Intensity-Based 2D-3D Spine Image Registration Incorporating One Fiducial Marker

Daniel B. Russakoff[1,2], Torsten Rohlfing[2], Ramin Shahidi[2],
Daniel H. Kim[2], John R. Adler, Jr.[2,3], and Calvin R. Maurer, Jr.[2]

[1] Department of Computer Science
[2] Image Guidance Laboratories, Department of Neurosurgery
[3] Department of Radiation Oncology
Stanford University, Stanford, CA, USA
dbrussak@stanford.edu,
calvin.maurer@igl.stanford.edu

Abstract. In this paper, we propose a hybrid similarity measure for
2D-3D image registration that is a weighted combination of an intensity-
based image similarity measure and a point-based measure incorporating
a single fiducial marker. We evaluate its accuracy and robustness using
gold-standard clinical spine image data. The use of one fiducial marker
substantially improves registration accuracy and robustness.

1 Introduction

In order to use preoperatively acquired three-dimensional (3D) images for intra-
operative therapy guidance, the images must be registered to a patient coordi-
nate system defined in the operating room. Image-to-physical registration is one
of the fundamental steps in all image-guided interventions. Surgical navigation
systems use the image-to-physical registration transformation to track in real
time the changing position of a surgical probe on a display of the preoperative
images or to direct a needle to a surgical target visible in the images. Stereotactic
radiotherapy and radiosurgery systems use the image-to-physical transformation
to direct radiation to a surgical target visible in the images.

A promising method for obtaining the image-to-physical transformation is the
registration of a 3D x-ray computed tomography (CT) image to one or more two-
dimensional (2D) x-ray projection images (e.g., fluoroscopy images, amorphous
silicon detector images). This approach has applications in image-guided spine
surgery [3,19] and radiosurgery [10,16]. The 2D-3D registration problem involves
taking one or more x-ray projection (2D) images of the patient's anatomy and
using those projections to determine the rigid transformation **T** (rotation and
translation) that aligns the coordinate system of the CT (3D) image with that of
the x-ray projection images and the operating room. Figure 1 shows a schematic
representation of the 2D-3D registration process. In general, most of the proposed
solutions to this problem fit in this framework.

For image-guided radiosurgery of spinal lesions, the best currently available
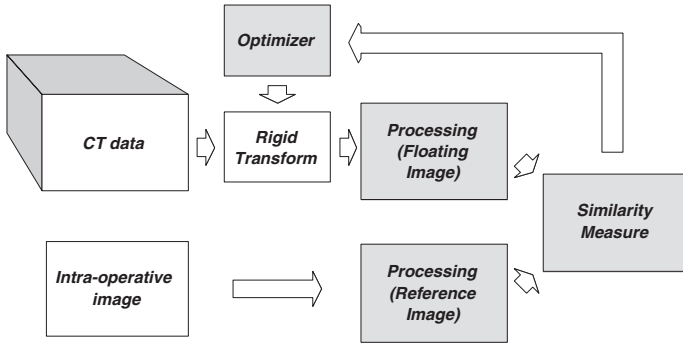2D-3D image registration approach uses metal fiducial markers implanted in

Fig. 1. Schematic overview of the 2D-3D registration process. For intensity-based 2D-3D registration, the reference image is an intra-operative x-ray projection (2D) image. It is used as is with little or no processing. The floating image is a CT (3D) image. It is processed by generating DRRs (synthetic x-ray projection images) for various orientations of the CT image relative to the x-ray imaging system. The optimizer searches for the rigid transformation **T** that produces the DRR that is most similar to the real x-ray projection image. The optimal transformation is used to align the CT coordinate system with that of the operating room.

the spine for point-based alignment. There are several fiducial-based systems in use [10,16] and under development [9]. Artificial fiducial markers, while less invasive than a rigid frame, still require a surgical implantation procedure. This entails risk, especially in the cervical spine where the vertebral structures are small and fragile.

We are particularly interested in intensity-based 2D-3D image registration [4, 11,12,19]. In this case, the reference image is one or more x-ray projection images and the floating image is a CT image. The method involves computing synthetic x-ray images, which are called digitally reconstructed radiographs (DRRs), by casting rays using a known camera geometry through the CT image. The DRR pixel values are simply the summations of the CT values encountered along each projection ray. The pose (position and orientation) of the CT image (given by the transformation **T**) is adjusted iteratively until the DRR it produces is most similar to the x-ray projection image. A variety of similarity measures have been used, including cross correlation, pattern intensity, gradient correlation, gradient difference, entropy, and mutual information [12,15]. Intensity-based 2D-3D image registration appears to be more accurate than feature-based registration (not including metal fiducial markers) [8]. But for spine images, intensity-based 2D-3D registration methods are very susceptible to local minima in the cost function and thus need initial transformations that are close to the correct transformation in order to converge reliably [12]. Also, intensity-based registration methods are basically untested on gold-standard clinical spine image data.

In this paper, we propose a hybrid similarity measure for 2D-3D image registration that is a weighted combination of an intensity-based image similarity measure and the distance of a fiducial marker. We evaluate its accuracy and robustness using gold-standard clinical spine image data.

2   Methods

2.1   2D-3D Image Registration Algorithm

The algorithm searches for the six parameters of the rigid transformation that produces the DRR (synthetic projection x-ray image) that is most similar to the real projection x-ray image. The algorithm performs four main functions corresponding to the four shaded boxes in Fig. 1: processing of the reference image, processing of the floating image, computation of a similarity measure, and optimization.

Processing of the Reference Image. We crop the reference image to include a specific region of interest (ROI). The ROI includes the anatomy that will be treated. Restricting the registration to a ROI has several advantages. First, this speeds up the registration process, as the DRRs are generated only for the ROI. Also, the similarity measure is computed only for the ROI. Second, the registration should be more accurate within the ROI. The smaller the ROI, the less likely that structures within the ROI have moved relative to each other between the time the preoperative CT is acquired and the time the procedure is performed. The definition of the ROI is performed manually and requires minimal effort. We generally specify an ROI that includes a vertebra of interest plus the two adjacent vertebra.

Processing of the Floating Image. A DRR (synthetic x-ray projection image) is generated for each transformation considered during the iterative search process. We use a light field rendering method to generate the DRRs. Using light fields allows most of the computation to be performed in a preprocessing step. After this precomputation step, very accurate DRRs can be generated quickly (about 50 ms for a 256×256 DRR). The details can be found in Ref. [14].

Similarity Measure. We perform registration using two similarity measures. The first is mutual information (MI) [5,18], which is an intensity-based similarity measure. There are two x-ray projection images and two corresponding DRRs. The similarity measure is the sum of the MI for each real-synthetic image pair. The second is a similarity measure that incorporates the position of a single fiducial marker. The hybrid similarity measure is the sum of the MI for each of the real-synthetic image pairs plus the distance d between the 3D position of the fiducial in the transformed CT image and the 3D position of the fiducial reconstructed by backprojecting its 2D positions in the x-ray images:

$$E_{\text{hybrid}} = \text{MI}_1 + \text{MI}_2 - \gamma d, \tag{1}$$

where the subscripts 1 and 2 refer to the two orthogonal real-synthetic image pairs and γ is a positive weighting factor.

Optimization Strategy. We currently use a fairly simple best neighbor search strategy similar to that in Ref. [17]. Basically the search process takes an initial transformation $\mathbf{T}_0$ as input. The twelve closest neighbors in parameter space are computed by varying each transformation parameter by some given step size. There are twice as many neighbors as parameters because the step size is both added and subtracted in order to look in both directions. Each neighbor $\mathbf{T}_i$

is itself a transformation and is evaluated by generating DRRs using $\mathbf{T}_i$ and the geometry of each x-ray camera and computing the similarity between the DRRs and the reference x-ray projection images. The neighbor with the best value of the cost function is picked, its neighbors examined, and so on until no further improvement in the value of the cost function can be made for the current step size. The process is repeated using a smaller step size until some predetermined resolution. The parameter step sizes are normalized using a scaling factor such that for a given step size, the average motion of all projected voxels in the projection plane is approximately equal for all parameters [13].

The search is performed in two passes, the first with smoothed versions of the reference images, and the second with the actual reference images. The reference images are smoothed in the first pass using a Gaussian filter with $\sigma = 1.5\,\mathrm{mm}$. This procedure has the effect of smoothing the cost function in order to help avoid local minima and to produce a good initial transformation for the second step. This is a multi-scale search strategy rather than a multi-resolution search strategy, which we have used in previous 3D-3D image registration work. Because the 2D images have a relatively limited number of pixels, we opt to blur the images and use all of the pixels rather than subsample the images in order to better estimate the joint probability density function and joint entropy. The initial step size for the first pass corresponds to an average motion of projected voxels of 5 mm. This is successively decreased to a final step size of 0.5 mm. The initial and final step sizes for the second pass are 2 mm and 0.1 mm, respectively.

During the iterative search process, we use a CT image origin that is centered in the region of interest. The selection of this origin is performed manually and requires minimal effort. For spine image registration, we generally specify an origin that is a point in the center of the vertebral body of interest.

2.2   Gold-Standard Clinical Spine Image Data

The CyberKnife Stereotactic Radiosurgery System (Accuray, Inc., Sunnyvale, CA) is an image-guided frameless robotic stereotactic radiosurgery system that was developed as a noninvasive means to precisely align treatment beams with targets [1]. Two orthogonal x-ray cameras in the treatment room establish a coordinate frame to locate the patient's target site with respect to the therapy beam directions for the robotic manipulator. A pair of images from the camera system determines the patient's position during treatment. Because the treatment position can differ from the position in the CT planning study, a 2D-3D image registration process is used to find the rigid-body transformation that relates the CT position to the treatment position. This transformation is communicated through a real-time control loop to a robotic manipulator that points a compact 6 MV x-band linear accelerator (LINAC). By taking images throughout the treatment process, shifts in patient position can be detected and the beams can be redirected accordingly.

Patients are currently undergoing treatment of spinal cord lesions with the CyberKnife through an extended FDA treatment protocol for use of the device. Before treatment, each patient is fitted with a simple immobilization de-

vice. The cervical spine patients are fitted with a molded Aquaplast face mask (WFR/Aquaplast Corp., Wyckoff, NJ) that stabilizes the head and neck on a radiographically transparent headrest. Thoracic and lumbar spine patients rest in a conformal alpha cradle during CT imaging and treatment. These supports maintain the general orientation of the anatomy and minimize patient motion.

For each patient, a contrast CT scan of the region of interest is acquired for treatment planning and also for reference in the image-guidance process. For spinal radiosurgery of thoracic and lumbar vertebrae, and most cervical vertebrae, fiducial markers are implanted percutaneously before CT scanning in the posterior bony elements of the vertebral levels adjacent to the lesions to provide radiographic landmarks. Because these implanted fiducials have a fixed relationship with the bone in which they are implanted, any movement in the vertebrae is detected and compensated by the CyberKnife. Implantation of fiducial markers occurs in the operating room under conscious sedation. The fiducial markers are 2×6 mm surgical stainless steel self-retaining tacks. Four or more fiducials are placed in a noncoplanar pattern and spaced approximately 25 mm apart. Each fiducial is implanted through stab wounds in the skin and guided with intraoperative fluoroscopy. These are implanted in the lamina or facet of the spine around the lesion of interest.

We obtained archived CyberKnife spinal image data for four patients. Two of these patients have cervical vertebrae lesions (C3 and C5) and two have thoracic vertebrae lesions (T1 and T8). For each patient, we obtained: 1) A pretreatment CT image with slice thickness 1.25 mm and a field of view sufficiently large to image the entire cross section of the body. 2) Approximately 20–30 pairs of orthogonal projection x-ray images obtained at intervals of approximately 60 seconds for the duration of treatment with the two Flashscan 20 flat-panel amorphous silicon x-ray cameras (dpiX, LLC, Palo Alto, CA). The x-ray images have 512×512 pixels with pixel size 0.4 mm and 12-bit intensity values. Only one randomly chosen pair of x-ray images is used for the work reported in this paper. 3) The camera calibration model and parameters for the two x-ray cameras. These parameters are obtained by scanning a calibration phantom as part of regular quality assurance testing. 4) Positions (3D) of the four fiducial markers in the CT image. 5) Positions (2D) of the four fiducial markers in the projection x-ray images.

2.3   Assessment of Registration Accuracy and Robustness

A gold-standard reference transformation is determined as follows. Each pair of corresponding 2D projection x-ray fiducial positions is backprojected to reconstruct the 3D fiducial position. The rays do not generally intersect. We take as the 3D coordinate the midpoint of the shortest line segment between the two rays. Then we perform a point-based registration by finding the rigid transformation that aligns the 3D fiducial positions from the CT image with the 3D backprojected fiducial positions from the x-ray images, such that the distance between corresponding points is minimized in the root-mean-square sense. The target registration error (TRE) of a registration transformation being evaluated

Table 1. 2D-3D Spine Image Target Registration Error

| Similarity | TRE (mm) | | | | | | Unsuccessful |
Measure	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Mean	Max	Registrations
MI	1.5	1.7	1.0	1.1	1.3	1.7	14%
MI plus one fiducial	1.4	1.4	0.8	0.8	1.1	1.4	1%

is computed as the difference between the positions of a target mapped by the evaluated transformation and the gold-standard transformation. The TRE values are computed for each voxel inside a rectangular box bounding the vertebra.

3   Results

Initial transformations were generated by perturbing the gold-standard reference transformation by adding randomly generated rotations and translations. The initial transformations were characterized by computing the TRE for the transformation and grouped into eight initial TRE intervals: 0–2, 2–4, 4–6, 6–8, 8–10, 10–12, and 12–14 mm. For each patient and each similarity measure, 240 registrations were performed, 30 in each of the eight misregistration intervals. The TRE value was computed for each registration transformation. The registrations were characterized as either "successful" if the TRE < 2.5 mm or "unsuccessful" if the TRE > 2.5 mm. The results are listed in Table 1. The TRE values that are listed are the mean TRE for all successful registrations. The similarity measure that incorporates the position of a fiducial marker produces slightly more accurate registrations than MI for each patient. Importantly, the hybrid similarity measure almost always produced a successful registration. Figure 2a shows how the percentage of successful registrations depends on the accuracy of the initial transformation for MI. Both similarity measures always produce successful registrations as long as the initial transformation has a TRE < 6 mm. But the robustness of MI registrations quickly decreases as the initial transformation gets further from the correct transformation.

   An important issue is the value of γ for the hybrid similarity measure in Eq. 1. There is for each patient a value of γ that produces an optimal transformation. This is illustrated for one patient in Fig. 2b. For this particular patient, the optimal value of $\gamma \sim 0.2$–0.4. We used a value of $\gamma = 0.05$, which is the average of the optimal values of γ over the four patients, but which is not the optimal value for any particular patient, including the one illustrated in Fig. 2b. The TRE values of registrations using the hybrid similarity measure are generally about 0.2–0.3 mm less than the values listed in Table 1 if the patient-specific optimal value of γ is used instead of the group average as was used. Although the accuracy is sensitive to the choice of γ, the robustness appears to be relatively insensitive.
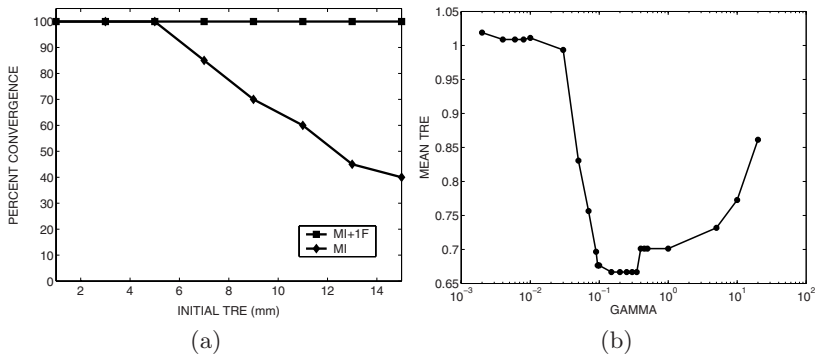
Fig. 2. (a) Percentage of successful registrations for initial transformations with different initial TRE values. Each data point represents a 2 mm range of initial TRE values. The abbreviation "MI" denotes registration performed using the intensity-based similarity measure mutual information; "MI+1F" denotes the hybrid similarity measure that incorporates the position of a fiducial marker. (b) Accuracy of registrations produced by the hybrid similarity measure for different values of γ.

4   Discussion

Previous work has reported using points to improve 3D-3D surface-based registration [6,7] and intensity-based registration [2]. This is the first work we are aware of that incorporates the position of a marker for 2D-3D image registration. Obviously the hybrid similarity measure in Eq. 1 is easily extended to include multiple points and other types of geometrical features such as contours and surfaces. It is also possible to use points other than fiducials, e.g., anatomic landmark features determined using differential operators [2].

The findings in this paper are potentially clinically relevant. When using four markers for 2D-3D registration, they must be placed relatively far apart for both accuracy and ease of identification in x-ray projection images. In practice, the markers are placed in the lamina or facets, which means they are implanted through muscle under fluoroscopic guidance. If one marker is sufficient to obtained clinically acceptable accuracy and robustness, the marker can be placed on a spinous process, which could easily be accomplished without penetrating muscle or using fluoroscopic guidance.

References

1. JR Adler Jr, *et al.* Image-guided robotic radiosurgery. *Neurosurgery* 44: 1299–1307, 1999.
2. T Hartkens, *et al.* Using points and surfaces to improve voxel-based non-rigid registration. *MICCAI 2002*, pp. 565–572, 2002.

3. S Lavallee, *et al.* Computer-assisted spinal surgery using anatomy-based registration. In: *Computer-Integrated Surgery: Technology and Clinical Applications*, RH Taylor, *et al.*, Eds., Cambridge, MA: MIT Press, pp. 425–449, 1996.
4. L Lemieux, *et al.* A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Med. Phys.* 21: 1749–1760, 1994.
5. F Maes, *et al.* Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* 16: 187–198, 1997.
6. CR Maurer Jr, *et al.* Registration of 3-D images using weighted geometrical features. *IEEE Trans. Med. Imaging* 15: 836–849, 1996.
7. CR Maurer Jr, *et al.* Registration of head CT images to physical space using a weighted combination of points and surfaces. *IEEE Trans. Med. Imaging* 17: 753–761, 1998.
8. RA McLaughlin, *et al.* A comparison of 2D-3D intensity-based and feature-based registration for neurointerventions. *MICCAI 2002*, pp. 517–524, 2002.
9. PM Medin, *et al.* Investigation of a minimally invasive method for treatment of spinal malignancies with LINAC stereotactic radiation therapy: Accuracy and animal studies. *Int. J. Radiat. Oncol. Biol. Phys.* 52: 1111–1122, 2002.
10. MJ Murphy, *et al.* Image-guided radiosurgery for the spine and pancreas. *Comput. Aided Surg.* 5: 278–288, 2000.
11. GP Penney, *et al.* Validation of a 2D-3D registration algorithm for aligning preoperative CT and intraoperative fluoroscopy images. *Med. Phys.* 28: 1024–1032, 2001.
12. GP Penney, *et al.* A comparison of similarity measures for use in 2D-3D medical image registration. *IEEE Trans. Med. Imaging* 17: 586–595, 1998.
13. T Rohlfing, *et al.* An intensity-based registration algorithm for probabilistic images and its application for 2-D to 3-D image registration. *Medical Imaging 2002: Image Processing*, Proc. SPIE 4684: 581–591, 2002.
14. DB Russakoff, *et al.* Fast calculation of digitally reconstructed radiographs using light fields. *Medical Imaging 2003: Image Processing*, Proc. SPIE 5032: 684–695, 2003.
15. DB Russakoff, *et al.* Evaluation of intensity-based 2D-3D spine image registration using clinical gold-standard data. *Workshop on Biomedical Image Registration*, 2003 (in press).
16. SI Ryu, *et al.* Image-guided hypo-fractionated stereotactic radiosurgery to spinal lesions. *Neurosurgery* 49: 838–846, 2001.
17. C Studholme, *et al.* Automated 3D registration of MR and PET brain images by multiresolution optimization of voxel similarity measures. *Med. Phys.* 24: 25–35, 1997.
18. P Viola, *et al.* Alignment by maximization of mutual information. *Int. Jour. Comp. Vis.* 24: 137–154, 1997.
19. J Weese, *et al.* Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery. *IEEE Trans. Inform. Technol. Biomedicine* 1: 284–293, 1997.