine-tuning of the divergence measure such that emphasis can vary between robustness and sensitivity for application-specic tailoring **?**.

Each point-set is represented as a PDF via a Gaussian mixture model (GMM). Assuming K point-sets denoted by $\{X_k, k \in \{1, \ldots, K\}\}$, the kth point-set is denoted by $\{x_1^k, \ldots, x_{|X_k|}^k\}$. The corresponding k^{th} PDF is calculated from the kth point-set as

$$\mathbf{P}_k(s) = \frac{1}{|X_k|} \sum_{i=1}^{|X_k|} G(s; x_i^k, C_i^k) \tag{18}$$

where $G(s; x_i^k, C_i^k)$ is a normalized Gaussian with mean x_i^k and covariance C_i^k evaluated at s. For each point, x_i, the associated weighted covariance matrix, $C_{\mathcal{K}_i}$, is given by

$$C_{\mathcal{K}_i} = \frac{\sum_{x_j \in \mathcal{N}_i, x_j \neq x_i} \mathcal{K}(x_i; x_j)(x_i - x_j)^{\text{T}}(x_i - x_j)}{\sum_{x_j \in \mathcal{N}_i, x_j \neq x_i} \mathcal{K}(x_i; x_j)} \tag{19}$$

where $\mathcal{N}_i$ is the local neighborhood of the point x_i and $\mathcal{K}$ is a user-selected neighborhood weighting kernel. We use an isotropic Gaussian for $\mathcal{K}$ with variance $\sigma_{\mathcal{K}_i}^2$ as well as a k-d tree structure for efficient determination of $\mathcal{N}_i$ (**?**). Calculation of the gradient requires the inverse of each covariance matrix. To avoid ill-conditioned covariance matrices, we use the modified covariance $C_i = C_{\mathcal{K}_i} + \sigma_n^2 I$ where I is the identity matrix and σ_n is a parameter denoting added isotropic Gaussian noise.

We designate the number of sample points generated for each of the K probability density functions as $\{M_1, \ldots, M_K\}$ and the kth set of points as $\{s_1^k, \ldots, s_{M_k}^k\}$. The JHCT divergence is then calculated using the K sets of points and the formula

$$\text{JHCT}_\alpha(\mathbf{P}_1, \ldots, \mathbf{P}_K) = \frac{1}{1-\alpha}$$
$$\left[\frac{1}{M} \left(\sum_{k=1}^{K} \sum_{j=1}^{M_k} \left[\mathbf{P}^*(s_j^k) \right]^{\alpha-1} - 1 \right) \right.$$
$$\left. + \frac{1}{N} \sum_{k=1}^{K} \frac{|X_k|}{M_k} \left(\sum_{j=1}^{M_k} \left[\mathbf{P}_k(s_j^k) \right]^{\alpha-1} - 1 \right) \right] \tag{20}$$

where

$$\mathbf{P}^*(X) = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{|X_k|} G(x; x_i^k, C_i^k), \tag{21}$$

$N = \sum_{k=1}^{K} |X_k|$, and $M = \sum_{k=1}^{K} M_k$. The prior weighting values are calculated from $\gamma_k = |X_k|/N$ such that the larger point-sets are weighted more heavily.

3. ANTS Implementation and Usage

ANTS, built upon an ITK foundation, maintains the same coding style as its base. For much of its functionality, ANTS requires ITK, necessitating the installation of ITK prior to installing ANTS. All ANTS source code is available via the online source code repository Source-Forge.[2] Binaries are also available from the same online location. For quality assurance and maintenance purposes we have established an ANTS test reporting open-source "dashboard"[3] on our lab website[4] to monitor compilation and testing of the ANTS program. A screenshot from a daily testing period is given in Figure **??**. Such a configuration facilitates reporting of problems encountered by users on a multitude of computing platforms.

Based on our experience with standard command line argument parsing packages (e.g. `getopt`), we developed our own set of classes for an intuitive command line interface. A summary of command line arguments are given in Table **??**. These ANTS argument parsing classes provide an intuitive compromise between parsers where every variable requires a unique flag and strict ordering requirements on the command line. A typical command line call to ANTS is given by

```
>ANTS 3
  --metric MSQ[fixedImage.nii,movingImage.nii,0.75]
  --metric PSE[fixedPoints.vtk,movingPoints.vtk,0.25]
  --transformation Exp[0.75]
  --regularization Gauss[3.0,0.5]
  --iterations 50x20x10x5
  --output-naming results.nii
```

The correspondence between the ANTS command line specification and the image normalization formulation with labeled point-sets X and Y illustrates the motivation for our command line interface,

$$\Pi(\mathcal{I}, \mathcal{J}, X, Y, \phi)$$
$$= \underbrace{\int_0^1 ||L\boldsymbol{v}|| dt}_{\text{-t Exp[·], -r Gauss[·,·]}} + \underbrace{\lambda_1 \int_\Omega ||\mathcal{I} \circ \phi(\mathbf{x}, t) - \mathcal{J}|| d\Omega}_{\text{-t Exp[·], -m MSQ[}\mathcal{J},\mathcal{I},\lambda_1\text{]}}$$
$$+ \underbrace{\frac{\lambda_2}{|Y|} \sum_{i=1}^{|Y|} \left\| y_i - \frac{1}{|X|} \sum_{j=1}^{|X|} G(y_i; x_j, \sigma_X) x_j \right\|^2}_{\text{-m PSE[}Y,X,\lambda_2\text{]}} . \tag{22}$$

---

[2] http://sourceforge.net/projects/advants/
[3] http://www.cdash.org
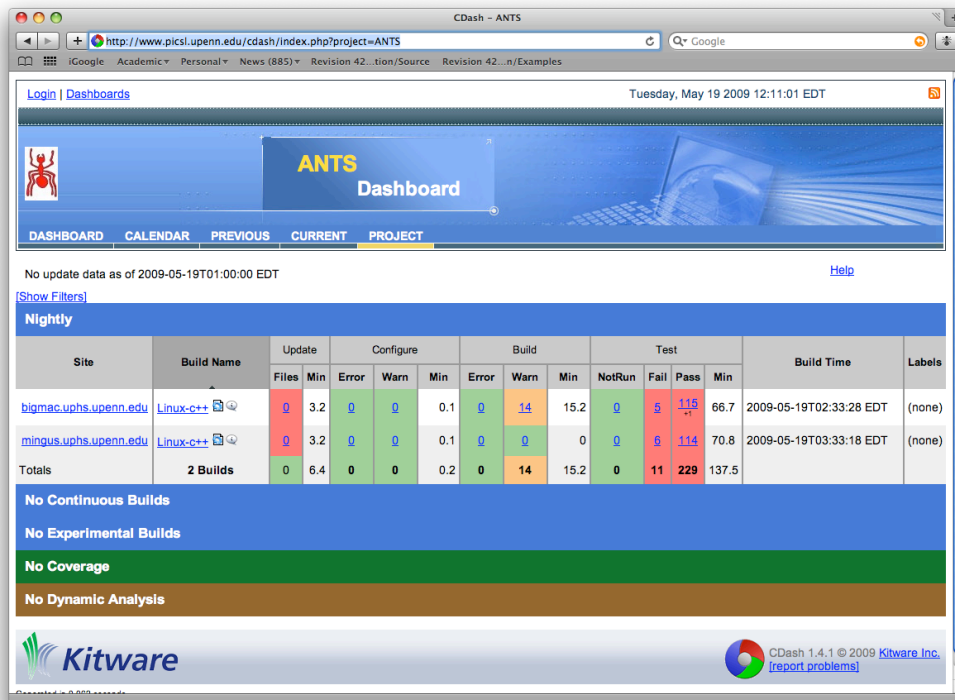[4] http://www.picsl.upenn.edu/cdash/index.php?project=ANTS

Figure 2: The ANTS dashboard, which is hosted on the PICSL website, reports daily building and testing of the ANTS software. It also allows any user to submit their own building and testing configurations to help with debugging issues and maintenance for a variety of computing platforms.

In Table **??** we give a brief summary of the arguments available for the normalizations offered by the ANTS package. This includes the corresponding variable specification. More information can be found on the ANTS website.

4. Experimental Evaluation

There are many avenues for exploration of the various components of ANTS. However, due to space constraints, we limit experimental analysis within this paper to an extension of the normalization assessment carried out by **?**. As previously mentioned, this large-scale assessment encompassed evaluation of 14 popular registration algorithms which were optimized, in terms of their parameters, by their respective authors before a thorough brain image normalization study. Although the Greedy SyN algorithm, outlined in an earlier section, was consistently one of the top two performers in Klein's study, for the benefit of the users of ANTS, we explore the other transformation model possibilities within ANTS and compare them with Greedy SyN. In terms of data, we utilize the NA0 evaluation database of the Non-Rigid Image Registration Evaluation Project (NIREP) for future comparison with evaluation studies that have been proposed by the NIREP initiative.[5]

As outlined in the introduction, in addition to the transformation, the optimization strategy and similarity metric form the image normalization schematic troika. Since our optimization strategy is limited to gradient descent, experimental analysis includes an exploration of optimal gradient steps within a sensible window where the steps are scaled according to the voxel spacing. In terms of similarity metric, we limit exploration to cross-correlation while varying the radius within reasonable values. Other metrics were not explored since the labeled brain images entail a simple intensity relationship between image pairs obviating the need for similarity metrics for more complex intensity relationships (e.g. MI) in addition to the fact that the consistently top two performers in **?** used cross-correlation.

Briefly, two experiments were performed. The first experiment consisted of a more extensive parameter search over both the transformation model space and the cross-correlation metric radius using the exhaustive pairwise combination of the eight 2-D images illustrated in Figure **??**. Based on the results of the first experiment, the parameter space was pruned and subsequently used for the registrations performed during the second experiment involving 10 randomly selected image pairs from the 16

---

[5]http://www.nirep.org/

7

	Argument	Flag	Variables	Sample Parameters
Linear	Iterations	`--linear-iterations`		N_1xN_2xN_3x...
	Similarity	`--linear-metric`	`MI,MSQ`	[N_{bins},$N_{samples}$]
	Affine or Rigid	`--do-rigid`		0
Deform.	Image Similarity	`--metric,-m`	`MI,CC,PR,MSQ`	[$\mathcal{I},\mathcal{J}$,radius]
	Point-Set Similarity	`--metric,-m`	`PSE,JHCT`	[$\mathcal{I},\mathcal{J},X,Y$]
	Iterations/Level	`--iterations,-i`		N_1xN_2xN_3x...
	Regularization	`--regularization,-r`	`Gauss,DMFFD`	[$\sigma^2_{gradient},\sigma^2_{total}$], [1x1,3x3]
	Transformation	`--transformation,-t`	`Elast,SyN,Exp`	[$\Delta_{gradient}$]
Misc.	Histogram Match $\mathcal{I},\mathcal{J}$	`--use-histogram-matching`		1
	NN Interpolation	`--use-NN`		0
	Mask Image	`--mask,-x`		`mask.nii`
	Output Naming	`--output-naming,-o`		`filename.nii`

Table 2: The various flags and variables for a variety of image registration possibilities. Additional information can be found on the ANTS website (**?**).
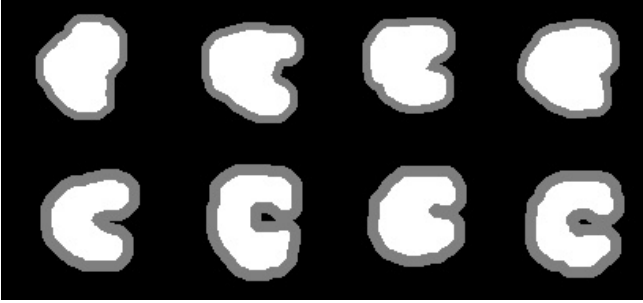

Figure 3: The eight 2-D simulated images used for the initial parameter search. Each image is of size $[102, 95]$ and are available with the ANTS source distribution.

labeled NA0 NIREP brain images.

4.1. 2-D Simulated Image Normalization Evaluation

4.2. 3-D NIREP Brain Image Normalization Evaluation

5. Discussion

Acknowledgment