Wolters Kluwer Health | Lippincott Williams & Wilkins

JOURNAL OF COMPUTER ASSISTED TOMOGRAPHY

 Volume 22(1), January/February 1998, pp 139-152

# Automated Image Registration: I. General Methods and Intrasubject, Intramodality Validation

[Image Processing]

Woods, Roger P.; Grafton, Scott T.; Holmes, Colin J.; Cherry, Simon R.; Mazziotta, John C.

From the Division of Brain Mapping (R. P. Woods, C. J. Holmes, and J. C. Mazziotta) and Departments of Neurology (R. P. Woods, C. J. Holmes, and J. C. Mazziotta), Pharmacology (S. R. Cherry and J. C. Mazziotta), and Radiology (J. C. Mazziotta), UCLA School of Medicine, Los Angeles, CA, and Departments of Neurology and Nuclear Medicine, Emory University School of Medicine, (S. T. Grafton), Atlanta, GA, U.S.A.

Address correspondence and reprint requests to Dr. R. P. Woods at Department of Neurology, 710 Westwood Plaza, Rm. 3-145, Los Angeles, CA 90095, U.S.A.

## Abstract

Purpose: We sought to describe and validate an automated image registration method(AIR 3.0) based on matching of voxel intensities.

Method: Different cost functions, different minimization methods, and various sampling, smoothing, and editing strategies were compared. Internal consistency measures were used to place limits on registration accuracy for MRI data, and absolute accuracy was measured using a brain phantom for PET data.

Results: All strategies were consistent with subvoxel accuracy for intrasubject, intramodality registration. Estimated accuracy of registration of structural MRI images was in the 75 to 150 μm range. Sparse data sampling strategies reduced registration times to minutes with only modest loss of accuracy.

Conclusion: The registration algorithm described is a robust and flexible tool that can be used to address a variety of image registration problems. Registration strategies can be tailored to meet different needs by optimizing tradeoffs between speed and accuracy.

---

In 1992 Woods et al. (1) described a method for aligning PET images using a calculus-based minimization procedure and voxel intensities. This technique also proved useful for intramodality registration of MR images (2) and was extended to allow cross-modality registration of PET and MR images (3). A cost function based on the uniformity of the ratio of one image to the other served to guide registration through iterative univariate calculus-based minimization. The method has compared favorably with other registration techniques (4,5) and has been distributed to many laboratories as part of the Automated Image Registration (AIR) package.

The AIR package has subsequently been revised to increase its speed and accuracy and to expand its ability to address a broader range of registration problems. The univariate minimization algorithm has been replaced by more robust multivariate methods; the cost function and rigid body spatial transformation models have been supplemented with alternative approaches; and interpolation routines more appropriate for MR images have been added. Consequently, earlier references (1,3) no longer accurately characterize the mathematics or the performance of the current version of the AIR package.

The purposes of this article are (a) to describe the mathematical basis of the registration strategy used by AIR 3.0 and its relationship to several similar techniques; (b) to systematically compare different minimization procedures, smoothing strategies, editing strategies, sampling strategies, and interpolation strategies for intrasubject, intramodality registration; (c) to describe a method for combining all possible pairwise registrations of a set of images to generate more accurate registration results; and (d) to describe the use of internal inconsistencies among redundant pairwise registrations to place limits on true registration accuracy in the absence of known gold standards. Intermodality registration will not be addressed, and intersubject registration is described and validated separately (6).

## REGISTRATION ALGORITHM

Figure 1 provides a schematic overview of the registration strategy used by AIR 3.0. After optional smoothing or interpolation to cubic voxels, one image, which will be referred to as the reslice image, is resampled to match the other image, referred to as the reference image. Resampling is based on the current parameters of the spatial transformation model and also requires an interpolation model to compute voxel intensities. After thresholding to exclude voxels outside the head and optional editing to exclude voxels outside the brain in the reference image, a cost function reflecting the similarity of the two images is computed. For linear spatial transformation models, biases are avoided by reversing the roles of the reslice and reference image and inverting the spatial transformation to compute a second estimate of the cost function, which is then averaged with the first. This bias elimination procedure can be optionally omitted in AIR but was always used here. To improve speed, the cost function is initially computed for only a limited sampling of the voxels (the default is every 81st voxel) and sampling is increased with subsequent iterations (the default is by factors of three to reach a final sampling of every voxel). The derivatives of the cost function with respect to the parameters of the spatial transformation model are computed and are used to compute new parameters and iteratively minimize the cost function. Termination criteria are tested with each iteration to decide whether to continue iterating, to increase sampling, or to stop. The spatial transformation that produced the lowest value of the cost function is stored and can be used to produce registered images. If desired, the parameters that produce the optimal transformation can also be stored independently. Substantial modifications as compared with the original AIR algorithm (1) are described in the following sections.

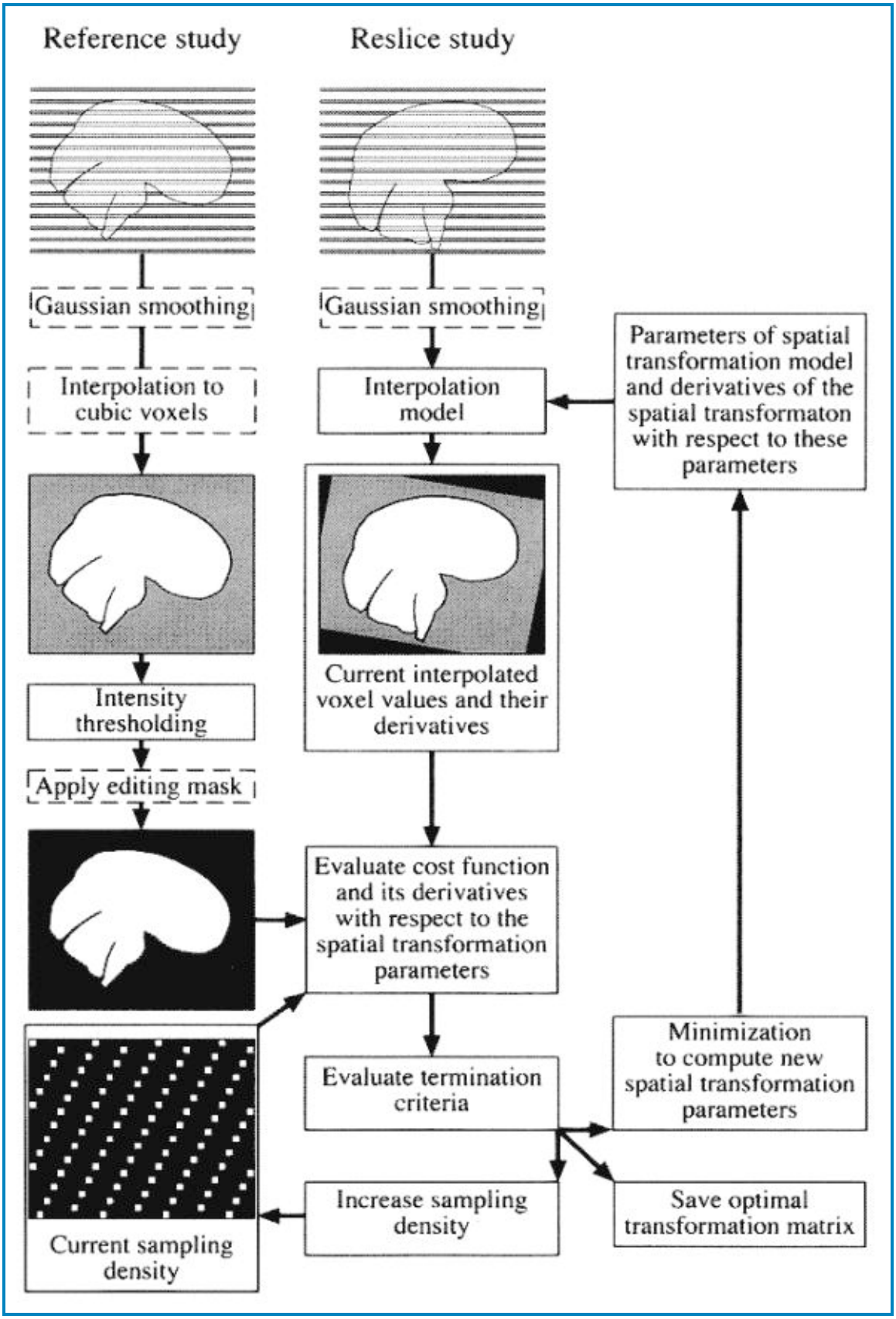


**FIG. 1.** Schematic diagram of the registration algorithm. Boxes shown in dashed lines represent optional procedures. One of the images is initially designated as the reference study and the other as the reslice study. Although not shown in the diagram, the cost function is computed a second time with the roles of reference and reslice study interchanged, and it is the average of these two cost function estimates that is minimized. Regions of the schematic images shown in solid black are excluded from analysis (a) because they are below the specified threshold, (b) because they are excluded by the optional editing mask, (c) because they are not part of the current sampling set, or (d) because they are outside the field of view of the reslice study. These exclusions actually occur before or during interpolation of the reslice study, but

this is not indicated in the diagram to preserve conceptual clarity of the other steps. When the sampling density is increased, the spatial transformation parameters that gave the optimal value for the cost function at the prior sampling density serve as the new starting point for minimization.

## Optional Image Smoothing

Smoothing noisy PET images before registration improves accuracy (1). Whether to smooth MRI data will be addressed in Validation Studies. An optional fast Fourier, Gaussian convolution routine replaces the box smoothing strategy used previously. The width of the smoothing kernel is specified independently for each image axis. If applied, smoothing is performed before resampling, thresholding, or application of any editing masks.

## Optional Interpolation of Reference Study to Cubic Voxels

In the original AIR algorithm, the reference volume always consisted of voxels that had been linearly interpolated to make them cubic. This is now an option, but not the default. Instead, anisotropic voxel sizes are taken into account mathematically while still ensuring the integrity of the selected spatial transformation model in real world coordinates. The tradeoffs between speed and accuracy as a function of whether the reference volume is interpolated will be addressed by the PET validation studies.

## Interpolation Model

To compare the images being registered, one image must be resampled according to the parameters of the spatial transformation model. This requires interpolation of intensities at locations between the voxel locations represented in the original image. After registration is complete, final images must be created, which again requires interpolation to compute resampled voxel intensities. It is possible, but not necessary, to use the same interpolation model in both contexts. In the distribution version of AIR 3.0, trilinear interpolation remains the only model used to compute the cost function during minimization, whereas any of several models including nearest neighbor, trilinear, sinc, and chirp-*z* interpolation (7) can be used to create the final images. Windowing (8) and scan line decomposition (9)[with precautions to avoid aliasing (10)] are also available in the final resampling algorithm to improve speed. Hybrid models are also included for multislice data sets that are not bandlimited along the axes between planes (8). A special implementation of AIR that uses windowed sinc interpolation to compute the cost function is described in Validation Studies.

## Thresholding, Optional Editing, and Bias Elimination

Simple thresholding can be used to exclude voxels from outside the body that provide no useful spatial information, but it may also be advantageous to exclude voxels from the scalp, skull, and dura when computing the cost function for MRI data (11). Exclusion of these structures could be accomplished by simply registering edited images, but tendencies to align the artificial edges created by the editing process could be problematic. This can be avoided by allowing the algorithm to apply user-generated mask files to the images. When an image is serving as the reference image, only voxels that are non-zero in the associated mask file contribute to the cost function, and the mask associated with the other image is ignored. The second mask file is used when the roles of the reference and reslice files are exchanged to compute an unbiased cost function. Since edited versions of both images are never compared directly with one another, alignment of artificial edges created by editing cannot lower the cost function. The effect of editing nonbrain structures on registration accuracy will be addressed in Validation Studies.

## Cost Functions

The cost function provides the algorithm with a quantitative measure of how well the images are registered. AIR 3.0 allows a choice of three different cost functions. The first cost function, referred to here as the ratio image uniformity (RIU) cost function, is identical to the one described previously (1). A resampled image is divided by the image to which it is being registered on a voxel-by-voxel basis to create a ratio image, and the uniformity of this ratio image is measured by computing its standard deviation. The standard deviation is then divided by the mean ratio to provide a normalized cost function value. Minimization of the cost function increases the uniformity of the ratio image, which is independent of global intensity scaling of the original images, and improves registration.

The second cost function assumes that the images being registered have already been properly adjusted for global intensity differences and uses a least-squares approach similar to that described by Hajnal et al.(8), and subsequently

adopted by Friston et al.(12). This cost function will be referred to as the least-squared difference image (LS) cost function. If no intensity rescaling is needed, the spatially resampled reslice image should be almost identical to the reference image when the images are well registered. The difference between the resampled reslice image and the reference image is computed at each voxel, and the square of this difference is averaged across voxels to generate this cost function.

The third cost function is similar to the LS cost function but adds an extra parameter to the minimization that allows global intensity rescaling of the images relative to one another. This approach has been advocated by Alpert et al. (13) and by Snyder (14) and will be referred to as the scaled least-squared difference image (SLS) cost function.

### Minimization Procedure and Spatial Transformation Model

The objective of the minimization procedure is to conduct an efficient search of the parameter space defined by the mathematical spatial transformation model to minimize the cost function. By default, AIR starts this search with parameters that will align the exact centers of the two image sets without additional real world rotation, translation, or scaling. Alternatively, any set of starting parameters can be explicitly declared. All of the work described herein uses a rigid body spatial transformation model, which has six parameters, so the minimization procedure must search a 6D parameter space to find the optimal rigid body registration. The original AIR algorithm conducted this multidimensional search by sequentially performing undimensional minimizations, each of which corresponded to only one of the six parameters. The AIR 3.0 algorithm replaces this strategy with two variants of a multivariate calculus-based minimization procedure. The first variant, which will be referred to as full Newton-type minimization, is similar to one briefly described for 12 parameter intersubject registration (15) but has been generalized so that it is applicable to any spatial transformation model. The method assumes that the cost function(as a function of the spatial transformation parameters) can be approximated near its minimum by a parabolic surface. This parabolic surface can be fully characterized by a vector **b**, consisting of the first partial derivatives of the cost function with respect to each parameter at a given point in parameter space, and a Hessian matrix **A**, consisting of the second partial derivatives of the cost function with respect to each pair of parameters at the same point in parameter space. If **x** is a vector representing the adjustment of each spatial transformation parameter needed to reach the minimum of the parabolic surface, **x** can be derived from**b** and **A** by solving the simple matrix equation (16)

$$\mathbf{A}*\mathbf{x} = -\mathbf{b}$$

The first and second derivatives of the cost function with respect to the transformation parameters are all calculated analytically and are a function of the spatial transformation model, the interpolation model, and the particular cost function. New estimates of the minimum are calculated iteratively until the termination criteria described in the next section are met. In the event that the Hessian matrix **A** is not positive definite(indicating a saddle point or tendency toward a maximum rather than a minimum), the algorithm either increases the sampling density or, if sampling is already maximal, generates a warning message and terminates.

The second minimization procedure is an approximation to the first one. It assumes that the second derivatives of the interpolated voxel values with respect to the sampling coordinate locations are all zero. This decreases the computation time per iteration and reduces the likelihood of encountering problematic local maxima or saddle points. Such approximations are commonplace in minimization algorithms. For example, the Levenberg-Marquardt algorithm (16) makes this same assumption and also assumes that the second derivatives of the spatial transformation model with respect to its parameters are all zero. This second minimization method will be distinguished from full Newton-type minimization by appending LM after the cost function(i.e., RIU-LM, LS-LM, and SLS-LM) and for the sake of brevity will be referred to as Marquardt-like minimization. Full Newton-type minimization should be assumed as the default.

### Termination Criteria

If the associated parabolic approximations were exact, the full Newton-type minimization method would proceed to the minimum of the cost function in a single iteration. However, they are not exact, and multiple iterations are required. Criteria are necessary to decide when iteration should terminate. For the original AIR algorithm (1), the criteria were based on the fact that the first derivative of the cost function with respect to each parameter should be very close to zero near the minimum. The primary termination criterion of the new algorithm takes advantage of the fact that the

derivatives used to compute the location of the minimum of a multidimensional parabolic surface can also be used to predict the change in the cost function associated with moving from the current point in parameter space to the predicted minimum. The predicted change ([DELTA]*c*) is given by the matrix equation. Equation

$$\mathbf{A} * \mathbf{x} = -\mathbf{b}$$

Equation 4C

where **b, x**, and **A** are defined as before (16, p 414). So long as **A** is positive definite, the predicted change will be negative and should get progressively closer to zero as the true minimum of an arbitrary differentiable surface is approached. Thus, the predicted cost function change provides a single numerical value that simultaneously incorporates information about all of the parameters in units that are independent of the particular spatial transformation model. Termination occurs when the predicted cost function change drops below a prespecified small value.

To provide finer control, the new algorithm also has secondary termination criteria. The two secondary termination criteria are (a) the total number of iterations performed at a given sampling interval and (b) the number of iterations performed without improvement in the actual cost function at a given sampling interval. In another modification of the original methodology, the new algorithm always retains a copy of the parameters that resulted in the lowest actual value of the cost function at the current sampling density and uses this best value, rather than the most recent value, when initializing the next sampling density or saving the final results.

Selection of appropriate primary termination criteria will be addressed in Validation Studies. For secondary termination criteria, a default value of 25 total iterations or 5 iterations without any improvement in the actual cost function should be assumed unless otherwise specified.

### Implementation and Distribution

AIR 3.0 is written entirely in C and requires no proprietary code or third party packages. The registration algorithm can be compiled to use either 8 or 16 bit internal image representation. The benchmarking work described here for PET images was done on a SunSPARC 10 workstation, compiled for 8 bit internal representation using the standard Sun C compiler. The MRI validation work was performed on a Power Macintosh 8500/120 running the MachTen UNIX environment, compiled for 16 bit images. Source code for the algorithm and supporting software are available to the research community for research purposes free of charge. Documentation and information about downloading the software are available on the World Wide Web at the Universal Resource Locator (URL):http://bishopw.loni.ucla.edu/AIR3/index.html

## VALIDATION STUDIES

Two separate sets of validation studies will be described. The first uses a previously described PET phantom data set (1) that allows registration accuracy to be established with independent gold standards. The second uses high resolution MR scans of a single subject to evaluate intrasubject MR registration. As often happens with real data sets, gold standards are not available for the MRI validation study, and special attention is given in both validation studies to the use of internal inconsistency measures as an alternative approach to validation in this setting. Before describing the individual validation studies, some terms and methods common to both studies are described.

### Discrepancies, Errors, and Internal Inconsistencies

Several terms will be used repeatedly and warrant explicit definition. The local discrepancy between two different transformations for registering a pair of images will be defined as the 3D distance between the locations to which a given voxel is mapped by the two transformations. The mean discrepancy between two different transformations for registering a pair of images will be defined as the average of the local discrepancies over all voxels that constitute those parts of the brain that are present in both images. The corresponding plural term mean discrepancies will refer collectively to a set that specifically includes the mean discrepancy of every possible unique pairwise registration of a group of images of the same object. The term global mean discrepancy will refer to the average mean discrepancy

across such a collective set. The terms maximum discrepancy, maximum discrepancies, and global maximum discrepancy will be defined analogously using the appropriate maximum, rather than the mean, across voxels or registrations. The unqualified term discrepancies will refer collectively to mean discrepancies and maximum discrepancies, and the term global discrepancies collectively to the global mean discrepancy and the global maximum discrepancy.

If, and only if, one of the two transformations (or sets of transformations) being compared is based on independent gold standard measurements, the terms error or errors will analogously replace the words discrepancy or discrepancies in all the above definitions. If, and only if, a set of transformations is being compared with reconciled mean transformations(defined in the next section) derived from that same set, the terms internal inconsistency or internal inconsistencies will analogously replace the word discrepancy or discrepancies in all the above definitions. Discrepancies will always be characterized by indicating the two distinct registration strategies to which the discrepancy applies, whereas errors and internal inconsistencies are properties of a single registration strategy.

## Derivation of Reconciled Mean Transformations from Pairwise Registrations

If a perfect registration method were used to perform all possible pairwise registrations of a rigid body, the resulting transformations would be completely internally consistent. For example, the direct pairwise registration of image A to image C would be identical to the combined results for registering image A to image B and for registering image B to image C. In this sense, a perfect set of all pairwise registrations should contain redundant information about the interrelationships between the images. With imperfect registration, direct pairwise registration of image A to image C will not be identical to the combined results for registering image A to B and image B to C due to errors. Conceptually, each imperfect pairwise result can be viewed as the combination of the correct result and an error associated with that particular pair. Better estimates of the correct results should be attainable by defining the minimal set of nonredundant parameters sufficient to derive all pairwise registrations and then adjusting this set to somehow minimize local discrepancies between the predicted pairwise results and those that are actually observed. For $N$ images, $N$ - 1 six parameter rigid body transformations suffice to derive all possible pairwise registrations. The completely internally consistent set of transformations defined by these six ($N$ - 1) parameters will be referred to here as the reconciled mean transformations.

AIR 3.0 includes an algorithm that computes reconciled mean transformations from an existing set of all possible pairwise registrations. The algorithm initializes the six ($N$ - 1) parameters by assuming that the images are all already perfectly registered and refines these initial estimates iteratively. From a computational stand-point, it is advantageous to minimize the summed squares of the distances defined by each local discrepancy between the observed and predicted transformations rather than the sum of the distances themselves. Squared local discrepancies are computed for all voxels that correspond to parts of the brain that are represented in all of the images, and these squared values are summed across voxels and across all possible pairwise registrations to generate a global summed squared discrepancy. Appropriate brain voxels need only be explicitly identified once in a single image since their locations can be remapped to other images using the current estimates of the reconciled mean transformations. Full Newton-type minimization is used to iteratively minimize the global summed squared discrepancy by adjusting the six ($N$ - 1) parameters until the predicted change for the next iteration is $<10^{-10}$. The algorithm reports the final optimized global discrepancy, normalized for the number of image pairs and the number of landmark locations, and writes out the reconciled mean transformations.

## Statistical Comparisons

Distributions of mean (or maximum) errors or internal inconsistencies among all possible pairwise registrations generated by two different registration strategies will be compared with one another using the two sided two sample Kolmogorov-Smirnov test implemented in the statistical package, S-Plus(MathSoft, Seattle, WA, U.S.A.). This test evaluates whether two samples are drawn from the same distribution by comparing their empirical cumulative distribution functions (16) and does not presuppose any particular shape for the underlying distribution. All unqualified or implicit references to significance testing will pertain specifically to Kolmogorov-Smirnov tests applied to mean or maximum errors or internal inconsistencies. Results reported as significant will indicate a p value of$<0.05$. All results to be reported as significant have been verified graphically to involve a consistent shift in one of the empiric cumulative distribution functions rather than an isolated change in shape of one of the empiric cumulative distribution functions.

## PET Phantom Validation

AIR 3.0 was validated for intramodality, intrasubject registration of PET data using the same brain phantom data set used to validate the original AIR algorithm (1). It consists of 31 PET scans of the Hoffman brain phantom (17) acquired at various known scanner gantry and bed positions. Image pairs with rotational misalignment of>30° and translational misalignment up to 10 mm are included. The amount of radioisotope in the phantom was selected to simulate the poor counting statistics of 2D PET $H_2{}^{15}O$ studies. Inflatable balloons containing higher levels of activity were deflated partway through the data acquisition to simulate focal changes in activity. The data set was acquired in 2D mode using a Siemens/CTI 831-08 tomograph (Siemens, Hoffman Estates, IL, U.S.A.). Images were reconstructed with a Shepp reconstruction filter with a roll-off frequency of 0.16 $mm^{-1}$ to generate images with a full width at half-maximum (FWHM) in-plane resolution of 6.1 mm. Voxels in the reconstructed images were 1.745 × 1.745 × 6.75 mm, and the image matrix dimensions were 128 × 128 × 15 planes. All images were smoothed with a 2D isotropic Gaussian filter to an in-plane resolution of 10 mm. During smoothing, images were scaled to map the hottest voxel in an image to the highest representable 8 bit value.

The goals of the phantom validation studies were (a) to compare AIR 3.0 with the original version of AIR: (b) to compare the different cost functions implemented in AIR 3.0 with one another; (c) to determine whether interpolation of the reference image to cubic voxels improves registration accuracy in data sets with highly anisotropic voxel sizes; (d) to determine whether sparse sampling of the data can be used to increase registration speed without adversely influencing registration accuracy; (e) to test the hypothesis that reconciling all possible pairwise registrations improves registration accuracy; and (f) to evaluate internal inconsistencies as a metric for registration accuracy.

To address these issues, all 465 unique pairs of the 31 phantom images were registered using several different strategies (see Table 1). These strategies also included the original AIR algorithm. In all cases, an 8 bit threshold value of 55 served to segment the reference image into brain and nonbrain values. Gold standard registration parameters were computed for each pair of images based on the known scanner gantry and bed positions: All 31 of the original data sets were resampled into a single common space using the gold standard transformation parameters and averaged(with weightings necessary to adjust for missing data from outside the field of view) to produce a single mean phantom image set. This image set was manually edited to remove all peripheral nonbrain voxels. The resulting brain mask was projected back onto the original 31 data sets using the gold standard registration parameters and used to distinguish brain from nonbrain voxels when measuring errors and internal inconsistencies. For each registration strategy, mean and maximum errors were computed. The mean registration times were recorded for each strategy. Reconciled mean transformations and mean internal inconsistencies were derived, and the mean errors of the reconciled mean transformations were computed for each strategy.

| Method | Convergence threshold | Final sampling (voxels) | Cubic voxels | Mean registration time (s) | Global errors (mm) mean (maximum) | Global mean internal inconsistency (mm) | Global reconciled mean errors (mm) |
|---|---|---|---|---|---|---|---|
| RIU | $10^{-6}$ | 1 | Yes | 185 | 0.323 (1.557) | 0.11 | 0.297 |
| RIU | $10^{-6}$ | 3 | Yes | 90 | 0.323 (1.557) | 0.111 | 0.297 |
| RIU | $10^{-6}$ | 9 | Yes | 51 | 0.324 (1.557) | 0.112 | 0.298 |
| RIU | $10^{-6}$ | 27 | Yes | 34 | 0.327 (1.609) | 0.117 | 0.299 |
| RIU | $10^{-5}$ | 1 | Yes | 141 | 0.326 (1.609) | 0.114 | 0.299 |
| RIU-LM | $10^{-5}$ | 1 | Yes | 82 | 0.325 (1.604) | 0.115 | 0.298 |
| RIU | $10^{-4}$ | 1 | Yes | 138 | 0.344 (1.667) | 0.133 | 0.309 |
| SLS | $10^{-6}$ | 1 | Yes | 323 | 0.333 (1.629) | 0.112 | 0.306 |
| SLS | $10^{-2}$ | 1 | Yes | 93 | 0.334 (1.619) | 0.114 | 0.307 |
| SLS-LM | $10^{-2}$ | 1 | Yes | 88 | 0.334 (1.621) | 0.114 | 0.307 |
| SLS | $10^{-1}$ | 1 | Yes | 87 | 0.337 (1.639) | 0.12 | 0.308 |
| SLS | $10^{-2}$ | 27 | Yes | 26 | 0.336 (1.639) | 0.117 | 0.308 |
| LS | $10^{-6}$ | 1 | Yes | 303 | 0.349 (1.639) | 0.116 | 0.323 |
| RIU | $10^{-5}$ | 1 | No | 47 | 0.369 (1.618) | 0.172 | 0.311 |
| SLS | $10^{-5}$ | 1 | No | 68 | 0.379 (1.711) | 0.169 | 0.323 |
| SLS | $10^{-2}$ | 1 | No | 23 | 0.378 (1.757) | 0.169 | 0.323 |
| LS | $10^{-5}$ | 1 | No | 73 | 0.395 (1.977) | 0.176 | 0.337 |

Results are based on 465 unique pairwise comparisons. Yes indicates that the reference volume was interpolated to cubic voxels before computing the corresponding reslice volume voxels and the cost function. See text for abbreviations.

**TABLE 1.** *Average registration times, global errors, and global internal inconsistencies for PET phantom validation*

To evaluate whether the focal activity changes simulated by inflating or deflating the balloons within the phantom affected registration accuracy, transformations derived using a given strategy were subdivided into those in which the balloons were in the same state in both scans (either both inflated or both deflated) and those in which the balloons were in different states. The error distributions were then compared across these two subgroups using Kolmogorov-Smirnov tests. This was repeated for each registration strategy independently.

### Intrasubject MRI Validation

A single normal subject was scanned eight times consecutively on a Phillips 1.5 T MR scanner. Sagittal volumes of 140 1 mm slices were acquired with a field of view of 256 × 204 mm. A 3D spoiled GRASS sequence [TR/TE = 18/10 ms, flip angle 30°, NSA (NEX) 1, flow compensation] was used with a total scan time per volume of 10 min 50 s. The subject (one of the authors) was highly motivated to prevent any head movements during the course of any given acquisition. Head packing was used to help prevent movements during acquisition, but no precautions were taken to prevent small head movements between acquisitions. The data were stored at 12 bit precision in 16 bit format with quantitative preservation of absolute scaling across studies. A set of 160 planes covering the brain from the bottom of the cerebellum to the top of the brain was selected in the first data set. The same planes (with respect to the scanner, not with respect to the anatomy) were selected in the other seven data sets, so any apparent movement between scans accurately reflects true movement by the subject. The images were similarly reduced along the right-left axis of the brain so that only the ears were visible in the most extreme sagittal planes of the first data set. The resulting data sets consisted of isotropic 1 mm voxels with dimensions of 256 × 160 × 160 voxels. All 28 possible unique pairwise registrations of the data were performed using several different registration strategies. The goals were (a) to compare the speed, internal consistencies, and results of the three cost functions; (b) to investigate the effects of smoothing on registration; (c) to investigate the effects of editing nonbrain structures on registration; (d) to determine the tradeoffs associated with different convergence criteria, minimization procedures, and data sampling strategies; and (e) to use internal inconsistencies to establish limits for true registration accuracy. The effects of cost function, smoothing, and editing of the images were investigated first. Isotropic, 3D smoothing filters with a FWHM of 0, 2.0, 4.0, and 8.0 mm were applied. The two images being registered were always filtered identically. To ensure consistency, manual editing to remove nonbrain structures was performed only once on an average of the eight unedited, unsmoothed images after co-registration with the LS cost function. The results of manual editing were then transformed back into the eight native image spaces (using the inverse of the transformations used for co-registration) in the form of masks. The registration algorithm masked each unedited image internally to avoid any tendency to align artificial edges created by the editing process. The default sampling intervals, convergence criteria, and minimization procedure were used. A 16 bit voxel value of 480 served to separate the head (brain and surrounding tissues) from background. For each registration strategy, the mean and maximum internal inconsistencies were measured after computing the corresponding set of reconciled mean transformations. Mean and maximum discrepancies between selected pairs of strategies were also computed.

The effects of different primary convergence thresholds, sampling densities, and the two different minimization procedures for each of the cost functions were investigated using unedited and unsmoothed data. A modified version AIR was also used to investigate whether computation of the LS cost function using windowed sinc interpolation would improve the internal inconsistency of the results. A cosine half-bell windowing technique identical to that described by Hajnal et al. (8) was used with full 3D sinc interpolation. Because sinc interpolation is extremely slow, the sinc version of the algorithm was always initialized with optimum parameters obtained using trilinear interpolation.

## RESULTS

### PET Phantom

Table 1 shows the average registration times, global mean errors, and global maximum errors for the various PET registration strategies. It also shows the global internal inconsistencies associated with each registration strategy and the global mean errors of the reconciled mean transformations. All methods resulted in subvoxel accuracy with global maximum errors of <2.0 mm.

As compared with using an uninterpolated reference image, interpolation of the reference image to cubic voxels universally resulted in significantly smaller errors and internal inconsistencies, independent of all other factors. This was achieved at the expense of registration time, which increased by a factor roughly proportional to the increase in the number of voxels through interpolation. Because of this consistent and significantly poorer performance, registrations

without interpolation to cubic voxels were excluded from the remainder of the analyses reported here. A convergence threshold of $10^{-4}$ for the RIU cost function gave significantly larger mean errors, maximum errors, and internal inconsistencies than more stringent thresholds of $10^{-5}$ or $10^{-6}$. Similarly, a convergence threshold of $10^{-1}$ for the SLS cost function resulted in significantly higher internal inconsistencies than more stringent thresholds of $10^{-2}$ or $10^{-6}$. These least stringent thresholds only modestly reduced registration times. The most stringent thresholds resulted in longer registration times with no significant improvement in accuracy compared with intermediate thresholds.

The LS cost function produced significantly larger mean errors than the RIU cost function. No other significant differences between the RIU, SLS, and LS cost functions were identified when using stringent or moderate convergence thresholds. The SLS cost function tended to be faster than the RIU cost function. The original AIR algorithm (1) produced a global mean error of 0.323 mm and a global maximum error of 1.568 with a mean registration time of 180 s. The errors are similar to those produced by the new algorithm using the RIU cost function.

Increasing the final sampling interval from every voxel to every 27th voxel for the RIU and SLS cost function did not significantly increase errors or internal inconsistencies and reduced computation times by a factor of 3-5. Speed improvements without significant loss of accuracy were also achieved by using Marquardt-like rather than full Newton-type minimization. The reconciled mean transformations invariably resulted in significantly smaller errors than the original transformations. Internal inconsistencies, reflecting deviations from these reconciled mean transformations, were always smaller than the errors based on the gold standards. Larger errors were correlated with larger internal inconsistencies for any individual cost function, but the internal inconsistency measures were insensitive to the differences between cost functions.

The sites of focal differences in activity resulting from inflation or deflation of the balloons had no significant effect on accuracy. The distributions of errors was invariably statistically indistinguishable for registrations in which the balloons were in the same state as for registrations in which the balloons were in different states.

### Intrasubject MRI

Visual inspection of the resampled MR images from the single subject confirmed that the registrations were qualitatively correct. Images of a transverse brain section through the anterior commissure before and after registration are shown in Fig. 2. The estimated maximum initial misalignment in the brain between pairs of images ranged from 0.26 to 3.80 mm. The estimated average initial misalignment across all brain regions and image pairs was 0.76 mm. Based on results from the SLS cost function, which were confirmed by direct inspection, the global intensity scaling varied to a small extent from one image to the next with the largest pairwise differences being [almost equal to]6%.

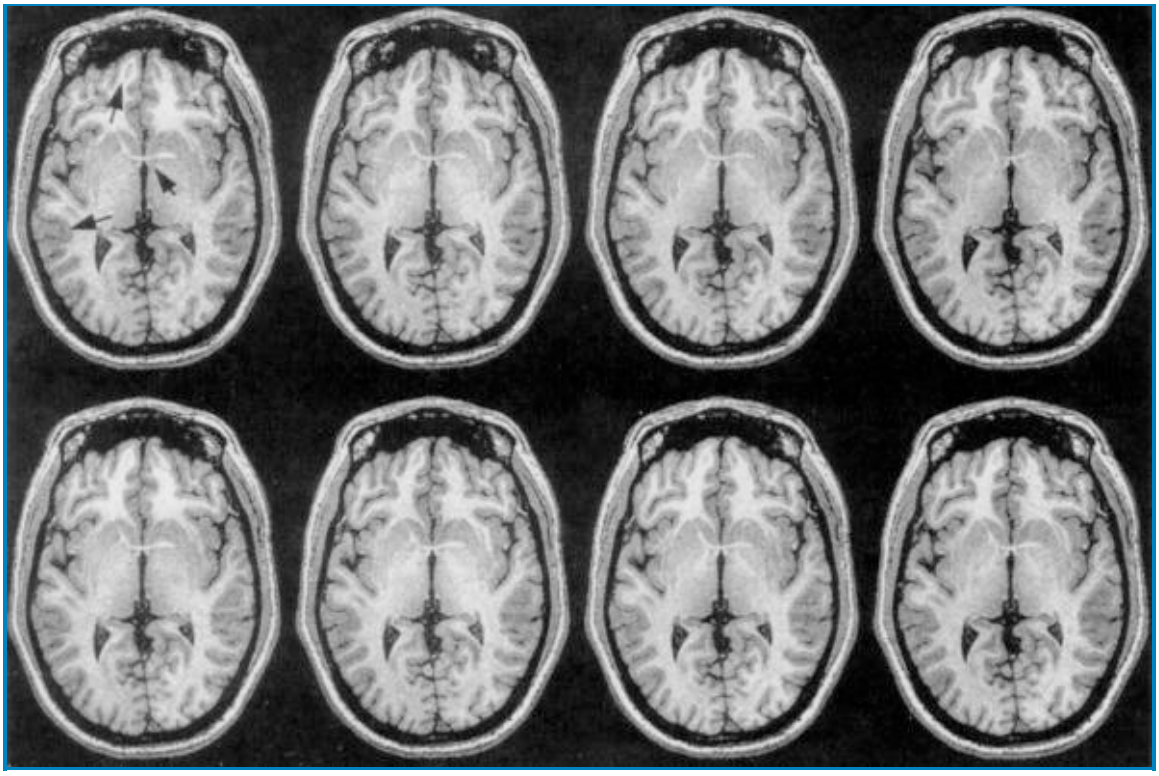

**FIG. 2.** A single transverse section from four of the eight image sets used for MRI validation. The first row shows the section before registration. Subtle differences due to misregistration can be seen and are highlighted by black arrows. The second row shows the same section after registration of the corresponding volumes using the LS cost function with sparse sampling at every 81st voxel, requiring <70 s per registration. Data resampling was performed using windowed sinc interpolation. All registration strategies investigated resulted in images virtually indistinguishable from those on the second row.

Table 2 shows the average registration time and internal inconsistencies of the three different cost functions with and without smoothing and editing of the data.Tables 3-5 show discrepancies between results obtained when cost function, smoothing, or editing was varied while keeping all other factors constant.

| Cost function | Smoothing filter (mm FWHM) | Edited | Mean time per registration (min) | Global mean (maximum) internal inconsistency (μm) |
|---|---|---|---|---|
| LS | 0 | No | 12.3 | 22 (75) |
| LS | 2 | No | 13.0 | 6 (19) |
| LS | 4 | No | 13.0 | 4 (15) |
| SLS | 0 | No | 4.9 | 22 (74) |
| SLS | 2 | No | 6.7 | 6 (19) |
| SLS | 4 | No | 6.7 | 4 (15) |
| RIU | 0 | No | 7.5 | 27 (96) |
| RIU | 2 | No | †[a] | 6 (23) |
| RIU | 4 | No | †[a] | 4 (20) |
| LS | 0 | Yes | 6.6 | 40 (125) |
| LS | 2 | Yes | 8.3 | 5 (20) |
| LS | 4 | Yes | 8.5 | 4 (17) |
| SLS | 0 | Yes | 2.8 | 40 (124) |
| SLS | 2 | Yes | 5.3 | 4 (16) |
| SLS | 4 | Yes | 5.2 | 5 (24) |
| RIU | 0 | Yes | 4.3 | 34 (111) |
| RIU | 2 | Yes | †[a] | 4 (14) |
| RIU | 4 | Yes | †[a] | 5 (16) |

Editing indicates whether masks were applied to exclude voxels outside the brain. Primary convergence thresholds were 1.0 for the LS and SLS cost functions and $10^{-5}$ for the RIU cost function. See text for abbreviations.

[a] Times are not listed because the registrations were initialized with optimal parameters obtained from the corresponding unsmoothed RIU registration.

**TABLE 2.** *Effects of cost function, smoothing filter, and data editing on MRI registration time and global internal inconsistency*

| Gaussian smoothing filter (mm FWHM) and editing | Cost function comparison | | |
|---|---|---|---|
| | LS vs. SLS global discrepancies mean (maximum) (μm) | LS vs. RIU global discrepancies mean (maximum) (μm) | SLS vs. RIU global discrepancies mean (maximum) (μm) |
| 0 | 1 (5) | 22 (123) | 22 (121) |
| 2 | 4 (13) | 15 (62) | 13 (59) |
| 4 | 10 (32) | 25 (69) | 18 (50) |
| 0(edited) | 1 (6) | 15 (59) | 15 (59) |
| 2(edited) | 3 (23) | 10 (33) | 10 (33) |
| 4(edited) | 7 (32) | 19 (62) | 19 (60) |

Registration results from the different cost functions were compared with one another. The Gaussian smoothing filters were applied before any editing to exclude nonbrain voxels. Corresponding registration times can be compared in Table 2. See text for abbreviations.

**TABLE 3.** *Global mean and maximum discrepancies between MRI registration results obtained with the three different cost functions as a function of smoothing filter and editing*

| Registration method | Gaussian smoothing filter comparison | | |
|---|---|---|---|
| | 0 vs. 2 mm FWHM global discrepancies mean (maximum) (μm) | 0 vs. 4 mm FWHM global discrepancies mean (maximum) (μm) | 2 vs. 4 mm FWHM global discrepancies mean (maximum) (μm) |
| LS | 109 (327) | 140 (383) | 40 (121) |
| SLS | 108 (326) | 137 (375) | 39 (121) |
| RIU | 98 (283) | 125 (331) | 38 (149) |
| LS (edited) | 101 (245) | 159 (388) | 72 (171) |
| SLS (edited) | 101 (263) | 160 (400) | 73 (167) |
| RIU (edited) | 104 (259) | 170 (391) | 78 (184) |

Registration results after applying various Gaussian smoothing filters to the data were compared with one another. Editing excluded nonbrain voxels. Corresponding registration times can be compared in Table 2. See text for abbreviations.

**TABLE 4.** *Global mean and maximum discrepancies between MRI registration results obtained with different smoothing filters as a function of registration method (cost function and editing)*

| Cost function | Gaussian smoothing filter: 0 mm FWHM Edited vs. unedited global discrepancies mean (maximum) (μm) | 2 mm FWHM Edited vs. unedited global discrepancies mean (maximum) (μm) | 4 mm FWHM Edited vs. unedited global discrepancies mean (maximum) (μm) |
|---|---|---|---|
| LS | 99 (312) | 74 (274) | 108 (406) |
| SLS | 99 (311) | 75 (273) | 110 (401) |
| RIU | 84 (261) | 74 (258) | 118 (390) |

Registration results obtained when using editing masks to exclude nonbrain voxels were compared with results obtained without using editing masks. Corresponding registration times can be compared in Table 2. See text for abbreviations.

**TABLE 5.** *Global mean and maximum discrepancies between results obtained from edited versus unedited MRI data as a function of registration method and Gaussian smoothing filter*

None of the three cost functions produced significantly better internal inconsistencies than either of the others. The SLS cost function was fastest and produced results within 32 μm of the LS cost function. The RIU cost function was unexpectedly faster than the LS cost function for unsmoothed images and gave results that always agreed with the other two cost functions to within 123 μm. The mean discrepancy between cost functions was always<=25 μm for a given smoothing and editing strategy.

Data smoothed with a 2 mm Gaussian filter invariably produced significantly smaller mean and maximum internal inconsistencies than unsmoothed data. This was true even for the RIU cost function, which required initialization with the optimal registration parameters obtained from the corresponding unsmoothed RIU registration when using smoothed data. Without such initialization, the RIU (and likewise the RIU-LM) cost function made excessively large steps at the first iteration and failed to find any way to improve the default initialization parameters. For unedited data (but not for edited data), even heavier smoothing with a 4 mm Gaussian filter produced additional significant improvements in internal inconsistencies. However, when smoothing with an 8 mm Gaussian filter was investigated for the LS and SLS cost functions, internal inconsistencies were always worse than with 4 mm smoothing but still superior to those with unsmoothed data. Smoothing increased registration time.Table 4 shows that smoothing altered registration results substantially. Results from unsmoothed images differed from those obtained with smoothed images and otherwise identical strategies by distances as large as 400 μm, and even global mean discrepancies were at least 98 μm.

To verify that these differences between the smoothed and unsmoothed results were not simply due to local minima, the unsmoothed, unedited registrations were repeated, using the corresponding optimal registration parameters obtained with 4 mm Gaussian smoothing as the starting parameters for minimization. Likewise, the unedited registrations using 4 mm Gaussian smoothing were repeated, using the optimal parameters obtained with unsmoothed data as starting parameters. To eliminate sparse sampling as a potential confound, both sets of repeat registrations utilized full sampling at every voxel from the first iteration. Despite being initialized with a different set of parameters, the final results were unchanged. For all three cost functions, global mean discrepancies between the original and repeat registrations never exceeded 10 μm.

The effects of editing the data to exclude scalp, skull, and dura varied depending on the smoothing strategy. In the absence of smoothing, editing led to significantly worse internal inconsistencies for the SLS and LS cost functions. In the presence of 2 mm Gaussian smoothing, editing led to significantly better internal inconsistencies for the SLS and RIU cost functions, and in the presence of 4 mm Gaussian smoothing, editing had no significant effect. Editing always reduced registration times, but by a factor of <2. Editing also substantially changed the registration results with global maximum discrepancies between edited and unedited results in Table 5 as large as 406 μm and global mean discrepancies always larger than 74 μm.

Table 6 shows the effects of sampling intervals, convergence thresholds, and minimization procedures for each cost function using unsmoothed, unedited data. Baseline measurements for each cost function were performed using full Newton-type minimization with a primary convergence threshold of zero, forcing as many iterations as the secondary termination criteria allow. Different sampling intervals and minimization strategies were then compared against the corresponding baseline values to compute the discrepancies shown in the table. The RIU convergence threshold of $10^{-5}$ is perhaps not optimally stringent for MRI data since the results with this threshold differed by as much as 83 μm from

those obtained with a convergence threshold of zero. The SLS and LS thresholds of 1.0 gave results within 10 μm of those obtained with a threshold of zero. For all three cost functions, sparser sampling was associated with a modest increase in global discrepancies from the baseline, an increase in global internal inconsistencies, and a substantial reduction in registration time. Marquardt-like minimization was also associated with shorter registration times, giving results almost identical to full Newton-type minimization for the LS and SLS cost functions, but more divergent results for the RIU cost functions. Both minimization strategies proved several times faster than the strategy used in the original version of AIR (data not shown).

| Cost function | Full Newton-type minimization | Convergence threshold | Sampling | Mean time per registration (min) | Global mean (maximum) discrepancies (μm) compared with convergence threshold of zero |
|---|---|---|---|---|---|
| LS | Yes | 0 | 1 | 118.7 | 0 (0) |
| LS | Yes | 1.0 | 1 | 12.3 | 2 (10) |
| LS | Yes | 1.0 | 9 | 2.8 | 3 (22) |
| LS | Yes | 1.0 | 81 | 1.1 | 10 (37) |
| LS | No | 1.0 | 1 | 4.1 | 2 (10) |
| LS | No | 1.0 | 9 | 1.3 | 3 (22) |
| LS | No | 1.0 | 81 | 0.8 | 11 (35) |
| SLS | Yes | 0 | 1 | 20.9 | 0 (0) |
| SLS | Yes | 1.0 | 1 | 4.9 | 2 (10) |
| SLS | Yes | 1.0 | 9 | 1.4 | 3 (22) |
| SLS | Yes | 1.0 | 81 | 0.8 | 10 (36) |
| SLS | No | 1.0 | 1 | 4.4 | 2 (10) |
| SLS | No | 1.0 | 9 | 1.4 | 3 (22) |
| SLS | No | 1.0 | 81 | 0.8 | 11 (38) |
| RIU | Yes | 0 | 1 | 81.9 | 0 (0) |
| RIU | Yes | $10^{-5}$ | 1 | 7.5 | 6 (83) |
| RIU | Yes | $10^{-5}$ | 9 | 1.9 | 7 (83) |
| RIU | Yes | $10^{-5}$ | 81 | 0.9 | 13 (83) |
| RIU | No | $10^{-5}$ | 1 | 4.0 | 7 (113) |
| RIU | No | $10^{-5}$ | 9 | 1.3 | 8 (113) |
| RIU | No | $10^{-5}$ | 81 | 0.8 | 14 (117) |

For each cost function, the first registration strategy with a convergence threshold of zero served as the basis for comparison when computing discrepancies. Unsmoothed, unedited data were used in all cases. See text for abbreviations.

**TABLE 6.** *The effects of minimization method, convergence threshold, and final sampling interval on MRI registration*

With use of sinc interpolation to compute the LS cost function with a very stringent convergence threshold of $10^{-5}$ and a final sampling interval of every voxel with unedited, unsmoothed data resulted in a global mean internal inconsistency of 20 μm and a global maximum internal inconsistency of 63 μm. The distributions of internal inconsistencies were not significantly better than with comparable use of trilinear interpolation and a less stringent convergence threshold.

Within the ranges investigated, the sampling interval and minimization method made the least difference, the primary convergence threshold, the cost function, and the interpolation model had an intermediate effect, and smoothing and editing had the greatest impact on the final results. All methods were in general agreement with global maximum discrepancies always<500 μm. The results are compatible with, but not conclusive proof of, typical registration accuracies on the order of 75-150 μm. It is possible that one of the strategies included here may be more accurate than the others, possibly with typical accuracies <10 μm, but this clearly cannot be the case for all of the strategies simultaneously.

## DISCUSSION

The AIR 3.0 registration algorithm is robust, fast, accurate, and applicable to a diverse range of registration problems. Except for the difficulties with registration of smoothed MRI data using the RIU cost function, no registration failures were identified, and the algorithm consistently achieved subvoxel accuracy. Reconciling internal inconsistencies among all possible pairwise registrations further improved registration accuracy, and quantification of these internal inconsistencies provided a basis for evaluating registration accuracy in the absence of gold standards.

### Estimating Registration Accuracy

When sufficiently accurate gold standards are available, quantification of registration accuracy is straight-forward. For

real data from living humans, gold standards sufficient to quantify subvoxel accuracy are difficult to achieve, especially when even movements between the brain and the skull cannot be disregarded as negligible (11). Accurate gold standards can be produced for phantoms or for simulated data sets, but the resulting assessments may be unrealistically optimistic since these methods generally fail to model all of the factors that contribute to scan-to-scan variability in real human data. These factors include noise, non-rigid body movements due to respiratory or cardiac cycles (18), artifacts due to rigid body movement during scan acquisition, distortions due to field inhomogeneities (2,19), and inaccuracies in distance calibration (20). For PET data, noise and spatial resolution are probably the main factors limiting registration accuracy. The PET phantom models these particular factors reasonably well, and the accuracies on the order of 2 mm can reasonably be extrapolated to humans.

For MRI data, phantom studies and simulations suggest that registration accuracies of a fraction of a millimeter are possible, but this may have little bearing on real human data. Hajnal et al.(8,11) have suggested that a statistical method (16) applicable to least-squares minimization can be used to assess registration accuracy for real human data. However, the errors estimated by this method are only the errors in identifying the true minimum of the cost function. Errors in the underlying assumption that the images are correctly registered when the cost function is minimized are not evaluated. When applied to the MRI least-squares minimizations reported here, this method confirms that the true minimum of the cost function should be identified within 10 μm, but the associated internal inconsistencies and discrepancies between strategies indicate that the true accuracy cannot be uniformly this good.

Even in the absence of gold standards, internal inconsistencies place an irrefutable limit on the degree of registration accuracy that can be achieved since the true inaccuracy as measured against any gold standard is guaranteed to be at least as large as the internal inconsistency. However, true accuracy can always be worse than the internal inconsistencies imply, so a strategy that performs worse in terms of internal consistency might nonetheless turn out to be the more accurate strategy as measured against independent gold standards. All other things being equal, it would seem prudent to prefer a method that produces significantly smaller internal inconsistencies over some less consistent method; but so long as the internal inconsistencies do not indicate intolerably poor accuracy, any valid advantage could provide sufficient justification for use of a less consistent method.

### Choice of Minimization Procedure

The performance of the multivariate calculus-based minimization procedures in AIR 3.0 matched and sometimes substantially exceeded the univariate minimization procedure used by the original AIR algorithm. The full Newton-type minimization procedure is a direct implementation of the theoretical foundation of all calculus- based minimization procedures for image registration (16). Other minimization procedures, such as those used by Hajnal et al. (8), Alpert et al.(13), and Friston et al. (12) are based on approximations to full Newton-type minimization, as is the Marquardt-like approximation described here. All these approaches are valid from a theoretical standpoint (16), and in many contexts the differences in speed and accuracy between them are probably unimportant. Use of the predicted change of the cost function as the primary termination criterion for minimization avoids the registration failures that can be seen in the presence of large movements when using a small, fixed number of iterations (12). For the RIU cost function, a primary convergence criterion of $10^{-5}$ is generally sufficient for intrasubject registration of PET data, but a more stringent criterion might be better for MRI data. For the least-squares cost functions with 16 bit data (12 significant bits), a primary termination criterion around 1.0 is effective. For 8 bit data, a lower value around 0.1 may be appropriate. The ideal termination criteria may vary as a function of the nature of the data, so these values should be viewed as only approximate guidelines.

### Choice of Cost Function and Interpolation Model

The LS cost function performed significantly worse than the RIU cost function for PET data. This probably relates to the fact that the LS cost function was unable to take differences in global intensity scaling of the images into account. The SLS cost function, which includes an explicit intensity scaling factor, did not perform significantly worse than the RIU cost function for PET data. For MRI data, no significant differences among cost functions were found, except that the RIU cost function failed with smoothed data unless initialized with parameters already very close to the correct answer. The SLS cost function is probably the best overall choice for MRI data since it is insensitive to global intensity differences by design and was robust and generally the fasted in all contexts.

Snyder (14) has suggested that the SLS cost function should be superior to the RIU cost function on theoretical grounds and has presented phantom PET data using the LS cost function in support of this suggestion. The discrepancy between Snyder's PET results and those presented here may be due to differences in methodology since Snyder kept the phantom's position stationary in all scans and quantitatively scaled the image intensities to a consistent value. This approach does not simulate interpolation errors or positional distortions related to image acquisition, nor does it evaluate performance when image intensity is variable. Eberl et al. (21) have also performed PET phantom studies and concluded that a cost function similar to the LS cost function is more accurate than the RIU cost function. However, interpretation of their results is clouded by the fact that their independent implementation of the RIU cost function produced errors far larger than have been reported previously (1) or than were identified here.

Although sinc interpolation is theoretically the ideal interpolation function for resampling hand-limited MRI data when a fully 3D acquisition technique is used, the results presented here suggest that trilinear interpolation is adequate for registration. The results thus obtained can then be used to resample the data using sinc or chirp-*z* interpolation to generate high quality final images.

### Image Smoothing and Editing

For PET registration, smoothing has been shown previously to improve registration accuracy of noisy images (1). For the MRI data presented here, modest smoothing with a 2 to 4 mm isotropic Gaussian filter improved the internal consistency of registration results. Smoothing systematically changed the registration results more than expected based on the associated internal consistencies, showing that optimal minimization of the cost function does not necessarily imply optimal registration of the images. In the absence of gold standard MRI data that would allow definitive determination of which resolution gives the most accurate results, smoothing of MRI images with a 2 to 4 mm Gaussian filter is a reasonable approach.

Hajnal et al. (11) recommend routine editing of nonbrain structures for intrasubject registration since the brain can move with respect to the skull between image acquisitions. This is presumably less likely to be problematic during a single imaging session than with images acquired on separate days. If no relative movement has actually occurred, such editing effectively throws away sharply defined boundaries that could have served to improve registration accuracy. No evidence was found to suggest that movement of the brain relative to the skull occurred in MRI data sets used here, and editing sometimes significantly worsened internal inconsistencies. Nonetheless, editing of the images is certainly appropriate if movements of the brain relative to the scalp and skull are known or likely to be present. Editing might also be helpful in situations where the rigid body model is significantly violated by imaging artifacts.

### Spatial Transformation Model

For intrasubject registration, a rigid body model is generally assumed to be the spatial transformation model of choice. However, if distances are poorly calibrated, a more general linear model with 7-11 parameters, depending on which distances are known to be inaccurate, may give better results. In data not presented here, use of a 12 parameter affine model to register the MRI data sets resulted in internal inconsistencies that were worse than those obtained with the rigid body model. Alternatively, calibration errors estimated from anatomic landmarks within the images can be used to adjust voxel dimensions before registration with a rigid body model (20). It is advisable to correct significant spatial or intensity distorations associated with magnetic susceptibility artifacts (19) or with the use of local gradient coils (22) before registration since these distorations are not readily modeled with linear scaling of spatial distortions or intensity. For extremely fast MR techniques such as echo planar imaging, it may be appropriate to consider the use of nonlinear spatial transformation models (6) to compensate for distorations associated with respiratory and cardiac cycles. For slower imaging techniques, it should be kept in mind that, at best, the acquired images represent a temporal average of an object that is subject to constant small scale physiologic fluctuations. To the extent that this temporal averaging varies from one image to the next, the very notion of an absolutely correct spatial transformation model may be unjustified.

### Registration Speed

If reduced accuracy can be tolerated, sparse sampling of the data can be used to substantially increase registration speed. For PET data with anisotropic voxels, sparse sampling is a more accurate way to increase speed than omitting the

step of interpolating the reference volume to cubic voxels. If initial misregistration is large, the secondary termination criteria may need to be adjusted when using sparse sampling to allow a sufficient total number of iterations.

### Generality of Results

Because of its generality, AIR is not restricted to any specific modality, species, or anatomic structure. The systematic investigation of various factors that might alter speed or accuracy reported here provides insights that are applicable not only to AIR, but also to other similar registration algorithms. To the extent that the results may depend on the specific data sets investigated here, the general approaches described for validation are nonetheless broadly applicable. This is particularly true of the internal consistency measures, which require only that three or more images be available for pairwise registration. We hope that the methods and results described here will help users of a variety of registration methods ensure the accuracy of their results while maintaining practical levels of performance.

**Acknowledgment:** This work was supported by National Institute of Neurological Disorders and Stroke grant 1 K08 NS-01646, Department of Energy contract DE-FCO3-87ER60615, NIH grant 5P01MH52176 to the International Consortium for Brain Mapping, the Canadian Medical Research Council SP-30, generous gifts from the Pierson-Lovelace Foundation, the Ahmanson Foundation, the Tamkin Foundation, North Star Fund, and the Brain Mapping Medical Research Organization.

## REFERENCES

1. Woods RP, Cherry SR, Mazziotta JC. Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr* 1992;16:620-33. [Context Link]

2. Jiang AP, Kennedy DN, Baker JR, et al. Motion detection and correction in functional MR imaging. *Hum Brain Map* 1995;3:224-35. [Context Link]

3. Woods RP, Mazziotta JC, Cherry SR. MRI-PET registration with automated algorithm. *J Comput Assist Tomogr* 1993;17:536-46. [Context Link]

4. Strother SC, Anderson JR, Xu XL, Liow JS, Bonar DC, Rottenberg DA. Quantitative comparisons of image registration techniques based on high-resolution MRI of the brain. *J Comput Assist Tomogr* 1994;18:954-62. [Context Link]

5. West J, Fitzpatrick JM, Wang MY, et al. Comparison and evaluation of retrospective intermodality image registration techniques.*J Comput Assist Tomogr* 1997;21:554-66. **Ovid Full Text** | [Context Link]

6. Woods RP, Grafton ST, Watson JDG, Sicotte NL, Mazziotta JC. Automated image registration: II. Intersubject validation of linear and nonlinear models. *J Comput Assist Tomogr* 1998;22:155-67. [Context Link]

7. Rabiner LR, Schafer RW, Rader CM. The chirp z-transform algorithm and its application. *Bell System Tech J* 1969;48:1249-92. [Context Link]

8. Hajnal JV, Saeed N, Soar EJ, Oatridge A, Young IR, Bydder GM. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *J Comput Assist Tomogr* 1995;19:289-96. [Context Link]

9. Unser M, Thevenaz P, Yaroslavsky L. Convolution-based interpolation for fast, high-quality rotation of images. *IEEE Trans Image Proc* 1995;4:1371-81. [Context Link]

10. Fraser D, Schowengerdt RA. Avoidance of additional aliasing in multipass image rotations. *IEEE Trans Image Proc* 1994;3:721-35. [Context Link]

11. Hajnal JV, Saeed N, Oatridge A, Williams EJ, Young IR, Bydder GM. Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. *J Comput Assist Tomogr* 1995;19:677-91. [Context Link]

12. Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images.*Hum Brain Map* 1995;3:165-89. [Context Link]

13. Alpert NM, Berdichevsky D, Levin Z, Morris ED, Fischman AJ. Improved methods for image registration. *Neuroimage* 1996;3:10-8. [Context Link]

14. Snyder AZ. Difference image versus ratio image error function forms in PET-PET realignment. In: *Quantification of brain function using PET*. San Diego: Academic Press, 1996;131-7. [Context Link]

15. Woods RP, Mazziotta JC, Cherry SR. Automated image registration. In: Uemura K, Lassen NA, Jones T, Kanno I, eds.*Quantification of brain function: tracer kinetics and image analysis in brain PET*. Amsterdam: Excerpta Medica, 1993;391-8. [Context Link]

16. Press WH, Teukolsky SA, Vetterling WT, Flannery BP.*Numerical recipes in C: the art of scientific computing*. 2nd ed. Cambridge: Cambridge University Press, 1992. [Context Link]

17. Hoffman EJ, Cutler PD, Guerrero TM, Digby WM, Mazziotta JC. Assessment of accuracy of PET utilizing a 3D phantom to simulate the activity distribution of [$^{18}$F]fluorodeoxyglucose uptake in the human brain. *J Cereb Blood Flow Metab* 1991;11:A17-25. **Bibliographic Links** | [Context Link]

18. Poncelet BP, Wedeen VJ, Weisskoff RM, Cohen MS. Brain parenchyma motion: measurement with cine echo-planar MR imaging.*Radiology* 1992;185:645-51. **Bibliographic Links** | [Context Link]

19. Chang H, Fitzpatrick JM. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities.*IEEE Trans Med Imag* 1992;11:319-29. [Context Link]

20. Freeborough PA, Woods RP, Fox NC. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *J Comput Assist Tomogr* 1996;20:1012-22. **Ovid Full Text** | [Context Link]

21. Eberl S, Kanno I. Fulton RR, Ryan A, Hutton BF, Fulham MJ. Automated interstudy image registration technique for SPECT and PET.*J Nucl Med* 1996;37:137-45. **Bibliographic Links** | [Context Link]

22. Zeffiro TA, VanMeter JW, Eden GF, Woods RP. Signal gradient normalization improves head motion correction in surface-coil functional magnetic resonance imaging. *Neuroimage* 1996;3:S110. [Context Link]

Index Terms: Image registration; Magnetic resonance imaging; Emission computed tomography; Brain mapping

*Accession Number: 00004728-199801000-00027*