

Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain

B.B. Avants^{a,*}, C.L. Epstein^b, M. Grossman^c, J.C. Gee^a

^a Department of Radiology, University of Pennsylvania, 3600 Market Street, Philadelphia, PA 19104, United States

^b Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, United States

^c Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, United States

Received 11 November 2006; received in revised form 23 May 2007; accepted 6 June 2007

Available online 23 June 2007

Abstract

One of the most challenging problems in modern neuroimaging is detailed characterization of neurodegeneration. Quantifying spatial and longitudinal atrophy patterns is an important component of this process. These spatiotemporal signals will aid in discriminating between related diseases, such as frontotemporal dementia (FTD) and Alzheimer's disease (AD), which manifest themselves in the same at-risk population. Here, we develop a novel symmetric image normalization method (SyN) for maximizing the cross-correlation within the space of diffeomorphic maps and provide the Euler–Lagrange equations necessary for this optimization. We then turn to a careful evaluation of our method. Our evaluation uses gold standard, human cortical segmentation to contrast SyN's performance with a related elastic method and with the standard ITK implementation of Thirion's Demons algorithm. The new method compares favorably with both approaches, in particular when the distance between the template brain and the target brain is large. We then report the correlation of volumes gained by algorithmic cortical labelings of FTD and control subjects with those gained by the manual rater. This comparison shows that, of the three methods tested, SyN's volume measurements are the most strongly correlated with volume measurements gained by expert labeling. This study indicates that SyN, with cross-correlation, is a reliable method for normalizing and making anatomical measurements in volumetric MRI of patients and at-risk elderly individuals.

Published by Elsevier B.V.

Keywords: Diffeomorphic; Deformable image registration; Human cortex; Dementia; Morphometry; Cross-correlation

1. Introduction

Frontotemporal dementia (FTD) prevalence may be higher than previously thought and may rival Alzheimer's disease (AD) in individuals younger than 65 years (Ratnavalli et al., 2002). Because FTD can be challenging to detect clinically, it is important to identify an objective method to support a clinical diagnosis. MRI studies of individual patients are difficult to interpret because of the wide range of acceptable, age-related atrophy in an older cohort susceptible to dementia. This has prompted MRI studies that

look at both the rate and the anatomic distribution of change (Chan et al., 2001; Fox et al., 2001; Studholme et al., 2004; Kertesz et al., 2004; Avants et al., 2005a; Ballmaier et al., 2004).

Manual, expert delineation of image structures enables *in vivo* quantification of focal disease effects and serves as the basis for important studies of neurodegeneration (Studholme et al., 2004). Expert structural measurements from images also provide the gold-standard of anatomical evaluation. The manual approach remains, however, severely limited by the complexity of labeling 256^3 or more voxels. Such labor is both time consuming and expensive to support, while the number of individual experts available for such tasks is limited. A third significant difficulty is the

* Corresponding author.

E-mail address: avants@grasp.cis.upenn.edu (B.B. Avants).

problem of inter-rater variability which limits the reliability of manual labeling (Sparks et al., 2002). While rarely available for large-scale data processing, an expert eye remains valuable for limited labeling tasks that give a basis for algorithmic evaluation.

Deformable image registration algorithms are capable of functioning effectively in time-sensitive clinical applications (Dawant et al., 2003) and high-throughput environments and are used successfully for automated labeling and measurement research tasks. One challenge is reliable performance on non-standard data, as in studies of potentially severe neurodegenerative disorders. These types of images violate the basic assumptions of small deformations and/or simple intensity relationships used in many existing image registration methods.

Diffeomorphic image registration algorithms hold the promise of being able to deal successfully with both small (Bajcsy et al., 1983; Gee et al., 1993; Gee and Bajcsy, 1999; Peckar et al., 1998; Rueckert et al., 1999; Rogelj and Kovacic, 2006; Ashburner et al., 2000) and large deformation problems (Trounev, 1998; Christensen et al., 1997; Dupuis et al., 1998; Younes, 1998; Joshi and Miller, 2000; Miller et al., 2002; Beg et al., 2005; D’Agostino et al., 2003; Lorenzen et al., 2006; Vaillant et al., 2004). State of the art methods also give full space–time optimizations, are symmetric with respect to image inputs and allow probabilistic similarity measures (Avants et al., 2005b). Inverse consistent image registration (ICIR) is an additional common alternative to diffeomorphic mapping. Inverse consistency was first introduced by Thirion as an extension to his Demons algorithm (Thirion, 1998) but was popularized by Christensen and Johnson (2001) and others (Shen and Davatzikos, 2002). Symmetric methods are distinct from ICIR in that symmetric algorithms, first, guarantee that results are identical regardless of the order of input data and, second, use exact inverse transformations guaranteed by diffeomorphisms. Inverse consistency approximates symmetry by including variational penalties in the normalization optimization algorithm. Depending on the weights of the various data, regularization and inverse consistency terms, consistency may be satisfied (or not) at the expense of the other matching criterion. Furthermore, inverse consistent algorithms use approximate inverse transformations (Christensen and Johnson, 2001). Because the inverse transformations themselves are approximate, the consistency term, as well, is compromised.

Here, we develop a novel symmetric diffeomorphic optimizer for maximizing the cross-correlation in the space of topology preserving maps. The cross-correlation measure has been used in medical image registration before (Bajcsy et al., 1983; Gee et al., 1993; Hermosillo et al., 2002) and more extensively in computer vision. However, this measure has not been investigated for the diffeomorphic case. Furthermore, it has not been used in symmetric normalization or “inverse consistent” studies. Applying our novel normalization formulation to cross-correlation provides

the advantage (or option) of symmetrizing the cross-correlation Euler–Lagrange equations. We show that these symmetric Euler–Lagrange equations can be computed with only minor additional computational costs. We then give a careful evaluation of the performance of our symmetric diffeomorphic algorithm for high dimensional normalization of elderly and neurodegenerative cortical anatomy. We compare the method to an elastic cross-correlation optimizer as well as the Demons algorithm which was shown to outperform other methods in a careful evaluation of inter-subject brain registration (Hellier et al., 2003).

2. Registration methods

2.1. Demons

Thirion’s Demons algorithm (Thirion, 1996) is known to perform well in inter-subject deformable image registration. The method uses an approximate elastic regularizer to solve an optical flow problem, where the “moving” image’s level sets are brought into correspondence with those of a reference or “fixed” template image. In practice, the algorithm computes an optical flow term which is added to the total displacement (initially zero). The total displacement is then smoothed with a Gaussian filter. The process repeats for a set number of iterations for each resolution in a multi-resolution optimization scheme. The method is freely available in the Insight ToolKit and has been optimized by the ITK community (www.itk.org).

Dawant et al. used the Demons algorithm for segmenting the caudate nucleus, the brain and the cerebellum for a morphometric comparison of normal and chronic alcoholic individuals (Dawant et al., 1999). Their evaluation of the algorithm found reasonable agreement between automated and manual labeling. They also showed results on the automated labeling of hippocampus but did not evaluate performance. Their comparison used the Dice statistic (overlap ratio):

$$S(R1, R2) = \frac{2\#(R1 \cap R2)}{\#(R1) + \#(R2)}, \quad (1)$$

which measures both difference in size and location between two segmentations, $R1$ and $R2$. The $\#(R)$ operator counts the number of pixels in the region, R . This sensitive measure varies in the range $[0, 1]$ where values greater than 0.6 for smaller structures and 0.8 for larger structures are considered good by some authors (Dawant et al., 1999; Sparks et al., 2002).

The range of acceptable values for the Dice statistic are, of course, highly dependent upon the application. Both the amount of certainty that one has in the “gold standard” dataset and, also, the specific use of the segmentations determine a reasonable operating range. Our goal, in this paper, is to use manually segmented structures as a foundation for comparing automated normalization methods. In this respect, it is relative performance (measured with respect to the Dice statistic) that is of critical value.

2.2. Symmetric diffeomorphisms

A diffeomorphism is a differentiable map with a differentiable inverse (Arnold and Khesin, 1992; Arnold, 1991). We typically restrict our solutions to the diffeomorphic space $Diff_0$ with homogeneous boundary conditions. That is, we assume that rigid and scaling transformations have been factored out and the image border maps to itself.¹ Shortest paths between elements in this space are termed *geodesic*. Diffeomorphic methods were introduced into medical computer vision (Trounev, 1998) for the purpose of providing a group theoretical, large deformation space–time image registration framework. Current developments in large deformation computational anatomy by Miller, Trounev and Younes extended the methods to include photometric variation and to use Euler–Lagrange equations (Miller et al., 2002). However, the standard version of these methods, Beg’s Large Deformation Diffeomorphic Metric Matching (LDDMM) (Beg et al., 2005), do not formulate the transformation optimization symmetrically. They are only symmetric *in theory* and their implementation requires parallel computation. Through personal communication, however, we understand that a symmetrization operator, based on the transformation Jacobian, is being included in current developments by Trounev and Younes (2005) and Younes (preprint). However, these methods do not guarantee symmetry for similarity metrics other than the intensity difference.

Our current work extends the Lagrangian diffeomorphic registration technique described in Avants et al. (2006a). This new formulation has symmetry properties required for a geodesic connecting two images, I and J , in the space of diffeomorphic transformations and guarantees symmetry regardless of the chosen similarity measure. This formulation accounts for the natural symmetry in the problem: both images move along the shape (diffeomorphism) manifold. Symmetric diffeomorphisms guarantee two properties that are intrinsic to the notion of a geodesic path: the path from I to J is the same as it is when computed from J to I , regardless of similarity metric or optimization parameters. Symmetry is required for distance estimates and makes results independent of arbitrary decisions about which image is “fixed” or “moving.”

Our method is also unique in that it guarantees sub-pixel accurate, invertible transformations in the discrete domain by directly including invertibility constraints in the optimization. While diffeomorphisms are theoretically guaranteed to be invertible, interpolation errors can lead to invertibility errors that increase linearly with the number of interpolation steps. Our solution, on the other hand, directly minimizes this error by exploiting the invertibility guaranteed by diffeomorphisms. Finally, the method is efficient enough to use on single-processor machines and in processing large datasets.

¹ Note that extending the background space of an image allows almost any diffeomorphism of the image to be captured in $Diff_0$.

We define a diffeomorphism ϕ of domain Ω , generally, for transforming image I into a new coordinate system by $\phi I = I \circ \phi(\mathbf{x}, t = 1) = I(\phi(\mathbf{x}, t = 1))$, which indicates that I is warped forward by the map defined by $\phi(\mathbf{x}, 1)$. One may also use a more standard backward warping strategy, through $\phi^{-1}(\mathbf{x}, 1)$, to achieve the same deformation. The parameters of these transformations are time, t , a spatial coordinate, \mathbf{x} , and a velocity field, $\mathbf{v}(\mathbf{x}, t)$ on Ω , which is a square-integrable, continuous vector field (Arnold, 1991). The correspondence maps, ϕ , are gained by integrating the velocity fields in time, $\phi(\mathbf{x}, 1) = \phi(\mathbf{x}, 0) + \int_0^1 \mathbf{v}(\phi(\mathbf{x}, t), t) dt$, where \mathbf{v} is indexed at $\phi(\mathbf{x}, t) = \mathbf{y}$. The distance is then $D(\phi(\mathbf{x}, 0), \phi(\mathbf{x}, 1)) = \int_0^1 \|\mathbf{v}(\mathbf{x}, t)\|_L dt$, where L defines the linear operator regularizing the velocity. The functional norm, $\|\cdot\|_L$, induces regularity on the velocity field via a linear differential operator such as $L = a\nabla^2 + b\mathbf{Id}$ (a and b are constants). Such a diffeomorphism gives a dense map in both space and time and is illustrated in Fig. 1.

A basic fact of diffeomorphisms allows them to be decomposed into two parts, ϕ_1 and ϕ_2 . We exploit this fact to define a variational energy that explicitly divides the image registration diffeomorphisms into two halves such that I and J contribute equally to the path and deformation is divided between them. Assume that \mathbf{x} indicates the identity position of some anatomy in image I and \mathbf{z} indexes the identity position of the same anatomy in image J . We assume, also, that the diffeomorphism maps homologous anatomy in these images. The prior knowledge that this diffeomorphic map should apply evenly to both images can be captured by including the constraint $D(\mathbf{Id}, \phi_1(\mathbf{x}, 0.5)) = D(\mathbf{Id}, \phi_2(\mathbf{z}, 0.5))$ directly in the formulation of the problem. The result is a method that finds correspondences with equal consideration of both images. Note that below we will derive the equations assuming intensity difference as a similarity measure, for simplicity. However, in actuality, we have a variety of statistical image similarity measures (robust intensity difference, cross-correlation, mutual information) at our disposal, as in Hermsillo et al. (2002), or employ user landmarks as in Avants et al. (2006a). After this introductory section, we will develop our method for the cross-correlation.

Define the image registration optimization time, $t \in [0, 1]$ where t indexes both ϕ_1 and ϕ_2 , though in opposite directions. The similarity seeks ϕ_1 such that $\phi_1(\mathbf{x}, 1)I = J$. Recalling the basic definition of diffeomorphisms allows us to write any geodesic through composing two parts, each of which is a geodesic (any sub-part of a geodesic is a geodesic). We apply this fact in the second step of our derivation below:

$$\begin{aligned} \phi_1(\mathbf{x}, 1)I &= J, \\ \phi_2^{-1}(\phi_1(\mathbf{x}, t), 1 - t)I &= J, \\ \phi_2(\phi_2^{-1}(\phi_1(\mathbf{x}, t), 1 - t), 1 - t)I &= \phi_2(\mathbf{z}, 1 - t)J, \\ \phi_1(\mathbf{x}, t)I &= \phi_2(\mathbf{z}, 1 - t)J, \end{aligned} \tag{2}$$

which converts the similarity term from $|\phi_1(\mathbf{x}, 1)I - J|$ to $|\phi_1(\mathbf{x}, t)I - \phi_2(\mathbf{z}, 1 - t)J|^2$. A visualization of these

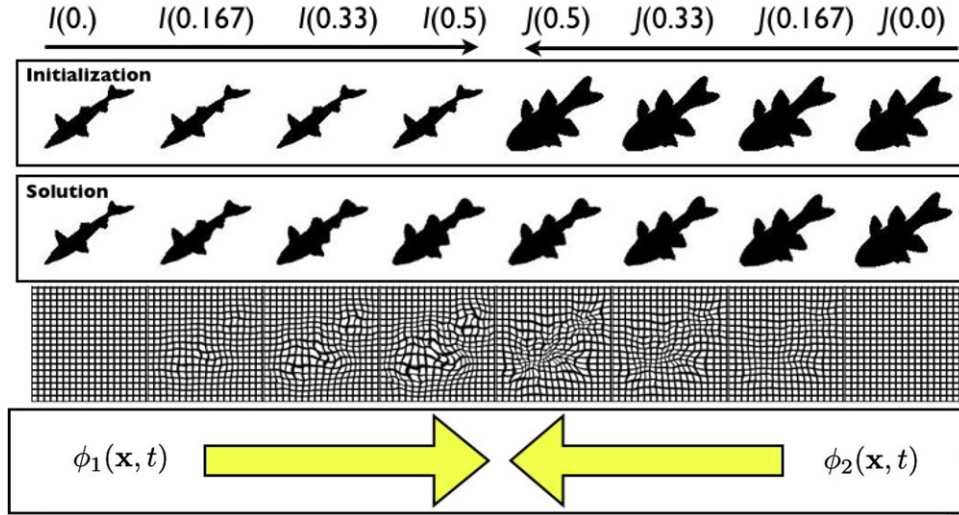


Fig. 1. An illustration of a SyN geodesic path between images. The images in the top row are the original images, I and J , at the initialization of the method. After the SyN solution converges (second row), these images deform in time along the series of diffeomorphisms that connect them. The deforming grids associated with these diffeomorphisms are shown in the bottom row. The maps, ϕ_1 and ϕ_2 are of equivalent length and map I and J to the mean shape between the images. The full path, ϕ and ϕ^{-1} , are found by joining the paths ϕ_1 and ϕ_2 . The symmetric nature of this problem (proven in Avants et al., 2006b) is due to the interchange-ability of the labels I and J in our problem formulation.

components of ϕ is in Fig. 1. The forward and backward optimization problem is then, solving to time $t = 0.5$

$$E_{\text{sym}}(I, J) = \inf_{\phi_1} \inf_{\phi_2} \int_{t=0}^{0.5} \left\{ \|v_1(\mathbf{x}, t)\|_L^2 + \|v_2(\mathbf{x}, t)\|_L^2 \right\} dt + \int_{\Omega} |I(\phi_1(0.5)) - J(\phi_2(0.5))|^2 d\Omega.$$

Subject to each $\phi_i \in \text{Diff}_0$ the solution of:

$$d\phi_i(\mathbf{x}, t)/dt = \mathbf{v}_i(\phi_i(\mathbf{x}, t), t) \text{ with } \phi_i(\mathbf{x}, 0) = \mathbf{Id} \text{ and } \phi_i^{-1}(\phi_i) = \mathbf{Id}, \phi_i(\phi_i^{-1}) = \mathbf{Id}. \quad (3)$$

Minimization with respect to ϕ_1 and ϕ_2 provides the *symmetric* normalization (SyN) solution and also solves a 2-mean problem. Landmarks may also be included, as in the Lagrangian Push Forward method (Avants et al., 2006a), by dividing the similarity term, as done with the image match terms above. The constraint $D(\phi_1(\mathbf{x}, 0.5)) = D(\phi_2(\mathbf{x}, 0.5))$ is built into the fact that we integrate the solution from 0 to 0.5. Because the velocities are of (approximately) constant arc length and are, at each iteration, of exactly the same length, the length of ϕ_1 , integrated over $[0, 0.5]$, is equivalent to the length of ϕ_2 integrated over $[0, 0.5]$.

As noted in Section 1, this method is quite distinct from inverse consistent image registration (ICIR) (Johnson and Christensen, 2002). ICIR uses vector fields, h and g , to define correspondence from I to J and J to I , respectively. Therefore, in total, ICIR uses four vector fields in its approach to normalization, h , g , h^{-1} and g^{-1} . The inverses of these fields are not guaranteed to exist (as the optimization is not performed in the diffeomorphic space) and no exact method is used to compute the inverses. Instead, judging from the brief discussion in Johnson and Christensen

(2002), the inverse is only guaranteed to be exact at a few points in the domain and the inversion algorithm itself is not well-specified or exact. This inexactness means that the variational term measuring the difference between two vector fields, $\|h(\mathbf{x}) - g^{-1}(\mathbf{x})\|^2$, only coarsely estimates consistency. Furthermore, this means that ICIR must include two terms to compute consistency, that is, both $\|h(\mathbf{x}) - g^{-1}(\mathbf{x})\|^2$ and $\|h^{-1}(\mathbf{x}) - g(\mathbf{x})\|^2$. If inverses were exact (or very close to exact) only one of the above terms would be enough as minimizing one would imply minimizing the other.

SyN, alternatively, provides an inverse that is guaranteed to be everywhere sub-pixel accurate. Furthermore, SyN avoids a basic computational redundancy present in ICIR – the solutions $h(\mathbf{x})$ and $g(\mathbf{x})$ overlap in time. SyN's pair of solutions, ϕ_1 and ϕ_2 , on the other hand, do not overlap in time. They are only two parts of a longer path. This distinction is shown in Fig. 2. However, SyN is still able to compute the measure of inverse consistency by simply composing our time 0.5 maps together and comparing $\phi_1(\mathbf{x}, 1)$ and $\phi_2^{-1}(\mathbf{x}, 1)$ in the I domain (a similar computation may be done in the J domain). Recall that, by definition and diffeomorphic computations, we are able to guarantee $\phi_1^{-1}(\phi_1) = \mathbf{Id}$ and $\phi_2^{-1}(\phi_2) = \mathbf{Id}$. We therefore have an inverse consistency error of zero, up to the inaccuracy induced by the single interpolation required in computing the time 1 maps, e.g. $\phi_1(\mathbf{x}, 1) = \phi_2^{-1}(\phi_1(\mathbf{x}, 0.5), 0.5)$. A key to the symmetry of our method is the ability to compute sub-pixel accurate inverses such that, for each $i = 1, 2$, we have (to a user-selected numerical precision) $\phi_i^{-1}(\phi_i) = \mathbf{Id}$ and $\phi_i(\phi_i^{-1}) = \mathbf{Id}$. Typical precision is 1.0×10^{-6} L_1 norm and 0.2 L_∞ norm, measured in terms of voxel/pixel coordinates. Further details on the numerical methods employed in optimizing this energy may be found in Avants et al. (2006a) and will be summarized below.

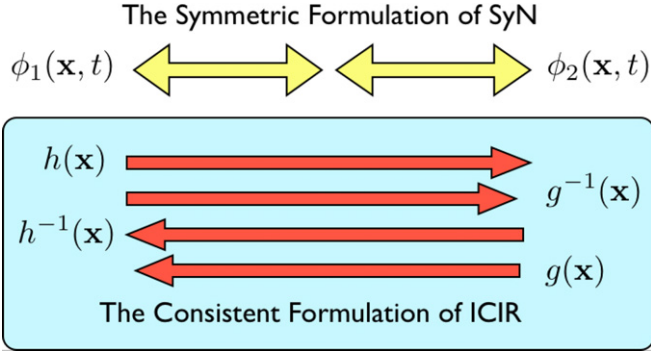


Fig. 2. The symmetric normalization method is represented, at top, by its two components, ϕ_1 and ϕ_2 , meeting at the middle of the normalization domain. Note that each sub-path may be traversed either from the middle to the end or from an end to the middle. Alternatively, the ICIR method is shown in a schematic at the bottom panel of the figure. The correspondence defining vector fields associated with ICIR are called h and g . In ICIR, all four deformation fields overlap in time and may, in fact, be different from each other. The inverse of h may not be its true inverse. Further, the inverse of h may not be equivalent to g .

2.3. The cross-correlation with symmetric diffeomorphisms

We now propose to symmetrically solve the following image matching problem: *Find a spatiotemporal mapping, $\phi \in \text{Diff}_0$, such that the cross-correlation between the image pair is maximized.* This formulation of the problem represents a change of philosophy when compared to Bajcsy et al. (1983) and, later, Gee et al.'s (1993) elastic matching, which also used cross-correlation. Elastic image registration methods seek to balance a regularization term and a similarity term allowing one to find a constrained deformable solution. The approach recommended here departs from this strategy by allowing an unconstrained optimization of the similarity term *within the space of diffeomorphisms*. This strategy is used when finding optimal mappings in lower-dimensional (e.g. affine) transformation groups. The main advantage of an unconstrained search within the space of diffeomorphisms is its simplicity: one allows the method to maximize the similarity until a local maximum or limit on computation time is reached.² The disadvantage is one requires a similarity metric that, when optimized in *Diff*, provides a useful solution. Therefore, design and choice of the similarity metric requires great care.

The mutual information (MI) similarity metric garnered significant interest in recent years (Maes et al., 1997; Wells et al., 1997; Rueckert et al., 1999; Studholme et al., 2006). Intuitively speaking, MI estimates the optimal matching between images by inferring how much global information is shared in the image pair, as estimated from the pair's joint histogram. At the same time, MI prevents over-fitting

by penalizing clustering of the marginal image probabilities (Hermosillo et al., 2002). The globality of this approach makes it extremely useful for robust rigid registration but may limit performance in deformable registration, in particular in cases where non-stationary noise patterns or intensity inhomogeneity requires a locally adaptive similarity. This problem has been addressed by Studholme et al. (2006) and in our previous work (Yoo, 2004). One difficulty with a locally varying estimate to the MI is that joint probabilities need a large number of samples for reliable statistics. Therefore, as locality in the MI estimate increases, its statistical reliability decreases.

Cross-correlation (CC), on the other hand, adapts naturally to situations where locally varying intensities occur and is suitable for some multi-modality problems. The CC depends only on estimates of the local image average and variance which may be accurately/exactly measured with relatively few samples. Furthermore, the cross-correlation has shown historically to perform well in many real-world computer vision applications where one requires robustness to unpredictable illumination, reflectance, etc. An example of the robustness of our method to strong MRI inhomogeneity (bias field) is shown in Fig. 3. For these reasons, we revisit the classical cross-correlation as a similarity metric for use in our emerging diffeomorphic image registration.

We now provide a symmetric formulation of the diffeomorphic image registration problem as driven by CC, along with the Euler–Lagrange equations for this problem. First, define $I_1 = I(\phi_1(\mathbf{x}, 0.5))$ and $J_2 = J(\phi_2(\mathbf{x}, 0.5))$. We also define a variable to represent each image with its local mean subtracted as $\bar{I}(\mathbf{x}) = I_1(\mathbf{x}) - \mu_{I_1}(\mathbf{x})$ and $\bar{J}(\mathbf{x}) = J_2 - \mu_{J_2}(\mathbf{x})$. We compute μ over a local n^D window centered at each position \mathbf{x} , where D is the image dimension. We usually choose $n = 5$. The cross-correlation is then

$$CC(\bar{I}, \bar{J}, \mathbf{x}) = \frac{\langle \bar{I}, \bar{J} \rangle}{\langle \bar{I} \rangle \langle \bar{J} \rangle} = A^2 / BC, \quad (4)$$

where the inner product is also taken over a n^D window. Note that our CC is implicitly a function of ϕ_1 and ϕ_2 through its dependence on I_1 and J_2 . We now are able to define the variational optimization problem in similar fashion to Eq. (3),

$$E_{CC}(\bar{I}, \bar{J}) = \inf_{\phi_1} \inf_{\phi_2} \int_{t=0}^{0.5} \left\{ \|\mathbf{v}_1(\mathbf{x}, t)\|_L^2 + \|\mathbf{v}_2(\mathbf{x}, t)\|_L^2 \right\} dt + \int_{\Omega} CC(\bar{I}, \bar{J}, \mathbf{x}) d\Omega.$$

Subject to each $\phi_i \in \text{Diff}_0$ the solution of:

$$\begin{aligned} d\phi_i(\mathbf{x}, t)/dt &= \mathbf{v}_i(\phi_i(\mathbf{x}, t), t) \text{ with } \phi_i(\mathbf{x}, 0) \\ &= \mathbf{Id} \text{ and } \phi_i^{-1}(\phi_i) = \mathbf{Id}, \phi_i(\phi_i^{-1}) = \mathbf{Id}. \end{aligned} \quad (5)$$

The problem, here, is the same as before but with the cross-correlation as the driving similarity term. We now take the variation of this function with respect to ϕ_1 at time 0.5 and ϕ_2 at time 0.5. Following Beg's derivation, adapted for our

² The contribution of the regularization term to the total energy is small compared to the similarity. However, the regularization term in the diffeomorphic matching problem remains important. It guarantees that one finds a path of minimal length.

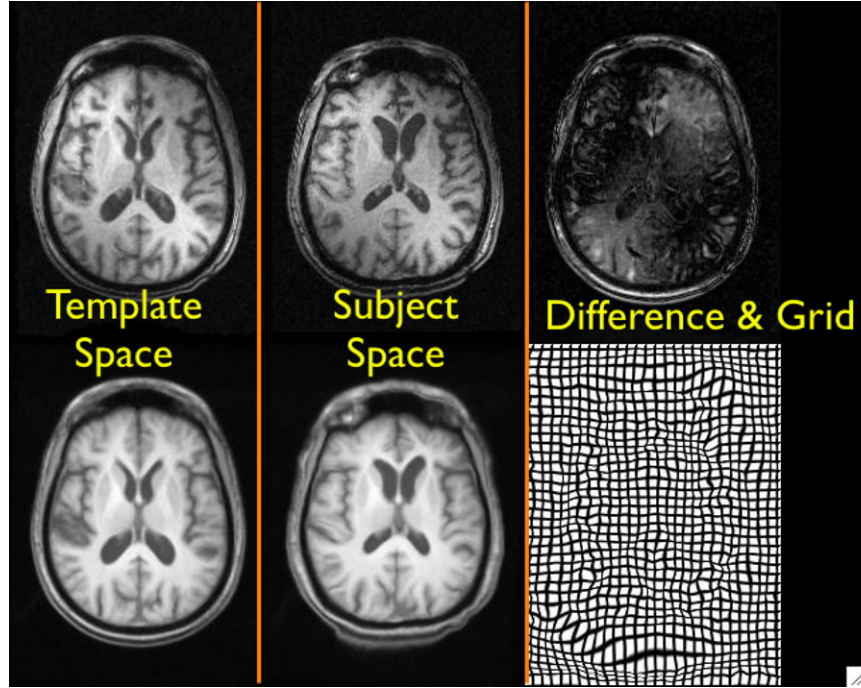


Fig. 3. The local cross-correlation measure allows robust matching of images despite the presence of a strong bias field affecting the image quality. The smoothness of the grid is also unaffected by the bias.

symmetric normalization and particular similarity metric, we find two Euler–Lagrange equations:

$$\begin{aligned} \nabla_{\phi_1(\mathbf{x},0.5)} E_{CC}(\mathbf{x}) &= 2L\mathbf{v}_1(\mathbf{x},0.5) + \frac{2A}{BC} \\ &\times \left(\bar{J}(\mathbf{x}) - \frac{A}{B}\bar{I}(\mathbf{x}) \right) |D\phi_1| \nabla \bar{I}(\mathbf{x}), \end{aligned} \quad (6)$$

$$\begin{aligned} \nabla_{\phi_2(\mathbf{x},0.5)} E_{CC}(\mathbf{x}) &= 2L\mathbf{v}_2(\mathbf{x},0.5) + \frac{2A}{BC} \\ &\times \left(\bar{I}(\mathbf{x}) - \frac{A}{C}\bar{J}(\mathbf{x}) \right) |D\phi_2| \nabla \bar{J}(\mathbf{x}). \end{aligned} \quad (7)$$

The gradients given in Eqs. (6) and (7) include that fact that the velocities exist in the space of smooth vector fields given by the linear operator L as well as the determinant of each ϕ_i transformation Jacobian, $|D\phi_i|$. This equation is similar to both that derived for LDDMM (Beg et al., 2005) and the derivative for the cross-correlation given by Hermosillo et al. (2002). This equation differs from Beg’s Euler–Lagrange equation in that we have an E–L equation for both ϕ_1 and ϕ_2 instead of just ϕ_1 . In addition, Beg used the intensity difference metric given in Eq. (3). The derivation of this equation differs from that in Hermosillo et al. (2002) by the presence of the derivative of the diffeomorphism $\partial \bar{I} / \partial \phi_i$, which leads to the Jacobian term, instead of a vector field (small deformation) derivative. Furthermore, we have represented the CC in a different arrangement of terms that suits our own derivation of the CC variation and computational implementation. This arrangement of terms allows one to precompute, for each iteration, the locally varying values of A , B , C and store them for use in the local computation of the derivative.

Our novel symmetric formulation, therefore, does not add significant additional cost to the normalization. The main additional cost is that of smoothing the derivative estimate to gain the \mathbf{v}_i , not in the actual estimate of the similarity term. Both the ϕ_1 and ϕ_2 derivatives require the terms A , B and C . Therefore, in estimating Eqs. (6) and (7), we use the procedure detailed in Algorithm 1.

Algorithm 1 (Computing cross-correlation derivatives).

- (1) Deform I by $\phi_1(0.5)$ and J by $\phi_2(0.5)$.
- (2) Calculate \bar{I} and \bar{J} from the result of step (1).
- (3) Calculate and store images representing A , B and C .

These steps enable us to loop over the image domain to rapidly compute Eqs. (6) and (7) at the same time from these precomputed variables. Note, however, that it could be possible to modify Lewis’s fast normalized correlation measure (Lewis, 1995) to speed up this process even further.

Eqs. (6) and (7) gives the update to the velocities at time 0.5 and consequently the diffeomorphisms ϕ_1 and ϕ_2 . However, we also require the ϕ_i to satisfy our *o.d.e.* and exact invertibility constraints. We satisfy these constraints by, given a new velocity estimate, first updating the ϕ_i through the *o.d.e.* and then integrating backward in time to find ϕ_i^{-1} and verify invertibility. This approach is a standard one and detailed, algorithmically, in the LPF method (Avants et al., 2006a). For completeness, we include an abbreviated explanation here.

First, assume arbitrary ϕ and \mathbf{v} , related by the standard *o.d.e.* $d\phi(\mathbf{x},t)/dt = \mathbf{v}(\phi(\mathbf{x},t),t)$ used to generate

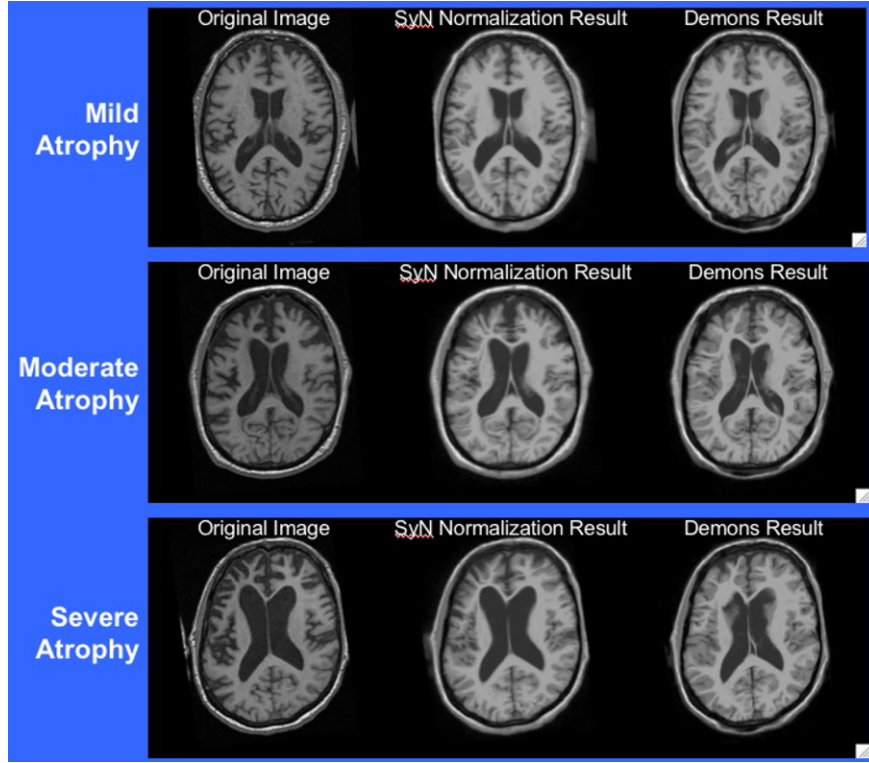


Fig. 4. The atlas was initially aligned to these FTD images via a rigid plus uniform scaling transformation. The subsequent Demons registration to each image, used for labeling, is in the right column. The SyN result is in the center column, while the corresponding original images are in the left column. The Demons method does a reasonable normalization, but leaves the ventricles and other smaller structures only partly normalized. The quadratic elastic penalty prevents the remaining shape differences from being captured. A similar loss of resolution in the mapping is seen in the elastic cross-correlation mappings. These are illustrative images from our previous study, [Avants et al. \(2006c\)](#), and were not used as actual study data in this exposition.

diffeomorphisms. The update method for our diffeomorphism comes from discretization of this *o.d.e.*, such that:

$$\phi(\mathbf{x}, t + \Delta t) \leftarrow \phi(\mathbf{x}, t) + \Delta t \mathbf{v}(\phi(\mathbf{x}, t), t). \quad (8)$$

This discretization is used to update both ϕ_i from time 0 to 0.5. Second, [Algorithm 2](#) gives a method for generating inverses when an arbitrary diffeomorphism ϕ is updated by a small time-step through a velocity field, as in the previous equation. The algorithm typically converges within one to a few iterations, particularly if the time step is small. **The existence of a solution is guaranteed by the integrability condition established for diffeomorphic image registration (Dupuis et al., 1998), while uniqueness comes from the uniqueness theorem of *o.d.e.s* (Arnold, 1991).**

Algorithm 2 (*Inversion method*).

The algorithm uses a fixed point method to push the inverse of ϕ forward by a small amount performing a gradient descent on $\phi^{-1}(\phi) = \mathbf{Id}$, enforcing $\phi^{-1}(\phi) = \mathbf{Id}$ at a sub-pixel level. The same approach was used in the Lagrangian Push Forward algorithm ([Avants et al., 2006a](#)). We use a temporary variable, ψ , to represent the input diffeomorphism that, on output, will be the inverse of ϕ . Input $\phi(\mathbf{x}) = \mathbf{y}$, $\psi^{-1}(\tilde{\mathbf{y}}) = \mathbf{x}$ where $\tilde{\mathbf{y}} \neq \mathbf{y}$ and output $\psi^{-1}(\mathbf{y}) = \mathbf{x} = \phi^{-1}(\mathbf{y})$. At convergence, $\tilde{\mathbf{y}} = \mathbf{y}$ and $\|\psi^{-1}(\phi) - \mathbf{Id}\|_{\infty} < \epsilon_2 r$ where r is the image resolution and ϵ_2 a small constant, typically 0.1.

- 1: **while** $\|\psi^{-1}(\phi(\mathbf{x})) - \mathbf{x}\|_{\infty} > \epsilon_2 r$ **do**
- 2: Compute $\mathbf{v}^{-1}(\mathbf{x}) = \psi^{-1}(\phi(\mathbf{x})) - \mathbf{x}$.
- 3: Find scalar γ such that $\|\mathbf{v}^{-1}\|_{\infty} = 0.5r$.
- 4: Integrate ψ^{-1} s.t. $\psi^{-1}(\tilde{\mathbf{y}}, t) + = \gamma \mathbf{v}^{-1}(\psi^{-1}(\tilde{\mathbf{y}}, t))$.
- 5: **end while**

We now summarize the SyN method in [Algorithm 3](#). Our implementation essentially updates each ϕ_i with the current estimate to the velocity, follows the update to each ϕ_i by calling [Algorithm 2](#) to generate inverse maps, and then re-estimates the velocity from the new estimate to the ϕ_i . This formulation and diffeomorphic representation guarantee SyN's sub-pixel invertibility and algorithmic independence to input permutations. These methods allow us to symmetrically match images to the degree that discrete diffeomorphisms are invertible. Furthermore, while extended here to cross-correlation, similar techniques may be used to efficiently symmetrize almost any other similarity measure. We now leverage ITK for a fair implementation and evaluation of SyN.

Algorithm 3 (*Overview of symmetric normalization method*).

- (1) Initialize $\phi_1 = \mathbf{Id} = \phi_1^{-1}$ and $\phi_2 = \mathbf{Id} = \phi_2^{-1}$.
- (2) Repeat the following steps until convergence:

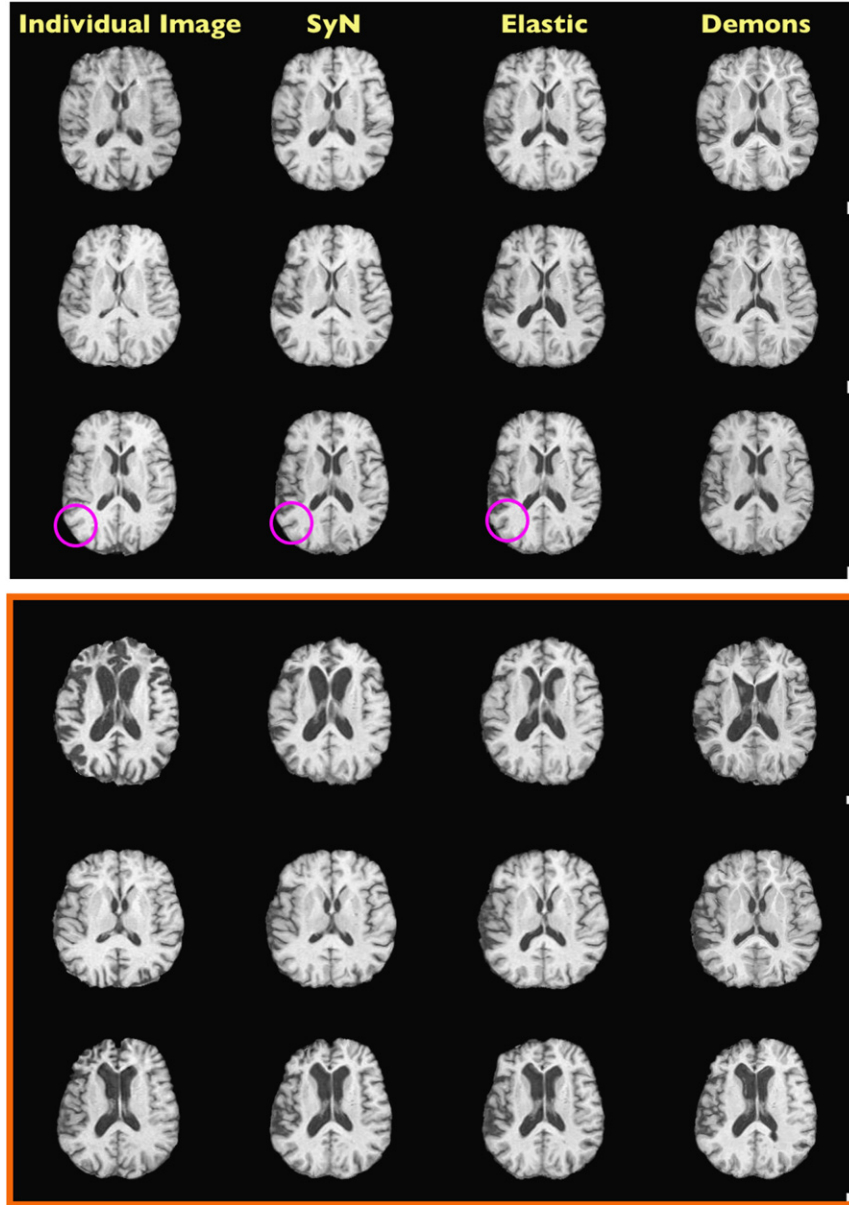


Fig. 5. Normalization results. All images in each row should look similar to the first image in the row. The atlas was initially aligned to these elderly (top three) and FTD (bottom three) images via a rigid plus uniform scaling transformation. These examples are from our current study. We have highlighted (with a circle) the type of small scale difference one may see in the registration quality. Larger scale differences are also clear. In particular, the Demons and elastic cross-correlation have problems both shrinking and expanding the ventricles (the second elderly image). In addition, the Demons intensity consistency assumption causes significant errors in the first FTD image, a case where this assumption does not hold.

- (3) Compute the cross-correlation as described in [Algorithm 1](#).
- (4) Compute each v_i by smoothing the result of step (3) in this table. One may also use the modified midpoint method for each velocity, as in the LPF algorithm (Avants et al., 2006a), to give smoothness in time.
- (5) Update each ϕ_i by v_i through the *o.d.e.* as described in Eq. (8). This step automatically adjusts the time step-size such that the maximum length of the updates to the ϕ_i is sub-pixel and approximately constant over iterations. We explicitly guarantee $\|\mathbf{v}_1(\cdot, t)\| = \|\mathbf{v}_2(\cdot, t)\|$. We also update the estimate to the geodesic distance by trapezoidal rule, as in the LPF method.
- (6) Use [Algorithm 2](#) to get the inverses of the ϕ_i .
- (7) Generate the time 1 solutions from $\phi_1(1) = \phi_2^{-1}(\phi_1(\mathbf{x}, 0.5), 0.5)$ and $\phi_1^{-1}(1) = \phi_2(1) = \phi_1^{-1}(\phi_2(\mathbf{x}, 0.5), 0.5)$.

2.4. Implementation in ITK

The Demons algorithm is freely available in the standard ITK distribution and has been quantitatively evaluated by the ITK community. We have implemented SyN within our extended version of the ITK deformable image registration framework, described in Yoo (2003). Therefore, we are in a position to measure performance gains

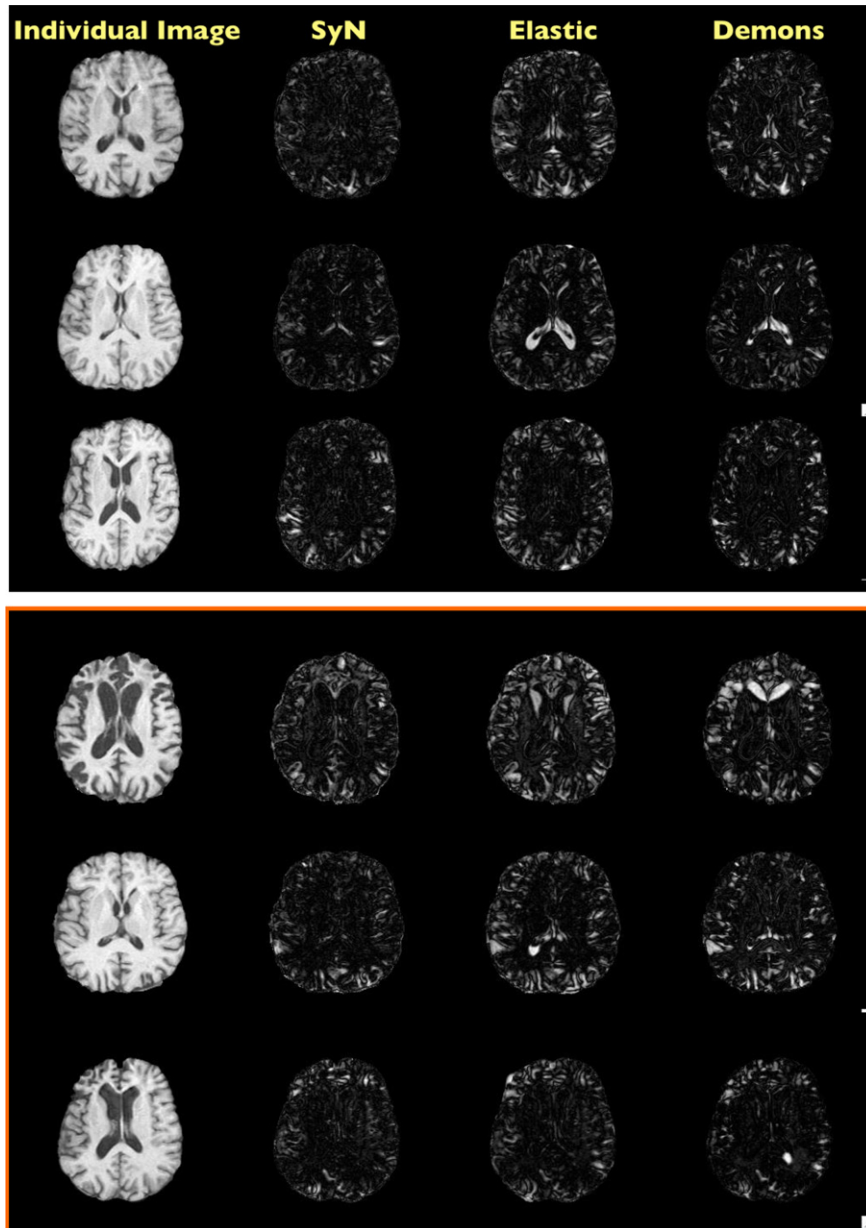


Fig. 6. Normalization results difference images. This figure shows the same results as Fig. 4, but with absolute value of the image difference, after normalization. One can observe that small sulci may not be captured by any of the methods. Under the intensity difference metric, the Demons method should be expected to perform best, as it explicitly minimizes this error.

by varying first the *similarity metric* and then the *transformation model*, keeping an identical underlying code base.

2.4.1. Switching the similarity metric

For this part of the study, we replace the `itkDemonsRegistrationFunction`, in the `itkDemonsRegistrationFilter`, with our own `itkCrossCorrelationRegistrationFunction`. The interface and operation of this modified deformable registration algorithm are identical to the Demons algorithm, except that the image forces come from the small deformation version of Eq. (6), as may be found in Hermosillo et al. (2002). In addition to switching the `itkDemons-`

`RegistrationFunction` for our cross-correlation image forces, we also modified the time step used during this elastic registration process. The time-step was increased from 1, for Demons, to 4 for cross-correlation. The Demons force, based on optical flow, is extremely aggressive compared to our analytical cross-correlation derivative. We found, for our dataset, that time-steps beyond 4 led to non-positive Jacobians in a significant fraction of our normalizations.

2.4.2. Switching the transformation model

As mentioned previously, we have implemented SyN in our extended Insight ToolKit. For this study, SyN will use

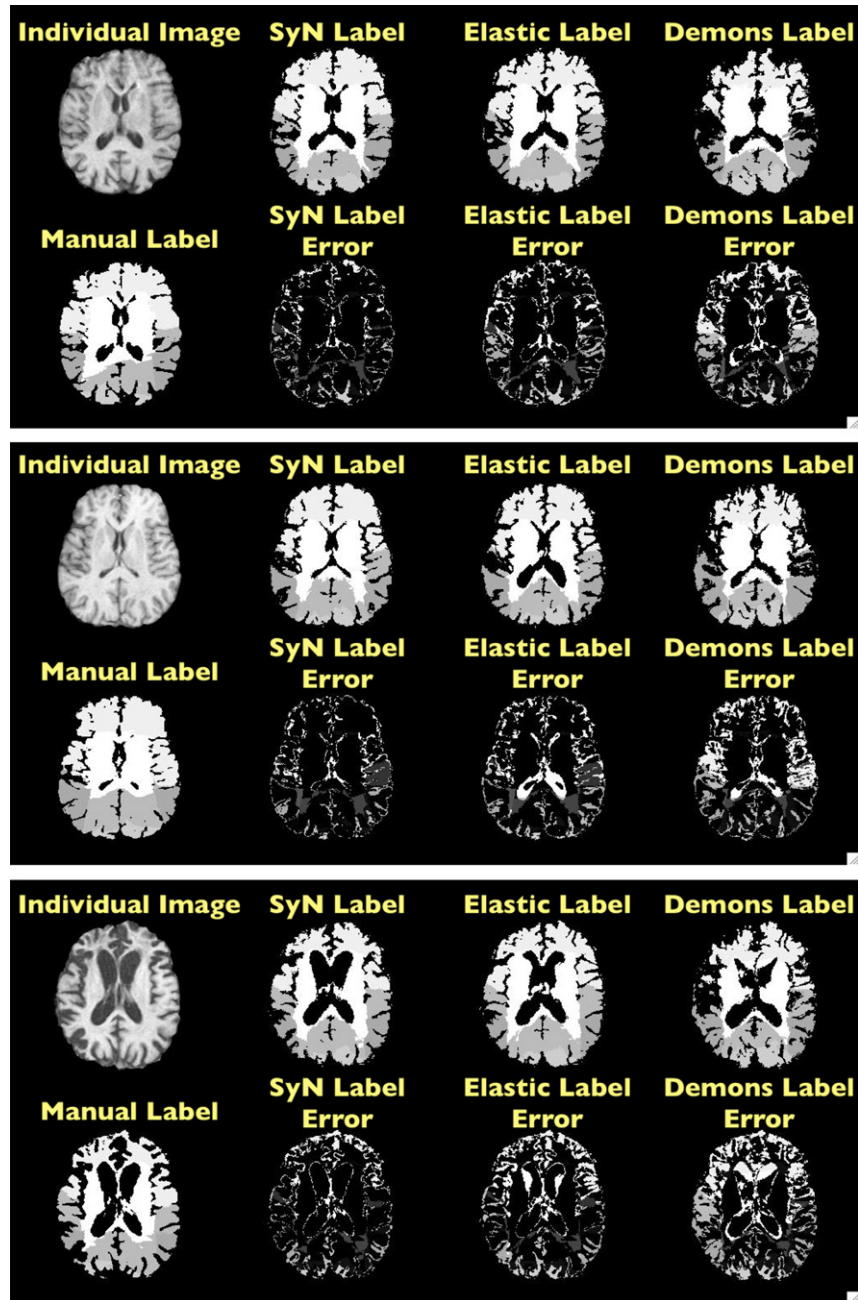


Fig. 7. Normalization results label error images. Here, we show three individual images along with the manual, SyN, elastic and Demons labels. This bottom row for each individual shows the difference of the algorithmic and manual labels. The majority of the errors are due to a shift or a curvature in the boundary definition gained algorithmically, with respect to the manual definition. Manually defined lobar boundaries tend to be planar. Further, when one compares the (in particular, parietal) boundaries determined by the labeler on these three images, it becomes apparent that there may be some inconsistency. Secondary errors are due to lack of sulcal definition in the template labels, with respect to the individual labeled ground truth.

the `itkPDEDeformableRegistrationFilter` as implemented in ITK for its Gaussian smoothing operator.³ This smoothing operator will be applied only to the incremental output of our `itkCrossCorrelationRegistrationFunction`. The Demons and elastic cross-correlation algorithms both

apply Gaussian smoothing to the *total* displacement field, in accordance with small deformation transformation models. The presence of a large deformation assumption, along with the use of a space–time parameterization of the transformation, are the main (not the only) differences between diffeomorphic and elastic registration algorithms.

The key contrast between the first pair of methods (Demons, elastic cross-correlation) is the similarity metric. The only difference between the second two (elastic cross-correlation, SyN cross-correlation) is in the transformation

³ The Gaussian kernel, G_σ , is an estimate to the Green's kernel for the linear operator $L = \nabla^2 + \text{Id}$ (Bro-Nielsen and Gramkow, 1996) and is adequate for providing the velocity field regularity necessary for integrating *o.d.e.s.*

model. We are therefore investigating, first, if cross-correlation provides better normalization than Demons optical flow, for our neurodegenerative and elderly T1 MRI dataset. The second comparison evaluates if changing from the elastic to the symmetric diffeomorphic transformation model yields additional improvement. Our hypothesis was that each change would yield benefits. This hypothesis was born out in our experimental data, as suggested by the preliminary study data shown in Fig. 4.

3. Data and experiments

3.1. Dataset

Our database consists of 20 T1 MRI images ($0.85 \times 0.85 \times 1.5$ mm, GE Horizon Echospeed 1.5 T scanner) from 10 normal elderly and 10 frontotemporal dementia patients. The 10 frontotemporal dementia individuals are a different set than used in our previous study (Avants et al., 2006c). Each of the 20 images, along with the elderly template, was manually labeled with the protocol described in Sparks et al. (2002). This protocol was shown to be highly reproducible for both small and large structures via six-month intra-rater reliability and inter-rater reliability measurements. Left hippocampus labeling, for example, showed a 0.92 intra-rater overlap ratio (Eq. (1)) and 0.83 average for inter-rater overlap. As the hippocampus is relatively small, these values are reasonable. Finally, as the labeling was intended to be only on the cortex (not cerebrospinal fluid), we masked the label images by the inverse of the cerebrospinal fluid segmentations. This also makes our evaluation more sensitive to differences in the accuracy of sulcal and gyral alignment.

3.2. Evaluation

We now use these two comparative pairings of three algorithms to study performance for characterizing the volumetric differences between elderly and frontotemporal dementia cortex, hippocampus, amygdala and cerebellum. An example comparison of the SyN and Demons methods' normalization abilities is in Fig. 4. This study reveals the ability of these methods, SyN, elastic cross-correlation and Demons, to reproduce results gained from an expert user's labeling of our 20 image dataset. This test is performed by using each method to map the template labels (an elderly individual also labeled by our expert user) to each individual. Twenty rigid registrations and 60 non-rigid registrations were required (20 subjects were each deformably registered by three algorithms). The atlas labelings are then warped by the composed rigid and non-rigid transformation into the space of the patient image. We then compute Dice overlap ratios between the manual and automatic structural segmentations for each structure.

Each method was applied in an identical four-level multi-resolution scheme and ran until convergence or a

fixed (maximum) number of iterations was reached. We allowed up to 100 iterations at the first level, 100 iterations at the second level, 100 iterations at the third level and 20 iterations at the full resolution. The relative running times for each algorithm varied depending upon the particular cases being run. However, the relative average running times (in terms of the average Demons running time) were 1 (Demons), 4.2 (elastic cross-correlation) and 5.5 (SyN). The mean runtime of the Demons method on a 2.0 GHz Intel Mobile Pentium processor was 20.4 min. The relative similarity of the elastic time cost to the cost of SyN also indicates that our symmetric approach does not create significant additional computational cost.

4. Results and discussion

A visualization of six individuals from our dataset and the normalization of the template to these individuals is shown in Fig. 5. Associated intensity error images are shown in Fig. 6. Label error images are shown in Fig. 7. The eight labeled individual structures are shown, along with the summary statistics for our results, in Fig. 8. Both cross-correlation algorithms produced segmentation results above the minimum threshold of 0.6 for all structures, as shown in Fig. 8. We also compared the minimum Jacobian of the elastic cross-correlation and SyN cross-correlation methods and did not find significant differences ($T = -1.67725$, $P < 0.101703$). This indicates that SyN's results are not significantly less constrained at a local level. The smallest difference in performance between the Demons and elastic cross-correlation methods was on the cerebellum. The highest difference was on the frontal lobe. The largest performance difference in SyN and elastic cross-correlation was on the temporal and frontal lobe. Performance gains expectedly focus on frontal and temporal lobe due to the known effect of FTD on these two areas. The shape changes caused by this complex disease are likely difficult to capture with a constrained method such as elastic normalization. The smallest differences between the elastic and diffeomorphic methods were found in the hippocampus and amygdala. This is likely to indicate that the similarity metric does not provide a rich enough feature space over which to optimize the correspondence of these structures. **Therefore, a better optimizer operating with that similarity metric will not be likely to provide an advantage.**

A likely improvement in performance would come from using a method such as the STAPLE algorithm (Warfield et al., 2004) to bootstrap results. Similarly, an optimal template would augment results (Avants and Gee, 2004). However, both of these enhancements would increase performance in a consistent manner across all our tested registration methods and would be unlikely to change the relative performance of our algorithms.

The second test we performed was to evaluate whether the volume measurements obtained by the normalization are strongly correlated with the measurements of the expert

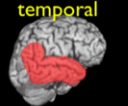





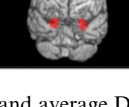
Structure	Demons	Elastic XCor > Demons	SyN XCor > Elastic
 temporal	Mean+Sigma: 0.76 +/- 0.021	0.81 +/- 0.02	0.84 +/- 0.019
	Min - Max : [0.69-0.79]	[0.76-0.84]	[0.79-0.87]
	Significance:	p< 0.0001	p< 0.0001
 parietal	0.69 +/- 0.034	0.74 +/- 0.03	0.78 +/- 0.027
	[0.62-0.73]	[0.68-0.79]	[0.70-0.83]
	-	p< 0.0001	p< 0.0001
 occipital	0.78 +/- 0.030	0.79 +/- 0.024	0.83 +/- 0.022
	[0.72-0.82]	[0.73-0.84]	[0.78-0.87]
	-	p< 0.011	p< 0.0001
 hippocampus	0.62 +/- 0.070	0.72 +/- 0.036	0.72 +/- 0.038
	[0.48-0.73]	[0.65-0.77]	[0.63-0.79]
	-	p< 0.0001	p<0.7
 frontal	0.74 +/- 0.026	0.81 +/- 0.026	0.85 +/- 0.024
	[0.65-0.77]	[0.73-0.84]	[0.79-0.88]
	-	p< 0.0001	p< 0.0001
 cerebellum	0.89 +/- 0.012	0.89 +/- 0.011	0.92 +/- 0.011
	[0.87-0.92]	[0.88-0.92]	[0.91-0.93]
	-	p<0.2	p< 0.0001
 amygdala	0.59 +/- 0.053	0.73 +/- 0.065	0.74 +/- 0.05
	[0.5-0.68]	[0.59-0.81]	[0.63-0.81]
	-	p<0.0001	p<0.24

Fig. 8. Performance comparison and average Dice statistic for each method and each structure. Example brain labelings mark the structures over which we evaluated the three algorithms. The final results showed that, overall, SyN > Elastic > Demons for automatically labeling these lobar structures. The P -values for the difference in performance were computed using non-parametric permutation testing based on the paired T -test on vectors of Dice statistics. We used 10,000 permutations per example, thus limiting our minimum P -value to 0.0001. Extremely small P -values indicate that the method uniformly and strongly outperformed the method with which it was paired.

user. We performed this analysis only on regions (frontal, temporal, parietal lobes) where one may reasonably expect FTD to induce a difference from the normal elderly population and where labeling performance was good. The purpose of this test is to determine if conclusions made by analyzing the output of our automated normalization methods are at all consistent with the expert user's analysis. As we are using the expert labeling as our gold standard, the "better" method should produce volume measures that correlate more strongly with the volume measures gained by the expert. The volume, in cubic centimeters, of each structure was calculated by summing the voxel volumes that were given the appropriate label for that structure. As shown in Table 1, SyN outperformed elastic cross-correlation (and the Demons method) according to the degree to which the automated volumes correlate with the manual volumes for each of three structures. SyN also had the least overall discrepancy in the measured volume, that is, $\sum_i |V_{Alg} - V_{Man}|$, where V_{Man} is manually measured volume and V_{Alg} refers to algorithmically measured volume. These results are shown in Table 2.

Table 1

Pearson correlations between manual and algorithmic volume measures

Structure	Corr(Man, Syn)	Corr(Man, Elas)	Corr(Man, Demons)
Temporal	0.86	0.69	0.79
Frontal	0.89	0.67	0.71
Parietal	0.71	0.42	0.66

Table 2

The average absolute volume error between manual volume and algorithmic volume measured over the dataset for each structure

Structure	VolErr(Man, Syn)	VolErr(Man, Elas)	VolErr(Man, Demons)
Temporal	8.4	9.2	8.7
Frontal	11.1	16.1	15.8
Parietal	7.9	9.3	7.9

We also plot the estimated volume versus the manually computed volume for the temporal lobe in Fig. 9, for frontal lobe in Fig. 10 and for parietal lobe in Fig. 11. We use linear regression to fit the estimated volume with a line in order to assess the closeness of the method's estimate to

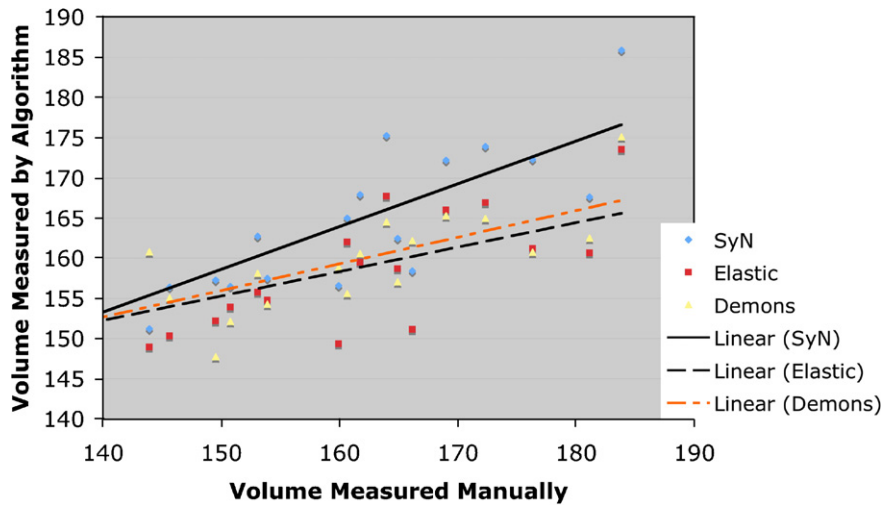


Fig. 9. Temporal lobe volume, in cubic centimeters, as measured by each of the three methods. The algorithmic measures are plotted against the manually measured volume.

the manual estimate. The Pearson correlation of SyN volumes with the manually measured volumes was 0.86 for temporal lobe, 0.89 for frontal lobe and 0.71 for parietal lobe. The correlation of the elastic method volumes with the manually measured volumes was 0.69 for temporal lobe, 0.67 for frontal lobe and 0.42 for parietal lobe. The correlation of Demons volumes with the manually measured volumes was 0.79 for temporal lobe, 0.67 for frontal lobe and 0.66 for parietal lobe. One interesting finding, here, is that the Demons method, compared to the elastic method, produces stronger correlations with manual labelings (in terms of volume measurements) but produces smaller overlap ratios.

Our final analysis tests the significance of the difference in volumes between the FTD and elderly members of our image population. The P-values of these results, obtained

by non-parametric permutation testing, are shown in Table 3. None of the structures showed a significant volumetric difference between groups. However, SyN appears to reflect the significance obtained by the manual rater more closely than the other methods. At the same time, we do not find these results to be strong enough to warrant definitive conclusions. Comparing a larger dataset (or possibly a different set of control and subject images) would likely show stronger differences than this dataset, as we found in our previous, preliminary study (Avants et al., 2006c).

4.1. Contrasting automated and manual results

Thus, we can see how an apparently small, *but consistent* difference in performance (as measured by overlap ratio) can have an impact on the validity of the study outcome.

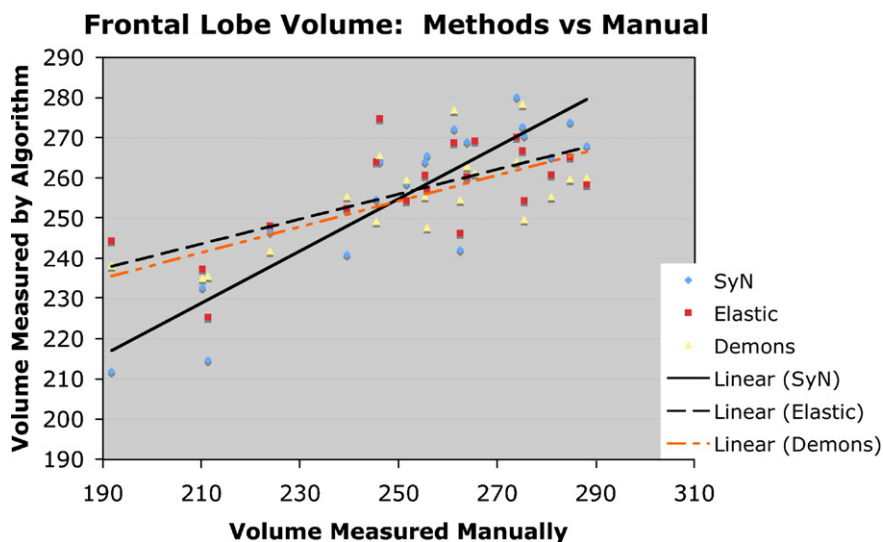


Fig. 10. Frontal lobe volume, in cubic centimeters, as measured by each of the three methods. The algorithmic measures are plotted against the manually measured volume.

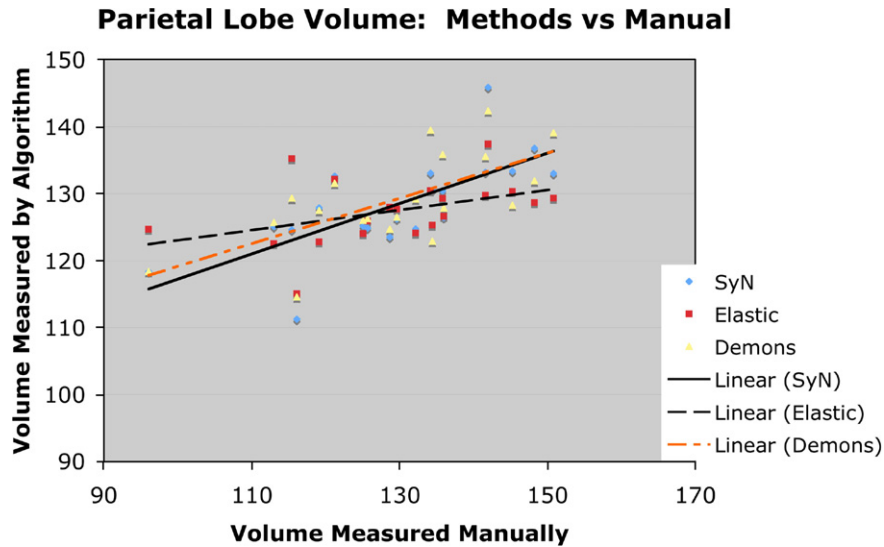


Fig. 11. Parietal lobe volume, in cubic centimeters, as measured by each of the three methods. The algorithmic measures are plotted against the manually measured volume.

Table 3
Significance of elderly-FTD volume differences as measured by manual, SyN, elastic and Demons methods

Structure	Eld-FTD Man.	Eld-FTD SyN	Eld-FTD Elas.	Eld-FTD Demons
Temporal	0.25	0.35	0.47	0.26
Frontal	0.18	0.13	0.49	0.52
Parietal	0.32	0.46	0.82	0.15

Significance was determined by a permutation test based on the unpaired *T*-test of elderly and FTD structure volumes.

That is, the performance gains in overlap ratio translate to more accuracy in making clinically meaningful measurements, such as volume. Although the quality of our results are reasonable, by some standards, these algorithms, as applied here, cannot claim to accurately reproduce manual labelings. The reasons for this are well-known. The primary reason is that expert knowledge is not directly encoded in these methods. Secondly, the uncertainty inherent to the problem of neuroanatomical labeling limits the accuracy and reproducibility of both manual and automated segmentation. Finally, it is not yet known the extent to which brains of different individuals, when represented as magnetic resonance images, are diffeomorphic to each other. This problem is even less well understood with elderly and patient brains. Note also that, when measuring the trends in the difference of elderly and FTD structure volumes, the Demons, elastic correlation and SyN significance tend to estimate larger volumes than the manual results. This is likely caused by two things: segmentation bias towards the template and the fact that registration-based segmentations are smoothed (elastic and Demons more than SyN), while the manual segmentations are not. Both of these types of errors may be observed in Fig. 7. Fig. 7 also shows that the labeler may have used different

“styles” of labeling or changed the decision-making process over the dataset. All of these confounds impact the overlap ratios and correlations that we find here. We accentuate that our goal, here, is not specifically to reproduce the expert labels, but to use these labeled structures as a neuroanatomically based evaluation of relative algorithmic performance.

5. Conclusion

We first described the symmetric normalization formulation. We then extended this formulation to use the cross-correlation similarity function providing, in addition, Euler–Lagrange equations for the variational problem in the symmetric diffeomorphic context. We contrasted our method with the popular inverse consistent image registration technique, which was outperformed by the Demons method in an unbiased comparative evaluation of brain segmentation and alignment (Hellier et al., 2003). We then provided algorithmic details of the SyN methodology. Finally, we leveraged our ITK implementation to compare the SyN cross-correlation optimizer with the Demons and an elastic cross-correlation optimizer. This enabled us to explore the effect of similarity metric and optimizer separately and purely, as we use an identical linear operator for regularization and an identical code base, the Insight ToolKit. The relative similarity of the elastic computation time to the computational cost of SyN also indicates that our symmetric approach does not require significant additional computational expense. We found, in short, that the cross-correlation behaves well on elderly and neurodegenerative data. In particular, it outperforms the Demons optical flow on our dataset. We also found that the symmetric diffeomorphic optimizer outperforms the elastic optimizer on the same dataset, while guaranteeing that the topology

of the images is preserved and, importantly, that the algorithm's performance does not depend upon the order in which one inputs the images.

This careful comparison establishes the distinct advantage of SyN for segmenting elderly and neurodegenerative cerebrum, cerebellum and, in particular, the temporal and frontal lobe. Note that, in addition to better performance, SyN provides a dense space–time map and transformation inverses. The differences in performance are consistent, statistically significant and have a major impact on study outcome. One can extrapolate even larger differences between SyN and algorithms with lower dimensionality than either Demons or SyN. For this reason, along with the theoretical advantages that translate into practical benefits, we promote diffeomorphic algorithms and the cross-correlation similarity metric in neuroimaging research, in particular when studying non-standard datasets, such as FTD and AD.

Acknowledgements

We thank the reviewers for greatly improving the contents of this paper. Much of this work was supported by NIH grant R01-EB006266.

References

- Arnold, V.I., 1991. Ordinary Differential Equations. Springer, Berlin.
- Arnold, V.I., Khesin, B.A., 1992. Topological methods in hydrodynamics. *Ann. Rev. Fluid Mech.* 24, 145–166.
- Ashburner, J., Good, C., Friston, K., 2000. Tensor based morphometry. *Neuroimage* 11, 465.
- Avants, B., Gee, J.C., 2004. Geodesic estimation for large deformation anatomical shape and intensity averaging. *Neuroimage (Suppl. 1)*, S139–S150.
- Avants, B., Grossman, M., Gee, J.C., 2005a. The correlation of cognitive decline with frontotemporal dementia induced annualized gray matter loss using diffeomorphic morphometry. *Alz. Dis. Assoc. Dis.* 19 (Supplement 1), S25–S28.
- Avants, B., Epstein, C.L., Gee, J.C., 2005. Geodesic image interpolation: parameterizing and interpolating spatiotemporal images. In: *ICCV Workshop on Variational and Level Set Methods*, pp. 247–258.
- Avants, B., Schoenemann, P.T., Gee, J.C., 2006a. Landmark and intensity-driven Lagrangian frame diffeomorphic image registration: application to structurally and functionally based inter-species comparison. *Med. Image Anal.* 10, 397–412.
- Avants, B., Epstein, C.L., Gee, J.C., 2006b. Geodesic image normalization in the space of diffeomorphisms. *Math. Found. Comput. Anat.*, 125–133.
- Avants, B., Grossman, M., Gee, J.C., 2006c. Symmetric diffeomorphic image registration: evaluating automated labeling of elderly and neurodegenerative cortex and frontal lobe. *WBIR*, 50–57.
- Bajcsy, R., Lieberman, R., Reivich, M., 1983. A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *J. Comput. Assist. Tomogr.* 5, 618–625.
- Ballmaier, M., O'Brien, J.T., Burton, E.J., Thompson, P.M., Rex, D.E., Narr, K.L., McKeith, I.G., DeLuca, H., Toga, A.W., 2004. Comparing gray matter loss profiles between dementia with Lewy bodies and Alzheimer's disease using cortical pattern matching: diagnosis and gender effects. *Neuroimage* 23, 325–335.
- Beg, F., Miller, M., Trounev, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision* 61, 139–157.
- Bro-Nielsen, M., Gramkow, C., 1996. Fast fluid registration of medical images. *Proceedings of Visualization in Biomedical Computing*. Springer, Hamburg, pp. 267–276.
- Chan, D., Fox, N.C., Jenkins, R., Scahill, R.I., Crum, W.R., Rossor, M.N., 2001. Rates of global and regional cerebral atrophy in ad and frontotemporal dementia. *Neurology* 57 (10), 1756–1763.
- Christensen, G., Johnson, H., 2001. Consistent image registration. *IEEE Trans. Med. Imaging* 20 (7), 568–582.
- Christensen, G.E., Joshi, S.C., Miller, M.I., 1997. Volumetric transformation of brain anatomy. *IEEE Trans. Med. Imaging* 16 (6), 864–877.
- D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P., 2003. A viscous fluid model for multimodal non-rigid image registration using mutual information. *Med. Image Anal.* 7 (4), 565–575.
- Dawant, B., Hartmann, S., Thirion, J.P., Maes, F., Vandermeulen, D., Demaerel, P., 1999. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations, part II: methodology and validation on severely atrophied brains. *IEEE Trans. Med. Imaging* 18, 926–971.
- Dawant, B., Li, R., Cetinkaya, E., Kao, C., Fitzpatrick, J., Konrad, P., 2003. Computerized atlas-guided positioning of deep brain simulators: a feasibility study. In: *Gee, J., Maintz, J.B. (Eds.), Workshop on Biomedical Image Registration*. Philadelphia, July, pp. 142–150.
- Dupuis, P., Grenander, U., Miller, M.I., 1998. Variational problems on flows of diffeomorphisms for image matching. *Q. Appl. Math.* 56 (3), 587–600.
- Fox, N., Crum, W., Scahill, R., Stevens, J., Janssen, J., Rossor, M., 2001. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet* 358, 201–205.
- Gee, J.C., Bajcsy, R.K., 1999. Elastic matching: continuum mechanical and probabilistic analysis. In: *Toga, A.W. (Ed.), Brain Warping*. Academic Press, San Diego, pp. 183–197 (Chapter 11).
- Gee, J.C., Reivich, M., Bajcsy, R., 1993. Elastically deforming a 3D atlas to match anatomical brain images. *J. Comput. Assist. Tomogr.* 17, 225–236.
- Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Le Goualher, G., Collins, D.L., Evans, A., Malandain, G., Ayache, N., Christensen, G.E., Johnson, H.J., 2003. Retrospective evaluation of inter-subject brain registration. *IEEE Trans. Med. Imaging* 22 (9), 1120–1130.
- Hermosillo, G., Chéfd'Hotel, C., Faugeras, O., 2002. A variational approach to multi-modal image matching. *Int. J. Comput. Vision* 50 (3), 329–343.
- Johnson, H.J., Christensen, G.E., 2002. Consistent landmark and intensity-based image registration. *IEEE Trans. Med. Imaging* 21 (5), 450–461.
- Joshi, S.C., Miller, M.I., 2000. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Image Process.* 9 (8), 1357–1370.
- Kertesz, A., Martinez-Lage, P., Davidson, W., Munoz, D.G., 2004. The corticobasal degeneration syndrome overlaps progressive aphasia and frontotemporal dementia. *Neurology* 55, 1368–1375.
- Lewis, J.P., 1995. Fast normalized cross-correlation. *Vision Interface. Canadian Image Processing and Pattern Recognition Society*, pp. 120–123.
- Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., Joshi, S., 2006. Multi-modal image set registration and atlas formation. *Med. Image Anal.* 19 (3), 440–451.
- Maes, F., Collignon, A., Meulen, D., Marchal, G., Suetens, P., 1997. Multi-modality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* 16, 187–198.
- Miller, M., Trounev, A., Younes, L., 2002. On the metrics and Euler–Lagrange equations of computational anatomy. *Annu. Rev. Biomed. Eng.* 4, 375–405.
- Peckar, W., Schnorr, C., Rohr, K., Stiehl, H.S., Spetzger, U., 1998. Linear and incremental estimation of elastic deformations in medical registration using prescribed displacements. *Mach. Graph. Vision* 7 (4), 807–829.

- Ratnavalli, E., Brayne, C., Dawson, K., Hodges, J.R., 2002. The prevalence of frontotemporal dementia. *Neurology* 58 (11), 1585–1586.
- Rogelj, P., Kovacic, S., 2006. Symmetric image registration. *Med. Image Anal.* 10 (3), 484–493.
- Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* 18 (8), 712–721.
- Shen, D., Davatzikos, C., 2002. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21 (11), 1421–1439.
- Sparks, B., Friedman, S., Shaw, D., Aylward, E., Echelard, D., Artru, A., Maravilla, K., Giedd, J., Munson, J., Dawson, G., Dager, S., 2002. Brain structural abnormalities in young children with autism spectrum disorder. *Neurology* 59, 184–192.
- Studholme, C., Cardenas, V., Blumenfeld, R., Schuff, N., Rosen, H.J., Miller, B., Weiner, M., 2004. Deformation tensor morphometry of semantic dementia with quantitative validation. *Neuroimage* 21 (4), 1387–1398.
- Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V., 2006. Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Trans. Med. Imaging* 25 (5), 626–639.
- Thirion, J.P., 1996. Non-rigid matching using demons, In: *IEEE Computer Vision and Pattern Recognition*. pp. 245–251.
- Thirion, J., 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.* 2 (3), 243–260.
- Trounev, A., 1998. Diffeomorphism groups and pattern matching in image analysis. *Int. J. Comput. Vision* 28 (3), 213–221.
- Trounev, A., Younes, L., 2005. Metamorphoses through lie group action. *Found. Comput. Math.* 5 (2), 173–198.
- Vaillant, M., Miller, M.I., Younes, L., Trounev, A., 2004. Statistics on diffeomorphisms via tangent space representations. *Neuroimage* 23, S161–S169.
- Warfield, S.K., Zou, K.H., Wells III, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Med. Image Anal.* 23 (7), 903–921.
- Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1997. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* 1, 35–51.
- Yoo, T., 2003. *Insight Into Images: Principles and Practice for Segmentation, Registration and Image Analysis*. A.K. Peters Ltd., Natick, MA.
- Yoo, T. (Ed.), 2004. *Insight Into Images: Theory for Segmentation, Registration and Image Analysis*. A.K. Peters Ltd., Natick, MA, Chapter Avants et al., Non-rigid registration.
- Younes, L., 1998. Computable elastic distance between shapes. *SIAM J. Appl. Math.* 58, 565–586.
- Younes, L., preprint. Jacobi fields in groups of diffeomorphisms and applications, *Q. Appl. Math.*