

# Phantom validation of coregistration of PET and CT for image-guided radiotherapy

William C. Lavelly<sup>a)</sup>

*Department of Radiology and Radiological Sciences, Vanderbilt University, Nashville, Tennessee*

Christopher Scarfone

*Department of Radiology and Radiological Sciences and Department of Radiation Oncology, Vanderbilt University, Nashville, Tennessee*

Hakan Cevikalp

*Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee*

Rui Li

*Department of Radiology and Radiological Sciences and Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee*

Daniel W. Byrne

*General Clinical Research Center, Vanderbilt University, Nashville, Tennessee*

Anthony J. Cmelak

*Department of Radiation Oncology, Vanderbilt University, Nashville, Tennessee*

Benoit Dawant

*Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee*

Ronald R. Price

*Department of Radiology and Radiological Sciences, Vanderbilt University, Nashville, Tennessee*

Dennis E. Hallahan

*Department of Radiation Oncology, Vanderbilt University, Nashville, Tennessee*

J. Michael Fitzpatrick

*Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee*

(Received 20 October 2003; revised 19 January 2004; accepted for publication 28 January 2004; published 14 April 2004)

Radiotherapy treatment planning integrating positron emission tomography (PET) and computerized tomography (CT) is rapidly gaining acceptance in the clinical setting. Although hybrid systems are available, often the planning CT is acquired on a dedicated system separate from the PET scanner. A limiting factor to using PET data becomes the accuracy of the CT/PET registration. In this work, we use phantom and patient validation to demonstrate a general method for assessing the accuracy of CT/PET image registration and apply it to two multi-modality image registration programs. An IAEA (International Atomic Energy Association) brain phantom and an anthropomorphic head phantom were used. Internal volumes and externally mounted fiducial markers were filled with CT contrast and <sup>18</sup>F-fluorodeoxyglucose (FDG). CT, PET emission, and PET transmission images were acquired and registered using two different image registration algorithms. CT/PET Fusion (GE Medical Systems, Milwaukee, WI) is commercially available and uses a semi-automated initial step followed by manual adjustment. Automatic Mutual Information-based Registration (AMIR), developed at our institution, is fully automated and exhibits no variation between repeated registrations. Registration was performed using distinct phantom structures; assessment of accuracy was determined from registration of the calculated centroids of a set of fiducial markers. By comparing structure-based registration with fiducial-based registration, target registration error (TRE) was computed at each point in a three-dimensional (3D) grid that spans the image volume. Identical methods were also applied to patient data to assess CT/PET registration accuracy. Accuracy was calculated as the mean with standard deviation of the TRE for every point in the 3D grid. Overall TRE values for the IAEA brain phantom are: CT/PET Fusion =  $1.71 \pm 0.62$  mm, AMIR =  $1.13 \pm 0.53$  mm; overall TRE values for the anthropomorphic head phantom are: CT/PET Fusion =  $1.66 \pm 0.53$  mm, AMIR =  $1.15 \pm 0.48$  mm. Precision (repeatability by a single user) measured for CT/PET Fusion: IAEA phantom =  $1.59 \pm 0.67$  mm and anthropomorphic head phantom =  $1.63 \pm 0.52$  mm. (AMIR has exact precision and so no measurements are necessary.) One sample patient demonstrated the following accuracy results: CT/PET Fusion =  $3.89 \pm 1.61$  mm, AMIR =  $2.86 \pm 0.60$  mm. Semi-automatic and automatic image registration methods may be used to facilitate incorporation of PET data into radiotherapy treatment planning in relatively rigid anatomic sites, such as head and neck. The overall accuracies in phantom and patient images are  $< 2$  mm and

<4 mm, respectively, using either registration algorithm. Registration accuracy may decrease, however, as distance from the initial registration points (CT/PET fusion) or center of the image (AMIR) increases. Additional information provided by PET may improve dose coverage to active tumor subregions and hence tumor control. This study shows that the accuracy obtained by image registration with these two methods is well suited for image-guided radiotherapy. © 2004 American Association of Physicists in Medicine. [DOI: 10.1118/1.1688041]

Key words: CT/PET image registration, radiotherapy treatment planning, target registration error, intensity-modulated radiotherapy (IMRT)

## I. INTRODUCTION

A quantitative assessment of positron emission tomography (PET) and x-ray computed tomography (CT) image registration is a logical prerequisite for precisely identifying and contouring target and normal tissues sites for radiotherapy treatment planning (RTP). Morphologic CT is the customary imaging technique in RTP, with applications ranging from target delineation and dose calculation to isocenter verification. The anatomical detail provided by CT is excellent, and with contrast enhancement, vascular structures may also be well visualized. However, there is a complete range of biochemical events associated with cancer that cannot be imaged using anatomical/geometrical techniques such as CT. For example, it is possible, using 99m-Technetium single photon emission computed tomography ( $^{99m}\text{Tc}$ -SPECT), to image apoptotic cells<sup>1</sup> and using 18-Fluorine fluorodeoxyglucose ( $^{18}\text{F}$ -FDG) PET, to visualize analogues to glucose metabolism<sup>2</sup> and help differentiate proliferating and nonproliferating regions within a tumor. Compared to these methods, CT has the additional limitation of not differentiating scar tissue, lung atelectasis, and edema from that of neoplastic tissues.<sup>3</sup>

Although molecular imaging techniques offer additional information for RTP, there are well-documented limitations to using these methods in quantitative applications. For example, molecular images have a relatively low signal-to-noise ratio, a fact that may be compounded by poor uptake in low grade tumors, and have a spatial resolution on the order of 10–20 times larger (poorer) than CT. In addition, there is potential for false positive findings related to infectious and inflammatory processes. Perhaps the greatest restriction to using molecular images in RTP, however, is the lack of anatomical correlates in the image data. Radiation treatment planning requires an accurate and precise assessment of the location of target and normal tissues for contouring purposes. Thus, accurate CT/PET registration is a chief requirement for RTP. Also, tissue density information obtained from CT data is needed for accurate dosimetry planning. In the absence of anatomical landmarks, as is the case with PET imaging, it is difficult to associate area(s) of high uptake with particular anatomy and therefore with radiation beam placement and the distribution of absorbed dose.

With the additional information provided by PET, the current trend in RTP is to incorporate PET imaging data to reduce or eliminate geographical misses and help guide dose

intensification to the most metabolically active areas within the neoplasm. Registration of CT and PET data arguably improves the ability of radiotherapy to adequately treat the tumor when compared to both CT and PET imaging modalities considered individually. An inherent problem in the registration of independently acquired CT and PET data sets is the accuracy of the final image alignment. If the registration algorithm cannot specify the accuracy of the registration, then contours based on the registered images may not accurately reflect tumor and normal tissue extent, and therefore may reduce the probability of tumor control and increase the probability of normal tissue complications. An earlier study investigated the possibility of using mutual information as a criterion for registering MR and PET images.<sup>4</sup> While it did not give a direct measure of the accuracy of the registrations, its results indicated that mutual information has the potential to provide a robust measure for these two modalities. A recent study showed that mutual information is also a good choice for CT and PET.<sup>5</sup> It gave a direct measure of registration accuracy but did not include an independent measure of ground truth. In this paper, we quantitatively evaluate CT/PET image registration based on internal structures in both phantom and patient data and compare the results to registration using fixed external landmarks (fiducials) as ground truth.

## II. METHODS

### A. Structure-based and fiducial-based registration

To measure target registration error (TRE), and therefore the accuracy of CT/PET image registration, a baseline or “gold standard” registration method representing the “ideal” alignment between CT and PET is required for comparison with the registration method under study. Based on previous research at this institution,<sup>6–8</sup> a fiducial-based registration (FBR) technique was developed as the “gold standard” to which structural-based registration (SBR) was compared.

Fiducial-based registration relies on the use of externally mounted fiducial markers. The fiducial markers used in this study are capsules that are filled with a mixture of both radio-opaque and radioactive materials and affixed to the external surface of an object (phantom). The capsules are then visible on CT and PET, respectively. Throughout this paper, the phrase “fiducial points” or “fiducials” refers to the physical position of the fiducial marker on the phantom.

Fiducial-based registration works by calculating the transformation that minimizes the mean square distance between corresponding fiducials in the two images.<sup>8</sup> The spatial location of a fiducial marker was reduced to a single point by calculating the centroid or, equivalently, the center-of-mass of the active image pixels defining the fiducial as determined using thresholding techniques. Fiducial-based registration was then accomplished by matching the centroids of the CT fiducials with the corresponding centroids in the post-registration PET images. By direct comparison of SBR to FBR, TRE for SBR was calculated.

The accuracy of structure-based CT and PET image registration was measured based on image data acquired from two rigid medical imaging phantoms. Following SBR and FBR comparison, a three-dimensional map of TREs was created from the reformatted PET images, which allowed the determination of registration error at any point in the phantom image. This map was then displayed as color-coded grid superimposed over the CT image data.

### B. Image registration software programs

Two different image registration programs were chosen to evaluate image registration accuracy: CT/PET Fusion (General Electric Medical Systems, Milwaukee, WI) and Automatic Mutual Information-based Registration (AMIR) (Vanderbilt University, Nashville, TN).<sup>9</sup>

### III. CT/PET FUSION

CT/PET Fusion employs a proprietary feature matching algorithm. First, both the CT and the PET images are segmented into foreground and background; then, the surfaces are examined to identify distinguishing geometrical features, followed by a search to establish a correspondence between the features in the two spaces. Finally, a transformation is determined to bring these points approximately into coincidence.

The program allows for either a semi-automatic setup with manual definition of common points on the two exams, or a completely automatic setup. It also allows for varying degrees of rigidity in the transforms. In this study, however, the semi-automatic setup and the completely rigid option were always used. Registration was a two-stage process: (1) semi-automatic registration followed by (2) a minimum of three landmark selection steps that may lead to the computation of a new registration. The application then calculates a rigid transform (translation and rotation) to register (align) the coordinates in the second exam (the "registered" exam), i.e., PET, with the coordinates of the first exam (the "reference" exam), i.e., CT. Once the first stage has been completed, the user defines a minimum of three pairs of matching points in the two image sets, using landmarks (e.g., bony structures, soft tissue boundaries) that can be clearly identified in both data sets. The system then calculates a new rigid registration. During this second stage, the user selects points in the reference exam and the software displays the corre-

sponding points as a marker in the registered exam. If necessary, the user then moves the marker to indicate the corresponding point in the registered exam. After the second stage has been completed, a visual inspection of the registration is done by blending the two images to assess correspondence of the CT and PET structures. If the user is dissatisfied with the registration, both stages are repeated. When the user is satisfied with the registration, contours can be drawn on the registered views and the result of the work stored for further study.

### IV. AUTOMATIC MUTUAL INFORMATION-BASED REGISTRATION

Automatic Mutual Information-based Registration provides fully automatic registration of multi-modality images. Mutual information has been shown to be a highly effective measure of registration quality.<sup>10,11</sup> Variants such as normalized mutual information<sup>12</sup> and mutual information augmented by gradient information<sup>13</sup> are popular as well, but we have found that mutual information itself produces excellent results in this application. A template configuration file was generated automatically by the software for both image data sets to be registered. The template configuration file contained information about the images, such as dimensions, voxel sizes, orientation, and data types, as well as the parameters for the registration, such as step sizes, number of bins, tolerance, etc., were filled in by the user. The above-mentioned parameters represent the total range of the image information required by the software. The program uses an emission PET image and a CT image as inputs. While a transmission PET might be more easily registered to a CT, which is itself a transmission image, we chose the emission image because in patient studies motion between the acquisition of transmission and emission scans may reduce the accuracy of the registration.

The program then finds the alignment of the two images in three-dimensional space by maximizing the registration criterion mutual information between them using the Powell algorithm. Powell's algorithm searches for an optimum along conjugate directions without requiring the computation of the gradient of the function to be optimized. In this work we have followed the implementation on pages 314-17 of *Numerical Recipes in C*.<sup>14</sup> For rigid-body registration, six degrees of freedom (three translations and three rotations) are involved in the searching. For every rigid body registration we have performed in this study, we have optimized the parameters in the following order: translation in row, column, slice direction, followed by rotation about the row, column and slice directions. A multi-resolution scheme is used to speed up the calculation, and lower the likelihood of converging into a local maximum position. Once the final alignment of the two images is reached, both the registered image and the transformation are output by the software. The transformation can be either saved directly in the form of matrices in an ASCII file, or converted to displacement fields for further use.

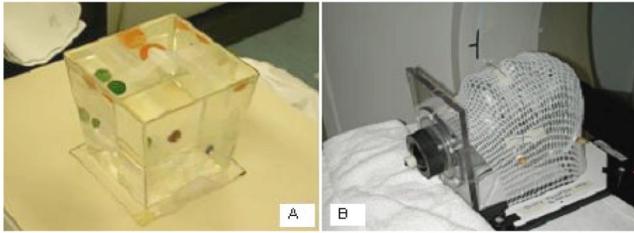


FIG. 1. (A) The IAEA brain phantom; (B) the anthropomorphic head phantom.

## V. PHANTOM DESCRIPTIONS

### A. International Atomic Energy Association Brain Phantom

The International Atomic Energy Association (IAEA) Brain Phantom is a 15 cm<sup>3</sup> hollow plastic three-dimensional trapezoid with a total liquid volume of approximately 2160 ml [Fig. 1(A)]. On two of the opposite internal sides are positioned four hollow spheres with pinpoint holes providing access for introduction of liquid materials into the spheres via needles. The spheres have volumes of 0.50 ml (blue), 1.75 ml (red), 4.00 ml (green), and 13.50 ml (orange), respectively. The blue and red spheres are located on one side and the green and orange spheres are located on the opposite side. The spheres on each side are located in opposite diagonal directions.

### B. Anthropomorphic head phantom

The Anthropomorphic Head Phantom (The Phantom Laboratory, Salem, NY) is a hollow plastic human head-shaped phantom which is sealable on the “neck” end [Fig. 1(B)]. The volume is 4130 ml. Extending into the phantom from the “neck” end on a 15 cm plastic rod is a 10 ml hollow plastic “tumor” which can be filled with contrast agent.

To obtain nearly identical phantom positioning between CT and PET studies, a Uni-frame head and neck immobilization system in the form of a patient-specific expanded plastic facemask (MED-TEC, Inc., Orange City, IA) was used during both CT and PET acquisition.

#### 1. Fiducial markers

The fiducials were made from the bulb end of a plastic transfer pipet (Fisher Scientific, Pittsburgh, PA) by cutting the tubular part off leaving a very short portion to seal after the contrast agent was infused into the bulb portion. The result is a capsule-shaped fiducial whose length and width are approximately 3 and 1 cm, respectively, and whose volume is approximately 4.5 ml. A silicone sealant was used to seal the opening of the fiducial. The fiducial markers were then placed in different orientations and at a variety of locations on the surface of each phantom (e.g., bridge of nose on head phantom) to provide several registration points for assessing the accuracy of the registration. One fiducial was

placed on each of four sides of the IAEA Brain Phantom; six fiducials were placed in different locations around the anthropomorphic head phantom.

#### 2. Contrast agent

The contrast agent consisted of 2.5  $\mu$ Ci/ml of FDG and 100 mg/ml Optiray 320 (Mallinckrodt, Inc., St. Louis). These agents were mixed to concentration with water and injected into the spheres and the fiducials. A trace amount of FDG was also added to the water used to fill the volume of the phantom to provide a minimal amount of background contrast.

#### 3. Image acquisition: CT imaging

CT scans of the phantoms were performed on a Picker PQ 5000 scanner (Picker International, Cleveland, OH). Scans consisted of a series of transverse views (512 $\times$ 512 acquisition matrix, 4.00 mm slice thickness) through the entire phantom.

## VI. FDG PET IMAGING

The images were acquired with a dedicated PET tomograph (GE Advance, General Electric Medical Systems, Milwaukee, WI). Both emission and transmission images were obtained over one standard 15 cm PET bed position in the IAEA phantom and over two standard 15 cm PET bed positions in the anthropomorphic head phantom. Emission and transmission images were acquired and reconstructed at 4.25 mm thickness. Both sets of FDG PET images with and without attenuation correction were reviewed on an interactive computer system with use of a linear gray scale and a continuous color scale with varying degrees of background subtraction. FDG PET studies were interpreted visually. Areas of FDG uptake were categorized based on location, intensity, shape, and size.

### A. Image registration

#### 1. CT/PET Fusion

Both CT and PET emission and transmission images were transferred via network to an Advantage Workstation Version 1.0.56 (General Electric Medical Systems, Milwaukee, WI). All three image sets (CT, PET emission, and PET transmission) were then loaded into the CT/PET Fusion program. An initial point of reference was chosen in the CT image; a corresponding point was chosen in either the PET emission or transmission image. This image registration program then used a mutual information-based algorithm to register the image sets. After a close registration based on the initial point of reference, a minimum of three additional points were then chosen in both the CT image and PET image (either transmission or emission) to improve the registration. To examine the effect of internal structure size on registration, the different size spheres in the IAEA brain phantom were chosen in order (smallest-to-largest, largest-to-smallest, and random) as the location of the initial point of reference. The

number of points used in the IAEA phantom was four; the number of points used in the anthropomorphic head phantom was six.

## 2. AMIR

The CT, PET emission, and PET transmission images were saved onto a CD and transferred to the AMIR workstation. Both CT and PET emission data sets were loaded into the program. The automatic registration was completed and the CT, PET emission and registered PET emission images were displayed in a visualization environment written in IDL (Interactive Data Language, Research Systems Inc., Boulder, CO).

## B. Number and qualification of registrants

Accuracy calculations were determined for each phantom in both registration algorithms. Precision calculations were calculated only in the CT/PET Fusion algorithm. Because the AMIR algorithm is fully automated and deterministic, the results are exactly reproducible, thus negating the need for precision calculations.

Ten medically familiar people were chosen to register both sets of phantom images: five radiology/radiation oncology Ph.D.'s, four radiology/radiation oncology M.D.'s, and one radiology technician. Only one of the Ph.D.'s and one of the M.D.'s were familiar with the registration programs at the time of registration.

The one M.D. (W.C.L.) who was familiar with the CT/PET Fusion registration program completed fifteen separate registrations on the IAEA Brain Phantom and five separate registrations on the anthropomorphic head phantom; precision calculations were determined from these registrations. Registration results for the IAEA brain phantom are classified according to sphere size and order of inclusion in the registration (i.e., smallest-to-largest, largest-to-smallest, and random) for the location of the initial reference point.

## C. Assessment of accuracy

The CT and registered PET emission images were transferred on disk to a separate computer for accuracy assessment. We evaluated CT/PET SBR accuracy for both CT/PET fusion and AMIR registration algorithms using image data from both phantoms. The accuracy of the final CT/PET registration from these algorithms was compared to the "gold standard" FBR, which was based on alignment of the calculated centroids of the fiducial markers in the CT and registered PET image sets. Any difference in the location of the PET data relative to the CT data when comparing SBR to FBR was recorded as the TRE.

In-house software was used to calculate the centroid of each fiducial marker on both the CT and registered PET images. All accuracy testing was completed in MATLAB, Version 6.1.0.450, Release 12.1 (The MathWorks, Inc., Natick, MA). The markers were designed to produce bright images in both CT and PET, but because of the finite resolution and the presence of noise in each modality, it is difficult to find their centroids visually. Instead, we adapted a method designed for

a similarly shaped marker.<sup>15</sup> There are two steps to this method. The first step finds a set of voxels that are likely candidates to lie inside a marker. The second step uses these candidate voxels to identify true markers and to find the centroids of the markers. The first step can be replaced by visual selection, but the second step cannot, and it is the second step that is critical to the accuracy of the method. For these reasons we describe only the second step, which itself has the following two parts.

(1) A succession of intensity thresholds is examined beginning from the maximum value in the image and descending. For each threshold the voxels that are three-dimensionally connected to the candidate voxel are identified as the "foreground" for that threshold. The foreground is examined to determine whether it is sufficiently similar in shape and size to a marker. The similarity is sufficient if and only if

- (a) The distance from the center of the candidate voxel to the center of the most distant voxel of the foreground must be less than or equal to the longest marker dimension.
- (b) The volume  $V_c$  of the foreground in cubic millimeters must be within the range  $[\alpha V_m, \beta V_m]$ , where  $V_m$  is the volume of the marker,  $\alpha$  is 0.65, and  $\beta$  is 1.65. The values of these parameters were determined empirically for the specific markers and imaging protocols used in this study. They represent the variation in apparent volume of markers at arbitrary orientations that results from the approximate representation of the cylindrical markers by finite sized, rectangular voxels.

(2) If no such threshold exists, the candidate point is discarded. Otherwise, the coordinates  $x_c$ ,  $y_c$ ,  $z_c$  of the marker centroid are calculated as follows:

$$x_c = \frac{\sum_i (I_i - I_0) x_i}{\sum_i (I_i - I_0)}, \quad y_c = \frac{\sum_i (I_i - I_0) y_i}{\sum_i (I_i - I_0)},$$

$$z_c = \frac{\sum_i (I_i - I_0) z_i}{\sum_i (I_i - I_0)},$$

where the summation is over the voxels in the foreground,  $I_i$  is the intensity of a voxel  $i$ ,  $I_0$  is the intensity of an empty voxel, and  $x_i$ ,  $y_i$ , and  $z_i$ , are the coordinates of voxel  $i$ .

Once localization (i.e., determination of the centroid of a marker) has been completed for all markers in both images, the resulting centroids were used to complete the registration. The registration was accomplished by finding the rotation matrix  $\mathbf{R}$  and translation vector  $\mathbf{t}$  that minimizes the sum of the squares of the distances between corresponding centroids,

$$\sum_i^N |\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i|^2,$$

where  $\mathbf{x}_i = (x_c, y_c, z_c)$  is the centroid of marker  $i$  in the CT image and  $\mathbf{y}_i$  is the centroid of the same marker in the PET image. Closed-form solutions exist for this problem. We

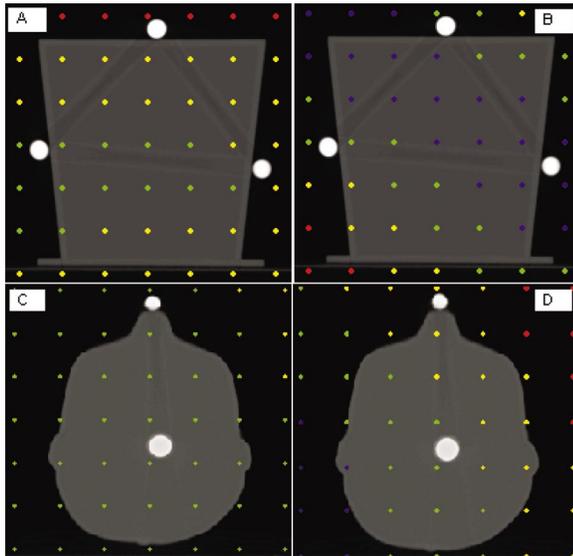


FIG. 2. (A) Slice 17 of the CT image of the IAEA brain phantom from the CT/PET Fusion program demonstrating color representation of the TRE grid. (B) Slice 17 of the CT image of the IAEA brain phantom from the AMIR program demonstrating color representation of the TRE grid. (C) Slice 29 of the CT image of the anthropomorphic head phantom from the CT/PET Fusion program demonstrating color representation of the TRE grid. (D) Slice 29 of the CT image of the anthropomorphic head phantom from the AMIR program demonstrating color representation of the TRE grid. The color legend is Blue < 0.5 mm, Green 0.5–1.0 mm, Yellow 1.0–1.5 mm, Red 1.5–2.0 mm, Magenta > 2.0 mm.

chose the solution based on the singular value decomposition of the cross-covariance matrix of the fiducial positions.<sup>16</sup>

This FBR algorithm is applied to the pair of images *after* registration has been carried out by the method being evaluated. The resulting rotation matrix  $\mathbf{R}$  and translation vector  $\mathbf{t}$  represent the disagreement between the evaluated method and the marker-based method. As a measure of the error represented by this disagreement, we used the average TRE, which is the discrepancy between points registered by the method under evaluation and the marker-based method. We evaluated TRE at each of a collection of points  $\mathbf{r}_n$ , as follows:

$$\text{TRE}_n = |\mathbf{R}\mathbf{r}_n + \mathbf{t} - \mathbf{r}_n|.$$

The points  $\mathbf{r}_n$  form a rectangular regular grid in each slice. This grid covers the object of our interest. There are totally 49 points on each grid, as shown in Fig. 2. The average TRE over the entire set of points for all slices for a given pair of images is then measured.

#### D. Statistical analysis

This study was designed to determine the overall accuracy of two CT/PET image registration using two different mutual information-based algorithms and to assess the general method of comparing SBR to FBR. Accuracy assessment was determined by the use of means and standard deviations of all TRE's for all registrants. Precision was calculated for the CT/PET Fusion program to establish the variability for one registrant; it is not calculated in the AMIR program be-

cause such an automatic algorithm is exactly reproducible. A repeated-measures analysis of variance (used for repeated registrations done in CT/PET Fusion) or a paired *t* test (used to compare AMIR to CT/PET Fusion) was used to compare the two algorithms for image registration of each phantom. Statistical analysis was performed on a personal computer with the statistical package SPSS for Windows, Version 11.0 (SPSS, Chicago, IL).

#### E. Patient imaging

Target registration error was also determined for one patient using the same process as that used on each phantom. The patient was a 54 year old male with T2N2aM0 squamous cell carcinoma of the soft palate with one enlarged lymph node. Prior to CT imaging, the patient was custom fitted with a Uni-frame head and neck immobilization system (MED-TEC Inc., Orange City, IA); the facemask was worn during both CT and PET scans. The fiducial markers were placed in five locations directly on the facemask. Four registration points were chosen in the CT/PET Fusion program: right base of the brain, medial aspect of the enlarged lymph node, inferior aspect of the enlarged lymph node, and left outermost point of brain. The AMIR program was fully automated, i.e., no initial or adjustment point(s) of registration were required. Target registration error was calculated for the entire 125 slice image set, a subset of image slices encompassing the tumor (slices 60–85) and a small 8 cm × 8 cm × 8 cm loco-regional area just around the tumor.

### VII. RESULTS

#### A. IAEA brain phantom

##### 1. CT/PET Fusion

Accuracy for the IAEA phantom was calculated for the largest-to-smallest order of registration because it demonstrated the lowest overall TRE from the precision testing. The overall accuracy for the IAEA brain phantom ( $n = 12$ ) was  $\text{TRE} = 1.71 \pm 0.62$  mm.

The overall precision (average TRE with standard deviation) for the CT/PET Fusion algorithm ( $n = 15$ ) was  $1.59 \pm 0.67$  mm. The overall precision for smallest-to-largest ( $n = 5$ ) was  $1.86 \pm 0.81$  mm (range 0.21–5.15 mm, 95% CI 1.84–1.87 mm); the overall precision for largest-to-smallest ( $n = 5$ ) was  $1.41 \pm 0.57$  mm (range 0.33–3.75 mm, 95% CI 1.41–1.42 mm); and the overall precision for random order ( $n = 5$ ) was  $1.50 \pm 0.51$  mm (range 0.32–3.50 mm, 95% CI 1.49–1.51 mm).

Table I gives the rotation and translation errors of the CT/PET Fusion system for each of the ten registrations done to evaluate the accuracy of the algorithm on the IAEA Brain Phantom. The axis of rotation passes through the center of each volume. These errors are determined by the fiducial marker system. The fiducial registration error (FRE) of the fiducial system is included for each registration as well. Based on the observed FRE, the fiducial localization error (FLE) of the fiducial markers is estimated to be 2.3 mm.<sup>16</sup>

TABLE I. The rotation and translation errors as determined by the fiducial registration system for each of the ten registrations done with CT/PET Fusion on the IAEA Brain Phantom. The axis of rotation passes through the center of each volume. The fiducial registration error (FRE) and fiducial localization error (FLE) of the fiducial system are also given. The averages are also noted.

IAEA registration	FRE (in mm)	Rotation (in deg)	Translation (in mm)
1	1.34	0.63	1.07
2	1.46	0.40	1.32
3	1.97	0.54	1.79
4	1.53	0.42	1.38
5	1.75	0.46	1.56
6	1.92	0.59	1.77
7	1.18	0.75	0.61
8	1.62	1.04	1.17
9	1.51	0.26	1.34
10	2.15	0.89	1.86
Average	1.64	0.60	1.39

Figure 2(A) demonstrates slice 17 in the IAEA brain phantom with TRE points noted in color. Figure 3(A) shows the CT/PET blended image of the same slice.

## 2. AMIR

Because AMIR is fully automated, the TREs for any given registration do not vary over multiple registrations of a given pair of images. The overall accuracy was TRE=1.13  $\pm$  0.53 mm (range 0.15–2.53 mm, 95% CI 1.11–1.15 mm).

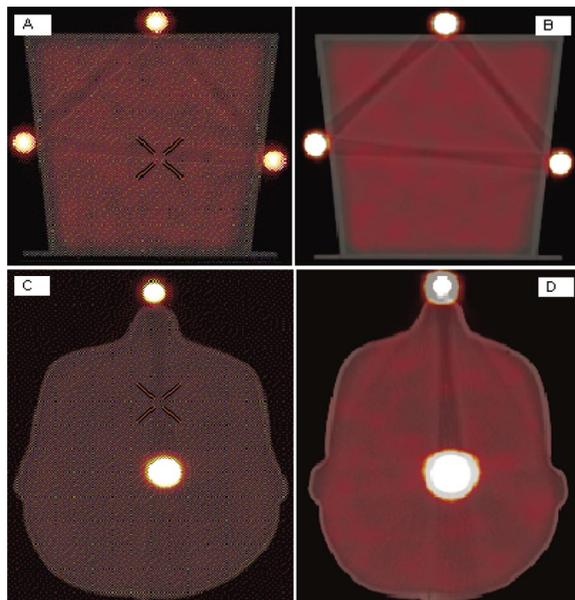


FIG. 3. (A) A blended CT/PET image (slice 17) of the IAEA brain phantom from the CT/PET Fusion program. It corresponds to the TRE grid in Fig. 2(A). (B) A blended CT/PET image (slice 17) of the IAEA brain phantom from the AMIR program. It corresponds to the TRE grid in Fig. 2(B). (C) A blended CT/PET image (slice 29) of the anthropomorphic head phantom from the CT/PET Fusion program. It corresponds to the TRE grid in Fig. 2(C). (D) A blended CT/PET image (slice 29) of the anthropomorphic head phantom from the AMIR program. It corresponds to the TRE grid in Fig. 2(D).

TABLE II. The rotation and translation errors as determined by the fiducial registration system for each of the ten registrations done with CT/PET Fusion on the anthropomorphic head phantom. The axis of rotation passes through the center of each volume. The fiducial registration error (FRE) and fiducial localization error (FLE) of the fiducial system are also given. The averages are also noted.

Head registration	FRE (in mm)	Rotation (in deg)	Translation (in mm)
1	1.62	0.42	1.43
2	1.87	0.78	1.48
3	0.87	0.17	0.74
4	1.23	0.09	1.16
5	1.09	0.11	1.00
6	1.31	0.35	1.13
7	1.11	0.36	0.90
8	1.63	0.17	1.56
9	1.45	0.39	1.29
10	0.87	0.34	0.62
Average	1.31	0.32	1.13

The rotation error was 1.45° and the translation error was 0.65 mm. For the fiducial system, the FRE was 1.99 mm; the estimated FLE was 2.81 mm.

Figure 2(B) demonstrates slice 17 in the IAEA phantom with TRE points noted in color. Figure 3(B) shows the CT/PET blended image of the same slice.

Comparison of CT/PET Fusion and AMIR Algorithms for IAEA Phantom

Utilizing repeated-measures analysis of variance as part of the general linear model, Table III demonstrates a statistically significant difference between the registration in the CT/PET Fusion program and the AMIR registration ( $P < 0.001$ ).

## B. Anthropomorphic head phantom

### 1. CT/PET Fusion

The anthropomorphic head phantom contained only a single cylindrical target “lesion,” therefore, surface structures were also used in the registration. Precision was calculated to be 1.26  $\pm$  0.52 mm (range 0.05–3.75 mm, 95% CI 1.25–1.26 mm). Overall accuracy was calculated to be TRE=1.66  $\pm$  0.53 mm.

Figure 2(C) demonstrates slice 29 in the anthropomorphic head phantom with TRE points noted in color. Figure 3(C) shows the CT/PET blended image of the same slice.

Table II gives the same errors as Table I for the registrations done on the anthropomorphic head phantom. Based on the observed FRE, the fiducial localization error (FLE) of the fiducial markers is estimated to be 1.6 mm.<sup>16</sup>

### 2. AMIR

The accuracy for AMIR was calculated to be TRE=1.15  $\pm$  0.48 mm (range 0.26–2.51 mm, 95% CI 1.15–1.16 mm). The rotation error was 0.62°; and, the translation error was 0.46 mm. For the fiducial system, the FRE was 1.07 mm; the estimated FLE was 1.31 mm

TABLE III. The accuracies of the two methods on the two phantoms are summarized. The entries are the mean and standard deviation of the target registration errors (TRE) for each method on each phantom. "AHP" is the anthropomorphic head phantom.

Phantom	CT/PET Fusion	AMIR
IAEA	$1.7 \pm 0.6$	$1.1 \pm 0.5$
AHP	$1.7 \pm 0.5$	$1.2 \pm 0.5$

Figure 2(D) demonstrates slice 29 in the anthropomorphic head phantom with TRE points noted in color. Figure 3(D) shows the CT/PET blended image of the same slice.

#### Comparison of CT/PET Fusion and AMIR Algorithms for Anthropomorphic Head Phantom

Utilizing a paired t-test for comparison of paired means, Table III demonstrates a statistically significant difference ( $P < 0.001$ ) between the CT/PET Fusion and AMIR registrations, indicating greater registration accuracy for AMIR.

### C. Patient results

#### 1. CT/PET Fusion

Figure 4(A) demonstrates the CT, PET emission and CT/PET blended images from the CT/PET Fusion program for the patient. Figure 4(B) shows a CT image (slice 77) with the TRE grid overlay from the entire image set (left) and in the region of the tumor (right).

The average TRE for the entire image set was calculated to be 7.01 mm; and the average TRE for slices which encompassed the tumor (slices 60–85) was calculated to be 6.66

mm. The regional TRE encompassing just the area around the tumor, however, demonstrated an average TRE of 3.89 mm.

### 2. AMIR

Figure 5(A) demonstrates the CT, PET emission, and CT/PET blended images from the AMIR program for the same patient. Figure 5(B) shows the same CT slice with the TRE grid overlays (entire image set-left, tumor region-right).

The average TRE was calculated to be 3.43 mm for the entire image set and 3.55 mm for slices 60–85. The TRE for the region encompassing the tumor was calculated to be 2.86 mm.

### VIII. DISCUSSION

In this work we examined the use of quantitative methods to validate the accuracy of multi-modality image registration for radiotherapy planning. Our study showed that the average registration accuracy of PET and CT images based on internal structures or mutual information algorithms was within 1.5 mm (phantom data) and 3.5 mm (patient data, region of tumor) of a fiducial-based centroid-matching method taken to be the "gold standard." Image registration accuracy on the order of  $< 0.5$  mm was achieved toward the center of each phantom, with larger inaccuracies occurring toward the periphery. In relatively rigid anatomic sites, such as head and neck, this work demonstrates that these registration algorithms may be used to facilitate incorporation of PET data into radiotherapy treatment planning to within an overall accuracy of  $< 4.0$  mm. This result agrees with an earlier evalu-

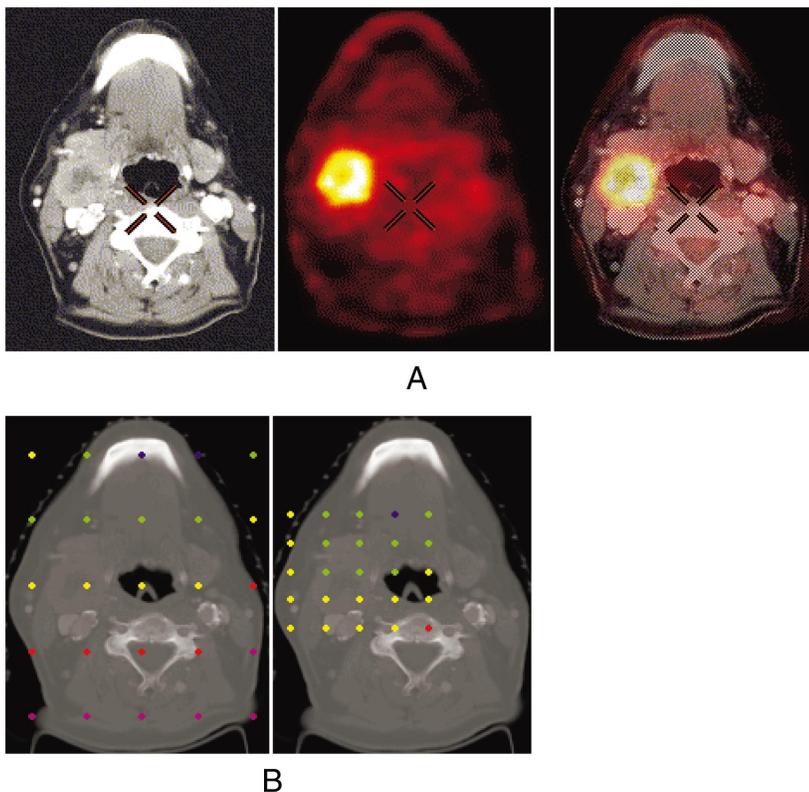


FIG. 4. (A) CT (left), PET emission (middle), and CT/PET blended (right) images of a representative slice (slice 77) of one patient from the CT/PET Fusion program. (B) Left-hand side: The TRE grid overlay from the same slice projected over the entire image set; (B) right-hand side: the TRE grid from the region of the patient's tumor. The color legend is Blue  $< 2.0$  mm, Green 2.0–4.0 mm, Yellow 4.0–6.0 mm, Red 6.0–8.0 mm, Magenta  $> 8.0$  mm.

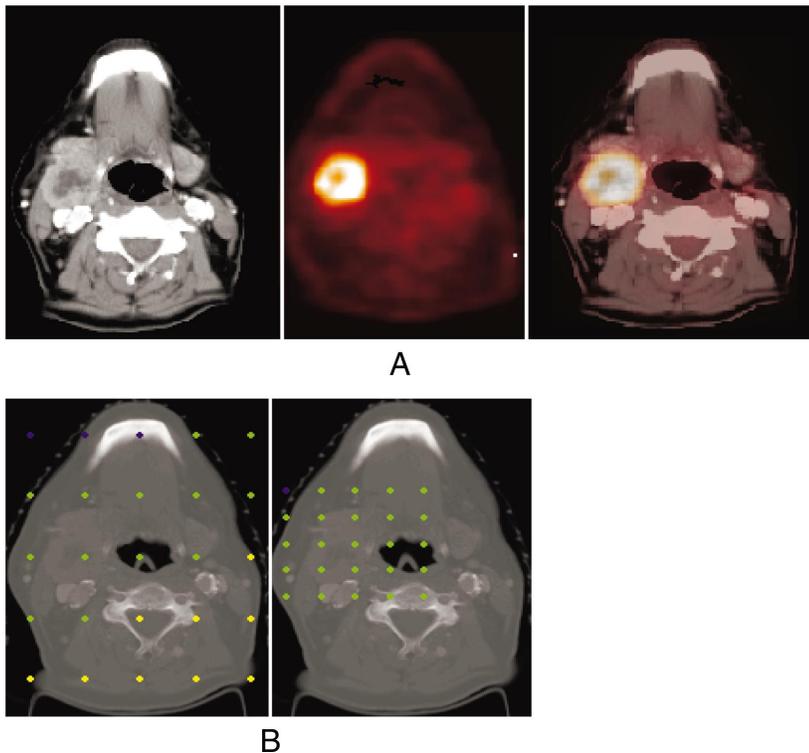


FIG. 5. (A) CT (left), PET emission (middle), and CT/PET blended (right) images of a representative slice (slice 77) of one patient from the AMIR program. (B) Left-hand side; the TRE grid overlay from the same slice projected over the entire image set; (B) right-hand side: the TRE grid from the region of the patient's tumor. The color legend is Blue < 2.0 mm, Green 2.0–4.0 mm, Yellow 4.0–6.0 mm, Red 6.0–8.0 mm, Magenta > 8.0 mm.

ation effort using a marker-based gold standard for MR/PET registration.<sup>17</sup> In that study, accuracies achieved by ten different research groups ranged from 2.5 to 5 mm. CT/PET registration accuracy may decrease, however, as distance from the location of the initial registration points (CT/PET Fusion) or the center of the image (AMIR) increases.

The fundamental premise behind the use of functional imaging modalities such as PET in RTP is that the PET data are accurately aligned with the CT data. The overall quality and accuracy of this alignment influences the definition of target contours, treatment volumes, treatment field size and angle of approach and ultimately the dose distribution and target coverage. For example, as a result of poor image registration, PET uptake may appear in nonphysical location such as air spaces. It may also be possible that PET uptake appears in normal tissue. This uptake may then be targeted leading to misplacement of radiation dose and possible damage to uninvolved tissues. The result is the targeting of radiation dose to areas where it does not belong. In both of these cases, the inclusion of PET degrades the quality of the radiation treatment. Accurate image registration is thus a unifying link between incorporation of PET data in RTP and the probability of accurate treatment and thus treatment success.

Manually selecting key anatomical landmarks via individual points as input data for the registration process is an important feature of the CT/PET Fusion SBR algorithm. Final image registration accuracy was found to depend on the location of these initial starting points and the physical size of the object within which the points were contained. We measured a higher degree of overall image registration accuracy associated with the selection of initial starting points located within large anatomical sites (e.g., large spheres) as

compared to smaller spheres. Using AMIR, we determined that image registration was more accurate toward the center or “center of mass” of the overall phantom image.

Ideally, the centroids of the CT and PET fiducials would align perfectly if identical reference points were chosen in the CT and PET emission images. This task is made difficult if not impossible primarily by the relatively poor resolution ( $\sim 5$ – $6$  mm) of the reconstructed PET image. Additionally, PET images processed without attenuation correction may contain more artifacts and could therefore lead to more dose-targeting errors. Registration would also be more difficult since boundaries may be blurred. PET with attenuation correction was used in this study since the images are considered to better visually represent where the activity truly is and to reduce artifacts, such as streaking.

It may be noted from the results of the phantom experiments that the typical FLE is of the same order as the typical TRE. It might seem necessary that the error associated with the fiducial system, which is being utilized as a standard to measure the accuracy of other methods, would need to be much less than the error to be measured. This requirement is satisfied here because the TRE of the fiducial system is much less than its FLE. While this relationship may seem paradoxical at first, it is in fact to be expected. The errors in localization of the individual fiducial markers, which lead to these FLE values, are independent in direction and therefore tend to “cancel.” As a result, their effect on errors on the determination target registration of target positions by the fiducial system (as opposed to the systems being measured) are diminished, especially when, as in the present case, the targets are located relatively close to the centroid of the marker configuration. Detailed examinations of the relation-

ship between FLE and TRE for fiducial registration systems are available.<sup>16,18</sup>

In the case of patient imaging, TRE values varied depending on location in a manner similar to that seen in the phantom studies. Although overall TRE values were greater in the patient data when compared to the phantom results, average TRE values in the region containing the tumor were, for this patient, less than the overall TRE value or TRE values based on a subset of image slices. The greater absolute error associated with the patient studies is attributed to several possible sources: (1) potential motion of the fiducial markers on the mask between PET and CT studies and/or slight differences in mask fitting between studies, (2) minor changes in soft tissue location between PET and CT scans, (3) general uncertainties associated with patient repositioning not present with the rigid phantoms, and (4) patient motion during scanning. The facemask should fit to within a few millimeters on repetitive treatments; however, there may be some fitting changes as a result of weight loss or tumor shrinkage over the course of treatment that may affect mask fitting. Further work is being undertaken to characterize CT/PET image registration in head and neck cancer patients.

There is one chief implication of registration errors for RTP: if the PET information is not accurately aligned with the CT data, then target contours drawn based on the PET data will erroneously overlap unconcerned regions of the CT image. Since ultimately these contours, including a treatment margin on the order of 5–10 mm, become the focus of large radiation doses, this situation may result in sub-optimal treatment of the intended target area with possible increased dose to intervening, normal tissues.

Image registration using PET and CT images is suitable for intensity modulated radiation therapy (IMRT). In that setting, multiple volumes can be defined by the physician. Therefore, a mass lesion can be treated to a typical prescribed dose while a metabolically active lesion on PET scan can receive a more intensified dosage. The accuracy and resolution obtained by image registration in this study is well suited for tumor subsets during IMRT.

## ACKNOWLEDGMENTS

This research was funded by National Cancer Institute Grant No. R25-CA92043-02, National Institute of Biomedical Imaging and Bioengineering Grant No. 8R01EB002124-03, and is supported in part by the Vanderbilt GCRC (RR 00095). This work supported in part by a contract from General Electric Medical Systems. Special appreciation is extended to Dawn Shone CNMT, Sara Moore CNMT, Janine Belote CNMT and Leigh Ann Wynans RT (T).

<sup>a)</sup> Author to whom correspondence should be addressed at. Department of Radiology and Radiological Sciences, Division of Nuclear Medicine, 601 N. Caroline Street, Baltimore, MD 21287; electronic mail: wlavelly1@jhmi.edu

<sup>1</sup> C. Lahorte, G. Slegers, J. Philippe, C. Van de Wiele, and R. A. Dierckx, "Synthesis and *in vitro* stability of 123I-labelled annexin V," *Biomol. Eng.* **17**, 51–53 (2001).

- <sup>2</sup> C. J. Hoekstra, S. G. Stroobants, O. S. Hoekstra, J. Vansteenkiste, B. Biesma, F. J. Schramel, N. van Zandwijk, H. van Tinteren, and E. F. Smit, "The value of [F-18]fluoro-2-deoxy-D-glucose positron emission tomography in the selection of patients with stage IIIA-N2 non-small cell lung cancer for combined modality treatment," *Lung Cancer* **39**, 151–157 (2003).
- <sup>3</sup> W. L. Wong, E. Chevetton, M. McGurk, and D. Croft, "Pet-FDG imaging in the clinical evaluation of head and neck cancer," *J. R. Soc. Med.* **88**, 469P–473P (1995).
- <sup>4</sup> C. Studholme, D. L. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," *Med. Phys.* **24**, 25–35 (1997).
- <sup>5</sup> B. M. Klabbbers, J. C. de Munck, B. J. Slotman, H. A. Langendijk, R. de Bree, O. S. Hoekstra, R. Boellaard, and A. A. Lammertsma, "Matching PET and CT scans of the head and neck area: Development of method and validation," *Med. Phys.* **29**, 2230–2238 (2002).
- <sup>6</sup> V. R. Mandava, J. M. Fitzpatrick, C. R. Maurer, Jr., R. J. Maciunas, and G. S. Allen, "Registration of multi-modal head images via attached markers," *Proc. SPIE* **1652**, 271–282 (1992).
- <sup>7</sup> 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 *Computer-assisted Radiology*, edited by H. U. Lemke, K. Inamura, C. C. Jaffe, and M. W. Vannier (Springer, Berlin, 1995), pp. 1197–1202.
- <sup>8</sup> C. R. Maurer, Jr., J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, Jr., R. J. Maciunas, and G. S. Allen, "Registration of head volume images using implantable fiducial markers," *IEEE Trans. Med. Imaging* **16**, 447–462 (1997).
- <sup>9</sup> R. Li, "Automatic placement of regions of interest in medical images using image registration," MS thesis in Electrical Engineering, Vanderbilt University, 2001.
- <sup>10</sup> W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.* **1**, 35–51 (1996).
- <sup>11</sup> F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging* **16**, 187–198 (1997).
- <sup>12</sup> C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recogn.* **32**, 71–86 (1999).
- <sup>13</sup> J. P. Pluim, J. B. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans. Med. Imaging* **19**, 809–814 (2000).
- <sup>14</sup> W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
- <sup>15</sup> 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.* **43**, 627–637 (1996).
- <sup>16</sup> J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer, Jr., "Image Registration," in *Medical Image Processing*, Handbook of Medical Imaging, Vol. II, edited by M. Sonka and J. M. Fitzpatrick (SPIE, Bellingham, WA, 2000), pp. 447–513.
- <sup>17</sup> 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. Hill, D. J. Hawkes, C. Studholme, J. B. 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 inter-modality brain image registration techniques," *J. Comput. Assist. Tomogr.* **21**, 554–566 (1997).
- <sup>18</sup> J. M. Fitzpatrick and J. B. West, "The distribution of target registration error in rigid-body, point-based registration," *IEEE Trans. Med. Imaging* **20**, 917–927 (2001).