A blinded evaluation and comparison of image registration methods

J. Michael Fitzpatrick*
Jay B. West*

Reprinted from "Empirical Evaluation Techniques in Computer Vision",
ed. Kevin W. Bowyer and P. Jonathon Phillips, IEEE Computer Society Press,
Los Alamitos, CA, 1998.

---

*Department of Computer Science, Vanderbilt University, Nashville, TN.

1: Introduction

During the period from March, 1995 through February, 1998 a project supported by the National Institutes of Health [1] was conducted at Vanderbilt University. The goal of the project was to perform an evaluation of a set of methods for registering medical images of the human head. The project was the first of its kind in medical imaging and has not of this writing been duplicated. The challenge was to provide a means for carrying out a blinded study of image processing algorithms on a common data set when the algorithms were being applied in laboratories that were scattered over the world. The term "blinded" here means that the people who were applying the registration methods were ignorant of the correct answers. By keeping the study blind and using a common data set we feel that we heightened the credibility of the evaluation considerably.

To carry out a study of this nature we had to solve many problems, some of which were specific to the particular problem of registration (*e.g.*, how to describe a registering transformation), some of which were specific to image processing (*e.g.*, how to describe an image format), and some of which would apply to any study involving the blinded application of algorithms at remote sites that were to be evaluated at some central site (*e.g.*, how to make sure that communication errors did not confuse the evaluation results). One problem that is universal to any such study is this: How are the correct answers determined? This problem can be a difficult one in general, but in our case it had been solved before this project was undertaken by the development of a highly accurate registration system at Vanderbilt. That system gets its accuracy by taking advantage of special preparation of the patients whose images were being used for the study.

This project encountered many problems, but they were all solvable, given enough attention and given sufficient cooperation among the many groups involved. After the first twelve groups' registrations had been evaluated we declared that the first study was completed. The results of the study were published immediately after its completion [6]; instead of concentrating here on the results, the emphasis will instead be on the procedures and potential pitfalls inherent in performing such a comparison study, in the belief that this is equally applicable to similar evaluation studies in any field.

Maintaining blindedness was a difficult challenge but we feel that the increased credibility is worth the effort.

1.1: The Registration Problem

The problem of image registration is as follows: given two different representations of the same object (*e.g.* Magnetic Resonance (MR) and Computerized Tomography (CT) images of a patient's head), to find a transformation which, when applied to one image, will align (or *register*) points in that image with corresponding points in the other. In this study, we considered three-dimensional MR, CT and Positron Emission Tomography (PET) images of patients who were due to undergo neurological surgery. Each image volume consisted of a stack of two-dimensional "slices". There were two tasks evaluated: the registration of CT images to the corresponding MR images, and the registration of PET to MR. The MR images could be subdivided into submodalities – Proton Density (PD) weighted, T1

---

[1] The images and the standard transformations were provided as part of the project "Evaluation of Retrospective Image Registration", National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.

weighted and T2 weighted, both with and without a geometrical distortion correction algorithm [3, 5] having been applied. As the human head behaves approximately as a rigid object, we evaluated only rigid-body registration methods, *i.e.* those that produced a transformation consisting solely of a rotation and translation. As we shall see later, this greatly simplified the evaluation step of this study.

1.2: The gold standard

In order to compare a set of registration techniques, it is necessary to have a single technique that is generally agreed to have superior accuracy to those in the evaluation set, and to use the results generated by this technique as the "ground truth" for comparison purposes. The Acustar^{TM} 1 Surgical Navigation System was used as the gold standard in this study. This system was developed at Vanderbilt in collaboration with Johnson & Johnson Professional, Inc. and relies on markers that are rigidly attached to the skull in order to derive the registration transform. The markers (called *fiducial* markers) are designed to show up brightly in the images, and may be replaced by devices called *localization caps* in the operating room in order to allow the surgeon to navigate interactively during surgery. Systems such as this one are labeled *prospective*: they require preparatory steps to be taken before imaging in order to perform the registration. The preparatory step in this case is the drilling of holes into the outer table of the patient's skull. This step is highly invasive and is considered a disadvantage of using this registration method. However, it is generally agreed that prospective techniques such as this one give more accurate results than *retrospective* techniques (*i.e.* those that do not require any special preparation to be performed before the images are acquired). Given that all the methods to be evaluated were retrospective in nature, we felt it appropriate to use as the gold standard the results generated by the Acustar system.

1.3: The techniques to be evaluated

Researchers at twelve separate sites performed a total of sixteen distinct retrospective registration techniques on the image data we provided. It is possible (with a few exceptions) to categorize these methods into two groups: volume-based and surface-based. The volume-based methods make use of a correlation between the intensity values of pixels in either image that correspond to the same anatomical location; the label stems from the fact that all the information in each image volume is used to compute the registration transform. The surface-based methods consist of a segmentation step that derives a representation of surfaces in each image that are either corresponding or closely fitting (*e.g.* the skull in CT and the brain surface in MR) and a registration step that finds the transformation minimizing some distance measure between the two surfaces.

2: Data preparation

The gold standard transformations were based on the positions of the fiducial markers in the images: the correct transformation is defined as the one that minimizes the root-mean-square (RMS) distance between corresponding fiducial points after registration. We used the closed-form solution developed by Arun et al. [1] to find this optimal mapping between the two images. All the registrations to be evaluated were retrospective in nature, and therefore were supposed to be able to register the images without recourse to the fiducial markers. It was therefore necessary for us to remove all trace of these markers from the images, while attempting to reconstruct the removed regions to make them appear as close as possible to the natural background pattern. Also present in the image was a stereotactic frame. This is a cage-like structure rigidly attached to the patient's head that was used for intrasurgical navigation before the advent of the marker-based system. The frame was present in addition to the markers in the images used in the study as these images were acquired during a testing phase of the fiducial markers, where the frame was used as a backup measure to the markers. Clearly the frame also had to be removed from the images in order to provide a fair test for the techniques being evaluated.

In order to prepare the images for use in this study, we developed a technique that we dubbed *air brushing*. This was performed by manually outlining the regions containing the structures to be removed, followed by approximately reconstructing 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 that 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 (Fig. 1):

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 Figure 8. The top row (Fig. 8*a–c*) shows slices of original volumes from each of the three modalities. The window and level have been set so that the background artifacts may be seen. The middle row (Fig. 8*d–f*) shows the same slices after the region R has been outlined and zeroed. This procedure is applied to each slice in the volume. The last row (Fig. 8*g–i*) 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.

3: Communication

The communication phase of this study may be divided into two phases: the transfer of the image data to the remote sites (those performing the registrations) and the transmission of the registration transformations back to the central site (Vanderbilt) where the evaluations were to be performed.

3.1: Transferring the images

When choosing the method by which the image volumes would be distributed to the remote sites there were several options available, including writing the data on to CD-ROMs and sending them by physical mail, and the use of FTP (File Transfer Protocol) to allow electronic transmission of data from the central site to the remote sites. We selected FTP for several reasons:

- It did not require extra hardware at the central site (*e.g.*, a writeable CD-ROM drive) or at the remote sites (*e.g.*, a CD-ROM drive connected to the computer being used to perform the registration algorithm).
- It allowed incremental changes in the image dataset to be made, *e.g.*, if the file containing one of the image volumes was found to be corrupted, the problem could be remedied with little additional work.
- The work contained in transmitting the information was largely distributed amongst the remote sites, rather than being concentrated at the central site.

In addition to the binary data corresponding the images themselves, it was also necessary that the remote sites have access to other pertinent data, *e.g.*, the resolution and physical dimensions of each voxel, and orientation of each image. This data was provided for each image by means of an ASCII header file that corresponded to the Interfile standard [2, 4]. This standard was developed specifically to facilitate the transmission of information describing three-dimensional medical images.

Another concern was the protection of patient confidentiality: the Vanderbilt Institutional Review Board deemed it permissible to distribute the patient image data to qualified researchers, but not to make them generally accessible to anyone who wished to view them. To this end, we created a login and password that allowed FTP access to the images; in order to be given this password, a responsible individual at each remote site was required to submit to the central site a signed document promising that the image data would not be redistributed.

3.2: Submission of the registration transforms

A three-dimensional rigid body transformation may be specified in terms of six parameters (three rotation angles and three translation components). Because of the fact that the amount of data involved in communicating a transformation is relatively small, Internet e-mail seemed the obvious choice as a protocol for transmission of results. There still remained the problem, however, that there are many different ways in which a rigid-body transform may be specified (*e.g.* rotation angles, a rotation matrix and translation vector, quaternions). Each remote site had its own internal representation format for the results; the task of analyzing the results was performed entirely at the central site, and we saw the danger of great confusion if each remote site were to be allowed to submit results in its own format. Hence we insisted on the adoption of a standard submission format that was to be used in all results submissions. This format was as follows: 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 (columns headed "x","y", and "z") "transformation table" (Fig. 4 and Table 1) for each pair of volumes.

After the retrospective registration transformation was determined, the transformed positions of these eight points *relative to the origin of the "To" modality* (Fig. 5) were computed by each site (see Table 1, "new_x","new_y", and "new_z"). As depicted in Fig. 5, 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 back from the transformation table to a rigid body transformation, a singular value decomposition (SVD) algorithm [1] was used, giving the rigid body transformation that minimized the mean square distance between the two sets of points. This was the same algorithm used to register the positions of the fiducial markers in the two image volumes.

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 nonrigid transformation would be approximated by a rigid one. The use of this protocol thus limited the scope of this project to an evaluation restricted to rigid-body transformations. However, by measuring the *fiducial registration error* (FRE), i.e., the distance between the pairs of corresponding points after the application of the SVD algorithm, it is possible to determine whether or not the retrospective registration uses nonrigid deformations, since the FRE in this case would be on 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 nonrigid 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 1).

4: Data Analysis

At Vanderbilt, after the transformation tables had been received from each site and the corresponding rigid-body transformations determined, the next step in the evaluation was 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) was chosen representing 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 was computed as follows (Fig. 3). The centroid pixel of the VOI was found, and its position was converted from a voxel index to a millimetric position $\mathbf{c}$ in the "To" modality using the known voxel size for the image volume. Let $\mathbf{R}_g$ and $\mathbf{t}_g$ be the rotation matrix and translation vector, respectively, of the gold-standard rigid-body transformation G, and $\mathbf{R}_r$ and $\mathbf{t}_r$ be the rotation and translation components of the retrospective transformation R. The point $\mathbf{c}'$ in the "From" modality was defined so that $\mathbf{c}$ is the mapping of $\mathbf{c}'$ under the gold-standard transformation. Thus,

$$\mathbf{c} = G(\mathbf{c}') = \mathbf{R}_g \mathbf{c}' + \mathbf{t}_g. \tag{1}$$

By inverting equation (1), we obtain

$$\mathbf{c}' = G^{-1}(\mathbf{c}) = \mathbf{R}_g^{-1}\mathbf{c} - \mathbf{R}_g^{-1}\mathbf{t}_g. \tag{2}$$

The point $\mathbf{c}''$ in the "To" modality is defined as the mapping of $\mathbf{c}'$ under the retrospective transformation. Thus,

$$\mathbf{c}'' = R(\mathbf{c}') = \mathbf{R}_r \mathbf{c}' + \mathbf{t}_r. \tag{3}$$

The difference between the registered target position of the retrospective method and that of the gold standard is $\mathbf{d} = \mathbf{c}'' - \mathbf{c}$. We define the *Target Registration Error* (TRE) of the retrospective registration at the anatomical location of the VOI to be the Euclidean distance d between $\mathbf{c}$ and $\mathbf{c}''$, i.e., TRE $= d = ||\mathbf{d}||$.

The particular anatomical positions corresponding to the VOIs used in this evaluation are as follows: 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, and 10) right occipital horn.

Just as there are many ways to specify a rigid-body transformation, there are also many formats in which the difference between the retrospective and gold standard registration may be written. We chose to give the TRE in millimeters, as specified at a set of major landmark locations in the brain, because we felt that this would allow physicians reading the results easily to assess how useful the registration techniques would be if they gave the same degree of accuracy when used in clinical practice.

5: Presentation of results

When we finished analyzing the registration transforms that were sent to us, we were left with error estimates for sixteen registration techniques on 76 image matching tasks. For each task, we estimated the error at ten VOIs, thus giving a total of 12,160 numbers. Clearly this was far too much data to present, so we had to devise a method of compiling the data into statistics that gave a good description of the accuracy of each registration method. We chose to present six tables: median, 75th percentile, and maximum values for PET-MR and CT-MR registration. Each table contained a column for each of the remote sites that submitted transforms for that modality pair (some sites performed only CT-MR or PET-MR registration); each row of the table contained statistics for matching PET or CT volumes to a particular MR submodality (PD, T1 and T2, with and without distortion correction having been applied). Each statistic was then compiled from the errors measured at all VOIs for all patient datasets containing that modality pair. A sample table of results is shown in Table 2.

The rationale for our choice of statistics was as follows. We presented the maximum value for each technique to give insight into the stability and reliability of the technique. We used the median rather than mean value to give an estimate of the typical registration accuracy, unweighted by large errors occurring in a small number of instances where the technique performed very poorly. The 75th percentile data was added to give extra insight into reliability, *i.e*, whether a particular method failed only in one instance or for a sizeable percentage of the registration tasks.

The question that everyone wanted our study to answer was "which of the registration techniques tested was the best?". Unfortunately, because of the fact that the number of techniques was large compared to the number of tasks on which they were tested, our data lacked the statistical power to proclaim a "winner". We did notice, however, that the volume-based techniques tended to perform better than the surface-based ones. This led us to perform another analysis of the data. We partitioned the registration methods into two groups: volume-based and surface-based (omitting the few that fell into neither group). We performed the same analysis on the aggregate results of each group. In this case, there was enough statistical power in the data for us to be able to proclaim that on many of the modality pairs, the volume-based methods performed significantly more accurate registrations than the surface-based ones [7]. We were also able to conclude that, for some modality pairs, mainly those including a CT volume, the application of geometrical distortion correction to the MR volume [3, 5] resulted in a significant reduction in registration error.

6: Blindedness

We use the adjective *blinded* to describe the remote sites, to reflect the fact that they remained ignorant of the gold standard solutions until the results had been submitted. The degree to which the gold standard solutions were to be kept secret, however, was a matter that needed detailed consideration. At one extreme was the possibility of making the fiducial-based registrations freely available throughout the study: this approach seemed to have few advantages and would lead many to dismiss the results as unreliable. It would also have been feasible to withhold the gold standard indefinitely from the remote sites, and to withhold the calculated errors of the retrospective transformations until publication of the results, as the evaluation step was carried out entirely at Vanderbilt. This protocol has the advantage that it would allow researchers at the remote sites to find and correct errors in their registration transformations after submission, without suspicion that the gold standard solutions (or the error statistics derived from them) had led them to discover mistakes that in clinical practice would have remained undetected. However, for reasons that will be explained in Section 7, some of the remote sites strongly felt that this approach was unacceptable, and we elected to adopt a compromise solution.

We withheld all results and gold standard transforms until all the remote sites had completed their submissions. At this point, we sent by e-mail to the researchers at each site a table containing the TRE values that we had calculated for their registration method. We also created a separate login and password that gave access to the gold standard information, and distributed this. We advised each site to examine carefully the numbers in that table, paying particular attention to any values that seemed larger than expected.

The blindedness of the study would clearly have been compromised if we had routinely accepted resubmissions of results after the gold standard had been released; it was of concern, however, that the assessment might be inaccurate because of mistakes in, for example, conversions between the internal representation format of remote sites and the format required for submission to the central site. These assessments would not give an accurate view of the clinical performance of a registration technique, as such conversions would not have to be performed for real-world registration tasks. For this reason, we allowed remote sites to review the results and if necessary to send a statement that an error had been made in the submission of a particular subset of the results. We did not allow resubmission of this subset; instead, we omitted the erroneous results from the calculation of the overall error statistics for that registration method and added a footnote describing the omission in the articles containing the results. We made only one exception to the "no resubmission" rule, as described in Section 7. We also allowed remote sites to send a declaration that the registration had failed in particular cases. This declaration would have allowed the omission of those cases from the overall error calculations, and a note would have been added stating how many of the registration tasks were declared to be failures for each method. We received no such declarations, however.

7: Errors

It is natural to expect that, because of the complexity of and number of participants in this study, there would be some degree of error and miscommunication. Rather than striving for the elimination of error, we tried to develop a protocol that would give a good opportunity for such errors, made by both the central and remote sites, to be detected and corrected before the final results were published.

7.1: The practice patient

In addition to the datasets that were used for the study, we provided an additional patient dataset that was similar to the others but different in two respects. First, the gold standard registration transforms for the dataset were immediately made accessible to the remote sites. Second, the images in this dataset had had randomly chosen gross rotations and translations applied to them, so that the image pairs were profoundly misregistered. The images used for the evaluations were only slightly misregistered.

The rationale for providing this "practice" dataset was that the remote sites would be able to perform registrations on this dataset and compare results against the gold standard. Inexplicably large errors would be a warning that the conversion between registration representation formats was possibly being performed incorrectly. The images were significantly misaligned so that such incorrect conversions would show up more clearly.

The practice patient rapidly proved valuable, but in a way that was quite surprising. A researcher at one of the remote sites complained that his registrations were significantly different from the gold standard, but that his results appeared to provide superior registrations on visual inspection. This led to the discovery of an error in the algorithm used at Vanderbilt to compute the gold standard transformations. Although this error was related to the preparatory reformatting steps taken in producing the practice dataset and would therefore not have affected the calculation of error statistics for the evaluations, it was imperative that the mistake be found and corrected lest the other remote sites be confused by the failure of their registration transforms to match the erroneous gold standard results on the practice patient. Because of this error one of the researchers urged us to release the gold standard transformations for all the image pairs before publication of the results, therefore allowing remote sites to check visually the accuracy of the transforms against which theirs were to be compared.

After we corrected the gold standard transformations, several remote sites still reported a large mismatch with their registration solutions. Communication with these sites revealed several conversion errors between transformation formats.

7.2: Large misregistrations

When we had completed the analysis of all the results submitted to us, we saw that, for some sites, the maximum registration errors were extremely large (greater than 40 mm). There were four possible explanations for a large error being observed for a particular matching task and remote site:

1. The gold standard was in error for that transform.
2. The remote site submitting the result had not correctly performed the conversion between transformation formats.
3. The retrospective registration technique in question had failed for that particular matching task.
4. An administrative error had been made by the remote site (*e.g.*, the images matched by the registration transform had been mislabeled).

In all instances, we ruled out the first and second possibilities. For the first possibility, an erroneous gold standard transform would have given consistently large error measures for all remote sites. We did not observe such consistent errors for any particular image pair. For the second possibility, the remote site would have submitted transforms all of which showed large error measures. We did not see this consistency either.

When the remote sites were presented with the results, several of them explained their large maximum errors by describing administrative mistakes that had been made. We considered these explanations and in each case this allowed us to omit the transforms in question from the overall calculation of error statistics (there were three such cases). There were still some large error measurements that remained unexplained, however. As stated in Section 6, we allowed the researchers performing the registrations to declare that their technique had failed for a particular task, but no such declarations were received. It is thus our belief that the large error values were because of registration failures that had remained undiagnosed because of a lack of visual inspection of the image matching. The conclusion of our publication of the results of the study highlighted the importance of using visual inspection to check the results of automated image registration techniques.

Finally, there was one case in which an exception was made to the rule that no registration transforms could be submitted once the gold standard had been released. Shortly before the results were due to be published, and after the gold standard had been made available, we received a statement from one remote site that they had made a mistake converting from their own transform representation to the submission format. Specifically, they claimed that the sign of one of the rotation angles of the transformation had been consistently negated during the conversion. As the orientation of the images being matched were fairly close, the rotation angles were small. The observed registration errors were changed by the sign negation on the order of 0.5 mm, which was about 50% of the median error. As we were able to independently verify that the new transforms were derived from the old ones simply by negating one of the rotation angles, we felt that the blindedness of the study was not compromised by allowing the resubmission of these results. In every case in which mistakes led to omission or changing of data a description of the nature of the mistake was included in our published results along with the evaluations both with and without the alterations.

8: Conclusion

By performing this study, we were able to publish what we believe to be a fair, quantitative comparison of many different retrospective image registration methods. Judging from the reprint requests we have received for the article containing the results [6] and from the number of research groups who have subsequently asked for and been granted access to the database so that their registration techniques may be similarly evaluated (sixteen), we believe our study was successful and hope it will provide encouragement to others to conduct blinded studies via the Internet. It appears to have provided both a service and information that was of great interest to many people active in this field. As a final note, we would point out that this study involved many people in many research groups in many countries. It could not have been successful unless these people felt strongly that it should be successful. As in any project in which the efforts of people are to be evaluated with the results to be published, there is a great potential for misunderstanding, impatience, and disagreement. These problems are especially critical when the people involved are competing in the same field for publications, funding, and peer approval. We would like to add a personal note of gratitude to the people with whom we interacted at the remote sites during this project for their professional approach. We have learned from our experience with this project, and we suspect that it applies to any such project, that while careful preparation and attention to detail are essential on the part of the evaluators, the project would fail without active, supportive cooperation on the part of those who are being evaluated.

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] 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.

[5] 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.*, 20:666–679, 1996.

[6] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire, Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comput. Assist. Tomogr.*, 21:554–566, 1997.

[7] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, and R. J. Maciunas. Retrospective intermodality registration techniques: Surface-based versus volume-based. In J. Troccaz, E. Grimson, and R. Mösges, editors, *CVRMed-MRCAS '97*, pages 151–160. Springer-Verlag, Berlin, 1997.

Table 1. Example 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. Median Errors for CT-to-MR Registration

MR	Technique														N
Modality	BA	CO	EL	HA	HE	HI‡	MAI	MAL	NO†	PE	RO1	RO2	RO3	RO4	
T1	1.6	1.5	1.6	3.4	1.4	1.2	5.1	4.3	3.3	2.7	4.2	5.2	5.7	5.4	7
PD	1.9	1.5	2.0	3.1	2.4	1.9	4.1	4.0	7.8	1.9	4.5	5.5	4.9	4.8	7
T2	2.5	1.5	1.6	4.2	4.7	1.5	3.9	5.0	3.9	2.5	4.5	4.5	5.4	4.7	7
T1 rect.	1.4*	0.7	0.9	3.3	1.0	0.7	4.9	5.4	3.4	2.2	5.9	5.9	6.3	5.9	6
PD rect.	1.7*	0.8	1.1	3.0	1.7	0.7	3.0	4.0	4.6	2.1	5.9	5.7	5.5*	5.5*	7
T2 rect.	2.1*	0.8	1.6	3.5	1.6	0.8	4.3	5.3	4.2	2.9	5.5	5.3	5.3	5.3	7

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.

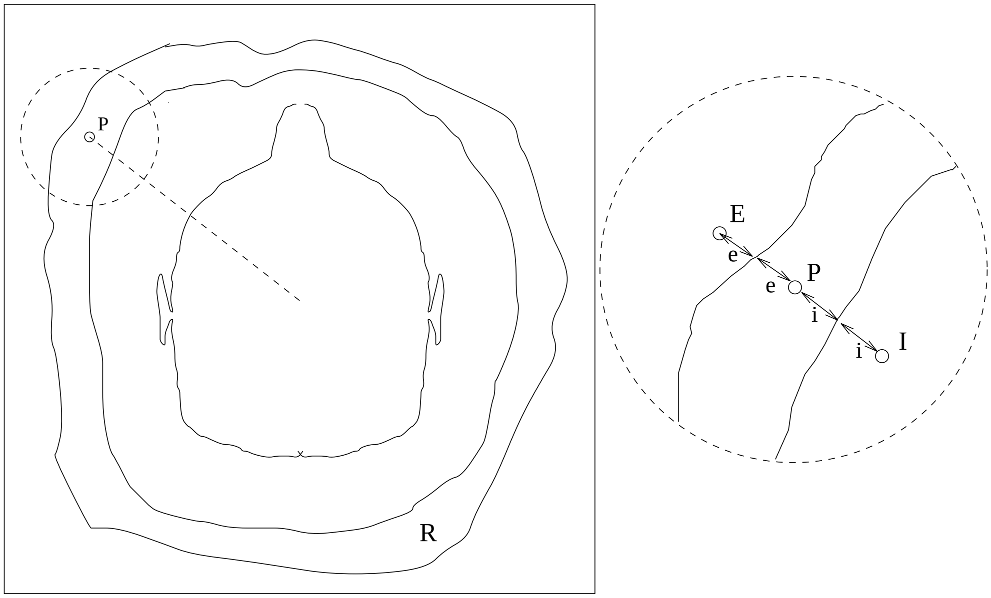
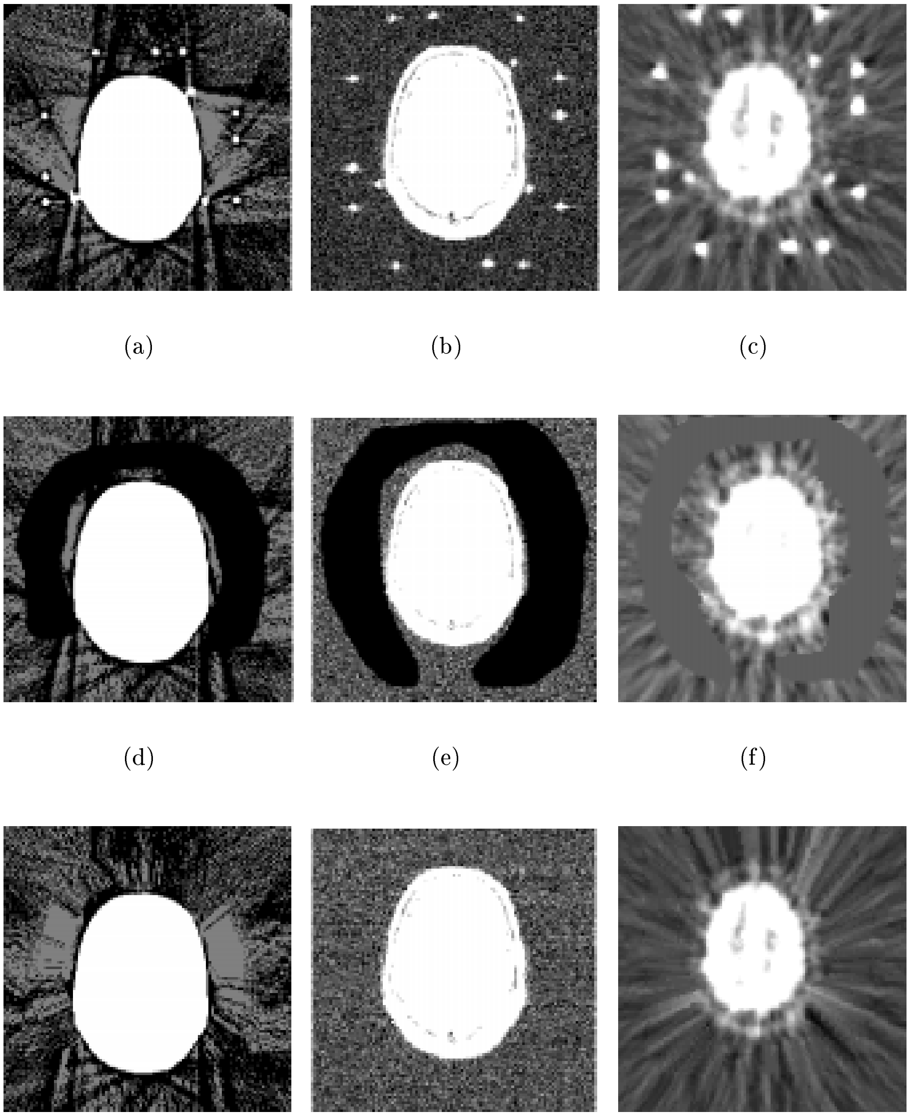


Figure 1. Calculation of points for interpolation of background patterns in CT and PET images. See text for an explanation.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 2. Removal of fiducial markers and stereotactic frame. The top row (a–c) shows sample original image slices from CT (a), MR (b), and PET (c). 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. The window and level have been set to show the background artifacts. The middle row (d–f) shows the same image slices after the region R has been outlined and zeroed. All trace of the stereotactic frame and fiducial markers has been removed. The bottom row (g–i) shows the image slices after reconstruction of the background in the region R. For MR, the replaced area is indistinguishable from the rest of the background. For CT and PET, there are slight discontinuities in the direction of the stripes, but the intensity changes relatively smoothly.

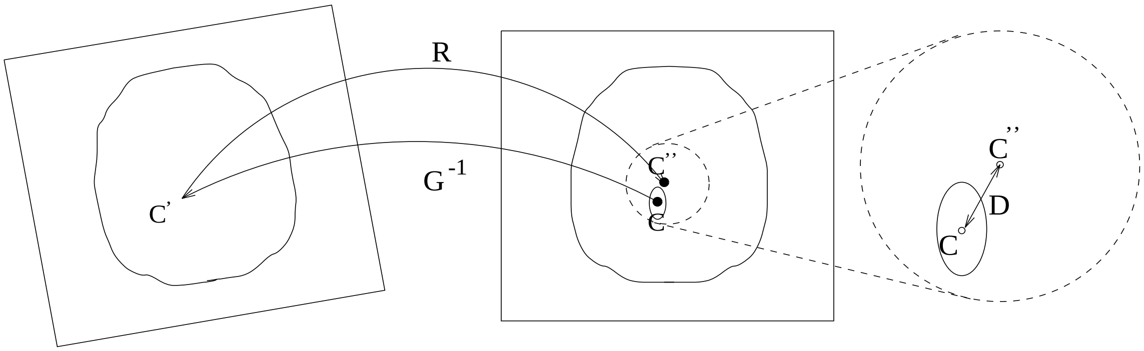
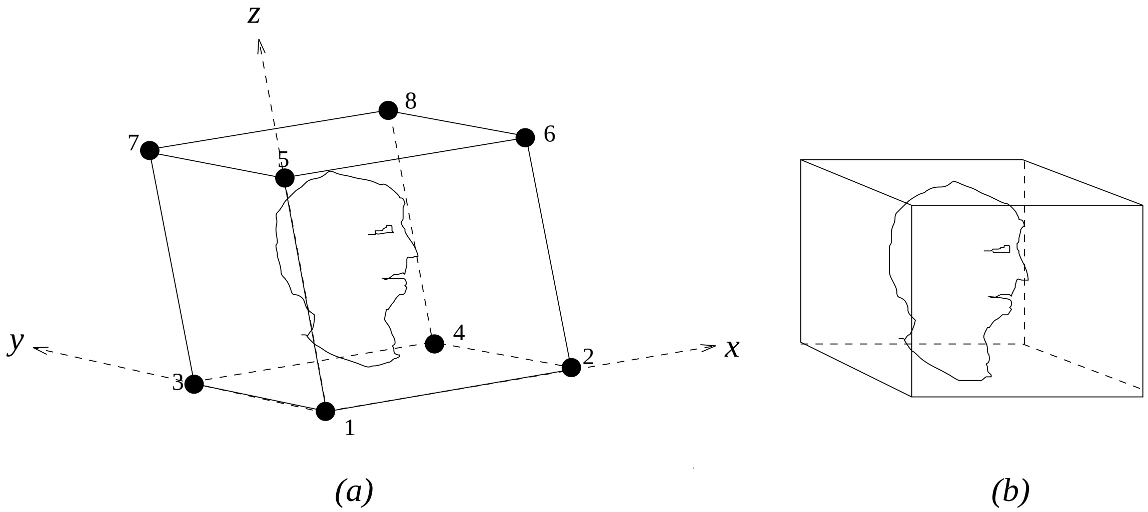


Figure 3. Calculation of the accuracy of a retrospective registration at a VOI. A VOI (represented here by an ellipse) is defined in the "To" modality (right image). The centroid voxel of the VOI is found and converted from a voxel index to a millimetric position c **using the known voxel size for the image volume. The inverse of the gold-standard rigid-body transformation** G **is applied to the point** c**, giving point** $\mathbf{c}' = G^{-1}(\mathbf{c})$ **in the "From" modality (left image). Then, the retrospective transformation** R **is applied to** c'**, giving point** $\mathbf{c}''$**. The registration error of the retrospective transformation at the centroid of the VOI is taken to be the Euclidean distance** d **between the points** c **and** $\mathbf{c}'' = R(G^{-1}(\mathbf{c}))$**.**



(a)

(b)

Figure 4. 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 1.**

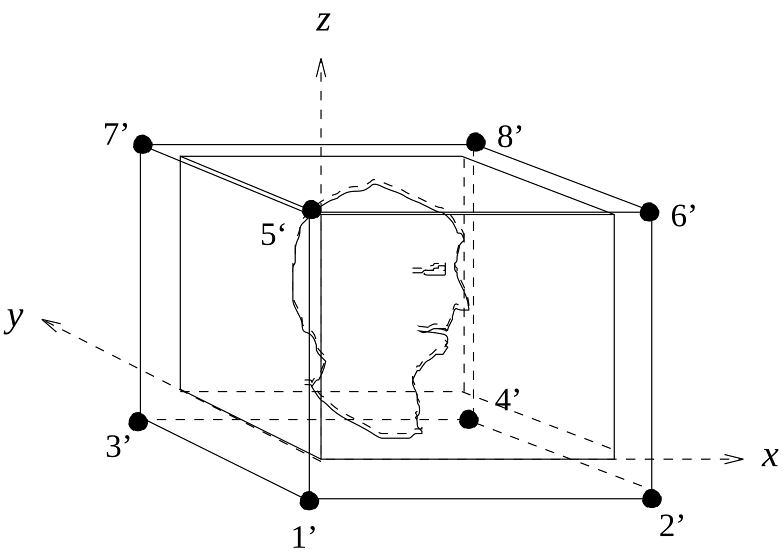


Figure 5. Image volumes after registration, so that the heads in the two volumes are almost perfectly aligned with each other. The points $1'$ to $8'$ are used to specify the transformation and form the right-hand side of the transformation table.