[Presented at Medical Imaging 1996, Newport Beach, CA, February 10–15, 1996. Will appear in *Medical Imaging 1996: Image Processing*, SPIE Proc. Vol. 2710.]

Comparison and evaluation of retrospective intermodality image registration techniques

Jay West[1], J. Michael Fitzpatrick[1], Matthew Y. Wang[1], Benoit M. Dawant[2], Calvin R. Maurer, Jr.[3], Robert M. Kessler[4], Robert J. Maciunas[5], Christian Barillot[6], Didier Lemoine[6], André Collignon[7], Frederik Maes[7], Paul Suetens[7], Dirk Vandermeulen[7], Petra A. van den Elsen[8], Paul F. Hemler[8], Sandy Napel[8], Thilaka S. Sumanaweera[8], Beth Harkness[9], Derek L. G. Hill[10], Colin Studholme[10], Gregoire Malandain[11], Xavier Pennec[11], Marilyn E. Noz[12], Gerald Q. Maguire, Jr.[12], Michael Pollack[12], Charles A. Pelizzari[13], Richard A. Robb[14], Dennis Hanson[14], Roger P. Woods[15]

[1]Department of Computer Science, [2]Department of Electrical Engineering, [3]Department of Biomedical Engineering, [4]Department of Radiology, [5]Department of Neurosurgery, Vanderbilt University, Nashville, TN; [6]Université de Rennes, Rennes Cedex, France; [7]Katholieke Universiteit, Leuven, Belgium; [8]Stanford University Medical Center, Stanford, CA; [9]Bowman Gray School of Medicine, Winston-Salem, NC; [10]Guy's Hospital, London, U.K.; [11]INRIA, Sophia Antipolis, France; [12]New York University Medical Center, New York, NY; [13]University of Chicago Hospital, Chicago, IL; [14]Mayo Clinic, Rochester, MN; [15]UCLA School of Medicine, Los Angeles, CA

ABSTRACT

All retrospective image registration methods have attached to them some intrinsic estimate of registration error. However, this estimate of accuracy may not always be a good indicator of the distance between actual and estimated positions of targets within the cranial cavity. This paper describes a project whose principal goal is to use a prospective method based on fiducial markers as a "gold standard" to perform an objective, blinded evaluation of the accuracy of several retrospective image-to-image registration techniques.

Image volumes of three modalities—CT, MR, and PET—were taken of patients undergoing neurosurgery at Vanderbilt University Medical Center. These volumes had all traces of the fiducial markers removed, and were provided to project collaborators outside Vanderbilt, who then performed retrospective registrations on the volumes, calculating transformations from CT to MR and/or from PET to MR, and communicated their transformations to Vanderbilt where the accuracy of each registration was evaluated. In this evaluation the accuracy is measured at multiple "regions of interest", i.e. areas in the brain which would commonly be areas of neurological interest. A region is defined in the MR image and its centroid C is determined. Then the prospective registration is used to obtain the corresponding point C' in CT or PET. To this point the retrospective registration is then applied, producing C'' in MR. Statistics are gathered on the *target registration error* (TRE), which is the disparity between the original point C and its corresponding point C''.

A second goal of the project is to evaluate the importance of correcting geometrical distortion in MR images, by comparing the retrospective TRE in the rectified images, i.e. those which have had the distortion correction applied, with that of the same images before rectification. This paper presents preliminary results of this study along with a brief description of each registration technique and an estimate of both preparation and execution time needed to perform the registration.

Keywords: Medical imaging, image registration, fiducial markers, accuracy validation, MR distortion, CT, MR, PET.

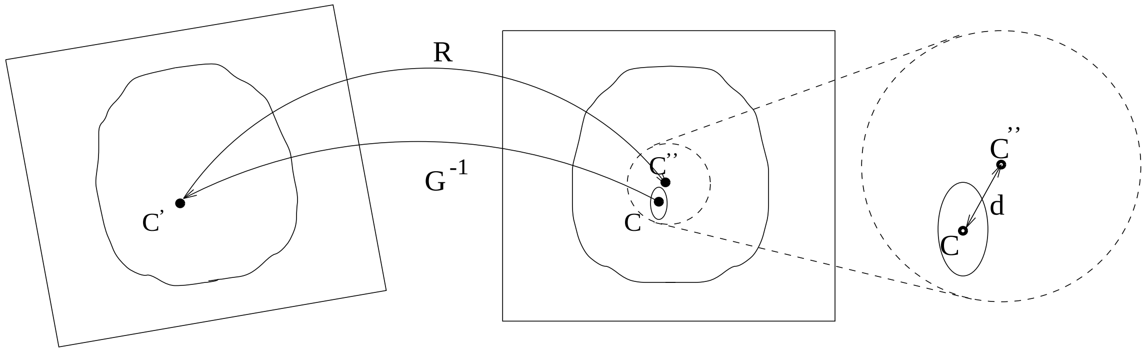

Figure 1. Calculation of registration error using VOIs. The Euclidean distance d between $C = G(C')$ and $C'' = R(C')$ is the error. The ellipse represents a VOI.

1. INTRODUCTION

All retrospective image registration methods have attached to them some intrinsic estimate of registration error. However, this estimate of accuracy may not always give a true representation of the distance between actual and estimated positions of targets within the cranial cavity, i.e. the *target registration error* (TRE). By using a fiducial marker based method[14,16] as a "gold standard", and by making the fiducial markers invisible, we are able to perform a detailed, blinded evaluation of retrospective techniques based on their TRE at several landmark locations within the brain.

There were two registration tasks evaluated in this study, CT to MR and PET to MR, and these tasks were broken into subtasks according to the type of MR and to whether or not the MR image was corrected for geometrical distortion. The image data sets of nine patients were used, seven of which contained both CT and MR scans, and seven with PET and MR. MR scans of three type—T1, PD, and T2—were included. Before imaging, each patient had four markers implanted and a COMPASS stereotactic frame attached. For some of the patients, scans were also used that had been corrected for geometrical distortion.[3,15] The first step toward evaluation is the calculation of the "answers": registrations derived with the aid of the fiducial markers. Next it is necessary to process the images by removing all traces of the markers and the frame in order to ensure that all subsequent registrations are truly retrospective in nature. This was done by manually segmenting the frame and markers out of the images, and replacing the removed regions by a simulated background pattern in each imaging modality. We call this process "air brushing."

These air-brushed image volumes were then made available on the Internet via FTP so that each site could apply its own retrospective technique. The resulting transformations were reported by specification of the motion of the eight corner points of the volume. A measurement of error is made by first determining from the given corner motions the rigid-body transformation for all points and then calculating the error in millimeters relative to the gold standard over a set of specified regions known as *volumes of interest* (VOIs). This is done by taking the centroid, C, of a VOI (defined in the "To" modality, i.e. MR), and applying the inverse of the gold standard transformation (see Figure 1), thus creating a point C' in the "From" modality.

To C' the retrospective registration is applied, giving result C'' in the "To" modality. The Euclidean distance between points C and C'' represents the TRE of the retrospective method in this VOI.

Each submission of retrospective transformations was accompanied by a statement indicating the cases in which the registration procedure was felt to have failed, i.e. was not good enough to be clinically useful. In all cases, the statements indicated that the registration was successful on every data set provided.

2. METHODS

2.1. Image acquisition

The specifications of the image volumes used in this study are listed in Table 1. The CT images were acquired using a Siemens DR-H scanner, the MR images using a Siemens SP 1.5 Tesla scanner, and the PET images with

Modality	Resolution			Voxel Size (mm)			Voxel data type
	x	y	z	x	y	z	
CT	512	512	28–34	0.65	0.65	4.0	16 bit, two's complement
MR	256	256	20–26	1.25–1.28	1.25–1.28	4.0	16 bit, two's complement
PET	128	128	15	2.59	2.59	8.0	16 bit, two's complement

Table 1. Description of the image volumes used in this project.

a Siemens/CTI ECAT 933/08–16 scanner. For MR, T1-weighted (T1), PD-weighted (PD), and T2-weighted (T2) Spin-Echo images were acquired. The T1 image volumes were acquired with an echo time (TE) of 15 ms and a repetition time (TR) of 650 ms (20 slices) or 800 ms (26 slices), the PD with TE of 20 ms and TR of 2550 ms (20 slices) or 3000 ms (26 slices), and the T2 with TE of 90 ms and the same TR as PD. Readout gradient strength for T1 was 2.45 mT/m and for PD/T2 was 1.47 mT/m, with four acquisitions in T1 and two in PD/T2. All MR images used half-Fourier reconstruction and a slice selection gradient of 6.8 mT/m. Three additional MR images were acquired for each patient with the identical imaging parameters except that the readout gradient was reversed. For PET, each patient was injected with 10 mCi of ^{18}F-fluorodeoxyglucose. Scanning was started 40–50 minutes after injection and continued for 25 minutes. Image reconstruction was performed using a Hamming reconstruction filter, resulting in images with a full-width at half-maximum resolution of 9 mm.

Internal Review Board authorization was obtained for the use of the patient data sets in this study. All patients whose images were to be used signed a release form indicating their consent.

2.2. Geometrical distortion correction

We correct MR images for static field inhomogeneity by using the image rectification technique of Chang & Fitzpatrick.[3,15] A new image, without inhomogeneity distortion, is generated from the pair of distorted images acquired with reversed readout gradients. The MR images are corrected for scale distortion by using the COMPASS stereotactic frame as an object of known shape and size. The scale factors are handled by changing the voxel dimensions in the image header.

2.3. Fiducial markers and fiducial localization

The markers are designed to be bright in CT and MR. They are constructed from hollow plastic cylinders with an inside diameter of 7 mm and an inside height of 5 mm. Plastic marker bases or posts are screwed into the outer table of the skull of the patient. The markers are attached to the posts during image acquisition. Additional detail about the markers, including pictures and image slices showing the typical appearance of the markers in CT and MR images, can be found in previous publications.[15,21]

We define the position of a marker as its centroid and call the determination of this position *fiducial localization*. We calculate an intensity-based centroid for each marker using the localization technique described in Wang et al.[21] This iterative, knowledge-based technique automatically finds the lowest threshold such that an object formed from voxels whose intensities are higher than the threshold and that are three-dimensionally connected to a selected voxel is neither too large nor too small to be a marker. The large point spread function increases the effective size of the marker substantially in PET. The published localization algorithm therefore was modified for this project to take advantage of the effective size in determining the centroid in the direction perpendicular to a slice.

2.4. Fiducial-based registration

When we use markers to register images, we call them *fiducial markers* and call their positions *fiducial points* or *fiducials*. We register CT to MR, and PET to MR, by calculating the rotation and translation parameters of the rigid body transformation that minimizes the mean square distance between corresponding fiducials in the two images.[14,16] We have implemented the closed-form solution developed by Arun, Huang & Blostein.[1] The rotation matrix is computed using the singular value decomposition of the covariance matrix of the centroid-subtracted position vectors in the two spaces. The translation vector is calculated as the difference between the centroids of the

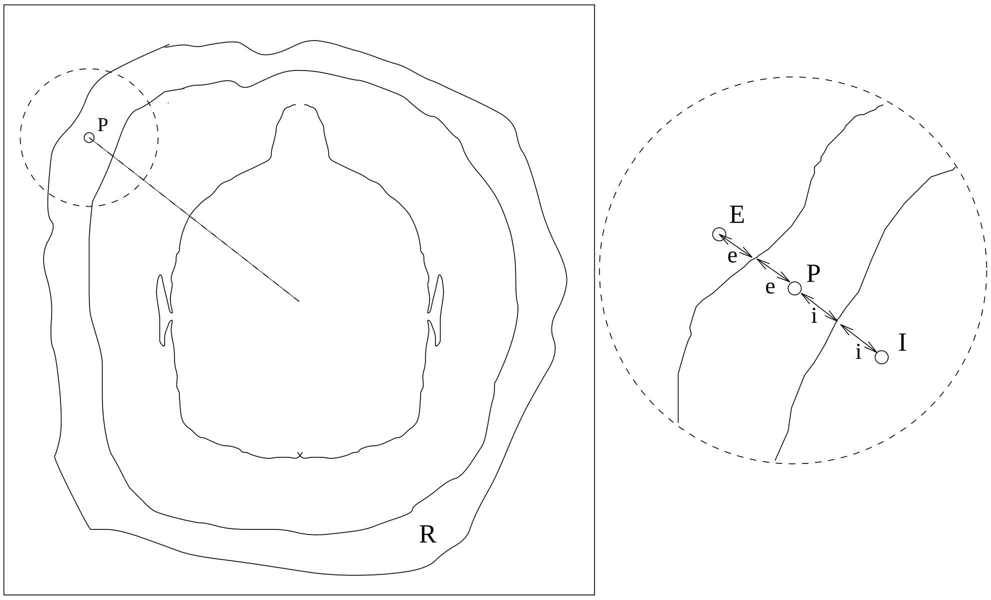

Figure 2. Calculation of points for interpolation of background patterns in CT and PET images.

two sets of fiducials. We define the *fiducial registration error* (FRE) as the root-mean-square (rms) distance between corresponding fiducials after registration and transformation.

This registration process was carried out for each patient data set and each MR modality. The derived transformations were tabulated, and used as a gold standard for evaluation of the retrospective registration methods.

2.5. Removal of fiducial markers and stereotactic frame

The next step is the removal of all traces of the fiducial markers and the stereotactic frame from each image. This was achieved by manual outlining of regions containing these structures, followed by an approximate reconstruction of the image background in each missing region R. In MR, where the background consists of unstructured noise, pixels at random positions between the edge of R and the lateral image boundary were sampled and placed in R. In CT, as the outer regions of the image are comprised mainly of reconstruction artifacts which take the form of quasi-radial stripes, the approach taken was to interpolate these stripes within the removed region. This was done at a given point P in region R by the following method (see Figure 2):

1. Calculate the radial distances e and i of the point P respectively from the external and internal boundaries of the region.
2. Identify points E and I at radial distances e and i respectively external and internal to the region. If E lies outside the image, set E to be on the border of the image. A similar precaution is taken to insure that I does not lie within the head.
3. Let the intensity at E be I_e, and at I be I_i. Assign P's intensity I_p as $I_p = I_e$ with probability ρ, or $I_p = I_i$ with probability $1 - \rho$, where $\rho = i/(i + e)$.

A similar technique was used for the PET images; in this case, however, pixels in region R were set to an intensity value linearly interpolated between the intensity of the internal and external boundaries of R.

This process is illustrated in Figures 3 through 5. Figure 3 shows slices of original volumes from each of the three modalities. The gray levels have been set so that the background artifacts may be seen. Figure 4 shows the same slices after the region R has been outlined and zeroed. This procedure is applied to each slice in the volume. Figure 5 shows the slices after reconstruction of the background in the region R. For the MR case, it can be seen that the replaced area is indistinguishable from the rest of the background. In CT and PET, there are slight discontinuities in the direction of the stripes, but the intensity changes smoothly.

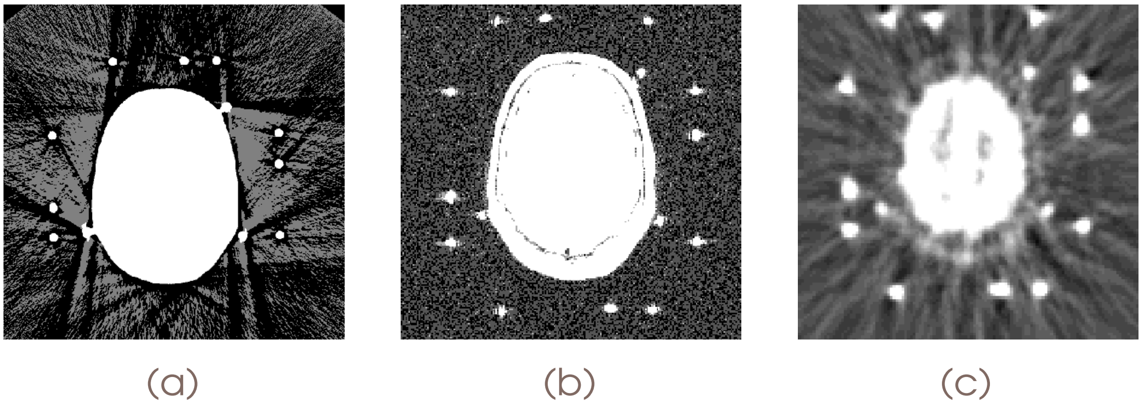

Figure 3. Original image volumes: (a) CT, (b) MR, and (c) PET. The stereotactic frame (bright circular spots towards the edge of the image) and fiducial markers (three bright spots near the head) are clearly visible in all three modalities.

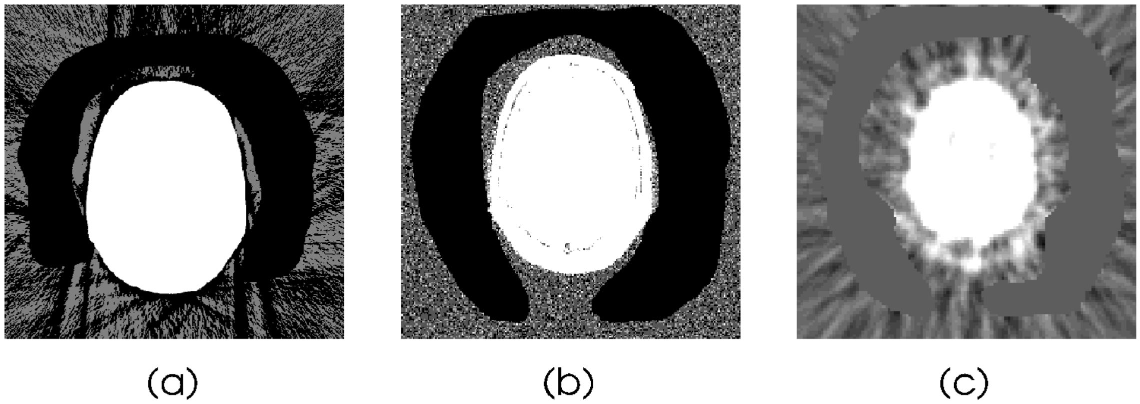

Figure 4. Image volumes after extraction of the region R: (a) CT, (b) MR, and (c) PET. Note that all trace of the stereotactic frame and fiducial markers has been removed.

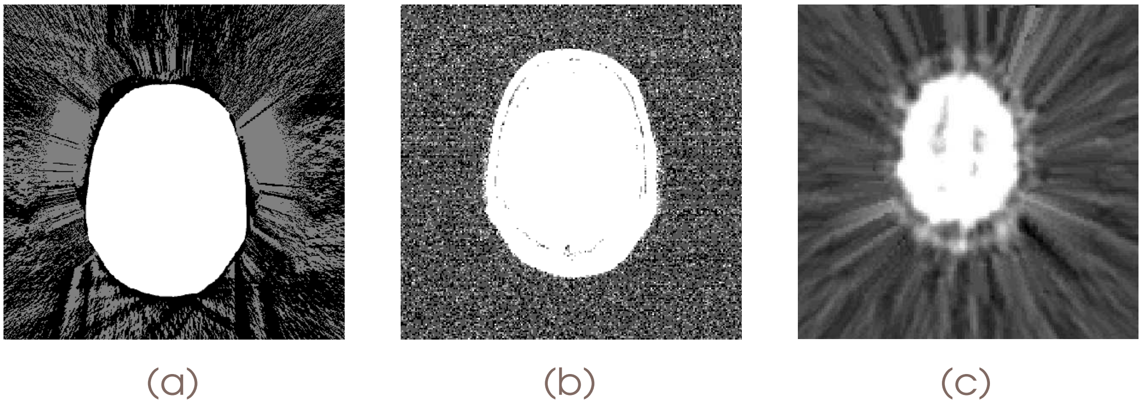

Figure 5. Image volumes after reconstruction of the background in R: (a) CT, (b) MR, and (c) PET.

For each air-brushed image volume an ASCII header for each volume was prepared. Each header contains a description of the image to which it relates, giving the pixel size, slice thickness, resolution, data format, and acquisition information. In the MR case the headers also contain a line stating the image to be either an unaltered volume, or a volume to which the distortion correction technique has been applied. The headers conform to the Interfile standard.[2,5]

2.6. Communications

After creation of the images and headers was complete, a login name, password, and Internet address were provided to all project participants by e-mail, so that they were able to transfer the images and headers to their own sites by means of FTP.

Once each retrospective registration had been completed, it was necessary that it be communicated to Vanderbilt in order to be evaluated. The various retrospective techniques to be evaluated employed a wide range of coordinate systems and transformation representation methods. Therefore, in order that a minimum of error and confusion be introduced into the data analysis, it was essential that a common coordinate system and format be used in all communications. The challenging step was to develop a protocol which allowed both simple interpretation during the analysis step, and easy conversion from other representation schemes.

The following scheme was adopted: In the "From" volume (e.g. CT in the case of CT-to-MR registration), the positions of eight points are calculated: taking the origin to be the center of the first voxel in the volume (i.e. the top left pixel of slice zero), the x, y and z coordinates of the centers of the eight corner voxels in the volume were derived. These positions were provided via FTP by Vanderbilt for every CT and PET volume in the form of a partially completed (left hand side) "transformation table" (see Figure 6 and Table 2) for each pair of volumes.

After the retrospective registration tranformation was determined, the transformed positions of these eight points *relative to the origin of the "To" modality* (see Figure 7) were computed by each site (see Table 2, right hand side). As depicted in Figure 6 the field of view of the two volumes is typically different, so it is important to specify which volume provides the origin relative to which the transformed positions are calculated.

All coordinates were specified to at least four decimal places in units of millimeters. Such high precision insures that any round-off error inherent in converting between a registration transformation and the eight-point sets is negligible. In order to convert the transformation table to a rigid-body transformation, the two point sets are registered using the point-based registration algorithm described in Section 2.4. Only three points are necessary to uniquely specify such a transformation, but the full set of eight was used for reasons of symmetry, error reduction, and error prevention.

Clearly this method of data transmission allows only rigid-body transformations to be accurately communicated, since any non-rigid transformation would be approximated by a rigid one. The use of this protocol thus limits the scope of this project to an evaluation restricted to rigid-body transformations. However, by measuring the FRE (that is produced by the point-based registration algorithm) it is possible to determine whether or not the retrospective registration uses non-rigid deformations since the FRE would be of the order of millimeters, as opposed to hundredths of a millimeter, for a well-specified rigid-body transformation. This feature is facilitated by the use of a larger point set than necessary, and guards against the possibility of a non-rigid transformation being mistakenly supplied and evaluated as if it were rigid.

Each transformation was transmitted to Vanderbilt by e-mailing an ASCII file containing a completed transformation table (see Table 2).

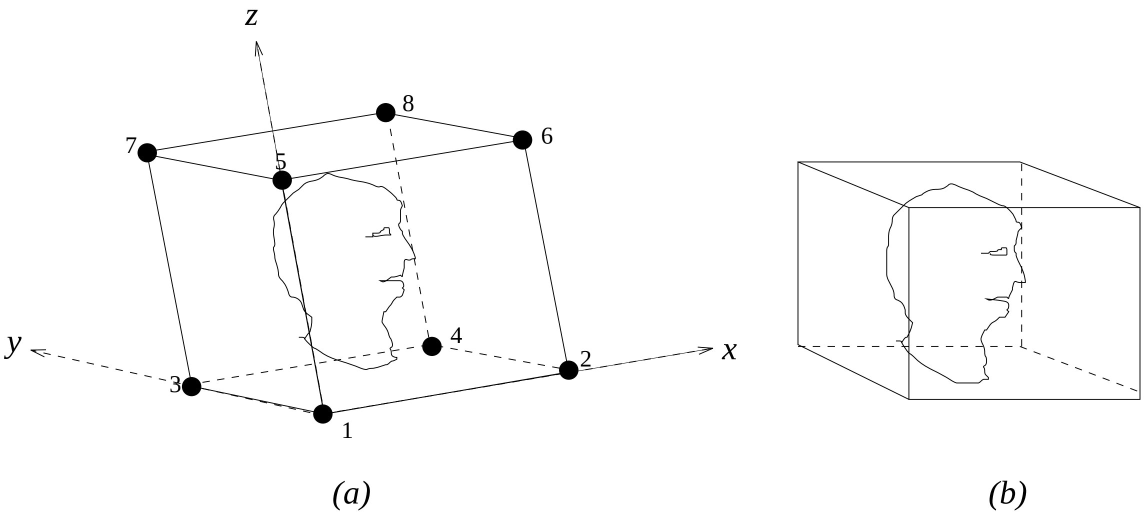

Figure 6. Image volumes prior to registration: (a) "From" volume and (b) "To" volume. The points 1 to 8 are defined as the corners of the volume, relative to the axes shown, and form the left hand side of the registration table shown in Table 2 of this paper.

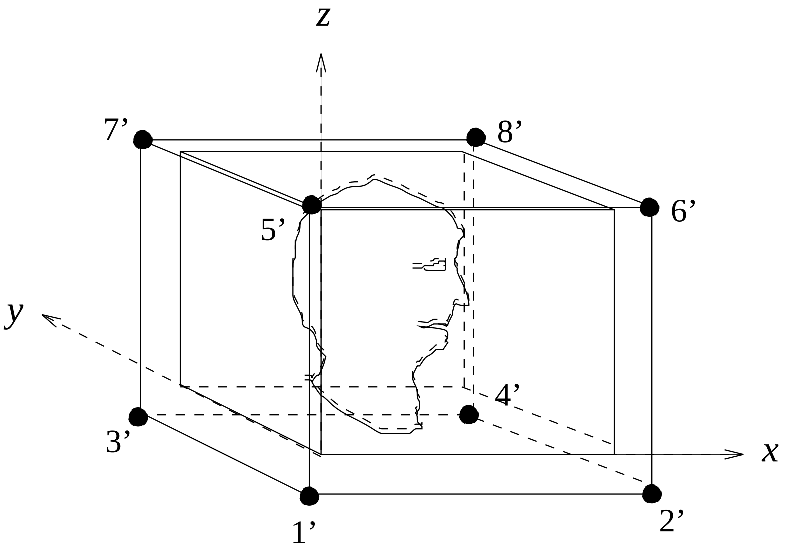

Figure 7. Image volumes after registration, so that the head in each volume is almost perfectly aligned with the other. The points $1'$ to $8'$ are used to specify the transformation, and form the right hand side of the transformation table.

Transformation Parameters						
Investigator(s): A. B. Cee and D. E. Eff						
Site: Extra University, Somewhere, New Country						
Method: 1						
Date: 23 August 1995						
Patient number: 001						
From: CT						
To: MR-PD						
Point	x	y	z	new_x	new_y	new_z
1	0.0000	0.0000	0.0000	-5.7884	-29.5052	-23.9565
2	333.9870	0.0000	0.0000	326.4905	-62.6706	-30.1350
3	0.0000	333.9870	0.0000	27.4767	302.7901	-19.3302
4	333.9870	333.9870	0.0000	359.7436	269.6247	-25.5087
5	0.0000	0.0000	112.0000	-3.8810	-31.2550	88.0136
6	333.9870	0.0000	112.0000	328.3798	-64.4204	81.8351
7	0.0000	333.9870	112.0000	29.3721	301.0404	92.6399
8	333.9870	333.9870	112.0000	361.6509	267.8750	86.4614
(All distances are in millimeters.)						

Table 2. Example transformation table.

2.7. The retrospective techniques

The retrospective registrations were performed in parallel at several sites outside Vanderbilt. Some methods were used that were applicable only to CT-to-MR or PET-to-MR registration, and some were suitable for both cases.

1. Barillot and Lemoine[9] used a two-stage technique both for CT-to-MR and PET-to-MR registration. The first step is to perform an approximate registration of objects which are similar in each image, e.g. the head or the surface of the brain, with the choice of object being made *a priori* by examination of the images for the most suitable candidate. After automatically segmenting the object in the image volumes (the algorithm used depending on the feature to be segmented), a principal component matching method is applied, bringing the objects into an approximate alignment. The second stage is the application of a multiresolution Powell[18] algorithm which minimizes the Euclidean distance between the two surfaces given by a $5 \times 5 \times 5$ chamfer mask. The total time taken for this procedure was approximately fifteen minutes using a Sun SPARCstation 20.

2. Collignon, Maes, Suetens, and Vandermeulen[4] used a fully automatic technique for CT-to-MR and PET-to-MR registration. The technique employs a Powell algorithm to search the space of registration parameters, judging the images to be in perfect registration when the mutual information of the intensities of corresponding voxels is maximized. The registration took ten to thirty minutes in the CT-to-MR case, and two to four minutes in the PET-to-MR case, using an IBM RS6000/3AT.

3. Van den Elsen, Napel, and Sumanaweera[20] used a fully automated multiresolution gray-value correlation technique for registering CT volumes to MR. The CT images are preprocessed using a "triangular" intensity mapping that reduces CT bone intensity and results in images that more closely resemble MR. The preprocessed CT and original MR volumes are then registered using gray-value correlation. In order to decrease the computational load, a multiresolution pyramid is created for each image, in which each level contains a lower resolution version of the image in the level below. In each level of the hierarchy, the matching transformation is determined by an exhaustive search of a small number of regions in parameter space. These regions are largest at the level of lowest resolution, and are reduced at higher resolutions. All promising extrema at each level are used as starting points around which the parameter space in the next level was scanned. The total time taken for this procedure was approximately one to two hours using an SGI Indigo 2.

4. Harkness applied a surface-matching technique developed by Pelizzari et al.[17] to CT-to-MR and PET-to-MR registration. There are two steps, the first of which is the manually aided segmentation of the required surfaces in the two image volumes: the scalp surface for CT to MR and the brain surface for PET to MR. The segmentation is achieved by drawing lines perpendicular to the boundary of the surface and applying a thresholding method along these lines to define the surface contours. Having defined the surfaces, a Powell algorithm is used to minimize the residual distance between the points in one modality and the surface in the other. The typical time required for each registration was approximately twenty minutes, most of which was human interaction time.

5. Hemler, Napel, and Sumanaweera[6] employed a surface-based method for registering from CT to MR. In this system the corresponding surface to be registered is first identified in each image set. For this study, the skin surface was used. It is identified using the full-width at half-maximum threshold selection technique. A recursive two-dimensional boundary-following technique then identifies and connects boundary pixels of each thresholded region. The boundary of the skin surface is then identified using size and shape information. To reduce redundancy, the two-dimensional skin surface boundary is approximated with a polygon. A three-dimensional surface is constructed between adjacent boundaries of the higher in-plane resolution image set, using a triangulation scheme. This process results in a triangular mesh representing the surface in one image set. The two-dimensional polygon points are used to represent the surface in the other image set. A least-squares minimization technique is then used to determine the rigid body transformation that minimizes the total sum squared perpendicular distance between each surface point in one image set and the triangular mesh in the other. (An estimate of registration time was not provided by these investigators.)

6. Hill and Studholme[19] registered CT to MR and PET to MR using a fully automated voxel similarity measure algorithm. The mutual information coefficient of the joint probability distribution of the intersection of the two image volumes was optimized using a multiresolution technique. This approach builds on the independent work of Wells[22] and Collignon.[4] Using a 0.125 mm voxel size for the final iteration, the time taken for registration was approximately twenty minutes on a Sun SPARCstation 20/61.

7. Malandain and Pennec[11–13] registered CT to MR and PET to MR. The method was originally designed for PET-to-MR registration and is used for CT-to-MR registration without modification. The first step is the segmentation of the brain in both images by means of morphological and topological operators. This is done automatically, but with human inspection and possible changes in the parameters if a satisfactory result is not achieved. The second step is the construction of a potential energy field of one volume, and the computation of the global force and torque acting on the other volume due to this field. The image pair is brought into registration by modeling the movement of the second volume under the forces acting on it, using a series of decreasing time steps. The procedure typically took approximately three minutes for CT-to-MR registration and forty seconds for PET-to-MR registration using a DEC Alpha 3000.

8. Noz, Maguire, and Pollack[10] used a two-step method for registration, both for CT to MR and PET to MR. A set of corresponding landmarks is chosen in each image volume, and an oblique projection performed along axes determined by visual inspection in order to bring the image slice planes into coincidence. Each landmark in the "To" modality is then superimposed on the neighborhood of its counterpart in the "From" modality, and moved in small increments until the cross correlation of the two areas is maximized. This provides a set of coordinate pairs that represent corresponding points in the two modalities. The mapping between the two spaces is approximated by a first- or second-order polynomial transformation whose coefficients minimize the root-mean-square distance between the point sets. The coefficients are derived using a linear regression technique followed by a Gauss-Jordan matrix inversion. The time typically required for registration was between fifteen and thirty minutes using a Sun SPARCstation 10/51, most of which was spent finding corresponding landmarks.

9. Pelizzari[17] used a two-step method for CT-to-MR and PET-to-MR registration. The first stage was the segmentation of contours using a combination of manual and threshold-based boundary following. The contours in the "From" modality were represented as a set of points, and in the "To" modality as a stack of disks. The second stage was the minimization, by means of a Powell algorithm, of the mean square distance of the points from the surface, measured by finding the intersection of the surface model with a ray drawn from each point

to the centroid of the surface model. The typical time taken for registration of one pair of volumes was twenty minutes, almost all of this spent in the segmentation step; the CPU time required for registration was approximately five seconds on an SGI Indigo 2.

10. Robb and Hanson[7,8] used a surface matching algorithm based on parametric chamfer matching for CT-to-MR and PET-to-MR registration. Four variations were performed using two different surfaces and two numbers of points per surface (RO1/RO2 were performed by matching skin to skin, RO3/RO4 brain—or inner surface of skull in CT—to brain; RO1/RO3 were performed using 100 points, RO2/RO4 1,000 points). Surface segmentation was accomplished using an automatic three-dimensional mathematical morphological algorithm conditioned by region-of-interest bounding on a single image slice, with manual intervention needed in a few cases (e.g. for removal of tissue around the eyes in the MR images). The segmented surfaces were represented as a stack of contours, and a shape-based interpolation used to give spatially isotropic voxels in the "base" image (i.e. the target for registration). A chamfer algorithm is used to approximate the Euclidean distance of points on one contour set from the other, and a multi-resolution, multi-threshold iterative approach taken to find the registration parameters which mimimize the root mean square value of the distance transformation. The total time for a registration on an SGI Challenge workstation ranged from three to twenty minutes, with morphologic-based segmentation taking two to three minutes, manual editing in selected cases ten to fifteen minutes, and actual registration thirty seconds (100 points) to two minutes (1,000 points).

11. Woods[23] registered PET to MR using a multi-sampling density method based on the correlation of intensity values between PET and MR. Two methods were used. In the first method (WO1), non-brain regions (i.e. skin, skull, meninges) are manually removed. After this manual editing, automated registration is performed by dividing the MR image into 256 partitions based on intensity. A Newton-Raphson method is used to find the transformation parameters that minimize a weighted average of the standard deviations of PET voxel intensities corresponding to each MR partition. Weighting factors for a given partition are proportional to the number of MR voxels in that partition. The second method (WO2) is identical to the first except that the MR images are edited more extensively to remove areas of focal brain pathology. The total time for a registration was typically between twenty and thirty minutes using a Sun SPARCstation IPX, with manual editing taking fifteen to twenty minutes and cost function minimization less than five minutes.

2.8. Data analysis

At Vanderbilt, after the transformation tables have been received from each site and the corresponding rigid-body transformations have been determined, the next step in the evaluation is to perform a comparison between these registrations and the fiducial-based ones. In collaboration with a neurological and a neurosurgical expert, a set of *volumes of interest* (VOIs) were chosen that represent areas of neurological and/or surgical interest. Each VOI was manually segmented within one of the MR image volumes; this procedure was repeated for each patient data set used. The VOIs were stored as sets of x, y and z voxel coordinates.

An estimate of the accuracy of the retrospective registration at the position of each VOI is computed (see Figure 1) using the following technique:

1. The centroid pixel of the VOI is found, and its position converted to a millimetric point C in the "To" modality, using the known voxel size for the image volume.

2. The inverse of the gold standard transformation G^{-1} is applied to C, thus giving the corresponding position $C' = G^{-1}(C)$ in the "From" modality.

3. The retrospective transformation, R, to be evaluated, is applied to C', with result $C'' = R(C')$.

4. The error of the retrospective registration at the anatomical location determined by this VOI is taken to be the Euclidean distance between C and C''.

The anatomical positions (see Figure 8) corresponding to the VOIs used in this evaluation are: 1. Maximum aperture of fourth ventricle, 2. Junction of fourth ventricle with aqueduct, 3. Right globe, 4. Left globe (not shown), 5. Optic chiasm, 6. Apex of left Sylvian fissure, 7. Apex of right Sylvian fissure, 8. Junction of central sulcus with midline, 9. Left occipital horn, 10. Right occipital horn.

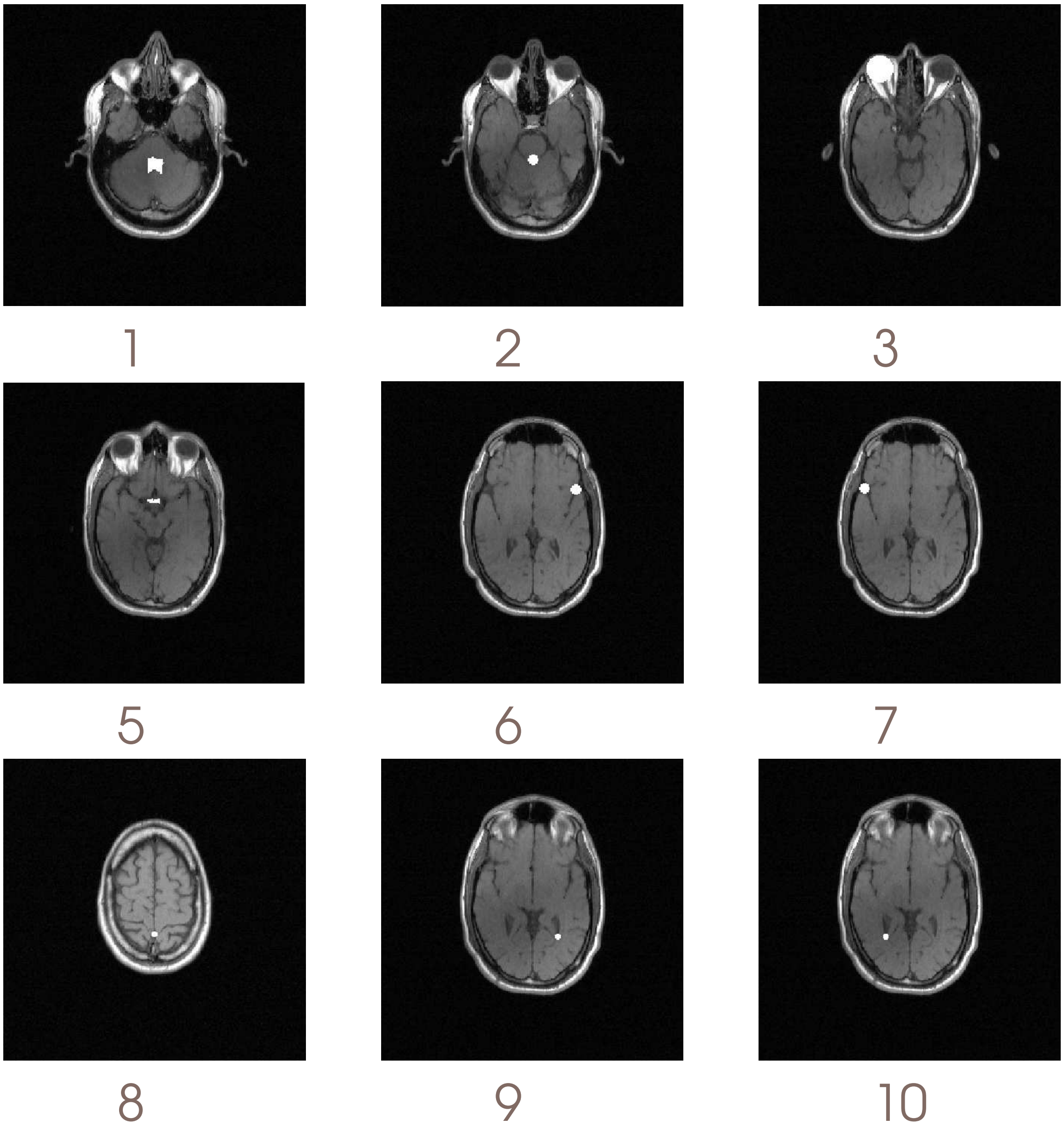

Figure 8. VOIs used for accuracy evaluation.

3. RESULTS

For a given technique, the TREs in all of the VOIs are very similar. The similarities are such that we are reporting statistics only for the pooled VOIs. A complete listing of the results will be made available via FTP (see the project web homepage `http://cswww.vuse.vanderbilt.edu/~image/registration/` for details).

Tables 3 through 6 present the median and maximum TREs for the CT-to-MR and PET-to-MR registrations. The median and maximum TRE listed for each technique and registration modality pair are the median and maximum of all individual TREs for all applicable patients and VOIs for that technique and modality pair. Each row represents a registration modality pair, and each column a registration technique. The sites are represented by abbreviation of the participants' names as follows: 1. Barillot et al. (BA); 2. Collignon et al. (CO); 3. van den Elsen et al. (EL); 4. Harkness (HA); 5. Hemler et al. (HE); 6. Hill et al. (HI); 7. Malandain et al. (MA); 8. Noz et al. (NO); 9. Pelizzari (PE); 10–13. Robb et al. methods 1–4 (RO1, RO2, RO3, RO4); 14–15. Woods methods 1 and 2 (WO1, WO2).

Modality	Technique													N
	BA	CO	EL	HA	HE	HI‡	MA	NO†	PE	RO1	RO2	RO3	RO4	
CT-T1	1.6	1.5	1.5	3.0	1.4	1.1	4.1	3.2	2.5	4.1	5.1	5.9	5.7	7
CT-PD	1.8	1.4	1.9	3.1	2.3	2.0	3.8	7.8	1.9	4.4	5.2	4.9	4.8	7
CT-T2	2.4	1.4	1.5	4.2	4.5	1.5	4.8	3.8	2.5	4.5	4.4	5.4	4.8	7
CT-T1 rect.	1.4*	0.7	0.9	3.3	1.0	0.8	5.3	3.2	2.2	5.9	5.9	6.3	5.9	6
CT-PD rect.	1.7*	0.8	1.1	2.9	1.7	0.8	4.0	4.5	1.9	5.3	5.5	5.5*	5.5*	7
CT-T2 rect.	2.0*	0.8	1.5	3.5	1.6	0.7	5.3	4.2	2.9	5.4	5.3	5.5	5.3	7

Table 3. Median errors for CT-to-MR registration. The label "rect." indicates that the MR image was corrected for geometrical distortion before registration. See text for technique abbreviations. *One patient omitted. †Non-rigid transformations. ‡Results resubmitted after gold standard released. All errors are in units of mm.

Modality	Technique											N
	BA	CO	HA	HI‡	MA	NO†	PE	RO3	RO4	WO1	WO2	
PET-T1	4.5	3.6	2.8	3.2	4.1	3.6	2.9	4.0	3.4	2.4	3.1	7
PET-PD	5.1	3.3	4.2	3.2	3.7	4.1	3.0	4.3	3.4	3.0	3.2	7
PET-T2	4.7	2.8	2.7	2.5	4.6	4.6	3.2	3.9	3.7	3.4	3.2	7
PET-T1 rect.	3.5	2.8*	3.6	3.0	2.5	4.0	2.9	4.0	3.6	2.0	2.0	4
PET-PD rect.	4.5	3.0	3.1	3.0	3.5	4.4	2.6	3.1*	3.2*	2.6	2.4	5
PET-T2 rect.	3.8	2.1	3.3	2.2	3.6	5.1	2.7	3.8	3.4	2.5	2.3	5

Table 4. Median errors for PET-to-MR registration. See notes in Table 3.

Modality	Technique													N
	BA	CO	EL	HA	HE	HI‡	MA	NO†	PE	RO1	RO2	RO3	RO4	
CT-T1	7.1	6.6	6.0	51.8	11.0	2.8	62.9	10.4	7.2	25.9	21.8	17.8	18.9	7
CT-PD	7.6	3.6	6.5	49.6	10.6	4.1	60.3	13.9	4.0	25.9	22.2	23.8	20.2	7
CT-T2	8.7	3.4	4.1	50.7	13.8	4.2	60.8	9.7	6.7	26.6	22.0	19.4	19.8	7
CT-T1 rect.	6.7*	3.8	2.6	48.2	2.0	2.3	62.3	9.6	5.8	27.4	22.3	18.1	18.2	6
CT-PD rect.	6.8*	2.5	5.3	45.9	3.7	2.3	64.2	11.5	4.5	27.4	22.1	24.9*	20.2*	7
CT-T2 rect.	7.7*	4.2	5.2	49.1	15.9	3.0	64.4	10.2	8.9	27.0	22.5	19.8	21.5	7

Table 5. Maximum errors for CT-to-MR registration. See notes in Table 3.

Modality	Technique											N
	BA	CO	HA	HI‡	MA	NO†	PE	RO3	RO4	WO1	WO2	
PET-T1	11.5	12.7	11.6	9.0	8.9	11.2	9.8	9.6	5.9	5.5	5.6	7
PET-PD	11.4	9.5	45.1	8.2	9.5	8.7	11.3	8.7	7.1	7.0	6.4	7
PET-T2	12.4	7.1	17.2	8.1	12.4	7.2	13.4	9.1	7.3	8.2	7.1	7
PET-T1 rect.	6.2	3.6*	17.7	5.9	8.7	14.2	8.2	7.3	8.9	4.2	5.0	4
PET-PD rect.	11.2	7.0	10.0	7.4	9.3	7.4	10.8	6.2*	5.7*	5.5	5.5	5
PET-T2 rect.	9.9	6.9	10.3	8.9	12.4	11.1	15.1	5.7	7.1	6.1	6.2	5

Table 6. Maximum errors for PET-to-MR registration. See notes in Table 3.

We investigated the effects of geometrical distortion correction in MR on registration accuracy by comparing the CT-to-MR and PET-to-MR registration errors with and without correcting for distortion. First, we pooled the results of all registration techniques and used the median TRE values listed in Tables 3 and 4. In all cases the differences are not significant (two-tailed paired t-test, $P < 0.05$). Second, we examined each technique individually and used as data for comparison the median TRE values obtained for each patient with that technique. These differences are significant (two-tailed paired t-test, $P < 0.05$) in only one case: CT-to-T2 registration by Collignon. The differences are "marginally" significant ($P < 0.10$) in four other cases: CT-to-T2 registration by Harkness, and CT-to-T1 by Collignon, by van den Elsen, and by Hill.

4. DISCUSSION

The principal goal of this project was to determine the accuracy of retrospective image registration techniques. It should be noted that because this study assesses only image-to-image registration, and not image-to-physical-space registration, its direct clinical application lies in intermodality image correlation. Clinical applications might include, for example, the assignment of anatomic specificity to functional activation studies with fMRI and PET or the longitudinal cross-correlation over time of imaging studies to follow tumor growth and response to therapy. In using our results to guide such applications, it is important to consider the validity of our approach and the accuracy of our gold standard.

4.1. Submission errors

This study was carried out in a blinded fashion, in the sense that the investigators at sites outside Vanderbilt did not know the correct transformations. Because this study involved multiple sites, each with its own method for storing images and for specifying geometrical transformations, and because all communication was carried out by electronic transmission, errors were inevitable. Our protocol included a test of communications with a practice data set and also included redundant information (eight corner points) in the submission of transformation. Thanks to these precautions and to the checking of images and headers by the investigators outside Vanderbilt and the checking of submitted transformations at Vanderbilt, many errors were discovered and corrected before the standard transformations were at last revealed.

After all transformation tables had been submitted, the gold standard tables and the original image volumes with visible fiducials were made available so that the correctness of the evaluation could be verified by all participants. In addition each site was provided with error statistics for its technique(s). At this point the blinded study was over. After this point, despite our precautions, four errors were discovered:

- Hill reported that a systematic error was made while converting his transformation format to the format of the Vanderbilt transformation table and submitted revised transformation tables. Specifically, Hill claimed that the angle of rotation about the x axis was accidentally negated. Vanderbilt independently verified this error by taking the original and revised transformation tables, computing the rotation matrices, converting the rotation matrices to rotations about orthogonal axes, and confirming that the original and revised transformations differed only by the sign of the angle of the rotation about the x axis. After verification, the error was corrected (noted by the double dagger in Tables 3 through 6). Because the correction is independently verifiable, it is felt that the blindedness of the study is not compromised by the inclusion of the corrected data.

- Barillot reported that for one patient the wrong set of tables was submitted for the rectified CT-to-MR registrations. In this case resubmission would not be appropriate. Instead, the three erroneous transformations were dropped from the study (noted by the asterisks in Tables 3 and 5).

- Collignon reported that a manual entry error was made while creating one transformation table for one patient. Again resubmission was not appropriate, and the single erroneous transformation was dropped from the study (noted by the asterisks in Tables 4 and 6).

- Robb reported that for one patient the wrong set of tables was submitted for the rectified CT-to-PD skull-to-brain and PET-to-PD brain-to-brain registrations. In this case resubmission would not be appropriate. Instead, the four erroneous transformations were dropped from the study (noted by the asterisks in Tables 3 through 6).

We include in the Appendix all results that were obtained before correcting (Hill) or dropping (Barillot, Collignon, and Robb) erroneous transformations.

4.2. Gold standard accuracy

The validity of these evaluations depends on the accuracy of the gold standard registrations. The TREs presented in this study are exact measures of error only if the gold standard provides perfect registration. This standard is of course not perfect. We cannot measure the accuracy of the gold standard directly, but we can estimate it using numerical simulations.[14,16] Briefly, the simulations are performed as follows:

> The head is modeled as a hemisphere of radius $R = 10$ cm. In a given simulation we specify the number of fiducials, and the *fiducial localization error* (FLE) in each of two image spaces, where FLE is the root-mean-square error in determining the positions of the fiducials. We repeatedly place the fiducials on the surface of the hemisphere and the targets randomly within the hemisphere, perturb the fiducials randomly and independently in each space to simulate FLE, register the perturbed fiducials, and compute FRE and TRE as the distances between the corresponding fiducials and targets, respectively. We performed each simulation 100,000 times for a single target. We repeated the simulations with different values of FLE in each space ranging from 0.2 mm to 2.0 mm.

The simulations show that the distribution of the ratio, TRE/FRE, is independent of the values chosen for FLE. For four markers TRE/FRE is 0.99 ± 0.61 (mean $\pm$ SD). Thus we can approximate the accuracy of the gold standard by multiplying the observed FRE value by 0.99. The FRE for marker-based CT-to-MR registration using the three types of MR images corrected for geometrical distortion is 0.40 ± 0.10 mm. The FRE for PET-to-MR registration is 1.67 ± 0.52 mm. Thus the accuracy of the gold standard is approximately 0.4 mm for CT-to-MR and 1.7 mm for PET-to-MR. The larger TRE for registrations involving PET is to be expected because of the larger voxels in this modality. The simulations do not account for geometrical distortion. Thus these simulation results apply only to the registrations obtained using MR images corrected for geometrical distortion, assuming that any remaining distortions are uncorrelated. The gold standard TRE can be expected to be somewhat larger for the uncorrected MR images.

Regardless of the true TRE of a given retrospective registration technique, the TRE observed in this study can be expected to be higher. The observed TREs can also be expected to be higher than the TRE of the gold standard. If the transformation errors of the fiducial marker and a retrospective technique are independent, then the TRE that we observe can be expected to be on the order of

$$\sqrt{(\mathrm{TRE_t})^2 + (\mathrm{TRE_g})^2},$$

where $\mathrm{TRE_t}$ is the true TRE of the retrospective technique, and $\mathrm{TRE_g}$ is the TRE of the gold standard. If a retrospective technique is approximately as accurate as the gold standard, then its observed TRE will be on the order of $\mathrm{TRE_o} \approx \sqrt{2}\ \mathrm{TRE_g}$, which is approximately 0.6 and 2.4 mm for the CT-to-MR and PET-to-MR cases, respectively.

4.3. Observations

As mentioned at the beginning of the Results Section, the variation among VOIs is so small that we decided to report statistics on only the pooled VOIs. It can be seen also that, while there is considerable variation among the techniques with regard to the medians and maximums calculated over these VOIs, there is within a given technique little variation in these statistics among the three MR types—T1, T2, and PD. Thus, it would appear that the retrospective techniques are insensitive to variation among these types. As discussed at the end of the Results section there is a significant variation with regard to MR rectification for only one technique, indicating that in most registrations the level of improvement, if any, is too small to be seen. It is clear on the other hand that the accuracy of registrations involving CT is better than that involving PET.

We note in particular that Noz et al. performed registration using non-rigid transformations. Our evaluation method approximated their transformations with rigid-body transformations. Thus the results reported in this study may not be a fair evaluation of the accuracy of their technique.

It appears from the medians that retrospective image registration has the potential of providing satisfactory results much of the time. The maxima are much larger, however, and indicate that visual inspection should play an important role in its clinical application.

We note finally that, though some registration techniques produced smaller errors than others in this study, it is not possible to draw statistically meaningful conclusions regarding the superiority of a technique for a given task (or subtask). That was not a goal of this project and, because of the large number of techniques (15) and the relatively small number of independent observations (maximum of 7 patients) used in their evaluation, it is not statistically feasible.

5. CONCLUSION

We have shown that it is possible by means of Internet communication to conduct from a central site a meaningful, blinded evaluation of retrospective image registration techniques that are applied at remote sites. Eleven groups of investigators applied fifteen techniques to selected registration tasks involving the registration of CT and/or PET to MR. Our results indicate that retrospective techniques have the potential to produce satisfactory results much of the time but that visual inspection is necessary to guard against large errors.

6. APPENDIX

We list here all results that were obtained before correcting (Hill) or dropping (Barillot, Collignon, and Robb) erroneous transformations (see Section 4.1).

- Hill. Table 3 (column HI, row 1 to row 6): 1.6, 2.0, 1.7, 1.2, 1.2, 1.7; Table 4: 2.9, 3.5, 2.2, 3.2, 3.4, 2.4; Table 5: 3.4, 4.5, 4.7, 3.0, 3.4, 4.2; Table 6: 8.8, 9.8, 8.3, 9.5, 7.4, 8.6.
- Barillot. Table 3 (column BA, row 4 to row 6): 1.8, 2.0, 2.4; Table 5: 22.0, 39.4, 25.8.
- Collignon. Table 4 (column CO, row 4): 3.0; Table 6: 107.0.
- Robb. Table 3 (column RO3 to column RO4, row 5): 5.7, 5.8; Table 4: 4.5, 4.3; Table 5: 173.0, 172.9; Table 6: 176.2, 61.6.

ACKNOWLEDGMENT

Funding for the this project was provided by the National Institutes of Health, Project Number 1 R01 NS33926-01.

The Vanderbilt collaborators on this project would like to thank Yu Shyr, Ph.D. for the time and effort he spent advising us about the statistical methods used in this study; J. B. Antoine Maintz, Utrecht University, for writing the Perl script provided for FTPing images; John Votaw, Ph.D., Emory University, for answering many questions about PET imaging; and Srikanth Gadamsetty for helping collect and process patient images. They would also like to thank Nicholas Ayache, Ph.D., INRIA; Frederic Fahey, D.Sc., Bowman Gray School of Medicine; and David Hawkes, Ph.D., Guy's Hospital for their support and encouragement. The investigators from Mayo Clinic would like to thank Vanessa Murrie for her help performing registrations. The investigators from New York University Medical Center would like to thank Elissa Kramer, M.D. for her help and advice.

REFERENCES

1. K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:698–700, 1987.
2. B. S. Baxter, L. E. Hitchner, and G. Q. Maguire, Jr. A standard format for digital image exchange. *American Association of Physicists in Medicine*, 1982.
3. H. Chang and J. M. Fitzpatrick. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Trans. Med. Imaging*, 11:319–329, 1992.
4. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging 1995*, pages 263–274. Kluwer Academic, Dordrecht, Netherlands, 1995.

5. T. D. Cradduck, D. L. Bailey, B. F. Hutton, F. de Conninck, E. Busemann-Sokole, H. Bergmann, and U. Noelpp. A standard protocol for the exchange of nuclear medicine image files. *Nucl. Med. Comm.*, 10:703–713, 1989.
6. P. F. Hemler, T. S. Sumanaweera, P. A. van den Elsen, S. Napel, and J. R. Adler. A versatile system for multimodality image fusion. *J. Image Guid. Surg.*, 1:?–?, 1995.
7. H. Jiang, K. S. Holton, and R. A. Robb. Image registration of multimodality 3-D medical images by chamfer matching. *Biomedical Image Processing and Three-Dimensional Microscopy 1992*, Proc. SPIE 1660:356–366, 1992.
8. H. Jiang, R. A. Robb, and K. S. Holton. A new approach to 3-D registration of multimodality medical images by surface matching. *Visualization in Biomedical Computing 1992*, Proc. SPIE 1808:196–213, 1992.
9. D. Lemoine, D. Liegeard, E. Lussot, and C. Barillot. Multimodal registration system for the fusion of MRI, CT, MEG, and 3D or stereotactic angiographic data. *Medical Imaging 1994: Image Capture, Formatting, and Display*, Proc. SPIE 2164:46–56, 1994.
10. G. Q. Maguire, Jr., M. E. Noz, H. Rusinek, J. Jaeger, E. L. Kramer, J. J. Sanger, and G. Smith. Graphics applied to medical image registration. *IEEE Comput. Graphics Appl.*, 11:20–29, March 1991.
11. G. Malandain, S. Fernández-Vidal, and J. M. Rocchisani. Improving registration of 3-D medical images using a mechanical based method. In *3rd European Conference on Computer Vision (ECCV '94)*, pages 131–136, 1994.
12. G. Malandain, S. Fernández-Vidal, and J. M. Rocchisani. Rigid registration of 3-D objects by motion analysis. In *Proc. 12th Int. Conf. Pattern Recognition*, pages 579–581, 1994.
13. G. Malandain, S. Fernández-Vidal, and J. M. Rocchisani. Physically based rigid registration of 3-D free-form objects: Application to medical imaging. Technical Report 2453, INRIA, Sophia Antipolis Cedex, France, January 1995.
14. V. R. Mandava, J. M. Fitzpatrick, C. R. Maurer, Jr., R. J. Maciunas, and G. S. Allen. Registration of multimodal volume head images via attached markers. *Medical Imaging VI: Image Processing*, Proc. SPIE 1652:271–282, 1992.
15. C. R. Maurer, Jr., G. B. Aboutanos, B. M. Dawant, S. Gadamsetty, R. A. Margolin, R. J. Maciunas, and J. M. Fitzpatrick. Effect of geometrical distortion correction in MR on image registration accuracy. *J. Comput. Assist. Tomogr.*, (In press).
16. C. R. Maurer, Jr., J. M. Fitzpatrick, R. L. Galloway, Jr., M. Y. Wang, R. J. Maciunas, and G. S. Allen. The accuracy of image-guided neurosurgery using implantable fiducial markers. In H. U. Lemke, K. Inamura, C. C. Jaffe, and M. W. Vannier, editors, *Computer Assisted Radiology 1995*, pages 1197–1202. Springer-Verlag, Berlin, 1995.
17. C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C.-T. Chen. Accurate three-dimensional registration of CT, PET, and/or MR images of the brain. *J. Comput. Assist. Tomogr.*, 13:20–26, 1989.
18. M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.*, 7:155–163, 1964.
19. C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated registration of truncated MR and CT datasets of the head. pages 27–36, 1995.
20. P. A. van den Elsen, E.-J. D. Pol, T. S. Sumanaweera, P. F. Hemler, S. Napel, and J. R. Adler. Grey value correlation techniques used for automatic matching of CT and MR brain and spine images. *Visualization in Biomedical Computing 1994*, Proc. SPIE 2359:227–237, 1994.
21. M. Y. Wang, C. R. Maurer, Jr., J. M. Fitzpatrick, and R. J. Maciunas. An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head. *IEEE Trans. Biomed. Eng.*, (In press).
22. W. M. Wells, III, P. Viola, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. pages 55–62. Wiley-Liss, New York, 1995.
23. R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET registration with automated algorithm. *J. Comput. Assist. Tomogr.*, 17:536–546, 1993.