UCLA
Department of Statistics Papers

Title

Quatitative Comparison and Analysis of Brain Image Registration Using Frequency-Adaptive Wavelet Shrinkage

Permalink

https://escholarship.org/uc/item/6v0839bv

Authors

Dinov, Ivo, Department of Statistics, UCLA
Mega, Michael S
Thompson, Paul M.
et al.

Publication Date

2002-03-01

eScholarship.org Powered by the California Digital Library
University of California

Quantitative Comparison and Analysis of Brain Image Registration Using Frequency-Adaptive Wavelet Shrinkage

Ivo D. Dinov, Michael S. Mega, Paul M. Thompson, *Member, IEEE*, Roger P. Woods, De Witt L. Sumners, Elizabeth L. Sowell, and Arthur W. Toga

***Abstract*—In the field of template-based medical image analysis, image registration and normalization are frequently used to evaluate and interpret data in a standard template or reference atlas space. Despite the large number of image-registration (warping) techniques developed recently in the literature, only a few studies have been undertaken to numerically characterize and compare various alignment methods. In this paper, we introduce a new approach for analyzing image registration based on a selective-wavelet reconstruction technique using a frequency-adaptive wavelet shrinkage. We study four polynomial-based and two higher complexity nonaffine warping methods applied to groups of stereotaxic human brain structural (magnetic resonance imaging) and functional (positron emission tomography) data. Depending upon the aim of the image registration, we present several warp classification schemes. Our method uses a concise representation of the native and resliced (pre- and post-warp) data in compressed wavelet space to assess quality of registration. This technique is computationally inexpensive and utilizes the image compression, image enhancement, and denoising characteristics of the wavelet-based function representation, as well as the optimality properties of frequency-dependent wavelet shrinkage.**

***Index Terms*—Brain imaging, magnetic resonance imaging (MRI), positron emission tomography (PET), quality of image registration, wavelets.**

I. Introduction

A. General

IMAGE registration (warping) is the process of aligning data into reference space. The main purpose of this alignment is to identify common features of interest in the data by using the atlas associated with the reference volume. In the case of studying human brain data, one is often times interested in understanding the interplay between brain function and structure. This problem can be simplified by warping the data into a reference (atlas) space and applying the reference-induced brain maps to study the data. However, image registration introduces another nontrivial problem of controlling the amount of deformation required to transform the data into atlas space. One way to measure and account for this data deformation is to compute various functionals on the displacement (warping) field (e.g., local or global Jacobian of the vector field). Another is to look at the pre- and post-registration data and the target of the image alignment to assess quality of warping and degree of image distortion. In this paper, we will only discuss the latter approach.

The overall organization of this paper is as follows. After reviewing basic warping strategies, we discuss wavelet thresholding (shrinkage). In Section II, we then present some observations about the useful properties of the large magnitude wavelet coefficients in discrete wavelet function representation. Section II-B introduces three different schemes for quantitative analysis of image registration using the compressed wavelet coefficients of the data (pre- and post-warping). Each of these three classification schemes is applied in different situations based on the ultimate goals of the image alignment. In Section III, we apply the methodology of "optimal" wavelet coefficient selection to analyze and rank the performance of four different polynomial techniques, i.e., the Montreal Neurological Institute (MNI) fiducial-based spline warp [9], and statistical parametric mapping (SPM) [21] trigonometric nonaffine registration. A manually derived fiducial-based validation of the wavelet space quantitative warp classification is presented in the conclusion of Section III.

B. Various Warping Approaches

An increasingly important problem in the area of medical imaging is the comparison of images of the same or different subjects, image modalities, or cognitive stimuli. An accepted way of controlling for anatomic variability within a population is to find a spatial deformation field that warps brain images onto a reference atlas [36].

This section will attempt to sketch the overall idea behind developing a technique for image alignment (registration/warping). To illustrate the focal points of data registration, we discuss the organization of the different methods, in general, and only present details about the polynomial-type alignment.

Manuscript received October 9, 2000; revised January 18, 2001. This work was supported by the Human Brain Project under Contract NIMH/NIDA: P20MH/DA 52176, by the National Library of Medicine under Contract LM/MH05639, by the National Institute of Neurological Disorders and Stroke under Contract NS 38753, and by the National Science Foundation under Contract DMS-9403454. The work of M. S. Mega was supported by the National Institute on Aging (NIA) under Award K08AG100784. The work of A. R. Toga was supported by the National Center for Research Resources under Contract RR 13642. The work of I. D. Dinov and D. W. L. Sumners was supported by the National Institutes of Health under Contract P20 MH57180.

I. D. Dinov was with the Laboratory of Neuro Imaging, Division of Brain Mapping, Department of Neurology, University of California at Los Angeles School of Medicine, Los Angeles, CA 90095 USA. He is now with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA 90032 USA.

M. S. Mega, P. M. Thompson, R. P. Woods, E. L. Sowell, and A. W. Toga are with the Laboratory of Neuro Imaging, Division of Brain Mapping, Department of Neurology, University of California at Los Angeles School of Medicine, Los Angeles, CA 90095 USA (e-mail: toga@loni.ucla.edu).

D. L. Sumners is with the Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510 USA.

Publisher Item Identifier S 1089-7771(02)00029-8.

1089-7771/02$17.00 © 2002 IEEE

A *stereotaxic polynomial warp* [39] is a three-dimensional polynomial displacement vector field that brings into register two volumetric data sets (template and target). The field is usually obtained by minimizing certain (cost) functions (e.g., least squares, standard deviation, etc.). In general, the higher the degree of freedom of the field (higher order polynomial coefficients), the better the registration. However, the image distortion increases with increases in the polynomial order of the field. Here, we address the following question: Is there an ultimate upper bound on the number of parameters beyond which the distortion of the field is too large to tolerate? The answer to this question depends on the template, target, optimization procedure, warping approach, and measure of goodness of registration.

There are a number of warping techniques in the literature. Two major classes of warps can be identified based on the main driving force of the minimization procedure producing the final deformation field. The *density-based* techniques are ones where image intensities are the core of the optimization algorithms. They are subdivided into *elastic warps* [4], [7], *viscous warps* [8], trigonometric function-based alignment, SPM [20], using the eigenvectors of the Laplacian operator $\bigtriangledown^2$, and *polynomial warps* [41], [39].

The second main class of warping methods is known as *fiducial-based*. A collection of fiducial points [6], curves [3], or surfaces [34], [35] is used as a constraining device of the spatial deformation. Then various deformable templates [7] or fluid models [35] are used to drive the data-target image similarity function to a minimum, subject to the boundary conditions imposed by the fiducial landmarks. In such cases, the image-similarity function includes a regularization term (e.g., for fluid and elastic warps, one may choose the following regularization factor $\int_\Omega ||L(\Phi)||^2\,dx$, where L is an operator, determined by the kinetics of the elastic object, which assures that structures are not broken apart [7]).

In general, a density-based warping scheme is described as follows. If $T(v)$ and $D(v)$ are the template and data intensities at a voxel $v = (p, q, r)$, v is in the image domain Ω, the deformation field Φ_v^n at iteration n, and location v is obtained by minimizing a cost function

$$C(\Phi^n) = S\Big(T(v),\, D(v + \Phi_v^n)\Big) + R(\Phi_v^n) \tag{1}$$

where S is an image similarity measure and R is a regularization term. Common image similarity measures are standard deviation of ratio images, least squares, least squares with intensity rescaling automated image registration (AIR, [39]), mutual information [33], [37] or cross correlation, [24]. The regularization factor in (1), [32] insures the smoothness of the induced displacement field. The displacement field Φ_v^n is computed iteratively in a hierarchical manner by measuring the similarities between the data and target images using a coarsely sampled vector grid on sub-sampled or filtered instances of the data and template volumes. The resolution is then increased whenever the similarity measure is below a fixed threshold value. At every run, trilinear interpolation is used to determine Φ_v^n based on its eight nearest neighbors $\Phi_{p,\,q,\,r}^n$ [24]. A table of coefficients determines the warping field. For example, an affine field has the particular form

$$\Phi_v = \Phi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \\ f_3(x, y, z) \end{pmatrix}$$

where each $f_k(x, y, z) = a_{k,0} + a_{k,1}x + a_{k,2}y + a_{k,3}z$: $R^3 \longrightarrow R$ is an affine function. The (globally determined) 12 parameters $a_{k,l}$, $1 \le k \le 3$, $0 \le l \le 3$, minimizing the cost function $C(\Phi)$ define an affine spatial displacement field that brings the data and target into (affine) register, modulo the choice of a measure of image similarity (S) and a regularization factor (R).

The density-based warps have the advantage that, in general, they do not require human interaction and are applicable for a variety of data sets. On the other hand, the fiducial-based deformations can handle very difficult data and may allow for incorporating prior knowledge about the data structure into the registration model.

C. Wavelet Thresholding

We are now interested in developing a quantitative approach for analyzing the performance of various image registration techniques. This assessment tool should be independent of the value of the cost function employed in the iterative warping algorithm since this function always attains a minimal value at the final step of the alignment process. Our wavelet-space analysis appeals to the efficiency of wavelet data representation and the image denoising properties of the function estimators obtained by wavelet thresholding (shrinkage). The usefulness of wavelets as a basis for function representation is well documented [11], [17]. The foundation of this efficient representation is the fact that only a few of the wavelet coefficients adequately encode the energy of a signal. The wavelets corresponding to these few coefficients are well localized in both time and frequency domains. Thus, small local changes in the signals will affect only a few of the corresponding wavelet coefficients. The image denoising characteristics of wavelets are an additional property that will now be discussed in some detail.

Suppose our data observations $\{y_i\}_{i=0}^{N-1}$ are obtained by adding noise to some real signal $f(t_i)$ for each time t_i. That is, $y_i = f(t_i) + \epsilon_i$, where $\epsilon_i \sim \aleph(0, \sigma^2)$ are independent (and identically distributed) Gaussian random variables. To recover the unknown function f, having the data y alone, we construct function estimators $\hat{f}$. The *Risk* function measures estimator performance by the average quadratic loss at the sample points

$$R(\hat{f},\, f) = \frac{1}{N} E\left(||\hat{f} - f||^2\right) = \frac{1}{N} E\left(\sum_{i=0}^{N-1} \Big[\hat{f}(t_i) - f(t_i)\Big]^2\right) \tag{2}$$

where E is the expectation of the misfit between the function and its estimator. Notice the similarity between this *Risk* function and the variance of unbiased estimators. Preference will be given to estimators $\hat{f}$ having small *Risk* measures (i.e., small "variances").

Let W represent the matrix of the discrete wavelet transform (DWT) [10], [15]. The time-domain model $y = f + \epsilon$ becomes $w = \theta + z$ in wavelet space, where $w = Wy$, $\theta = Wf$, and $z = W\epsilon$. If U is a subset of the set of all wavelet coefficients $\{w\}$ of the data y [16], define *selective-wavelet (SW) reconstruction* estimators by

$$\hat{f} = \sum_{w_{j,k} \in U} w_{j,k} W_{j,k} \tag{3}$$

where $w = \{w_n\}_{n=0}^{N-1} = \{w_{j,k}, 0 \le j \le J, 0 \le k \le 2^j - 1\}$ (the j codes for the frequency and the k subscript indexes the location of the corresponding wavelet coefficients), and $W_{j,k}$ represents the nth, i.e., the (j, k)th, row of the orthogonal wavelet transform matrix W [15]. We are interested in expressing the data y in terms of (some of) its wavelet coefficients w in frequency space. The reason for that will be shown later, but for now, such decomposition allows compression, denoising, and "optimal" representation of the observed signals.

II. Methods

Transform methods for characterizing image registration use the concise image representation induced by various discrete transformations, e.g., the discrete fractal transform (DFT) and DWT to reduce the data complexity and dimensionality [5], [12], [16]. In this paper, we exclusively employ the DWT to evaluate image registration. Before we describe our three warp classification strategies, we will discuss the foundations and properties of our frequency-adaptive function estimators.

A. Frequency-Adaptive Wavelet Space Thresholding

As in Section I-C, let $y = f + \epsilon$ and $w = \theta + z$ be the decompositions of the observed data in time and wavelet space, respectively. The noise terms ϵ and $z = W\epsilon$ are normally distributed with mean of zero and variance-covariance matrix $\sigma^2 I$, where σ^2 is the noise level (the variance matrices for ϵ and z are identical because W is orthogonal). Let $n_j = 2^j$ be the number of wavelet coefficients at the jth frequency level. Define the frequency-adaptive wavelet thresholding by

$$\begin{aligned}\eta_{\lambda_j}(w_{j,k}) &= \operatorname{sgn}(w_{j,k})\Big[|w_{j,k}| - \lambda_j\Big]_+ \\ &= \operatorname{sgn}(w_{j,k}) \max\Big(\{|w_{j,k}| - \lambda_j\}, \{0\}\Big)\end{aligned} \tag{4}$$

where $\lambda_j = \sigma\sqrt{2\ln(2n_j + 4)}$ and $\operatorname{sgn}(\cdot)$ is the sign function. The threshold level defined by Donoho and Johnstone [17], $\lambda_j = \sigma\sqrt{2\ln(N)}$ is different from the one we use. Our threshold level varies with the frequency characteristics of the wavelet coefficient, the j index. One can show [15] that the induced function estimator $\hat{f} = W^{-1}\eta_\lambda W(y)$ is essentially uniformly optimal. That is, no other SW reconstruction estimator can uniformly, across all possible "nice" signals, outperform it in terms of minimizing the *Risk* functional. An SW reconstruction scheme is any function for choosing a subset of the wavelet coefficients and using these as signal representatives. How good this data representations is can be assessed by inverting the wavelet transform on the chosen subset of wavelet parameters. The notation $\hat{f} = W^{-1}\eta_\lambda W(y)$ means that the operators W (wavelet transform), η_λ (wavelet space shrinkage), and W^{-1} (inverse wavelet transform) are applied to the observed data y, in this particular order, to obtain a function estimator, in the time domain. All of the results for the analysis reported in compressed wavelet space are obtained by applying only the operators W and η_λ, in that order, to the observed data y. Another point of interest is that the wavelet transform (and its inverse) are orthogonal. Therefore, $\|W^{-1}(W(y))\| = \|W(y)\| = \|y\|$, and if all of the wavelet parameters are employed (no thresholding), then the wavelet-space and the time-domain analyses produce the same results.

Symbolically, let $\Re_{N,\sigma}(\mathrm{SW}, f)$ denote the minimal *ideal Risk* of an estimator under any SW reconstruction. Note that $E(\|\hat{f} - f\|) = E(\|\hat{\theta} - \theta\|)$, where $\hat{\theta} = \eta_\lambda(w)$ and $\Re_{N,\sigma}(\mathrm{SW}, f) = 1/N \sum_i \min(\theta_i^2, \sigma^2)$, [15]. We then have an upper bound for the *Risk* function, obtained by (4), of the estimator $\hat{f}$ as follows:

$$R(\hat{f}, f) \le \big(1 + 2\ln(N+4)\big)\left(\frac{\sigma^2 \ln N}{N} + \Re_{N,\sigma}(\mathrm{SW}, f)\right). \tag{5}$$

A sketch of the proof of this fact is attached in the Appendix. The entire proof appears in [15] and basically follows the ideas presented in [16] applied to the frequency-adaptive threshold level $\lambda_j = \sigma\sqrt{2\ln(2n_j + 4)}$. In a similar fashion, replacing the coefficient of n_j by β, we obtain the following approximation:

$$\inf_{\hat{\theta}} \sup_{\theta \in R^N} \frac{\frac{1}{N} E\left(\|\hat{\theta} - \theta\|^2\right)}{\sigma^2 \frac{\ln N}{\beta N} + \frac{1}{N}\sum_i \min(\theta_i^2, \sigma^2)} \sim \frac{2\beta}{\beta + 1} \ln N. \tag{6}$$

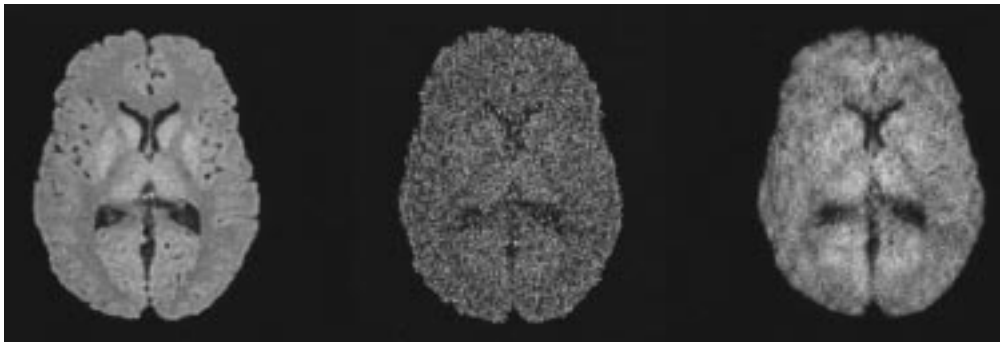
In other words, if $\lambda_j = \sigma\sqrt{2\ln(\beta n_j + 4)}$, $\eta_{\lambda_j}(w_{j,k}) = \operatorname{sgn}(w_{j,k})[|w_{j,k}| - \lambda_j]_+$, and $\hat{f} = W^{-1}\eta_\lambda W(y)$, we have the *Risk* functional of $\hat{f}$ bounded above by

$$R(\hat{f}, f) \le \big(1 + 2\ln(\beta N + 4)\big)\left[\frac{\sigma^2 \ln N}{\beta N} + \Re_{N,\sigma}(\mathrm{SW}, f)\right]. \tag{7}$$

This can be compared to the spatially adaptive shrinkage of Donoho and Johnstone [17], where $\lambda = \sigma\sqrt{2\ln(N)}$, $\eta_\lambda(w_{j,k}) = \operatorname{sgn}(w_{j,k})[|w_{j,k}| - \lambda]_+$, the function estimator $\hat{g} = W^{-1}\eta_\lambda W(y)$, and $R(\hat{g}, f) \le (1 + 2\ln(N))$ $[\sigma^2/N + \Re_{N,\sigma}(\mathrm{SW}, f)]$. Since $\Re_{N,\sigma}(\mathrm{SW}, f) \sim O(\ln N/N)$, the spatially adaptive estimators [16] have *Risk* measures of the order of $O(2(\ln N)^2/N)$. Our frequency adaptive function estimators have $R(\hat{f}, f) \sim O(2\ln(\beta N)\,((\beta+1)/\beta)\,(\ln N/N))$ $\sim O((2(\beta+1)(\ln N)^2)/\beta N)$ for $\beta \ll N$. Hence, for large β (but $\beta \ll N$), we have essentially optimal estimators.

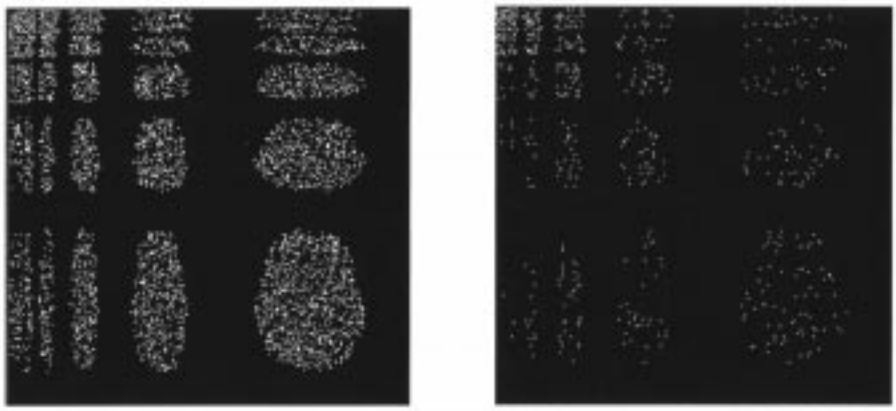
The differences in the upper bound of the *Risk* measures in (5) and (7) are caused by the differences of the threshold levels (λ_j values). The second one, i.e., (7), has a smaller upper bound for large β parameters. In practice, we have used $\beta = 2$, $\sigma = 10$, and the thresholding (shrinkage) scheme $\lambda_j = \sigma\sqrt{2\ln(\beta n_j + 4)}$.

To illustrate the power of frequency-adaptive wavelet shrinkage, we present an example involving degrading a two-dimensional (2-D) magnetic resonance imaging (MRI) image



(a) (b) (c)

Fig. 1. (a) Original MRI image, (b) original $+\aleph(0,\ 30)$, and (c) wavelet-space denoised recovered image. SNR of the wavelet space synthesized image (c) has increased to 13.7, compared to the noisy image (b) with $\mathrm{SNR} = 3.7$. This fourfold increase of image quality, measured by SNR, is indicative of the nice properties of the wavelet-space function representation.



(a) (b)

Fig. 2. (a) Wavelet transform of the original MRI image. (b) Frequency-adaptive compressed (thresholded) wavelet transform. All but about 6% of the wavelet coefficients survived the shrinkage process according to the scheme in (4).

by adding to it pixel-wise Gaussian noise of mean zero and variance $\sigma^2 = 30$. The signal-to-noise ratio (SNR) of the noisy image was reduced to 3.7 ($\mathrm{SNR} = 10\log_{10}\left(\sigma^2_{\mathrm{signal}}/\sigma^2_{\mathrm{noise}}\right)$). Following the compressed wavelet space analysis, the SNR is improved to 13.7, about a fourfold increase in image quality. Fig. 1, depicts the original image, noisy (observed) data, and recovered (denoised) image.

For this example, only about 6% of the wavelet coefficients were "selectively" chosen using (4), Fig. 2, and used to synthesize the new image estimate. Despite the fact that the synthesized image is not an exact replica of the original data, it still shows more fine detail compared to the observed signal. This is used as an empirical motivation for analyzing goodness of image registration in compressed wavelet space. The data representation in wavelet space is robust, concise, and computationally efficient. One may wonder if there are differences between the classical Fourier type of analysis and the proposed compressed wavelet space signal representation. First, the wavelet basis functions are well localized in time and frequency simultaneously [11]. Thus, local variations of the observed signals affect only a small number of the wavelet coefficients. Second, the frequency-adaptive wavelet shrinkage η_{λ_j} [see (4)] preserves wavelet parameters in the extremely high-frequency ranges, provided these have large amplitudes (are significant). In other words, wavelet compression does not necessarily induce image blurring. High-frequency information can be recovered, which implies that fine image details, e.g., edges, can be captured and preserved.

B. Wavelet-Based Warp Classification

There are many different goodness of registration criteria one can employ depending upon the specific aim and interpretation of the warping procedure. Here, we present three different approaches for evaluating warp performance in compressed wavelet space.

For notational convenience, suppose we are analyzing the performance of r image registration methods $\{\Phi_k\}_{k=1}^{r}$ having the volumetric data of s subjects $\{S_p\}_{p=1}^{s}$, each scanned a times under different "activation paradigms" (brain states) $\{A_q\}_{q=1}^{a}$. The data obtained by scanning subject p under the condition q is denoted by S_pA_q.

1) Triangle Scheme (Δ) for Warp Classification: The fundamental assumption of the *triangle* (Δ) warp classification is that a "good" warp is one that minimizes the destruction of self-similarities of the data, with respect to some metric, yet deforming (spatially) the data onto an image having self-symmetries resembling those of the target of the warp. In other words, we minimize the global distortion needed to transform the data exactly into the reference space. To identify image similarities and symmetries of stereotaxic data, we use a three-dimensional

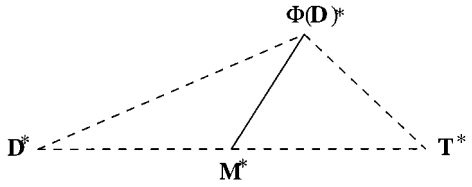


Fig. 3. Triangle warp classification scheme in compressed wavelet space. The distance between the warped resliced volume $\Phi(D)^*$ and the midpoint M^* between the data, D^* and the target T^* is used to quantify image registration in compressed wavelet space.

(3-D) DWT based on a 20-coefficient Daubechies wavelet filter [10], [26]. Signal "self-similarities" and "affine self-symmetries" for a function f are defined as any collection of contractive affine maps from the domain of f onto itself [23], [5]. By "contractive" maps, we mean maps that bring points closer together in the range space than they are in the domain space. The close relation between the discrete wavelet and the DFTs [12] justifies the use of the terminology "self-similarities" and "affine self-symmetries" in regard to the wavelet coefficients of the data.

To quantitatively evaluate a spatial deformation Φ of a data D to a target volume T, we use the following transform metrics evaluated on the compressed wavelet transforms of D, T, and $\Phi(D)$ (Fig. 3). The midpoint M of the *symmetry* line connecting the "points" D and T (in the compressed wavelet space[1]) is our "gold-standard" representing the best warp since it minimizes the total image distortion in registering D to T. If $\Phi(D)^*$ is far from the symmetry line connecting D^* and T^*, in compressed wavelet space, there would be a great deal of "unnecessary" deformation imposed on the data (Fig. 3). Also, the final result of the warping will not be close to the target. A fact that diminishes the purpose of image alignment. In essence, the total distance[2] $d(D^*, M^*) + d(M^*, T^*) < d(D^*, \Phi(D)^*) + d(\Phi(D)^*$, and $T^*)$, for $\Phi(D)^*$ lying away from the line connecting D^* and T^*. Thus, the amount of total warping [as measured by $d(D^*, \Phi(D)^*) + d(\Phi(D)^*, T^*)$] is unnecessarily larger than the ultimate total displacement $(d(D^*, M^*) + d(M^*, T^*) = d(D^*,$ and $T^*))$. The distance between M^* and $\Phi(D)^*$ is, therefore, a measure of goodness of the image registration. Why use the midpoint between D^* and T^*, which is computed as a regular arithmetic mean, instead of, say, the target T^* itself? After all, one of the ultimate goals of image alignment is to make the data D^* resemble the target T^*. As stated earlier, we base our analysis on the fact that there are different goals in data registration. For example, if comparing two groups of volumetric functional brain data, we may be inclined to look at the distance between the barycenters of the two groups of post-registered data instead of looking at how close each individual (or an entire group) is to the atlas (target of the alignment). We discuss such situations further in Sections II-B.2 and II-B.3. Choosing the actual target T^* as a gold standard addresses only one type of reasoning behind the process of image registration. Also, we are evaluating image alignment using the actual images and their corresponding reslices in target space, and not the induced displacement vector fields. A "perfect" warp of a data onto a target would then render the entire information content of the data lost in the registration process. Of course, regularization factors [see (1)] in the cost function employed in the optimization process of computing the warping field insure that the displacement field is not discontinuous and perhaps has no trivial Jacobian. Hence, in the case when we do not use the actual displacement vector fields, but only the native data, the target, and their resliced warped representations, we need to penalize against both loss of data self-similarities ("perfect" image matching) and imperfect reslice-target fit. Hence, the choice of the midpoint on the symmetry line.

For quantitative analysis of image registration based on a family of registered data sets, one uses the average distance between M^* and $\Phi(D)^*$ for every warping technique to obtain the desired warp ranking. The distances between the midpoint and warped resliced volume is inversely proportional to the cumulative degree of distortion of the "self-similarities" of the data by the warp. More precisely

$$\Delta(\Phi_k) = \frac{1}{s \times a} \sum_{p,q} d\Big(\Phi_{k_{p,q}}(S_p A_q)^*,\, M^*_{p,q}\Big) \qquad (8)$$

where $d(\bullet, \bullet)$ is a distance function, we have used the l^2 norm, $M^*_{p,q}$ is the midpoint between $S_p A^*_q$ and T^* in compressed wavelet space, and $\Phi_{k_{p,q}}(S_p A_q)^* = \eta_{\lambda_j} W(\Phi_{k_{p,q}}(S_p A_q))$ is the reslice of the data set $S_p A_q$ in target space using the kth alignment strategy and its corresponding displacement field $\Phi_{k_{p,q}}$ for this image. Note, that the measure $\Delta(\Phi_k)$ can also be computed in the image space by the formula

$$\Delta(\Phi_k) = \frac{1}{s \times a} \sum_{p,q} d\Big(\Phi_{k_{p,q}}(S_p A_q),\, M_{p,q}\Big).$$

Significant differences between these two measures, time versus compressed wavelet space, will be illustrated in the results. This warp classification approach is referred to as the "triangle" scheme because the data, target, and warped resliced volumes represent, in general, the vertices of a nontrivial triangle (two-cell) in compressed wavelet space (Fig. 3). In the case when a large number of volumes are to be co-registered, one may argue that the best *target* data set to use is one whose compressed wavelet space transform is the l^2-centroid (barycenter) of $\{\eta_{\lambda_j} W(S_p A_q)\}_{p,q}$.

This type of warp classification is useful for analyzing image registration based on small groups of data, when the goal of the warping is to bring all the data into a common reference space without significantly distorting the self-affine characteristics of the data, e.g., rigid-body-type registration without considerable local deformation. For example, this type of warp classification is extremely applicable for analyzing statistical differences between groups of functional brain data. In these cases,

[1]* denotes data in compressed wavelet space. Observe the duality between image and compressed wavelet space $X \leftrightarrow X^* = \eta_{\lambda_j} W(X)$ for each image X.

[2]If $D^* = \{a_{j,k}\} = \{\eta_{\lambda_j} W(D), 1 \le k \le 2^j\}$, $\Phi(D)^* = \{b_{j,k}\} = \{\eta_{\lambda_j} W(\Phi(D)), 1 \le k \le 2^j\}$ and $T^* = \{c_{j,k}\} = \{\eta_{\lambda_j} W(T), 1 \le k \le 2^j\}$ are the compressed wavelet representations of the data (D), the target (T), and the warped resliced data $(\Phi(D))$, then the l^2 distance between any pair is defined according to the usual rule

$$d(D^*, \Phi(D)^*) = \left(\sum_{j,k} |a_{j,k} - b_{j,k}|^2\right)^{1/2}$$

etc., and the midpoint in the compressed wavelet space is defined by $M^* = \{(a_{j,k} + c_{j,k}/2)\}_{j,k}$.

we prefer volume registration techniques that least distort the Gaussian field characteristics [13] so that of statistical models remain valid in atlas space, as they are in native data space.

2) Cluster Group Classification (CGC): The second type of warp classification is applicable when we want to compare a collection of registration algorithms applied to a group of volumetric data and we are primarily interested in identifying common regions, fiducial points, curves, or surfaces across the entire family of data sets. In this case, we allow a lot of local distortion, and penalize registration strategies that fail to cluster all the data. That is, a good warp in this sense is one that makes any data volume as similar to any other in the group as possible. Such a scheme for quantitative warp evaluation will be biased toward very local and highly nonlinear deformations and may or may not be always desirable. The CGC classifying functional is the *diameter* of the smallest convex set containing the warped data. The smaller the diameter, the better the warp.

As with the triangle scheme, distances between warped resliced volumes are measured in compressed wavelet space and the diameter measure is computed as the largest pairwise distance between two volumes in the group. Symbolically

$$\mathrm{CGC}(\Phi_k) = \mathrm{diam}\Big\{\Phi_{k_{p,\,q}}(S_pA_q)^*|p,\,q\Big\} \tag{9}$$

where $\Phi_{k_{p,\,q}}$ is the kth alignment method for the data set S_pA_q and diam denotes the diameter of the smallest convex set containing the warped resliced data in wavelet space. In practice, the diameter is computed as the maximum pair-wise distance $d(\Phi_{k_{p_1,\,q_1}}(S_{p_1}A_{q_1})^*,\ \Phi_{k_{p_2,\,q_2}}(S_{p_2}A_{q_2})^*)$, with d being the l^2 metric, as before. The analogous image space analysis may be done by evaluating d on the time-domain images: $d(\Phi_{k_{p_1,\,q_1}}(S_{p_1}A_{q_1}), \Phi_{k_{p_2,\,q_2}}(S_{p_2}A_{q_2}))$.

3) Spread Group Classification (SGC): Another common use of image alignment is to be able to identify differences between activation paradigms for functional and/or anatomical developmental effects for structural brain studies. The SGC warp evaluation scheme uses the goodness of warp criterion proposed by Kjems *et al.* [24], where the best warp is the one that "makes the differences between different brain states (activation paradigms, developments stages, etc.) most clear and profound." In other words, after the data sets are registered to a reference atlas, the differences between them stand out clearly and the locations of these variations can be easily identified. The SGC warp classifying functional we use in our studies is the distance between the barycenters (centers of mass) of the groups of warped data for a fixed brain state or activation paradigm. More precisely, the SGC functional is defined by the average

$$\mathrm{SGC}(\Phi_k) = \frac{1}{\binom{a}{2}} \sum_{1\le q_1<q_2\le a} d\Big(BC^*_{k,\,q_1}, BC^*_{k,q_2}\Big) \tag{10}$$

where $BC^*_{k,\,q} = (1/s)\sum_{p=1}^{s} \Phi_{k_p}(S_pA_q)^*$ is the barycenter of the group of data (in compressed wavelet space) for a fixed brain state q under warp k, $\binom{a}{2} = a!/((a-2)!2!)$ is the usual binomial coefficient, and d is the l^2 norm. The best warps chosen by this approach will have large SGC values, as opposed to the small Δ and CGC values of the first two classification schemes, indicating large between-group distances for different activation paradigms (brain states).

In summary, the following protocol describes the step-by-step process for obtaining a quantitative assessment of quality of volumetric image registration according to the wavelet-based approach presented here.

Step 1) Compute the wavelet transform of all data volumes, their warped-resliced representations, and the target data ($\hat{y} = Wy$).

Step 2) Apply a spatially adaptive, a frequency-adaptive, (4), or other wavelet shrinkage technique to the wavelet transformed data ($t_{\hat{y}} = \eta_\lambda(\hat{y})$).

Step 3) Evaluate the desired warp classification functionals (Δ, CGC, SGC), according to the formulas listed in (8)–(10).

All of the results presented in the following section were obtained by this protocol using a 20-coefficient-based wavelet Daubechies filter, i.e., Step 1). Our software contains about 17 different wavelet bases available (e.g., Spline, Daubechies, Coiflets, and others). We have looked for significant differences in the induced warp ranking using four- and 20-coefficient Daubechies filters and have found none. In general, the number of coefficients used to constrict the wavelet bases determines the time localization and smoothness of the wavelets [11].

III. Results

We now present the complete warp classification analysis on two types of data: functional [positron emission tomography (PET)] and structural (MRI). The two examples use different collections of warps; however, both include the triangle, CGC, and SGC warp rankings for the corresponding registration algorithms. For the second study, we perform a manual fiducial-based assessment of the wavelet-based warp ranking.

A. Warp Classification Based on Functional Brain Data

The functional PET data used for this first example consists of six subjects, each being scanned two times under different activation paradigms, i.e., right- and left-hand finger tracking tasks. Informed consent was obtained from all subjects, and the study protocol was approved by the University of California at Los Angeles (UCLA) Human Subject Protection Committee [41]. The top row of Fig. 4 represents axial views of the six left-hand volumes and the bottom row shows the corresponding right-hand scans. We "sequentially" used four polynomial warps (AIR 3.0, [39]) of 7, 12, 30, and 168 parameters, i.e., the output of low-order displacement fields were fed in as initialization parameters for higher order warps. The first two alignments are affine and the last two are second- and fifth-order polynomial warps.

Table I depicts the compressed wavelet space analysis, and the values of each of the three classifying functionals across the four different registration methods. The analogous image space (time-domain) analysis, where we compute Δ, CGC, and SGC values using all of the wavelet parameters, is shown in Table II. Recall in Section II-B that smaller values for Δ and CGC and larger values for SGC indicate better registration in terms of the corresponding *a priori* goodness of warp criteria.

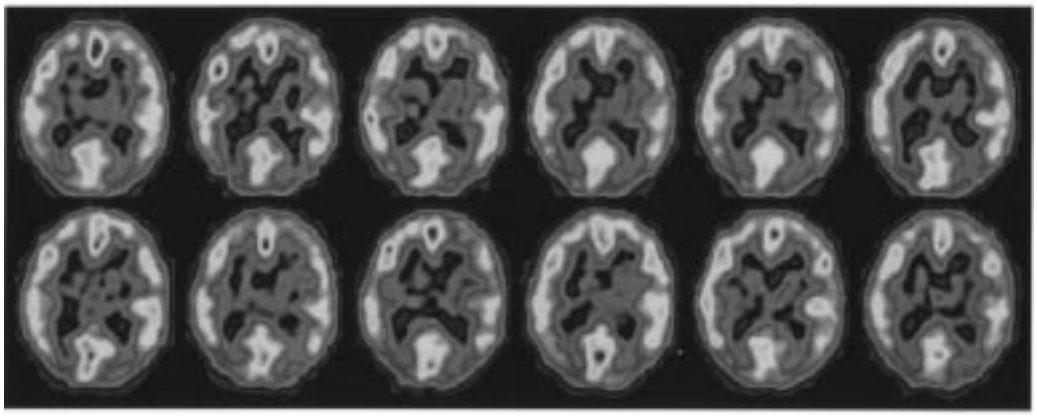


Fig. 4. Axial views of the PET data prior to warping (six subjects). (top row) Left-hand finger tracking task. (bottom row) Right-hand finger tracking task.

TABLE I
(COMPRESSED) WAVELET-SPACE WARP CLASSIFICATION USING TRIANGLE (Δ), CGC, AND SGC CLASSIFICATION SCHEMES. AIR 7-, 12-, 30-, AND 168-PARAMETER POLYNOMIAL WARPS ARE APPLIED TO FUNCTIONAL PET STUDIES, RIGHT- VERSUS LEFT-HAND FINGER TRACKING TASK

Warp / Funct.	7p	12p	30p	168p
Ranking Δ	4 0.8050	2 0.7534	1 0.6800	3 0.7944
Ranking CGC	1 7.5781	4 49.1949	3 49.0140	2 48.9079
Ranking SGC	4 3.3025	1 8.5697	2 8.4335	3 8.3436

TABLE II
IMAGE-SPACE WARP CLASSIFICATION USING TRIANGLE (Δ), CGC, AND SGC CLASSIFICATION SCHEMES. AIR 7-, 12-, 30-, AND 168-PARAMETER POLYNOMIAL WARPS ARE APPLIED TO FUNCTIONAL PET STUDIES, RIGHT- VERSUS LEFT-HAND FINGER TRACKING TASK

Warp / Funct.	7p	12p	30p	168p
Ranking Δ	3 0.7163	2 0.6751	1 0.6356	4 0.7534
Ranking CGC	1 7.7989	4 49.2595	3 49.0797	2 48.9898
Ranking SGC	4 3.3372	1 8.5793	2 8.4431	3 8.3560

There are some similarities and some differences between the wavelet- and image-space analyses. We will use the former for our quantitative warp ranking because warp ranking in compressed wavelet space is consistent with other types of registration ranking, e.g., canonical variate analysis (CVA) [24], and because of its image denoising, optimality, and compression properties (Section II-A).

For this study, one selects the 30-, seven-, or 12-parameter warp as the best alignment technique using Δ, CGC, or SGC, respectively, depending on the goal of the registration procedure. The three different goals and registration scenarios, listed in Section II-B, influence differently the final choice of the best image-alignment method. For example, if the aim of the study was to warp the two groups of functional data into an atlas space where the differences between the two groups are most clear and profound, one relies on a measure like the SGC scheme, and selects the 12-parameter polynomial alignment. If one's intentions are to simply have all data sets into a common anatomical space, without regard to possible decline in the between group variance, then the seven-parameter registration (under CGC) may be the method of choice. Last, if we want to minimize the global distortion done to the entire collection of data sets, in the process of aligning them into a reference frame, we select the 30-parameter polynomial warp (under Δ). Differences in the warp ranking between Δ, CGC, and SGC functionals should not be viewed as inconsistencies or weaknesses. On the contrary, these warp classifying functionals are systemic and complementary in nature. Their differences are indicative of the empirically clear thesis that the purpose of image alignment influences what we call a *good* or a *bad* warp. Before selecting a warping strategy, one needs to answer the question of why is image registration needed in the first place, i.e., what are the ultimate goals of image alignment of the real data at hand?

B. Warp Ranking Based on Structural Brain Data

A very challenging problem in the area of brain mapping and brain evolution is the question of brain development and how early anatomical changes may be affecting the complex brain function and metabolism. At the macroscopic level, we are interested in determining the evolution of cortical and subcortical interior brain regions over time.

To illustrate the usefulness of our wavelet analysis of image registration (WAIR) tool,[3] we discuss the following developmental example. The MRI volumetric data of two groups of nine subjects each are obtained. Group 1 contains children ages 7–9 and group 2 contains children ages 14–16 (see Figs. 5 and 6). All subjects were recruited as normal controls for a large multidisciplinary neurodevelopmental research center. All of the children were right handed and each was screened for neurological impairments and for any history of learning disability or developmental delay. Informed consent was obtained from all children and their parents. A gradient-echo (SPGR) T1-weighted series with $\mathrm{TR} = 24$ ms, $\mathrm{TE} = 5$ ms $\mathrm{NEX} = 2$, flip angle $= 45°$, field of view is 24 cm, section thickness of 1.2 mm, and no gaps. Imaging time for this series totaled approximately 20 min.

[3][Online]. Available: http://www.loni.ucla.edu/~dinov/WAIR.html

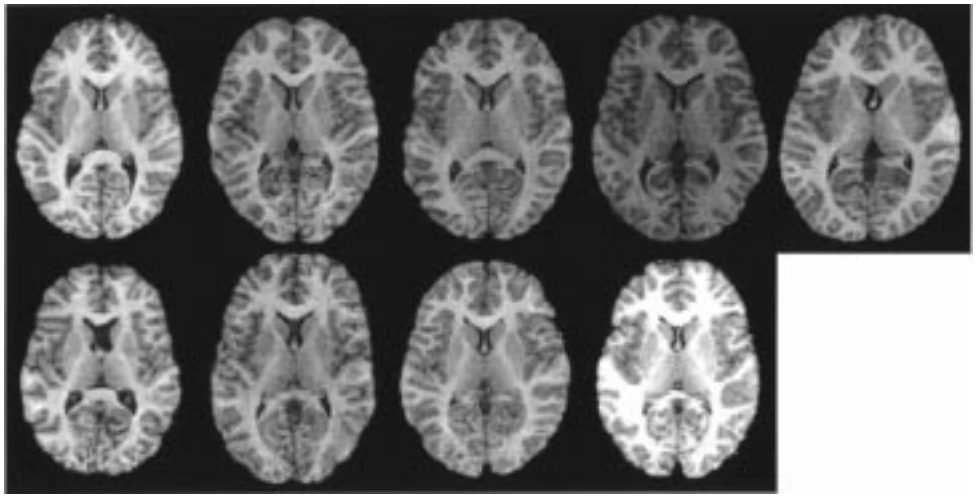


Fig. 5. Axial views of the structural MRI data for group 1, children ages 7–9.

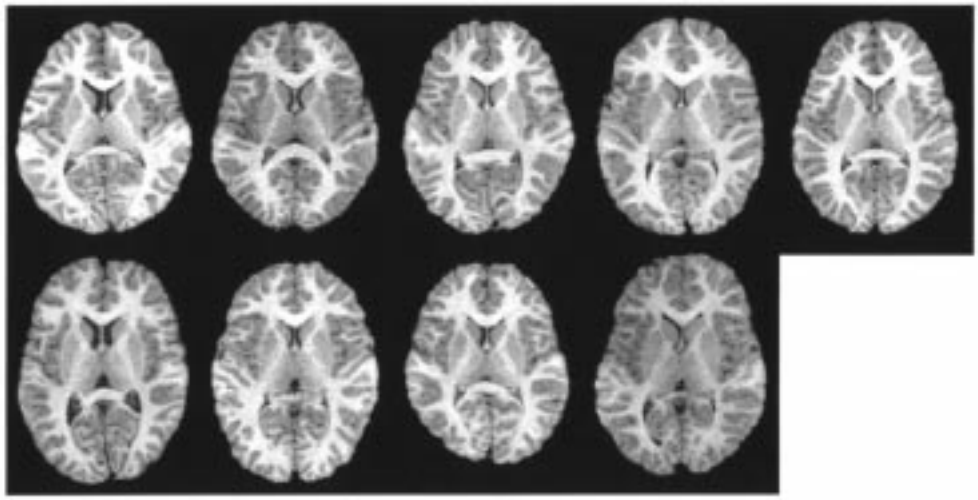


Fig. 6. Axial views of the structural MRI data for group 2, children ages 14–16.

The aim of this study is to identify the regions showing statistically significant variations between the two groups (developmental changes). There are several different ways to approach this problem [21], [42], however, we will be only interested in determining which of four available image registration methods should be used to warp all the data in a common anatomical reference space. The four warping techniques used here were: 1) polynomial AIR 12 [39]; 2) polynomial AIR 30 [39]; 3) MNI nonlinear (point-based) fiducial-based [9]; and 4) SPM nonlinear trigonometric [20] warps.

Tables III and IV show the compressed wavelet and corresponding image-space analysis on these data using the three warp classification schemes presented in Section II-B. Again, as in the previous example, smaller values of Δ and CGC and larger values of SGC indicate better registration. Depending on the prior goodness of warp hypothesis (Section II-B), one selects AIR 12 (using Δ), SPM (using CGC), or MNI (using SGC) warps. If *a priori*, the goodness of warp criterion chosen is to minimize the global deformation, across all data sets, one selects the 12-parameter polynomial alignment and the Δ classification scheme. In the case when the ultimate goal is to cluster the data together in an atlas space, without regard to the required total deformation, one selects the SPM trigonometric warp under the CGC classification scheme. Finally, if our aim is to investigate differences between the two age groups, we may choose the SGC scheme as a quantifier of goodness of registration. In this case, the MNI fiducial-based spline warp yields the largest

TABLE III
(COMPRESSED) WAVELET-SPACE WARP CLASSIFICATION USING TRIANGLE (Δ), CGC, AND SGC CLASSIFICATION SCHEMES. AIR 12- AND 30-PARAMETER POLYNOMIAL WARPS, MNI NONLINEAR SPLINE WARP, AND SPM NONLINEAR TRIGONOMETRIC WARPING ARE APPLIED TO STRUCTURAL MRI DEVELOPMENTAL STUDIES, NINE SUBJECTS SCANNED AT AGES 7–9 AND ANOTHER GROUP AT AGES 14–16

Warp / Funct.	AIR_12	AIR_30	MNI	SPM
Ranking Δ	1 0.4323	2 0.4330	4 0.8094	3 0.4980
Ranking CGC	3 6.3119	2 6.2800	4 24.5642	1 5.8295
Ranking SGC	3 0.7906	4 0.7326	1 6.3942	2 1.2897

TABLE IV
IMAGE-SPACE WARP CLASSIFICATION USING TRIANGLE (Δ), CGC, AND SGC CLASSIFICATION SCHEMES. AIR 12- AND 30-PARAMETER POLYNOMIAL WARPS, MNI NONLINEAR SPLINE WARP, AND SPM NONLINEAR TRIGONOMETRIC WARPING ARE APPLIED TO STRUCTURAL MRI DEVELOPMENTAL STUDIES, NINE SUBJECTS SCANNED AT AGES 7–9 AND ANOTHER GROUP AT AGES 14–16

Warp / Funct.	AIR_12	AIR_30	MNI	SPM
Ranking Δ	3 0.5197	2 0.5186	4 1.1666	1 0.5046
Ranking CGC	2 7.1874	1 7.1308	4 27.7210	3 7.5420
Ranking SGC	3 2.1656	4 2.0769	1 11.5150	2 2.8432

SGC value and it is, therefore, considered the best alignment method. Indeed, the MNI spline warp incorporates certain degree of objectiveness since it involves manual point-landmark delineations.

For this data, we also observe some differences in warp rankings between the time and wavelet space analyses. For example, the SPM trigonometric registration is ranked third in wavelet space by the triangle classifier Δ, whereas in time domain, it is ranked the best.

C. Quantitative Assessment of the Δ Scheme for the MRI Data

For the developmental example discussed in Section III-B, we conducted the following a biomedical validation of the results for the Δ classifier and the frequency-adaptive wavelet thresholding approach. The resliced volumes resulting from the four registration techniques (affine, second-order polynomial, MNI-spline, and SPM trigonometric warp) were realigned to the same target by a standard point fiducial-based 12-parameter affine alignment. For each of the (warped) resliced volumes, 18 (2×9) subjects under the four warping methods, nine clearly identifiable points [31] were manually selected on the target and the resliced data. These included points in the lateral cerebellum, corpus callosum, fourth ventricle, temporal lobe, mammillary body, and superial medial parietal lobe. Two neuroanatomists independently selected the fiducial points in a blinded design. The variations between the two delineators were small (within 10%) and did not affect the final warp rankings.

Two different quantitative measures were applied to assess the *residual* mismatch of the resliced data volumes and the target. First, the average root mean square (rms) distance (in millimeters) between the nine pairs of manually selected fiducial points was computed. Second, the *Jacobian* of the induced *residual affine* warps were calculated for each subject and alignment technique. The graphs of the RMS measure and the magnitude of the Jacobian across subjects for the 4 different alignment methods are shown in Figs. 7 and 8, respectively. Small mean values (across subjects) of the rms measure and Jacobian values about one (no contraction/expansion) indicate better initial fit and smaller *residual* deformation. With the exception of the SPM trigonometric warp, none of the measures appear statistically distinct from each other. However, if one is to choose a single alignment technique based on the results

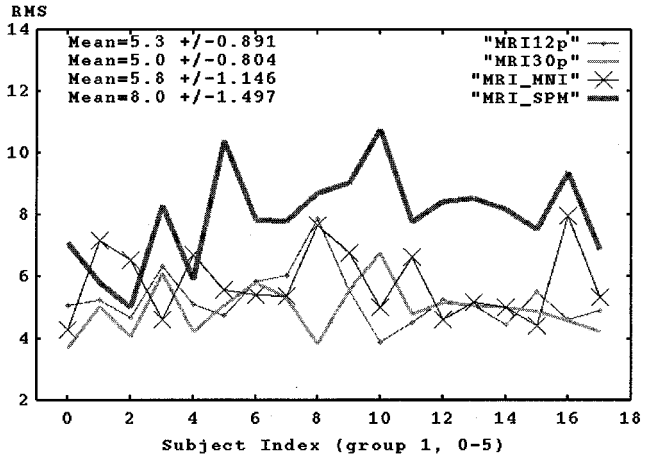


Fig. 7. Average stereotaxic fiducial point distances (rms).

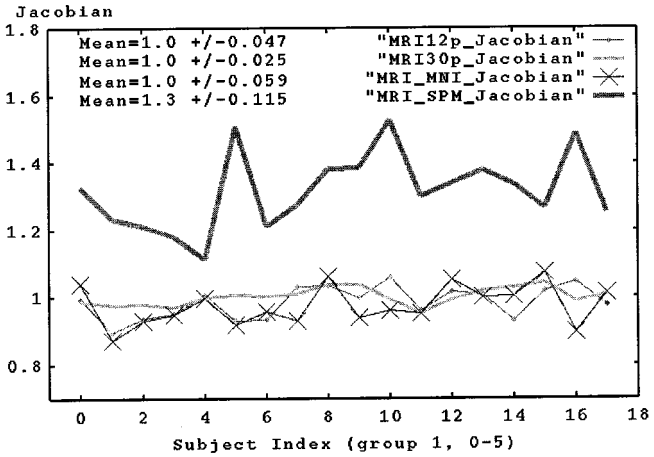


Fig. 8. Residual Jacobian magnitude.

presented in Figs. 7 and 8, clearly a polynomial alignment (first or second order) may be the final pick since they yield the smallest rms values and Jacobians of about one. Note the similarity between the rms/Jacobian measures and the ranking induced by the Δ wavelet-based warping classification scheme, which also gave preference to the affine and the 30 parameter polynomial warps (Table III). Similar anatomically driven validations for the CGC and SGC wavelet-based warp ranking schemes would be harder. These classification criteria are based on prior goodness of warp criteria, which do not require that the data perfectly matches the target (this is a condition imposed in minimizing the cost function [see(1)]), but rather that groups of resliced data are simultaneously clustered or spread apart (Section II).

IV. DISCUSSION AND CONCLUSION

There are optimal ways to select the "ideal" sub-collection of wavelet coefficients U with induced function estimators that optimize the *Risk* function $R(\hat{f}, f)$, [16], [14]. It is worth mentioning here that these SW estimators also have good image compression characteristics [11]. This means that a small number of the wavelet coefficients of the data efficiently encode the information context (energy) of the entire image.

The problem of finding good function estimators, based on the data alone, that have *Risk* functions close to the "oracle" based ideal *Risk* [17] was first approached from the classical theory point-of-view using Fourier analysis (e.g., [18]). Some of the inherited limitations of the Fourier basis functions, however, inclined Donoho and Johnstone to look at the problem of evaluating function estimators using wavelet analysis. They showed that, on discontinuous functions, their wavelet-based spatially adaptive approach produces better results and achieves rates of convergence of the order of $\ln^2 N/N$. Later, Efromovich [19] introduced an exponentially increasing frequency-adaptive wavelet thresholding having optimal characteristics and performing visually, and in terms of compression, at least as good as the Donoho and Johnstone's spatially adaptive wavelet shrinkage. The probability that a wavelet coefficient is purely due to noise increases with the increase in the frequency of the corresponding wavelet. That is why we have chosen steadily increasing wavelet thresholds [see (4)] instead of a static one, which depends heavily on the frequency component of the corresponding wavelet.

The idea of employing wavelets as efficient basis for function (signal) representation has recently gained increasing popularity. New methods exploiting the useful properties of wavelets for image compression [12], image enhancement, and denoising [17] and image registration [1] have been introduced, tested, and validated. In this paper, we present yet another application of the DWT to characterize stereotaxic image registration. This new approach appeals to the low computational complexity of the wavelet signal decomposition and the explicit compression, optimality of estimators, and denoising properties of the frequency-adaptive wavelet shrinkage.

A summary of the heuristic and theoretical properties of wavelet space shrinkage is in order. If $\{f_i\}_o^{N-1}$ are IID $\aleph(0, 1)$, the rationale behind the choice of the spatially adaptive threshold $\lambda^{DJ} = \sqrt{2\ln N}$ is motivated by the fact that the maximum order statistic $f_{(N)}$ is less than $\sqrt{2\ln N}$ with probability p, where $p \longrightarrow 1$ as $N \longrightarrow \infty$. Similarly, for the j-frequency band containing n_j wavelet coefficients and, therefore, that many IID noise variates, we use a frequency-adaptive thresholding level[4] $\sqrt{2\ln(4+n_j)}$. Hence, we are almost sure to have the pure noise terms shrunk to zero. In other words, the *soft* thresholding approach [see (4)] provides almost optimal function estimators—in terms of minimizing the risk function [mean square error (MSE)], over all possible wavelet coefficient selection schemes, including the ones where exact knowledge is given as to which are the *true* coefficients larger than the standard error of the noise.

Visually, wavelet shrinkage using (4) does not remove all the noise from the data (Fig. 1). This is a consequence of the fact that the MSE norm is used in minimizing the loss (risk function). MSE balances between the effects of the noise (*variability* measure) and the bias (*accuracy* measure) of the estimator, [see (2)]. If a uniform threshold level of $\sqrt{2\ln N}$ is applied across the frequency spectrum, an unnecessary bias may be introduced, but the noise level of the reconstructed image is minimal. On the other hand, if a low fixed uniform threshold value (say, two) is used, no bias is present; however, little denoising occurs in the process. Therefore, the tradeoff between variability and accuracy of the induced-function estimators is resolved by imposing optimality conditions [see (5) and (6)]. Hence, the derivation of the formula for the optimal threshold level. Thresholding approaches other than *soft* thresholding have been also investigated in [17]. Most intuitive is the *hard* threshold scheme, $\xi_\lambda(w) = wI_{\{|w|\geq\lambda\}}$, where I_A is the indicator (characteristic) function of the set A. Note the difference between η_λ and ξ_λ. The former shrinks *all*, whereas the later shrinks only the "*small*" wavelet coefficients. Therefore, η_λ is continuous and ξ_λ is not. Despite that, certain optimality properties could also be derived for the hard threshold scheme ξ_λ.

Other methods for quantitative analysis of image registration are based on estimating the local and global characteristics of the warping displacement fields. For example, the novel approach introduced by Strother *et al.* [32] measures accuracy of registration by explicitly using the scaling, rotation, and translation parameters of the alignment field. This numeric warp ranking approach measures performance of image registration by the mean ($\pm SD$) and the maximum voxel registration error calculated for all brain voxel locations within the reference (target) volume. Kjems *et al.* [24] developed another method for warp classification, which uses the multivariate functional SNR in PET/ single photon emission computed tomography (SPECT) data to assess registration error. The rms errors for anatomic landmarks can also be employed to assess quality of image registration [40]. Koole *et al.* [25] estimate the accuracy of the matching process for intramodality and intermodality registration via the residual mismatch between external markers. Various authors, e.g., [30] and [38], have also used phantom and simulated data to evaluate quality of different 3-D warping strategies based on decomposition of the warping fields into translation, rotation, affine scaling, and nonlinear displacement components separately.

The wavelet-based method presented here has the advantage that it is computationally inexpensive and allows the use of the resliced (post-warping) data alone or in combination with native-space (pre-warping) data in the process of quantitatively estimating the performance of different image-registration methods. In addition, we argue that different classification schemes for warp ranking should be employed under different circumstances based on the main goals of image alignment. The WAIR technique avoids certain statistical issues (e.g., low degrees of freedom) present in other methods for quantitative assessment of image registration using CVA, scale subprofile models, and other common statistical approaches.

There are three main steps in the WAIR toolkit: The first one is computing the discrete wavelet representation of the data (pre- and post-warping) and the target of the registration. The wavelet transform of a single 1-B per voxel volume of dimensions 256 $\times$ 256 $\times$ 256 is computed on a single processor SUN station (250 MHz) for about 50 s. Determining the optimal frequency-adaptive threshold curve of the wavelet coefficients takes about 20 min on the same machine. Finally, the warp ranking in compressed wavelet space is applied, which takes variable time depending on the number of volumes in the study and the warp clas-

[4]Note that, if the variance of the noise component $\sigma^2 \neq 1$, then a nontrivial factor σ comes into play in these threshold values. For details, see the Appendix.

sifier chosen. For example, CGC of our MRI data (two groups of nine subjects under four different warps) took about 20 min.

The development of new more accurate and robust application-specific methodologies for image registration will increase with our knowledge in the fields of biomedical research [36], [22], pattern recognition [29], video data analysis [2], stereo vision [27], and astrophysics [28]. The complexity and diversity of such new techniques, whether fiducial- or intensity-based, will augment the growth in number and quality of these registration tools. Employing various image-registration methods needs to be quantitatively assessed to produce valid and correct results. The present wavelet-based technique for numeric characterization and analysis of image registration shows promise in delineating and discriminating between separate warping methods for a variety of image modalities and study paradigms. However, further more-extensive testing and validation of this methodology need to be conducted on larger and more complex data sets (e.g., comparing new categories of warping strategies, intermodality versus intramodality alignment, etc.).

APPENDIX

We now sketch the derivation of the approximation [see (6)] and the induced upper bound of the *Risk* function [see (7)] of the function estimator shown in (4). The proof consists of three parts. Following the notation from Section I-C, we first show the result for a single observation with noise-level $\sigma = 1$. We then extend this to an arbitrary noise-level σ. Finally, we generalize the result to any collection of N observations.

1) Suppose $X \sim \aleph(\mu, 1)$, $\delta = (2n_j + 4)^{-1}$, $n_j = 2^j$, $t = \sqrt{2\ln(\delta^{-1})}$, $x \wedge y = \min\{x, y\}$, and $\eta_t(x) = \operatorname{sgn}(x)[|x| - t]_+$. Since $\eta_t(x) = \operatorname{sgn}(x)[|x| - |x| \wedge t] = x - \operatorname{sgn}(x)[|x| \wedge t]$ and $I = E((X-\mu)\operatorname{sgn}(X)[|X| \wedge t]) = P(|X| < t)$, we obtain an estimate of

$$\begin{aligned} R &= E\left(\left(\eta_t(X) - \mu\right)^2\right) \\ &= E\left(\left(X - \mu - \operatorname{sgn}(X)\left[|X| \wedge t\right]\right)^2\right) \\ &= 1 - 2P\left(|X| < t\right) + E\left(X^2 \wedge t^2\right). \end{aligned}$$

We can construct two different bounds on R and express the fact that R is less than the minimum of them. For one thing, $X^2 \wedge t^2 \le t^2$ and $R \le 1 - 2P(|X| < t) + E(t^2) \le 1 + t^2$.

Secondly, $X^2 \wedge t^2 \le X^2$ and $R \le 1 - 2P(|X| < t) + E(X^2) = 1 - 2P(|X| < t) + \operatorname{Var}(X) + \mu^2 = 1 - 2P(|X| < t) + 1 + \mu^2 = 2(1 - P(|X| < t)) + \mu^2 = 2P(|X| \ge t) + \mu^2$.

Note that $P(|X| \ge t)$ is implicitly a function of the (unknown) parameter μ [because $X \sim \aleph(\mu, 1)$]. If $g(\mu) = 2P(|X| \ge t)$, than using the Taylor expansion on g at zero and carefully estimating the $\max_\mu (\|g''(\mu)\|/2)$, we can derive an upper bound of $g(\mu) \le (t^2 + 1)e^{-t^2/2} + t^2\mu^2$. Therefore, the performance estimate, as measured by the *Risk* factor, is bounded above by

$$R = E\left(\left(\eta_t(x) - \mu\right)^2\right) \le \begin{cases} 1 + t^2 \\ e^{-t^2/2}\left(t^2 + 1\right) + t^2\mu^2 + \mu^2 \end{cases}$$

for all $t \ge 1$.

In other words,

$$R \le \begin{cases} \left(1 + t^2\right)\left(1 + e^{-t^2/2}\right) \\ \left(t^2 + 1\right)\left(e^{-t^2/2} + \mu^2\right) \end{cases} \le \left(1 + t^2\right)\left(e^{-t^2/2} + \mu^2 \wedge 1\right),$$

for all $t \ge 1$.

2) How would this bound change if we apply it to $\hat{\theta} \sim \aleph(\theta, \sigma^2)$?

Let $X \sim \aleph(\theta, 1)$. Denote $\delta = e^{-t^2/2}$. We then showed $R(X, t) = E\left((\eta_t(x) - \theta)^2\right) \le (1 + t^2)(\delta + \theta^2 \wedge 1)$. Define $\hat{\theta} = \sigma X \sim \aleph(\mu, \sigma^2)$, where $\mu = \sigma\theta$. Let

$$\begin{aligned} t &= \sigma\sqrt{2\ln(2n_j + 4)} \\ R(\hat{\theta}, t) &= E\left(\left(\eta_t(\hat{\theta}) - \mu\right)^2\right) \\ &= E\left(\left((\hat{\theta} - \mu) - \operatorname{sgn}(\hat{\theta})\left[|\hat{\theta}| \wedge t\right]\right)^2\right) \\ &= \sigma^2 R\left(X, \frac{t}{\sigma}\right) \\ &= \sigma^2 R(X, t_1) \\ &\le \sigma^2(1 + t_1^2)(\delta' + \theta^2 \wedge 1) \end{aligned}$$

where $t_1 = t/\sigma$ and $\delta' = e^{-t^2/2\sigma^2} = e^{-t_1^2/2}$. Then

$$\begin{aligned} R(\hat{\theta}, t) &\le (1 + t_1^2)\left(\sigma^2\delta' + (\sigma\theta)^2 \wedge (\sigma^2)\right) \\ &= \left(1 + 2\ln(N + 4)\right)\left(\sigma^2\frac{1}{(2n_j + 4)} + \mu^2 \wedge \sigma^2\right). \end{aligned}$$

3) Finally, we consider the situation when we have N observations $\hat{\theta} = \{\hat{\theta}_i\}_i = \{\hat{\theta}_{j,k}\}_{j,k}$, where $\hat{\theta}_i \sim \aleph(\theta_i, \sigma^2)$. In matrix notation $\hat{\theta} \sim \aleph(\Theta, \sigma^2 I)$,

$$R(\hat{\theta}, t) = \frac{1}{N} E\left(\left\|\eta_t(\hat{\theta}) - \Theta\right\|^2\right) = \frac{1}{N}\sum_{j,k} R(\hat{\theta}_{j,k}, t)$$

where $t = \lambda_j = \sigma\sqrt{2\ln[(2n_j + 4)]} = \sigma\sqrt{2\ln[1/\delta]}$.

Hence,

$$\begin{aligned} R(\hat{\theta}, \theta) &\le \frac{1}{N}\sum_j \left[\left(1 + 2\ln(2n_j + 4)\right)\right. \\ &\qquad \left.\cdot\left(\sigma^2\frac{2^j}{(2n_j + 4)} + \sum_k \theta_{j,k}^2 \wedge \sigma^2\right)\right] \\ &= \left(1 + 2\ln(N + 4)\right)\left[\sigma^2\frac{J}{2N} + \frac{1}{N}\sum_{j,k}\theta_{j,k}^2 \wedge \sigma^2\right] \\ &= \left(1 + 2\ln(N + 4)\right)\left[\sigma^2\frac{J}{2N} + \Re_{N,\sigma}(SW, \theta)\right]. \end{aligned}$$

ACKNOWLEDGMENT

The authors wish to thank the group of D. Rottenberg and S. Strother, VA Hospital and the University of Minnesota, Minneapolis, MN, for providing initial test data and for their support and encouragements.

REFERENCES

[1] Y. Amit, "A nonlinear variational problem for image matching," *SIAM J. Sci. Comput.*, vol. 15, no. 1, pp. 207–224, 1994.

[2] P. Anuta, "Digital registration of multispectral video data," *SPIE*, vol. 7, pp. 168–175, 1969.

[3] N. Ayache and B. Faverjon, "Efficient registration of stereo images by matching graph descriptions of edge segments," INRIA, inv. 119 103, 1986.

[4] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Comput. Vision, Graph. Image Process.*, vol. 46, pp. 1–21, 1989.

[5] M. Barnsley and L. Hurd, *Fractal Image Compression*. Natick, MA: AK Peters, 1992.

[6] F. Bookstein, *Morphometric Tools for Landmark Data*. New York: Cambridge Univ. Press, 1991.

[7] G. Christensen, M. Miller, and M. Vannier, "3D deformable magnetic resonance textbook based on elasticity," in *Proc. Amer. Assoc. Artificial Intell. Symp.* Stanford, CA, Mar. 21–23, 1994.

[8] G. Christensen, R. Rabbitt, and M. Miller, "A deformable neuroanatomy textbook based on viscous fluid mechanics," in *Proc. Inform. Sci. Syst.*, 1993, pp. 211–216.

[9] D. Collins, T. Peters, and A. Evans, "An automated 3D nonlinear image deformation procedure for determination of gross morphometric variability in human brain," in *Proc. Visual. Biomed. Comput.*, vol. 2359, Rochester, MN, 1992, pp. 541–551.

[10] I. Daubechies, "Orthonormal bases of wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.

[11] ——, "Ten lectures on wavelets," in *SIAM Series Appl. Math. Conf.*, 1992.

[12] G. Davis, "A wavelet-based analysis of fractal image compression," *IEEE Trans. Image Processing*, vol. 7, no. 2, pp. 141–155, 1998.

[13] I. Dinov, M. Mega, P. Thompson, L. Lee, R. Woods, C. Holmes, D. Sumners, and W. Toga, "Analyzing functional brain images in a probabilistic altas: A validation of sub-volume thresholding," *J. Comput. Assisted Tomography*, vol. 24, no. 1, pp. 128–139, 2000.

[14] I. Dinov, D. Sumners, S. Strother, U. Kjems, and D. Rottenberg, "Construction of a 'goodness of warp' measure," *NeuroImage*, vol. 7, no. 4, p. 735, 1998.

[15] I. D. Dinov, "Mathematical and statistical techniques for modeling and analysis of medical data," Ph.D. dissertation, Appl. Comput. Math. Dept., Florida State Univ., Tallahassee, FL, 1998.

[16] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[17] ——, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1996.

[18] S. Efroimovich and M. Pinsker, "A learning algorithm for nonparametric filtering" (in Russian), *Avtom. Telemekh.*, vol. 11, pp. 58–65, 1984.

[19] S. Efromovich, *Nonparametric Curve Estimation: Methods, Theory and Applications*, ser. Statistics. Berlin, Germany: Springer-Verlag, 1999.

[20] K. Friston, J. Ashburner, J. Poline, C. Frith, J. Heather, and R. Frackowiak, "Spatial registration and normalization of images," *Hum. Brain Mapping*, vol. 2, pp. 165–189, 1995.

[21] K. Friston, C. Frith, P. Liddle, and R. Frackowiak, "Comparing functional (PET) images: The assessment of significant change," *J. Cerebral Blood Flow Metabol.*, vol. 11, pp. 690–699, 1991.

[22] D. Hill and D. Hawkes, "Medical image registration using knowledge of adjacency of anatomical structures," *Image Vis. Comput.*, vol. 12, pp. 173–178, 1994.

[23] J. Hutchinson, "Fractals and self similarity," *Indiana Univ. Math. J.*, vol. 30, pp. 713–747, 1981.

[24] U. Kjems, S. Strother, J. Anderson, I. Law, and L. Hansen, "Enhancing the multivariate signal of $[^{15}o]$ water PET studies with a new nonlinear neuroanatomical registration algorithm," *IEEE Trans. Med. Imag.*, vol. 18, pp. 306–319, Apr. 1999.

[25] M. Koole, Y. D'Asseler, K. Van Laere, R. Van De Walle, C. Van De Wiele, I. Lemahieu, and R. A. Dierckx, "MRI–SPET and SPET–SPET brain co-registration: Evaluation of the performance of eight different algorithms," *Nucl. Med. Commun.*, vol. 20, no. 7, pp. 659–669, 1999.

[26] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.

[27] D. Maravall and E. Fernandez, "Contribution to the matching problem in stereo vision," in *Proc. Int. Conf. Pattern Recognition Conf.*, 1992, pp. I : 411–I : 414.

[28] D. Mink, "Wcstools: An image astrometry toolkit," presented at the Astronom. Data Analysis Software Syst. VIII, ASP Conf., 1999.

[29] R. Moss, C. Robinson, and W. Poppelbaum, "Online recognition (OLREC): A novel approach to visual pattern recognition," *Pattern Recognit.*, vol. 16, no. 6, pp. 535–550, 1983.

[30] D. Ros, M. Espinosa, J. F. Setoain, C. Falcn, F. Lomea, and J. Pava, "Evaluation of algorithms for the registration of 99Tcm-HMPAO brain SPET studies," *Nucl. Med. Commun.*, vol. 20, no. 3, pp. 227–236, 1999.

[31] E. Sowell, P. Thompson, C. Holmes, R. Batth, T. Jernigan, and A. Toga, "Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping," *NeuroImage*, vol. 9, pp. 587–597, 1999.

[32] S. Strother, J. Anderson, X. Xu, J. Liow, D. Bonar, and D. Rottenberg, "Quantitative comparison of image registration techniques based on high resolution MRI of the brain," *J. Comput. Asst. Tomography*, vol. 18, no. 6, pp. 954–962, 1994.

[33] C. Studholme, D. Hill, and D. Hawkes, "Incorporating connected region labeling into automated image registration using mutual information," in *Proc. IEEE Math. Methods Biomed. Image Anal. Workshop*, 1996, pp. 23–31.

[34] R. Szeliski and S. Lavallee, "Matching 3D anatomical surfaces with nonrigid deformations using octree-spline," in *Proc. Conf. Comput Vision Pattern Recognition, Biomed. Image Anal. Workshop*, 1994, pp. 144–153.

[35] P. Thompson and A. Toga, "A surface-based technique for warping three-dimensional images of the brain," *IEEE Trans. Med. Imag.*, vol. 15, pp. 402–416, Aug. 1996.

[36] A. Toga, *Brain Warping*. New York: Academic, 1998.

[37] W. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.

[38] J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. J. Maurer, R. Kessler, and R. Maciunas, "Retrospective intermodality registration techniques for images of the head: Surface-based versus volume-based," *IEEE Trans. Med.Imag.*, vol. 18, pp. 144–150, Feb. 1999.

[39] R. Woods, S. Cherry, and J. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images," *J. Comput. Assist. Tomography*, vol. 16, pp. 620–633, 1992.

[40] R. Woods, S. Grafton, C. Holmes, S. Cherry, and J. Mazziotta, "Automated image registration: I. General methods and intrasubject, intramodality validation," *J. Comput. Assist. Tomography*, vol. 22, pp. 141–154, 1998.

[41] R. Woods, J. Mazziotta, and S. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomography*, vol. 17, pp. 536–546, 1993.

[42] K. Worsley, A. Evans, S. Marrett, and P. Neelin, "A three-dimensional statistical analysis for CBF activation studies in human brain," *J. Cereb. Blood Flow Metab.*, vol. 12, no. 6, pp. 900–918, 1992.



Ivo D. Dinov received the Bachelor's degree in mathematics and informatics from the University of Sofia, Sofia, Bulgaria, in 1991, the M.S. degree in mathematics from the Michigan Technological University, Houghton, in 1993, the M.S. degree in statistics from the Florida State University, Tallahassee, in 1998, and the Ph.D. degree in applied and computational mathematics from Florida State University, in 1998.

From 1998 to 2001, he was a Postdoctoral Researcher of neurosciences in the Laboratory of Neuro Imaging, University of California at Los Angeles (UCLA). During that time, he was a Visiting Assistant Professor in mathematics at UCLA. In the summer of 2001, he joined the faculty of the Department of Statistics, UCLA. His publications includes fundamental research papers, as well as clinical research papers on neuroimaging and brain mapping of neurodegenerative diseases. His research interests include mathematical and statistical modeling and analysis of biomedical data, neuroimaging, diagnosis, prognosis and treatment planning for Alzheimer's Disease and other neurodegenerative disorders, fractal, wavelet and spectral signal representations, and geometrical and topological shape descriptors.

Dr. Dinov is a member of the American Mathematical Society, the Organization for Human Brain Mapping, and Pi Mu Epsilon. He was the recipient of an National Institute of Mental Health (NIMH)/National Institutes of Health (NIH) fellowship in neuroimaging and brain mapping (1998–1999)



Michael S. Mega received the B.S. degree in physiological psychology and the B.A. degree in philosophy from the University of California, Santa Barbara, in 1984, the M.D. degree from George Washington University, Washington, DC, in 1988, and the Ph.D. degree in neuroscience from the University of California at Los Angeles, in 1997.

From 1989 to 1992, was a Resident in Clinical Neurology at Boston University, Boston, MA. From 1994 to 1997, he was a Clinical Instructor in Neurology at UCLA. Since 1996, he has been the Director of the Memory Disorder Clinic and Alzheimer's Disease Clinic, UCLA School of Medicine. Since 1997, he has also been an Assistant Professor of neurology at UCLA. He has authored or coauthored more than 50 papers, book chapters, and manuscripts on a variety of topics ranging from geriatric medicine to neurooncology, brain imaging, and a broad spectrum of neuropsychiatric disorders. His research concerns the areas of human cognitive and behavioral abnormalities, neuroimaging, and brain mapping of structural and functional brain data of the elderly and Alzheimer's Disease populations.

Dr. Mega is a member of the American Academy of Neurology, the International Neuropsychological Society, the Behavioral Neurology Society, and the American Neuropsychiatric Association. He was the recipient of the 1994 Ralph Goldman Research Award.



Paul M. Thompson (M'98) received the B.A. degree in Greek and Latin languages and the M.A. degree in mathematics and philosophy from Oxford University, Oxford, U.K., in 1991 and 1993, respectively, and the Ph.D. degree in neuroscience from the University of California at Los Angeles (UCLA), in 1998.

He is currently an Assistant Professor of neurology at UCLA. His research is focused on development of new mathematical and computational algorithms for neuroimaging, structural and functional mapping of the cerebral cortex and corpus callosum, and construction of population specific probabilistic atlases of human brain anatomy. He has authored or coauthored more than 70 scientific papers in refereed journals.

Dr. Thompson is a member of various program committees for the IEEE and SPIE workshops. He was the recipient of the 1998 Di Chiro Outstanding Scientific Paper Award, the 1998 Eiduson Award for Neuroscience Research, and the 1997 SPIE Medical Imaging Award, Best Paper. He was a Fulbright Scholar from 1993 to 1998.



Roger P. Woods received the B.A. degree in chemistry from Washington University, Seattle, in 1980 and the M.D. degree from Harvard Medical School, Cambridge, MA, in 1984.

From 1984 to 1985, he was a Medical Intern at the University of Washington. From 1985 to 1987, he continued as a Resident in Internal Medicine at the University of Washington. From 1988 to 1993, he was a Fellow in Neuroscience Imaging and a Resident in Neurology at the University of California at Los Angeles (UCLA) Medical Center. Since 1993, he is an Assistant Professor of neurology at UCLA. Some of his techniques and his software are used by many brain-mapping centers around the world on a daily basis. He is an Associate Editor for the *Journal of Computer Assisted Tomography* and has authored or coauthored more than 40 research papers in peer-reviewed journals. His interests include stereotactic image registration, new brain-mapping techniques, image morphometrics, multidimensional modeling of brain data, sensory-motor integration, and cerebral hemodynamics. His research is internationally recognized.

Dr. Woods was the recipient of the 1993 S. Weir Mitchell Award for Research presented by the American Academy of Neurology and the 1991 Jean-Louis Riehl Award for Research in Neurology UCLA.



De Witt L. Sumners received the B.S. degree in physics from the Louisiana State University, Baton Rouge, in 1964 and the Ph.D. degree in mathematics from the University of Cambridge, Cambridge, U.K., in 1967.

He has been on the faculty of the University of Chicago, University of California at Berkeley, University of Minnesota, and Florida State University. Since 1999, he has been a Professor of mathematics and Department Chair at Florida State University (FSU), Tallahassee. He is a Robert O. Lawton Distinguished Professor of Mathematics, a member of the Institute of Molecular Biophysics at FSU, and Director of the multicenter Program in Mathematics and Molecular Biology (PMMB). His interests cover the broad spectrum of mathematical modeling of protein folding, stratification of DNA topology and geometry, polymer conformations, and analysis of stereotactic functional and anatomical human brain data. He has authored or coauthored a number of research papers and several books on fundamental mathematical developments, introducing new techniques for analyzing structural brain images, as well as knot theoretic and Monte Carlo models for studying the structure and topology of DNA.



Elizabeth L. Sowell received the B.A. degree from the University of California at Irvine, in 1987 and the Ph.D. degree from San Diego State University (SDSU)/University of California at San Diego (UCSD), in 1997, under the joint SDSU/UCSD Program in Clinical Neuropsychology.

From 1997 to 1999, she was a National Institute of Mental Health (NIMH) Research Fellow in the Division of Child and Adolescent Psychiatry, University of California at Los Angeles (UCLA). Since 2000, she has been an Assistant Professor of neurology at UCLA. She has authored or coauthored a number of papers on the effects of prenatal alcohol exposure, maturation in posterior perisylvian cortices, early onset schizophrenia, and post-adolescent brain maturation. Her interests include functional and structural imaging studies of brain development and development of learning and memory functioning between childhood, adolescence, and young adulthood as assessed with structural and functional imaging methods.



Arthur W. Toga received the B.S. degree in psychology from the University of Massachusetts, Amherst, in 1974 and the M.S. and Ph.D. degrees in neurosciences from St. Louis University, St. Louis, MO, in 1976 and 1978, respectively.

From 1978 to 1979, he was a Postdoctoral Fellow at the Washington University School of Medicine. Since 1983, he has been the Director of the Laboratory of Neuro Imaging. He is currently a Professor of neurology at the University of California at Los Angeles (UCLA), Co-Director, Division of Brain Mapping, Neuropsychiatric Institute, UCLA. His research includes the development of new algorithms and the computer-science aspects relevant to neuroimaging, and new visualization techniques and statistical measurement that are employed in the study of morphometric variability in humans, subhuman primates, and rodents. He has authored or coauthored more than 200 research papers in premiere scientific journals and several books on neuroimaging and brain mapping.

Prof. Toga was the recipient of the 1999 Smithsonian Award for Scientific Innovation, the 1999 Axiem Award, the 1999 Videographic Award for Excellence, the 1998 Communication Award for Scientific Video, the 1997 Giovanni DiChiro Award for Outstanding Scientific Research, and the 1997 Cum Laude Paper (First Prize) of SPIE Medical Imaging. He was the silver finalist for the National Medical Video and the finalist of the 1999 Telly Awards.