

Fast Generation of Digitally Reconstructed Radiographs Using Attenuation Fields With Application to 2D-3D Image Registration

Daniel B. Russakoff, Torsten Rohlfing, *Member, IEEE*, Kensaku Mori, *Member, IEEE*, Daniel Rueckert, *Member, IEEE*, Anthony Ho, John R. Adler, Jr., and Calvin R. Maurer, Jr.*, *Member, IEEE*

Abstract—Generation of digitally reconstructed radiographs (DRRs) is computationally expensive and is typically the rate-limiting step in the execution time of intensity-based two-dimensional to three-dimensional (2D–3D) registration algorithms. We address this computational issue by extending the technique of light field rendering from the computer graphics community. The extension of light fields, which we call attenuation fields (AFs), allows most of the DRR computation to be performed in a preprocessing step; after this precomputation step, DRRs can be generated substantially faster than with conventional ray casting. We derive expressions for the physical sizes of the two planes of an AF necessary to generate DRRs for a given X-ray camera geometry and all possible object motion within a specified range. Because an AF is a ray-based data structure, it is substantially more memory efficient than a huge table of precomputed DRRs because it eliminates the redundancy of replicated rays. Nonetheless, an AF can require substantial memory, which we address by compressing it using vector quantization. We compare DRRs generated using AFs (AF-DRRs) to those generated using ray casting (RC-DRRs) for a typical C-arm geometry and computed tomography images of several anatomic regions. They are quantitatively very similar: the median peak signal-to-noise ratio of AF-DRRs versus RC-DRRs is greater than 43 dB in all cases. We perform intensity-based 2D–3D registration using AF-DRRs and RC-DRRs and evaluate

registration accuracy using gold-standard clinical spine image data from four patients. The registration accuracy and robustness of the two methods is virtually identical whereas the execution speed using AF-DRRs is an order of magnitude faster.

Index Terms—Digitally reconstructed radiographs, image-guided therapy, intensity-based 2D–3D image registration, light fields.

I. INTRODUCTION

IN ORDER TO use preoperatively acquired three-dimensional (3-D) images for intraoperative therapy guidance, and plans (e.g., biopsy trajectory and radiation beam trajectories) constructed in the coordinate system of the preoperative image data, the images must be registered to a coordinate system defined in the operating room. One way to achieve this is to register the preoperative 3-D image to an intraoperative two-dimensional (2-D) image. Registration of an X-ray computed tomography (CT) image to one or more X-ray projection images (e.g., simulator images, portal images, fluoroscopy images, and amorphous silicon detector images) is a particularly interesting example of 2-D to 3-D (2D–3D) registration that has a number of clinical applications, including patient placement for radiotherapy planning and treatment verification [3], [4], radiosurgery [5], cranial neurosurgery [6], neurointerventions [7], [8], spinal surgery [9], [10], orthopedic surgery [11], and aortic stenting procedures [10], [12], [13].

The 2D–3D registration problem involves taking one or more X-ray projection (2-D) images of the patient’s anatomy and using those projections to determine the rigid transformation \mathbf{T} (rotation and translation) that aligns the coordinate system of the CT (3-D) image with that of the X-ray projection images and thereby the operating room. Fig. 1 shows a schematic representation of the 2D–3D registration process. In general, most of the proposed solutions to this problem fit into this general framework. We are interested in intensity-based 2D–3D image registration where the reference image is one or more X-ray projection images and the floating image is a CT image [6], [10], [13]–[21]. The method involves computing synthetic X-ray images, which are called digitally reconstructed radiographs (DRRs), by casting rays through the CT image using a known camera geometry. 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 \mathbf{T}) is adjusted iteratively until the

Manuscript received April 8, 2005; revised July 27, 2005. The work of D. B. Russakoff and C. Maurer was supported in part by the Bio-X Program at Stanford University through the Interdisciplinary Initiatives Program under the Grant “Image-Guided Radiosurgery for the Spine and Lungs.” The work of T. Rohlfing was supported in part by the National Science Foundation (NSF) under Grant EIA-0104114 “Integrating Soft Segmentation with Intensity-Based Matching for 2D/3D Image Data Registration.” The work of K. Mori was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology. Some of the material in this paper was presented at the Medical Imaging 2003 conference and the Second International Workshop on Biomedical Image Registration (WBIR 2003). The Associate Editor responsible for coordinating the review of this paper and recommending its publication was S. Aylward. *Asterisk indicates corresponding author.*

D. B. Russakoff is with the Department of Computer Science, Stanford University, Stanford, CA 94305 USA (e-mail: daniel.russakoff@cs.stanford.edu).

T. Rohlfing is with the Neuroscience Program at SRI International, Menlo Park, CA 94025 USA (e-mail: torsten@synapse.sri.com).

K. Mori is with the Graduate School of Information Science, Nagoya University, 464-8603 Nagoya, Japan (e-mail: kensaku@is.nagoya-u.ac.jp).

D. Rueckert is with the Department of Computing, Imperial College, SW7 2AZ London, UK (e-mail: d.rueckert@imperial.ac.uk).

A. Ho is with the Department of Radiation Oncology, Stanford University, Stanford, CA 94305 USA (e-mail: tonyho@reyes.stanford.edu).

J. R. Adler, Jr. is with the Departments of Neurosurgery and Radiation Oncology, Stanford University, Stanford, CA 94305 USA (e-mail: jradler@stanford.edu).

*C. R. Maurer, Jr. is with the Department of Neurosurgery, Stanford University, 300 Pasteur Drive, MC 5327, Stanford, CA 94305-5327 USA (e-mail: calvin.rmaurer@gmail.com).

Digital Object Identifier 10.1109/TMI.2005.856749

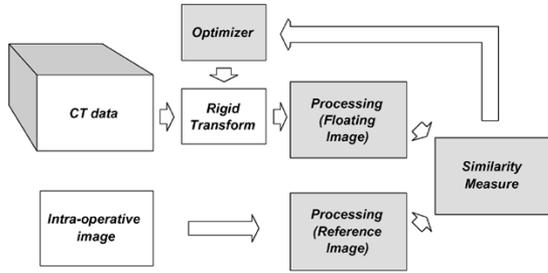


Fig. 1. Schematic overview of the 2D–3D registration process. For intensity-based 2D–3D registration, the reference image is an intra-operative X-ray projection (2-D) image. It is used as is, with little or no processing. The floating image is a CT (3-D) image. It is processed by generating DRRs (synthetic 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 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.

DRR it produces is most similar to the X-ray projection image. A variety of similarity measures have been used, including cross correlation, entropy, mutual information (MI), gradient correlation, pattern intensity, and gradient difference [19]. More recently similarity measures based on image intensity gradients have been used [18], [20].

Generation of DRRs is computationally expensive and is typically the rate-limiting step in the execution time of intensity-based 2D–3D registration algorithms. In this paper, we address this computational issue by extending the technique of light field rendering from the computer graphics community. The extension of light fields, which we call attenuation fields (AFs), allows most of the DRR computation to be performed in a preprocessing step; after this precomputation step, DRRs can be generated substantially faster than with conventional ray casting. We review light field rendering and discuss our extension to AFs for the generation of DRRs. Our AF is similar in concept to the Transgraph introduced by LaRose *et al.* [22]. This paper addresses several important issues. We derive expressions for the physical sizes of the two planes of an AF necessary to generate DRRs for a given X-ray camera geometry and all possible object motion within a specified range. An AF can require substantial memory, which we address by compressing it using vector quantization. We compare quantitatively DRRs generated using AFs (AF-DRRs) to those generated using ray casting (RC-DRRs) for a typical C-arm geometry and CT images of several anatomic regions. Finally, we perform intensity-based 2D–3D registration using AF-DRRs and RC-DRRs and evaluate registration accuracy using gold-standard clinical spine image data from four patients.

II. THEORY

A. Light Fields

Light fields were originally proposed by Levoy and Hanrahan [23]. A similar idea called the Lumigraph was simultaneously and independently presented by Gortler *et al.* [24]. Light fields were designed as a means for performing fast 3-D rendering. Essentially, light fields provide a method of parameterizing the set of all rays that emanate from a static scene. Each ray is represented by its intersection with two arbitrary planes in space

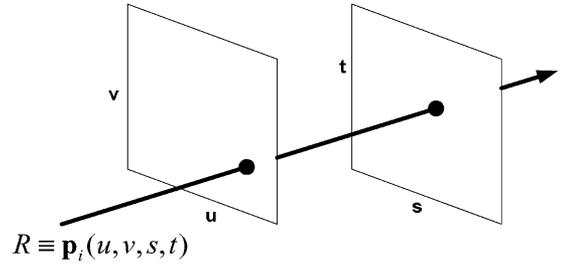


Fig. 2. Two-plane parameterization of a light field. Each ray R that emanates from a static scene is represented by its intersection with two planes in space. By convention, the coordinate system on the first plane is (u, v) and on the second plane is (s, t) . This two-plane parameterization means that each ray in space can be thought of as being represented by one point $\mathbf{p}_i = (u_i, v_i, s_i, t_i)$ in a 4-D space.

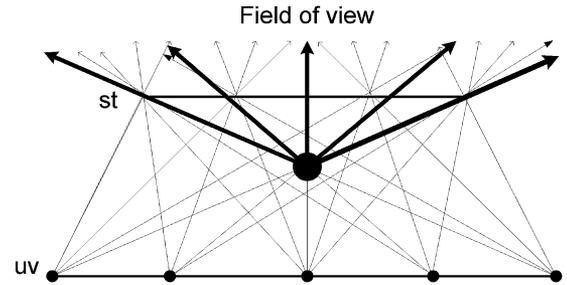


Fig. 3. View-based rendering using a light field. If one is able to calculate all rays inside a light slab, one can generate an image from a novel focal point (large solid circle inside light slab) by determining which rays are involved (thick lines) and associating them with their corresponding pixel value. Sampling the space of rays in the slab (thin lines) is a straightforward matter of generating images of the scene in question with the focal point on the (u, v) plane (small solid circles on (u, v) plane) and using the (s, t) plane as the image plane. In practice, there is a finite number of rays. Each pixel for the novel viewpoint image is calculated by computing its corresponding ray's 4-tuple (u, v, s, t) and either finding its nearest neighbor or performing quadrilinear interpolation among the neighboring samples.

(Fig. 2). By convention, the coordinate system on the first plane is (u, v) and on the second plane is (s, t) . This two-plane parameterization means that each ray in space can be thought of as being represented by one point $\mathbf{p}_i = (u_i, v_i, s_i, t_i)$ in a four-dimensional (4-D) space. As a practical matter, the planes are finite, representing a particular viewpoint of the scene. In practice $u, v, s,$ and t are restricted to values between 0 and 1 and, thus, points on each plane are restricted to lie within a convex quadrilateral. The (u, v) plane can be thought of as the focal plane and the (s, t) plane as the camera (image) plane. The shape created by connecting the (u, v) and (s, t) planes together is called a *light slab*, which represents all light that enters the restricted (s, t) plane and exits the restricted (u, v) plane.

If one is able to calculate all infinitely-many rays inside this light slab, one can recreate almost any image with a focal point inside the light slab simply by determining which rays are involved and associating them with their corresponding pixel value (Fig. 3). Practically, we cannot generate all rays in the light slab. We can, however, generate a large number of them and, provided we have reasonably dense coverage of the space, generate missing values by interpolating between existing discrete samples. Sampling the space of rays in the slab is a relatively straightforward matter of generating images of the scene in question with the center of projection (focal

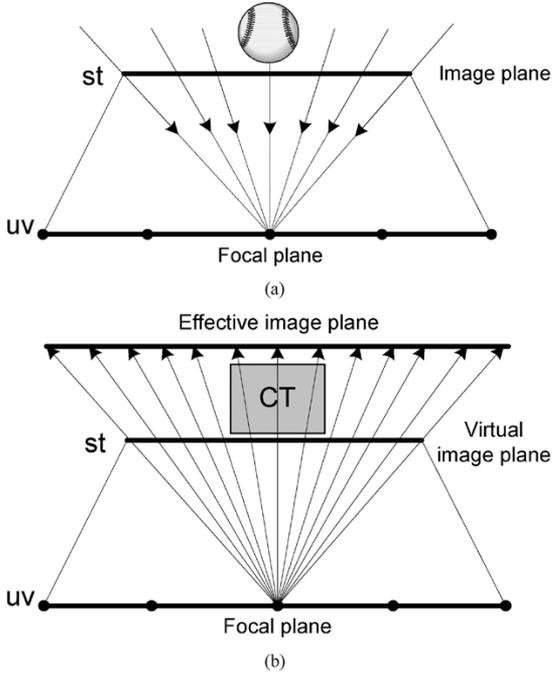


Fig. 4. (a) Standard light field. Each pixel on the image plane represents the amount of light reflected to the center of the projection from the surface its ray intersects. (b) Light field for DRR generation (AF). The virtual image plane allows the same two-plane parameterization as in standard light fields. In this case, however, the rays are extended to an effective image plane situated beyond the CT image (the scene) so that each pixel may be associated with the sum of linear attenuation coefficients along its ray.

point) on the (u, v) plane and using the (s, t) plane as the image plane (Fig. 3). These images are skewed perspective images and each one represents a 2-D slice through the 4-D space of all rays. Given a novel viewpoint within the slab, we can calculate each pixel value by computing its corresponding ray’s 4-tuple (u, v, s, t) and either finding its nearest neighbor or performing quadrilinear interpolation among the neighboring samples. These operations can be done in constant time, which greatly reduces the computational expense relative to ray casting.

B. Attenuation Fields (Light Fields for Generation of DRRs)

Light field rendering is not necessarily an obvious choice for DRR generation. A static “scene” is replaced by 3-D image data. Also, pixel values represent very different things in the two contexts. In both cases, a pixel corresponds to a ray through space. However, in traditional light field rendering, a vector-valued pixel represents the amount of light (and its spectral components, e.g., red, green, and blue channels) reflected from the first surface in the scene its ray intersects in the direction of that ray. By contrast, a scalar-valued DRR pixel is the integral of the linear attenuation coefficients its ray encounters along the path from the source to the destination.

To accommodate these differences, we modify the light field generation process by introducing the *virtual image plane* (Fig. 4). We call the extension of light fields for generation of DRRs *attenuation fields* (AFs). Essentially the virtual image plane is placed exactly where the (s, t) plane would be if

we considered the CT data to be a 3-D scene and performed standard light field rendering. The two-plane parameterization of the rays is, thus, unchanged from the normal case. In the creation of an AF, however, instead of generating images with the standard definition of pixels, we associate each sample $\mathbf{p}_i = (u_i, v_i, s_i, t_i)$ with a scalar function $\mathbf{p}_i \mapsto q(\mathbf{p}_i)$, which is the integral of the linear attenuation coefficients encountered along the ray $R_{\mathbf{p}_i}$

$$q(\mathbf{p}_i) = U_{\mathbf{p}_i} = \int_{R_{\mathbf{p}_i}} \mu(s) ds \approx \sum_{\mathbf{x}_j \in R_{\mathbf{p}_i}} \mu(\mathbf{x}_j) \Delta \mathbf{x} \quad (1)$$

where $U_{\mathbf{p}_i}$ is the total attenuation, $\mu(\mathbf{x})$ is the linear attenuation coefficient derived from the CT value at point \mathbf{x} , and $\Delta \mathbf{x}$ is the distance between discrete samples along the ray [25].¹ In standard ray casting, computation along a ray stops as soon as it intersects an opaque surface. In DRR generation by ray casting, the ray is traced through the entire CT image volume to determine its sum. To do so while maintaining the same parameterization of rays in space, we must cast the rays beyond the virtual focal plane onto the effective focal plane. The values we use to generate the light field are then those that lie on the image created on the effective focal plane. In both cases, the image created is a skewed perspective image. The main difference is that in regular light field generation the effective image plane remains fixed and between the scene and the focal plane. In AF generation, the virtual image plane remains fixed while the effective image plane can move and lies on the other side of the scene from the focal plane.

We can create an AF by generating DRRs from multiple viewpoints and then use the AF to generate DRRs from novel viewpoints using the same interpolation of 4-D ray space that Levoy and Hanrahan [23] used. Fig. 5 shows sample images used to create an AF. Although the AF is created by generating DRRs by ray casting from multiple viewpoints on the focal plane, the AF is really a ray-based data structure that stores the total attenuation for each ray in a dense set of rays. The total attenuation of each ray of a DRR from a novel viewpoint is determined by computing the ray’s 4-tuple (u, v, s, t) and either finding its nearest neighbor (total attenuation of the closest ray in the AF) or performing quadrilinear interpolation among the neighboring samples (Fig. 3). The substantial reduction in the computation time of generating DRRs using AFs relative to performing ray casting is that use of the AF essentially allows the total attenuation to be looked up in a set of precomputed values rather than computed using the integral or summation in (1). Because an AF is a ray-based data structure, it is substantially more memory efficient than a huge table of precomputed DRRs because it eliminates the redundancy of replicated rays.

¹To a good approximation, we can model the X-ray imaging process as a linear attenuation of X-rays as they pass through an object. This ignores the effects of scatter and beam hardening. The attenuation factor for ray $R_{\mathbf{p}_i}$ is $\exp(-U_{\mathbf{p}_i})$. The incident X-ray intensity is proportional to this attenuation factor and is transformed to an X-ray image intensity according to the characteristics of the imaging system and post-processing filters. Typically, this transformation includes a negative logarithm, in which case the X-ray image intensity is a linear function of $U_{\mathbf{p}_i}$. We create an AF using (1), but in principle other factors in the X-ray imaging process and post-processing transformations can be accounted for when creating the AF and/or when generating DRRs from the AF.

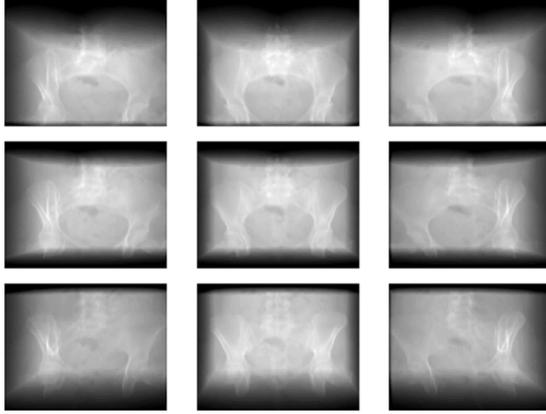


Fig. 5. Sample images used to create an AF. Each image, which is a DRR generated using conventional ray casting, represents a slice through the 4-D ray space keeping (u, v) constant. The DRRs are generated in the anterior-posterior direction from a CT image of the pelvic region. The (u, v) plane is a square with the length of a side equal to L_1 . The value of u is $-L_1/2, 0,$ and $+L_1/2$ in the left, middle, and right column images, respectively. The value of v is $-L_1/2, 0,$ and $+L_1/2$ in the bottom, middle, and top row images, respectively. Each pixel in an image corresponds to a particular value of (s, t) . The intensity gradients near the edges of these images are caused by rays that obliquely traverse the corners of a rectangular CT image.

We are interested in using AFs to generate DRRs for intensity-based 2D–3D image registration. In this case, the focal point (X-ray source) and detector are fixed and the CT image moves during the iterative registration process. This is equivalent to a fixed CT image and a moving viewpoint, and this is how the AF is used in our registration method.

We note that two of the principal problems often encountered with traditional light field rendering, occlusions and variations in illumination, are not issues in the extension to AFs for generation of DRRs.

C. Attenuation Fields: Geometry

Given a camera geometry and a range of motion of the object (CT image) relative to the camera, we need to know the physical sizes of the (u, v) and (s, t) planes necessary to generate DRRs for all possible motion within the specified range. In this section we derive an expression for the physical sizes of the (u, v) and (s, t) planes. First, we define some assumptions and terminology. Both planes are assumed to be square and parallel, with the length of a side of the (u, v) plane equal to L_1 , the length of a side of the (s, t) plane equal to L_2 , and the distance separating them given by f . We assume, without loss of generality, that the center of the (s, t) plane is at the origin and that the (u, v) plane is f units away in the positive y direction with its center at the X-ray source on the y -axis. The AF geometry and coordinate system is illustrated in Fig. 6.

The range of motion is a neighborhood of transformations about an initial transformation \mathbf{T}_0 , which places the CT image such that it is centered on the (s, t) plane (Fig. 6). Note that we are placing the (s, t) plane through the center of the initial (expected) position of the CT image. Because each pixel in the AF is associated with the sum of attenuation coefficients along its ray (Fig. 4), the placement of the (s, t) plane is arbitrary. This placement simplifies the analysis of the physical sizes of

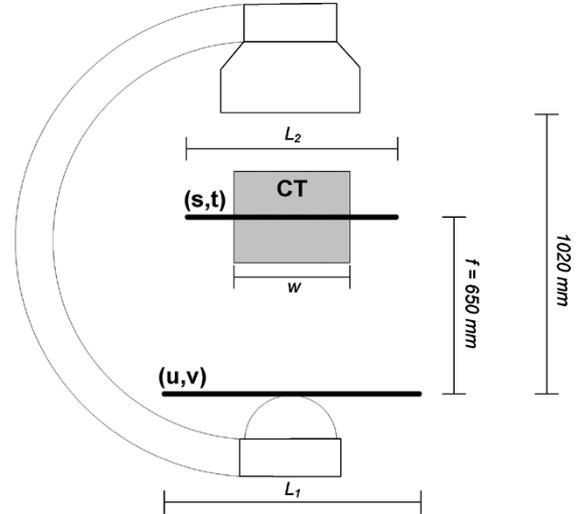


Fig. 6. The AF geometry and coordinate system for generation of DRRs similar to X-ray projection images that would be obtained with a 12-in C-arm fluoroscope with focal distance 1 020 mm. The (u, v) and (s, t) planes are assumed to be square and parallel. The length of a side of the (u, v) plane is L_1 , the length of a side of the (s, t) plane is L_2 , and the distance separating the planes is f . The (s, t) plane lies in the xz coordinate plane with its center at the origin. The (u, v) plane is f units away in the positive y direction with its center at the X-ray source on the y -axis. In this example, the distance $f = 650$ mm puts the (s, t) plane in the center of a typical CT image of a patient that would be lying on an operating room table. The CT image is assumed to be a cube with the length of a side equal to w .

the (u, v) and (s, t) planes. The neighborhood of transformations is defined by a maximum rotation Δr about each axis and a maximum translation Δt along each axis. Specifically, each transformation \mathbf{T} in the neighborhood is defined as

$$\mathbf{T} = \mathbf{T}_z^{(t)} \cdot \mathbf{T}_y^{(t)} \cdot \mathbf{T}_x^{(t)} \cdot \mathbf{T}_z^{(r)} \cdot \mathbf{T}_y^{(r)} \cdot \mathbf{T}_x^{(r)} \cdot \mathbf{T}_0 \quad (2)$$

where the superscripts (r) and (t) denote rotation and translation, respectively, and the subscripts $x, y,$ and z denote the coordinate axis. The rotations and translations are within the specified range. For example, $\mathbf{T}_x^{(r)}$ is a rotation about the x -axis where the rotation angle $r_x \in [-\Delta r, \Delta r]$.

Consider the set of all rays which can be generated by a light slab. Basically, we can generate any ray that intersects both the (u, v) plane and the (s, t) plane. It is important to notice the role played by symmetry. It suffices to consider this problem in the xy -plane since the symmetry of the construction implies that the same results must hold for the xz -plane as well. It is also instructive to observe that we need to consider only the transformations at the corners of the neighborhood defined by Δr and Δt because, if these transformations generate images from rays that intersect both planes, given that these transformations define a convex region, all intermediate transformations must also generate images from rays that intersect both planes.

We begin our analysis of the size of the (s, t) plane L_2 by starting with a camera with a fixed field of view (FOV) [Fig. 7(a)]. We define the most extreme ray as being the one that is cast farthest from the origin and, thus, is the ray closest to missing the L_2 plane. The main idea of our analysis is that if we consider the most extreme ray, and rotate and translate it so that it is cast as far from the origin as possible, we know that

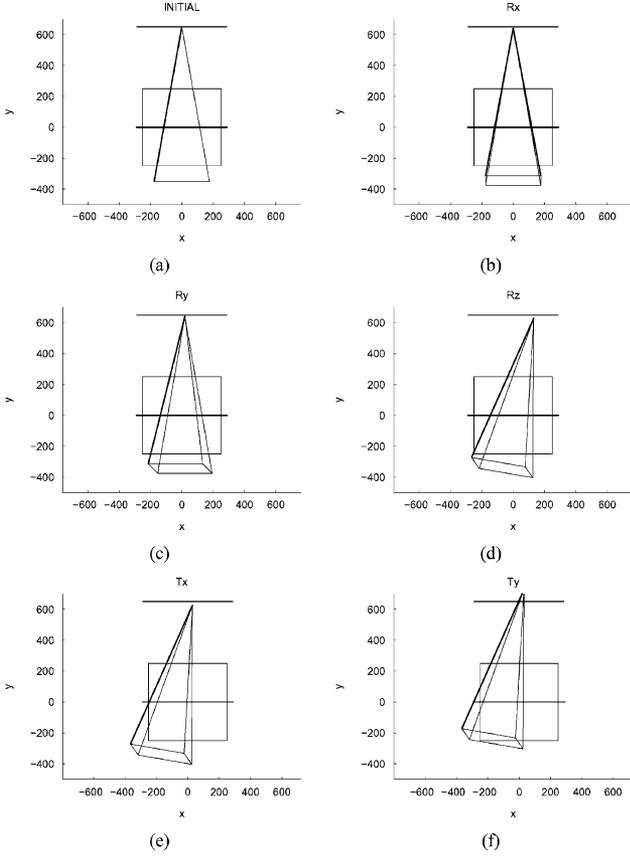


Fig. 7. Analysis of the size of L_2 . (a) Initial transformation. (b) Rotation about x -axis. (c) Rotation about y -axis. (d) Rotation about z -axis. (e) Translation along x -axis. (f) Translation along y -axis.

the point where this ray intersects the xz -plane represents the necessary extent of the L_2 plane.

In the initial configuration, all rays along the border of the viewing frustum of the camera are equally extreme. First, we rotate Δr about the x -axis as depicted in Fig. 7(b). Due to symmetry, we can rotate in either the positive or negative direction without loss of generality. As a result of this rotation, the two rays on the lower corners of the viewing frustum (shown in bold) are the most extreme. Next we rotate Δr about the y -axis [Fig. 7(c)]. This results in a single most extreme ray, which we then rotate $-\Delta r$ about the z -axis in such a way that it is cast even further from the origin [Fig. 7(d)]. Finally, we translate this ray by $-\Delta t$ along the x -axis [Fig. 7(e)] and Δt along the y -axis [Fig. 7(f)] so that it moves as far as possible from the origin. Note that a translation in the z -axis direction has no effect in this analysis, which is restricted to the xy -plane. The necessary length of the (s, t) plane L_2 is determined by observing where this extreme ray intersects the xz -plane.

Finally, we determine L_2 by computing where the extreme ray intersects the xz -plane. Let \mathbf{T}_e be the concatenation of all the transformations in this analysis, where the transformation is represented as a 4×4 homogeneous matrix, and \mathbf{R}_e be the 3×3 rotation matrix component of \mathbf{T}_e . We define $\hat{\mathbf{T}}_e$ and $\hat{\mathbf{R}}_e$ as the 2×4 and 2×3 truncations of \mathbf{T}_e and \mathbf{R}_e , respectively, that ignore the z component. Let $\mathbf{f} = (0, f, 0, 1)^t$ be the initial focal point and $\mathbf{e} = (-\tan(\alpha), -1, -\tan(\alpha))^t$ be a vector

that describes the direction of the most extreme ray in the analysis above in the initial configuration. The angle α is one-half of the FOV angle. The transformation in this analysis maps \mathbf{f} and \mathbf{e} to $\hat{\mathbf{T}}_e \mathbf{f}$ and $\hat{\mathbf{R}}_e \mathbf{e}$, respectively, where the z component is ignored after transformation. The line defined by the most extreme ray after transformation intersects the (s, t) plane at $\mathbf{l} = (-L_2/2, 0)^t$. Thus, we determine L_2 by solving

$$\mathbf{l} = \hat{\mathbf{T}}_e \mathbf{f} + \lambda \hat{\mathbf{R}}_e \mathbf{e} \quad (3)$$

After concatenating the transformations and performing some manipulation and simplification, we arrive at the following expression for L_2 :

$$L_2 = \frac{n_2}{d_2} \quad (4)$$

where

$$\begin{aligned} n_2 = & 2(S^5 T f + C^2 S^2 T f + 2C^2 S^3 T f + C^4 T f \\ & + C^4 S T f - S^2 T \Delta t - S^3 \Delta t + C S \Delta t \\ & - 2C S T \Delta t + C S^2 \Delta t - C S^2 T \Delta t + C^2 \Delta t \\ & + C^2 T \Delta t + C^2 S T \Delta t) \end{aligned} \quad (5)$$

and

$$d_2 = -S^3 - 2C S T - C S^2 T + C^2. \quad (6)$$

In these equations, the variables C, S , and T represent the trigonometric functions $C = \cos(\Delta r)$, $S = \sin(\Delta r)$, and $T = \tan(\alpha)$. If α and Δr are relatively small, then the expression for L_2 can be reduced using small-angle trigonometric approximations to

$$L_2 \approx 2(\alpha f + (1 + \alpha + \Delta r)\Delta t). \quad (7)$$

The analysis of the size of the (u, v) plane L_1 is similar to the analysis of L_2 above. We determine L_1 by solving (3). One important difference is that the translations of \mathbf{T}_e are in the opposite direction (in this case Δt along the x -axis and $-\Delta t$ along the y -axis). Also, the line defined by the most extreme ray after transformation intersects the (u, v) plane at $\mathbf{l} = (L_1/2, f)^t$. The resulting expression for L_1 is

$$L_1 = \frac{n_1}{d_1} \quad (8)$$

where the expression for d_1 is identical to that for d_2 in (6) and

$$\begin{aligned} n_1 = & 2(-S^2 T f - S^5 T f + C S f + C S^2 f + C^2 T f \\ & + C^2 S T f - C^2 S^2 T f - 2C^2 S^3 T f - C^4 T f \\ & - C^4 S T f - S^2 T \Delta t - S^3 \Delta t + C S \Delta t \\ & - 2C S T \Delta t + C S^2 \Delta t - C S^2 T \Delta t + C^2 \Delta t \\ & + C^2 T \Delta t + C^2 S T \Delta t). \end{aligned} \quad (9)$$

The expression for L_1 can be reduced using small-angle trigonometric approximations to

$$L_1 \approx 2(\Delta r f + (1 + \alpha + \Delta r)\Delta t). \quad (10)$$

D. Attenuation Fields: Memory

An important issue is the size of the AF. To effectively sample a 4-D space, one needs a large number of images and, thus, an AF can require substantial memory (Table I). For example, an

TABLE I
MEMORY SIZE AND CREATION TIME OF AFS

(u, v) Resolution (pixels)	(s, t) Resolution (pixels)	Memory Size (MB)		Creation Time (hours)
		Without Compression	With Compression	
32×32	256×256	128	8	0.4
32×32	512×512	512	32	1.7
64×64	256×256	512	32	1.7
64×64	512×512	2,048	128	6.8

The memory size of an AF was computed assuming each pixel represents attenuation with two bytes. The compression ratio was computed using Eq. 11. Using codewords that each represent a $2 \times 2 \times 2 \times 2$ -element tile of voxels in (u, v, s, t) space ($k = 16$ elements per vector), two bytes per element ($b = 2 \times 8 = 16$ bits per element), and a codebook consisting of up to 65,536 codewords each represented (without using bit packing) by two bytes ($\log_2(N) = 16$), the compression ratio is 16. The time required to create an AF was computed for a straightforward and relatively unoptimized implementation of ray casting. Using a PC workstation with one 2.2 GHz Intel Xeon processor, this implementation takes about 1.5 s to generate a 256×256 -pixel DRR from a $512 \times 512 \times 512$ -voxel CT image.

AF that represents attenuation with two bytes, has a (u, v) resolution of 64×64 pixels, and has a (s, t) resolution of 512×512 pixels requires 2 GB memory, which is more memory than typical workstations have. For a system with an orthogonal pair of X-ray cameras, two AFs are required, one for each camera, thus doubling the required memory. For a rotational X-ray system such as a C-arm fluoroscope, several AFs might be required. Fortunately, there is a great deal of redundancy in the data. There is redundancy in s and t corresponding to interpixel similarity and redundancy in u and v corresponding to interimage similarity. These redundancies can be exploited by compressing the data using vector quantization.

Vector quantization is a lossy compression technique wherein a vector of samples is quantized to one of a number of predetermined reproduction vectors [26]. A reproduction vector is called a codeword and the set of codewords available to encode a source is called the codebook. The codebook is constructed during a training phase. The training phase finds a set of codewords that best approximates a set of sample vectors, which is called the training set. We perform training on a subset of the AF rather than the entire AF to reduce the computation time and memory requirements of the training phase. After the codebook is constructed, encoding of the AF is performed by partitioning the AF into vectors and finding for each vector the closest codeword in the codebook. The encoded AF is stored as a set of codebook indices. Decoding of the AF is performed simply by looking up the the indices in the codebook and outputting the codewords corresponding to the indices. An advantage of vector quantization for our application is that decoding is a look-up procedure and, thus, is very fast. Depending on the size of the codebook and codevectors and the redundancy in the data, one can achieve large compression ratios. We partition the AF into $2 \times 2 \times 2 \times 2$ -element tiles, which are the vectors that are quantized. The use of 16-element tiles is possible because of the redundancy in the data. The compression ratio is

$$\text{Compression Ratio} = \frac{kb}{\log_2 N} \quad (11)$$

where k is the number of elements per vector, b is the number of bits per element, and N is the number of codewords. The

numerator (kb) is the number of bits per vector; the denominator ($\log_2 N$) is the number of bits per codeword. Using codewords that each represent a $2 \times 2 \times 2 \times 2$ -element tile of voxels in (u, v, s, t) space ($k = 16$ elements per vector), two bytes per element ($b = 2 \times 8 = 16$ bits per element), and a codebook consisting of up to 65 536 codewords each represented (without using bit packing) by two bytes ($\log_2 N = 16$), the compression ratio is 16.

III. EXPERIMENTS

A. Generation of DRRs Using Attenuation Fields

We generate DRRs computed using AFs (AF-DRRs) and compare these quantitatively to those generated using conventional ray casting (RC-DRRs). The RC-DRRs were computed using the summation approximation in (1) with $\Delta x = 0.1$ mm. Attenuation coefficients at the discrete sample points were estimated from the nearest voxels using trilinear interpolation. The important geometrical parameters are the angular FOV of the X-ray camera system and the maximum range of motion of the object relative to the camera. For these experiments we generate DRRs similar to X-ray projection images that would be obtained with a typical 12-inch C-arm fluoroscope. The AF geometry for the fluoroscope is shown in Fig. 6. The angular FOV = 17° ($\alpha = \text{FOV}/2 = 8.5^\circ$). The (u, v) plane is centered at the X-ray source. The (s, t) plane is parallel to the (u, v) plane and the distance between the planes is $f = 650$ mm. This distance puts the (s, t) plane in the center of a typical CT image of a patient that would be lying on an operating room table. We define the range of motion of the object (CT image) relative to the camera by a maximum rotation $\Delta r = 10^\circ$ about each axis and a maximum translation $\Delta t = 100$ mm along each axis. Using the expressions derived in Section II-C, the physical size of the (u, v) plane is $L_1 = 560$ mm and the size of the (s, t) plane is $L_2 = 525$ mm.

There are four parameters of an AF. Two of the parameters are the resolutions of the (u, v) and (s, t) planes. The other two parameters are related to compression of the AF. We compress the AF using vector quantization, so one of the compression parameters is the number of codevectors. Similar to Levoy and Hanrahan [23], we use a subset of the pixels for training, so the final parameter is the training set size (percentage of samples). In order to evaluate the importance of these four parameters on the quality of AF-DRRs, we created AFs using different values of the parameters. We created AFs using CT images of three anatomic regions—the skull, thorax, and pelvis. The pixel size of the CT images was 0.5–0.8 mm; the slice thickness was 1.0, 1.25, and 2.0 mm for the skull, thorax, and pelvis CT images, respectively. For each anatomic region and set of parameters, we created two AFs for different camera orientations, one for the anterior-posterior (AP) direction and one for the lateral direction. For each AF, we generated 600 AF-DRRs with resolution 256×256 pixels, 100 for each anatomic region and orientation, using random camera poses within the specified range of motion ($\Delta r = 10^\circ$, $\Delta t = 100$ mm). For each AF-DRR, a corresponding DRR was generated using conventional ray casting.

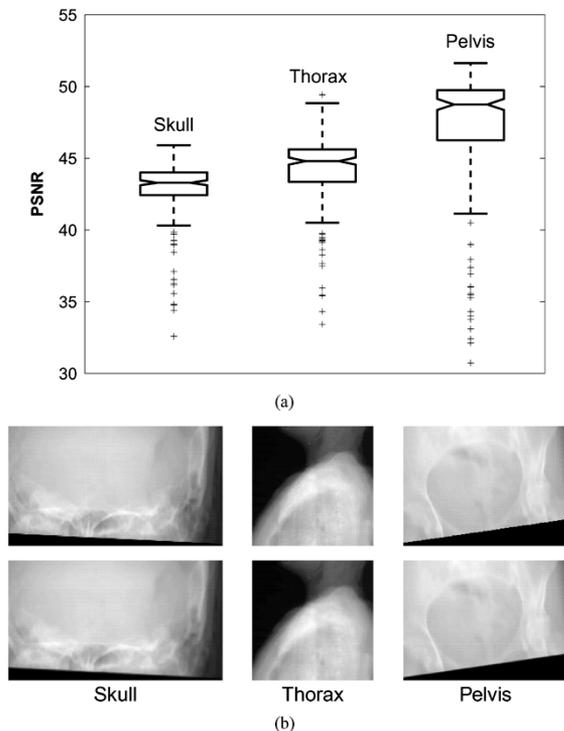


Fig. 8. (a) Box-and-whisker plot of PSNR values for AF-DRR images relative to corresponding RC-DRR images. Six AFs were generated, two orientations (AP and lateral) for each of three anatomic regions (skull, thorax, and pelvis). The AF parameters were: (u, v) plane resolution = 64×64 pixels, (s, t) plane resolution = 256×256 pixels, number of codevectors = 16 384, and training set size = 10% of samples. For each orientation and anatomic region, 100 AF-DRRs were generated from the corresponding AF using random camera positions and orientations within the specified range of motion (maximum rotation of 10° about each axis, maximum translation of 100 mm in each direction). For each AF-DRR, a corresponding RC-DRR was generated and the PSNR value between the two types of DRR images was computed. The central line in each box is the median PSNR value over all 200 DRRs for that anatomic region, the lower and upper edges of the box are the 25th and 75th percentile values, and the length of the vertical dashed lines are 1.5 times the interquartile range. Outliers, which are values that fall outside of 1.5 times the interquartile range, are designated by a cross (“+”). The median PSNR values are greater than 43 dB, almost all of the PSNR values are greater than 40 dB, and very few are less than 36 dB. (b) DRR images corresponding to the lowest PSNR value for each anatomic region. The top row are RC-DRRs; the bottom row are corresponding AF-DRRs. Qualitatively, even the image pairs associated with the lowest PSNR values are very similar.

The AF-DRRs are compared quantitatively to the corresponding RC-DRRs by computing the peak signal-to-noise ratio (PSNR)

$$\text{PSNR} = 20 \log_{10} \frac{S}{\sqrt{N}} \quad (12)$$

where S is the maximum pixel value and N is the root-mean-square pixel value difference between the two images. We ignore background pixels when computing PSNR because such pixels artificially inflate the value of PSNR.

We picked a set of AF parameters that produced AF-DRRs with median PSNR values above 43 dB, almost all PSNR values above 40 dB, and few PSNR values below 36 dB. The AF parameters are: (u, v) plane resolution = 256×256 pixels, (s, t) plane resolution = 64×64 pixels, number of codevectors = 16 384, and training set size = 10% of samples. The distribution of PSNR values for the three anatomic regions is shown in Fig. 8(a). There are few visible

differences between images with PSNR values above 30 dB [Fig. 8(b)]. The image compression literature considers a PSNR value above 36 dB to represent excellent image quality [27]. Difference images between AF-DRRs and corresponding RC-DRRs are shown for PSNR values of 36, 40, and 44 dB in Fig. 9. In all cases the pixel value differences are small. The difference image artifacts tend to correlate with image intensity gradients (object edges). There is a substantial reduction in image artifacts as the PSNR value increases from 36 to 44 dB.

The influence of each AF parameter is shown in Fig. 10. In each plot, one of the four AF parameters was varied. The resolutions of the (u, v) and (s, t) planes are the most important AF parameters.

An AF with (u, v) plane resolution = 256×256 pixels and (s, t) plane resolution = 64×64 pixels, which produces excellent AF-DRRs (Fig. 8), requires a memory size of 32 MB (Table I). Using a straightforward and relatively unoptimized implementation of ray casting, we can generate a 256×256 -pixel RC-DRR from a $512 \times 512 \times 512$ -voxel CT image in approximately 1.5 s on a PC workstation with one 2.2 GHz Intel Xeon processor. Thus, we can create an AF with (u, v) plane resolution = 256×256 pixels and (s, t) plane resolution = 64×64 pixels in approximately 1.7 h (Table I). Once the AF has been created, we can generate a 256×256 -pixel AF-DRR on the same PC workstation in approximately 50 ms.

B. Two-Dimensional to Three-Dimensional Registration Using AF DRRs

1) *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 [28]. 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 (Fig. 11). A pair of images from the camera system determines the patient’s position during treatment (Fig. 12). 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 Food and Drug Administration treatment protocol for use of the device [29], [30]. After informed consent is obtained and documented, a thorough medical history is obtained, and a physical examination and neurological evaluation are performed. Before treatment, each patient is fitted with a simple immobilization device. 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.

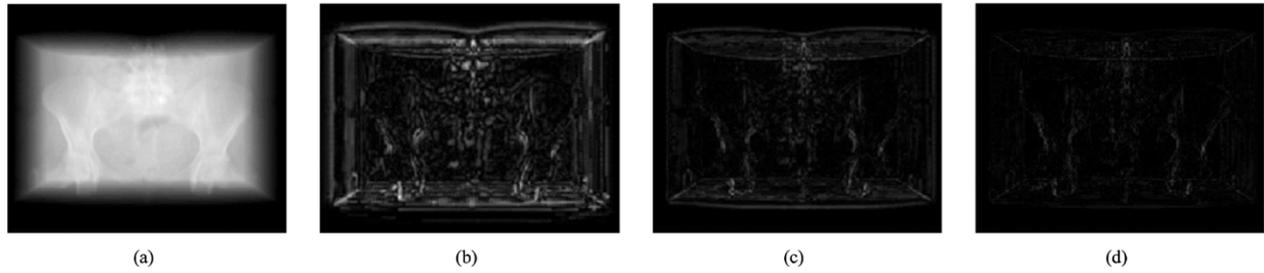


Fig. 9. (a) Sample AF-DRR image generated in the AP direction from a CT image of the pelvis. (b)–(d) Difference images between AF-DRR image and corresponding RC-DRR image with PSNR values of 36 dB (b), 40 dB (c), and 44 dB (d). The different PSNR values were obtained by generating AF-DRR images from AFs created with different resolutions. The difference images are displayed with a different window/level setting than the sample image to emphasize the pixel value differences, but all difference images are displayed with the same window/level setting to illustrate that there is a substantial reduction in image artifacts as the PSNR value increases from 36 to 44 dB.

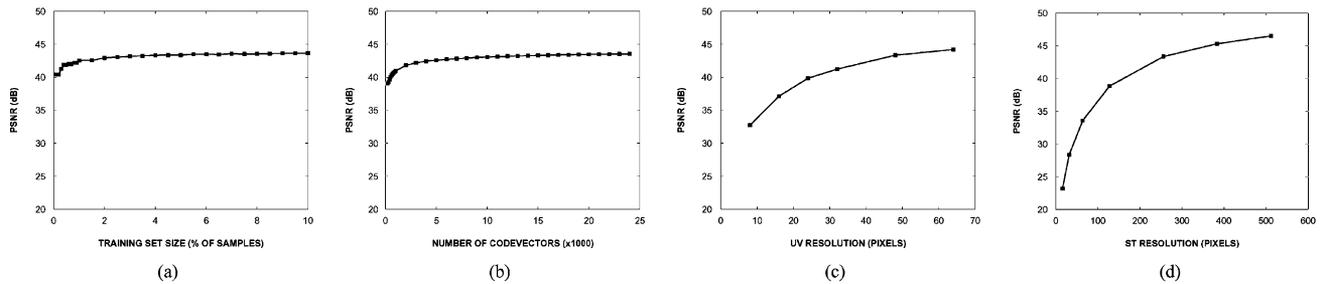


Fig. 10. Plots of PSNR values versus AF parameters. AFs were created for two orientations (antero-posterior and lateral) and each of three anatomic regions (skull, thorax, and pelvis). For each orientation and anatomic region, AFs were created using different parameters. For each set of parameters, 600 AF-DRRs (100 AF-DRRs for each orientation and anatomic region) were generated using random camera positions and orientations within the specified range of motion (maximum rotation of 10° about each axis, maximum translation of 100 mm in each direction). For each AF-DRR, a corresponding RC-DRR was generated and the PSNR value between the two types of DRR images was computed. The initial AF parameters were: (u, v) plane resolution = 64×64 pixels, (s, t) plane resolution = 256×256 pixels, number of codevectors = 16 384, and training set size = 10% of samples. In each plot, one of the four AF parameters was varied. Each symbol represents the mean PSNR value for the 600 AF-DRRs generated with that set of AF parameters.

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 (ROI) 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 (generally C3 and below), 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 of 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. Three noncollinear fiducials are required to define a rigid-body transformation. Four fiducials provide redundancy in the event that one of them is obscured or otherwise difficult to image. Each fiducial is implanted through stab wounds in the skin and guided with intraoperative fluoroscopy; they are implanted in the lamina or facet of the spine around the lesion of interest. No complications have been reported from this procedure, and all patients have been discharged home the same day.

We obtained 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 FOV 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 s for the duration of treatment with the two Flashscan 20 flat-panel amorphous silicon detector (ASD) 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 (they are stored as 16-bit integers, but the dynamic range is 12 bits). Only one randomly chosen pair of X-ray images per patient 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 (3-D) of the four fiducial markers in the CT image. 5) Positions (2-D) of the four fiducial markers in the projection X-ray images.

2) *Two-Dimensional Three-Dimensional Image Registration Algorithm*: The algorithm searches for the six parameters of the rigid transformation that produces DRRs (synthetic projection X-ray images) that are most similar to the real projection X-ray images. The algorithm performs four main functions corresponding to the four shaded boxes in Fig. 1: processing of

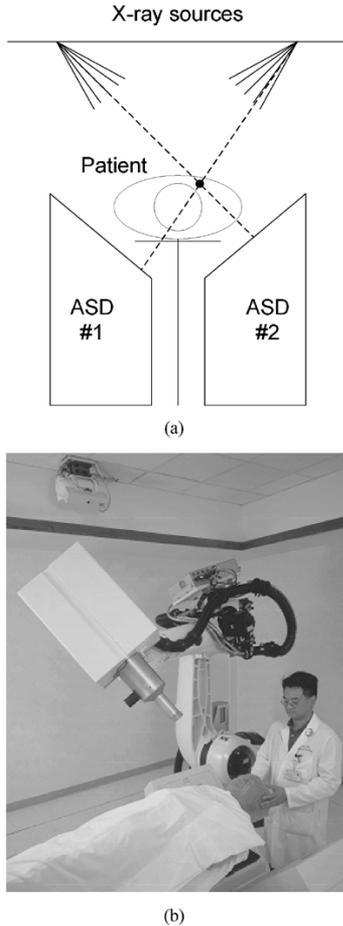


Fig. 11. (a) Schematic illustration of the CyberKnife Stereotactic Radiosurgery System (Accuray, Inc., Sunnyvale, CA), which is an image-guided frameless robotic stereotactic radiosurgery system that was developed as a noninvasive means to precisely align treatment beams with targets [28]. Two X-ray sources are mounted on the ceiling. Two ASD X-ray cameras are mounted on the floor. The orthogonal X-ray projection imaging system is calibrated so that the intrinsic and extrinsic camera parameters of both imaging devices are known. The orthogonal X-ray camera pair establishes 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. The dashed lines indicate the triangulation of the 3-D position of a fiducial marker that appears in both ASD X-ray projection images. (b) Photograph of the CyberKnife system with a patient lying on the treatment table. One of the ceiling mounted X-ray sources, both of the ASD X-ray cameras, and the robot with the LINAC are visible in this picture.

the reference image, processing of the floating image, computation of a similarity measure, and optimization. In summary, at each step of an iterative search process, the algorithm generates DRRs for each ASD X-ray camera viewpoint from the current pose of the CT image, measures the similarity of the real and synthetic X-ray images by computing the value of a similarity measure, and finds a new pose of the CT image that increases the value of the similarity measure. This process is repeated until no further improvement can be made.

Processing of the reference image. We use the ASD X-ray projection images as the reference images. We crop each reference image to include a specific ROI (Fig. 13). The ROI includes the anatomy that will be treated. Restricting the registration to

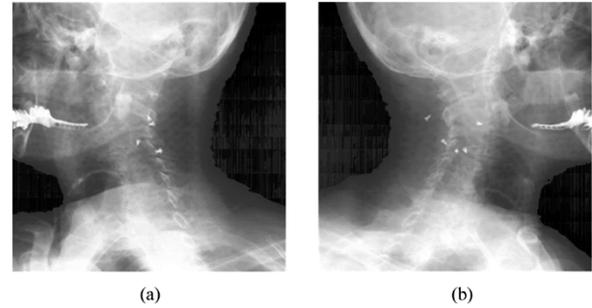


Fig. 12. A sample pair of images from the two ASD X-ray cameras in the CyberKnife radiosurgery system (Fig. 11). In these images, the implanted metal fiducial markers appear as small white objects in the cervical vertebrae near the base of the skull.

a ROI has several advantages. First, this speeds up the registration process. In particular, DRRs are computationally expensive to create, and their generation is typically a bottleneck in the execution of the registration process. 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 vertebrae. The size of the ROI is typically about 200×200 pixels.

Processing of the floating image: At each step of the iterative search process, DRRs are generated for each ASD X-ray camera viewpoint from the current transformation (pose) of the CT image. We generate AF-DRRs using two AFs, one for each of the X-ray cameras. The AF parameters are: (u, v) plane resolution = 256×256 pixels, (s, t) plane resolution = 64×64 pixels, number of codevectors = 16 384, and training set size = 10% of samples.

Similarity measure. We perform registration using the similarity measure MI [31], [32], which is an intensity-based image similarity measure that is commonly used in medical image registration [33] and has been used in 2D–3D image registration [13], [19], [21]. There are two X-ray projection images and two corresponding DRRs. The MI similarity measure is the sum of the MI for each real-synthetic image pair

$$E_{MI} = MI_1 + MI_2 \quad (13)$$

where the subscripts 1 and 2 refer to the two orthogonal real-synthetic image pairs.

Optimization strategy: We currently use a fairly simple best neighbor search strategy similar to that in [34]. Basically the search process takes an initial transformation 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 T_i is itself a transformation and is evaluated by generating DRRs using T_i and the geometry of each X-ray camera and computing the similarity between the

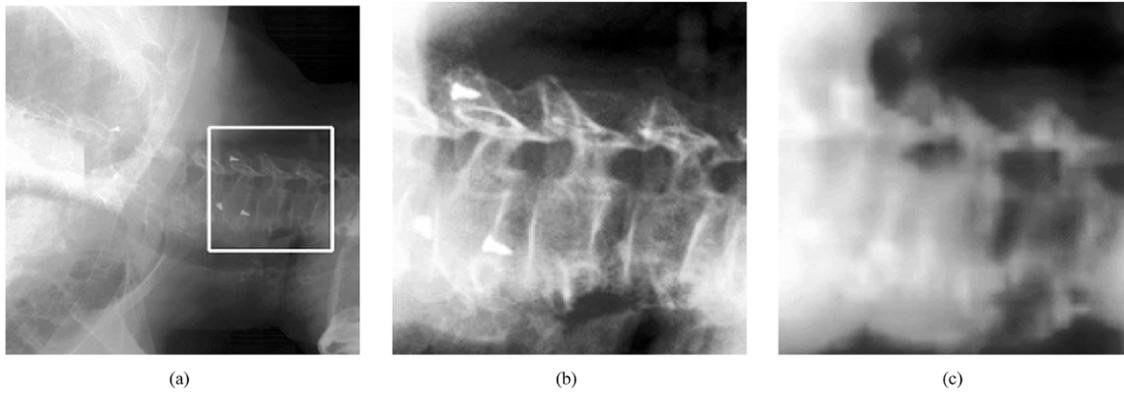


Fig. 13. (a) Example ASD X-ray projection image. The box delineates the manually selected ROI. We generally select an ROI that includes a vertebra of interest plus the two adjacent vertebrae. Three bone-implanted fiducial markers are visible in the ROI. (b) A magnified view of the selected ROI. The window/level settings in (a) and (b) are different. (c) An AF-DRR corresponding to the ROI in (b) generated from a registered CT image of the same patient using the camera geometry of the X-ray imaging system used to acquire the image in (a).

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 [35].

The search is performed hierarchically (coarse-to-fine) 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 $\sigma = 2$ mm. This has the effect of smoothing the cost function in order to help avoid local optima and to produce a good initial transformation for the second step. This approach 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 2-D 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, which are used to compute MI (the reader is referred to [33], [34] for details about how MI is computed from 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. Using an origin that is centered in the region of interest helps decouple the rotation and translation parameters during the iterative search.

3) *Assessment of Registration Accuracy and Robustness:* A gold-standard reference transformation is determined using the implanted metal fiducial markers. Each pair of corresponding

2-D projection X-ray fiducial positions is backprojected to reconstruct the 3-D fiducial position. The rays do not intersect exactly. We take as the 3-D coordinate the midpoint of the shortest line segment between the two rays. The length of this shortest line segment was <0.2 mm for most of the markers in this study. Then we perform a point-based registration by finding the rigid transformation that aligns the 3-D fiducial positions from the CT image with the 3-D backprojected fiducial positions from the X-ray images, such that the distance between corresponding points is minimized in the root-mean-square sense. This point-based registration problem has several closed-form solutions [36]. We use the method of Arun [37], which is equivalent to the first solution published by Schönemann [38].

We calculate the registration error of the transformations being evaluated by using as a reference gold standard the fiducial marker-based transformation. Specifically, we calculate the target registration error (TRE) of the registration transformation being evaluated as the distance between the position of a target point in CT mapped to physical space by the gold-standard reference transformation and its position mapped by the transformation being evaluated. The analysis is similar to the ones performed in [39]–[41]. West *et al.* [39] reported summary statistics of TRE at ten defined anatomic locations widely distributed in the head. In this paper we compute TRE for each registration as the mean value of TRE at all voxels inside a rectangular box bounding the vertebra of interest.

4) *Results:* We examined the shape of the image similarity cost function by varying the transformation parameters about the gold-standard transformation, one parameter at a time. Fig. 14 shows the results for the translation parameters for one patient. The value of MI is slightly higher using AF-DRRs than RC-DRRs, but the smoothness, shape, and location of the maximum are nearly identical for both types of DRRs.

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, 12–14, and 14–16 mm. For each patient and each

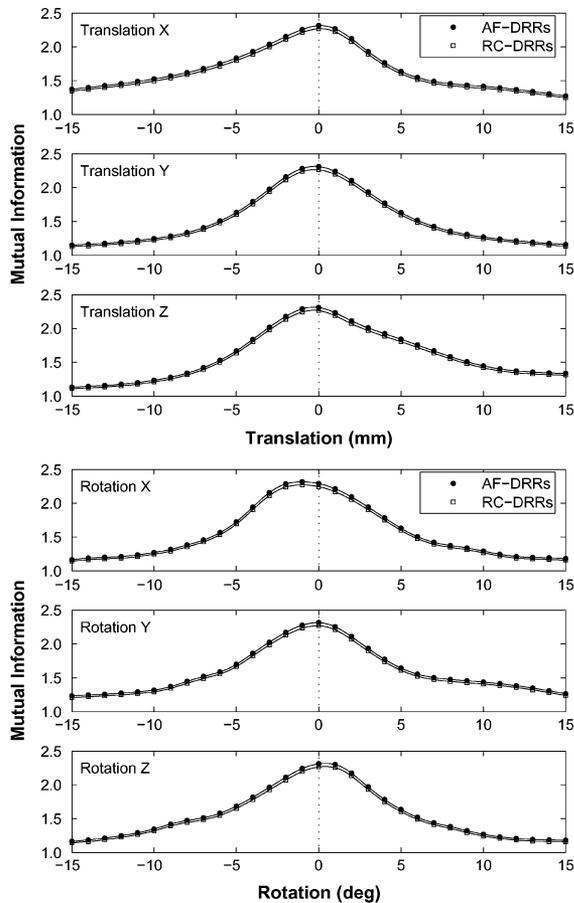


Fig. 14. Image similarity between a pair of real, clinical ASD images and DRRs generated from the corresponding CT image after translation (top three plots) and rotation (bottom three plots). The image similarity metric is MI. The translation and rotation parameters were varied about the gold-standard transformation, one parameter at a time, in steps of 0.1 mm and 0.1°, respectively. The curves are labeled with symbols every 1 mm and 1°, respectively. The solid circles represent results obtained using AF-DRRs; the open squares represent results using RC-DRRs. The AF parameters were: (u, v) plane resolution = 64×64 pixels, (s, t) plane resolution = 256×256 pixels, number of codevectors = 16384, and training set size = 10% of samples. The maxima occur at $\Delta t_x = 0.1$ mm, $\Delta t_y = -0.2$ mm, $\Delta t_z = -0.4$ mm, $\Delta r_x = -1.2^\circ$, $\Delta r_y = -0.1^\circ$, and $\Delta r_z = 0.6^\circ$. Rotation about the x -axis, which is oriented in the cranial-caudal direction of the patient and corresponds to out-of-plane rotation for both X-ray cameras, typically has the broadest peak and is the dominant component of registration error.

type of DRR (AF-DRR and RC-DRR), 240 registrations were performed, 30 in each of the eight misregistration intervals.² The TRE value was computed for each registration transformation.

Fig. 15 shows the histogram of TRE values produced using AF-DRRs. Most transformations produced TRE values <2.5 mm distributed in a relatively tight cluster (this histogram represents results from all four patients; the distribution is tighter for a single patient than for the combined results from all four patients). The remaining transformations produced widely scattered TRE values >5 mm. Such transformations represent the registration process getting stuck in a local optima

²This is stratified sampling. Initial transformations were randomly generated until there were at least 30 transformations in each of the eight misregistration intervals. At this point, most intervals contained more than 30 transformations. Then 30 transformations were randomly picked from the set of randomly generated transformations in each of the eight misregistration intervals.

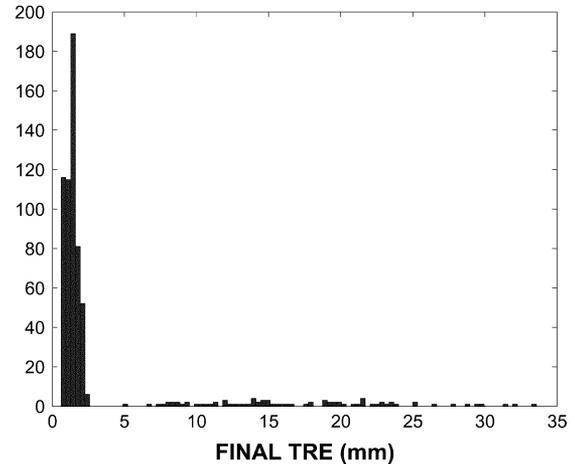


Fig. 15. Histogram of TRE values produced by the 2D–3D image registration algorithm using AF-DRRs. Most transformations produced TRE values <2.5 mm distributed in a relatively tight cluster (this histogram represents results from all four patients; the distribution is tighter for a single patient than for the combined results from all four patients). The remaining transformations produced widely scattered TRE values >5 mm. Such transformations represent the registration process getting stuck in a local optima during the iterative parameter search. Based on the histogram distribution, and the fact that a transformation with a TRE value <2.5 mm is potentially clinically useful, the registrations were characterized as either “successful” if the TRE <2.5 mm or “unsuccessful” if the TRE >2.5 mm. The histogram of TRE values using RC-DRRs is very similar to that obtained using AF-DRRs.

TABLE II
TWO-DIMENSIONAL (2-D) TO THREE-DIMENSIONAL (3-D) SPINE IMAGE TRE

Type of DRR	TRE (mm)						Unsuccessful Registrations		
	Individual Patients				Overall		Initial TRE (mm)		
	1	2	3	4	Mean	Max	0–6	6–16	Total
AF-DRR	1.0	1.6	1.4	1.3	1.3	1.6	1%	32%	20%
RC-DRR	0.9	1.5	1.4	1.2	1.3	1.5	1%	32%	20%

during the iterative parameter search. Based on the histogram distribution, and the fact that a transformation with a TRE value <2.5 mm is potentially clinically useful, the registrations were characterized as either “successful” if the TRE <2.5 mm or “unsuccessful” if the TRE >2.5 mm. The histogram of TRE values using RC-DRRs is very similar to that obtained using AF-DRRs.

The registration accuracy and robustness results are summarized in Table II. The TRE values that are listed are the mean TRE for all successful registrations. Fig. 16 shows how the percentage of successful registrations depends on the accuracy of the initial transformation. The registration algorithm almost always produces successful registrations as long as the initial transformation has a TRE <6 mm. However, the robustness quickly decreases as the initial transformation gets further from the correct transformation. The registration accuracy and robustness are virtually identical using AF-DRRs and RC-DRRs. The difference in overall mean TRE is 0.1 mm and is not statistically significant (two-tailed paired t -test, $P = 0.05$). The execution time using AF-DRRs is approximately 100 s whereas the execution time using RC-DRRs is approximately 3 000 s using a PC workstation with one 2.2-GHz Intel Xeon processor.

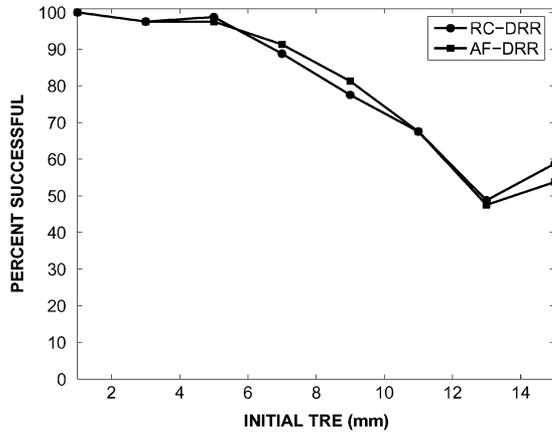


Fig. 16. Percentage of successful registrations for initial transformations with different initial TRE values. Each data point represents a 2 mm range of initial TRE values. The labels “AF-DRR” and “RC-DRR” denote registrations performed using AF-DRRs and RC-DRRs, respectively.

IV. DISCUSSION

A variety of fast volume rendering algorithms have been proposed, including the shear-warp method [42], frequency domain methods that use the Fourier slice projection theorem [43], space-leaping techniques using distance transforms [44], [45], and splatting methods [46], [47]. A special graphics board for volume rendering is commercially available (VolumePro 500, TeraRecon, Inc., San Mateo, CA). An important limitation of most of these methods, including this graphics board, is that they support only orthographic rendering. A more recent version of this graphics board (VolumePro 1000) supports perspective rendering, but unlike the case of orthographic rendering, perspective rendering using this board requires an additional software library running on the computer’s CPU to perform some of the computation, and generation of DRRs from a typical CT image takes slightly less than 1 s (personal communication with TeraRecon). A promising software-based volume rendering method that performs fast ray casting by highly optimizing the software and using single instruction multiple data instructions available in Intel processors has recently been reported; rendering of a $512 \times 512 \times 512$ -voxel image volume takes about 250 ms (4 fps) using a PC workstation with two 2.2 GHz Intel Xeon processors [48]. The use of splat volume rendering for generating DRRs has recently been reported; generation of a DRR takes about 100 ms (10 fps) if only voxels with intensity greater than 0 HU are used, but visual inspection revealed that quality of the splat rendering DRRs is lower than that of RC-DRRs (unfortunately no quantitative comparison using a measure such as PSNR was performed) [49]. Another promising volume rendering method generates 3D perspective renderings using texture-mapped volume rendering, which takes advantage of texture hardware (2-D or 3-D) on modern graphics boards [50], [51], but we are not aware of any published work using or evaluating hardware-based rendering to generate DRRs. The use of AFs is faster than these specialized rendering methods and the image quality of AF-DRRs is nearly identical to that of RC-DRRs (based on both visual and quantitative comparison). For example, using comparable Intel

processors, we generate AF-DRRs with resolution 256×256 pixels in approximately 50 ms with a single processor whereas [48] generates perspective renderings with the same resolution in approximately 250 ms with two processors. The generation of AF-DRRs is easily parallelized and should run approximately twice as fast with two processors. Nonetheless, these specialized rendering methods are still useful for our AF-based method. The AF is created from DRRs, and these specialized methods will speed up the precomputation of the AF.

We use the two-plane parameterization originally suggested by Levoy and Hanrahan [23]. In the 2D–3D registration application reported in Section III-B, the X-ray imaging system has two orthogonal cameras that are fixed in space. We use two AFs, one for each camera view. However, with C-arm fluoroscopes, the X-ray imaging system is rotated about the patient. One approach to deal with the large range of motion of the X-ray imaging system is to use multiple AFs. Another approach we are currently exploring is to use a cylindrical parameterization of the AF.

One issue discussed by Levoy and Hanrahan [23] that we do not address is anti-aliasing. When we generate AF-DRRs, we sample the 4-D (u, v, s, t) space and expose ourselves to the potential problem of aliasing. However, we perform quadrilinear interpolation, and the resulting smoothing applies an inherent anti-aliasing step. Also, we generate reasonably dense AFs. We have not yet observed any artifacts due to aliasing. If our AFs were particularly sparse, we would probably need to perform some prefiltering.

One potentially interesting aspect of AF-DRRs is the availability of fast image gradients. Interpolation in s and t is an inherent part of generating an AF-DRR. This bilinear interpolation, however, is simply a weighted combination of the nearest neighbors in s and t for a given ray. To calculate fast image gradients, the interpolation weights could be altered to mimic a gradient operator such as a Sobel filter. The output of an AF-DRR generation scheme modified in this way would be a gradient image that requires the same amount of computation time as a normal AF-DRR. This could be particularly useful for a registration algorithm using a gradient-based optimization.

The AF-DRRs we generate are substantially blurred relative to real X-ray projection images (Fig. 13). But they are visually and quantitatively very similar to the RC-DRRs we use to create the AF. The quality of the AF-DRRs is partially reduced by compression and interpolation between discrete samples, but is limited primarily by the quality of the DRRs used to create the AF. Any approach that can generate higher quality DRRs, such as the use of a better algorithm or a higher-resolution CT image, can be used to generate DRRs for the creation of the AF, and the resulting AF-DRRs will also be higher quality.

Levoy and Hanrahan [23] created both virtual light fields, which were generated from rendered images, and real light fields, which were generated from digitized images. In this paper, we created AFs using DRRs, which are synthetic X-ray images computed by casting rays through a CT image using a known camera geometry. It is also possible to create AFs using real X-ray images. For example, the X-ray images that are

acquired for a cone-beam CT image are essentially an AF with a cylindrical parameterization and with (u, v) space sampled in only one dimension (the arc of rotation). An AF could be generated by systematically moving an X-ray source and detector pair to various orientations in order to better sample (u, v) space. It might be possible to accomplish this with, for example, a motorized C-arm X-ray system such as is available in a modern angiography system. Radiation exposure could be minimized by limited the orientations to a small neighborhood and also by collimating the X-ray beam to a region of interest.

Execution time is an important consideration for the clinical application of 2D–3D registration algorithms. Generation of DRRs during the optimization search is the primary computational expense in the intensity-based 2D–3D registration process. Using a PC workstation with one 2.2 GHz Intel Xeon processor, computation of an AF-DRR with 200×200 pixels (which is the typical size of an ROI in our registration work) requires about 30 ms. Each iteration of the parameter search requires the generation of 24 DRRs (two DRRs per transformation parameter per image, six rigid transformation parameters, two orthogonal images) and, thus, each iterative step requires about 720 ms. With our simple best neighbor search strategy, the optimization requires approximately 100–150 iterations and, thus, the total execution time is about 100 s. Using a gradient-based search, we can reduce the number of iterations in the search to about 30–40 iterations and, thus, the total execution time can be reduced to about 25 s. Also, the generation of AF-DRRs is easily parallelized. By using two faster processors, such as the currently available 3.6 GHz Intel Xeon processor, we should realistically be able to generate an AF-DRR in about 10 ms and reduce the execution time of our intensity-based 2D–3D registration algorithm to about 5 s.

REFERENCES

- [1] D. B. Russakoff, T. Rohlfing, D. Rueckert, R. Shahidi, D. H. Kim, and C. R. Maurer Jr., "Fast calculation of digitally reconstructed radiographs using light fields," *Proc. SPIE—Medical Imaging 2003: Image Processing*, vol. 5032, pp. 684–695, 2003.
- [2] D. B. Russakoff, T. Rohlfing, A. Ho, D. H. Kim, R. Shahidi, J. R. Adler Jr., and C. R. Maurer Jr., "Evaluation of intensity-based 2D–3D spine image registration using clinical gold-standard data," in *Lecture Notes in Computer Science*, J. C. Gee, J. B. A. Maintz, and M. W. Vannier, Eds. Berlin, Germany: Springer-Verlag, 2003, vol. 2717, Proc. 2nd Int. Workshop Biomedical Image Registration (WBIR 2003), pp. 151–160.
- [3] J. Bijhold, "Three-dimensional verification of patient placement during radiotherapy using portal images," *Med. Phys.*, vol. 20, pp. 347–356, 1993.
- [4] K. G. A. Gilhuijs, P. J. H. van de Ven, and M. van Herk, "Automatic three-dimensional inspection of patient setup in radiation therapy using portal images, simulator images, and computed tomography data," *Med. Phys.*, vol. 23, pp. 389–399, 1996.
- [5] M. J. Murphy, "An automatic six-degree-of-freedom image registration algorithm for image-guided frameless stereotaxic radiosurgery," *Med. Phys.*, vol. 24, pp. 857–866, 1997.
- [6] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas, "A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs," *Med. Phys.*, vol. 21, pp. 1749–1760, 1994.
- [7] E. Kerrien, M.-O. Berger, E. Maurincomme, L. Launay, R. Vaillant, and L. Picard, "Fully automatic 3D/2D subtracted angiography registration," in *Lecture Notes in Computer Science 1679*, C. J. Taylor and A. C. F. Colchester, Eds. Berlin, Germany: Springer-Verlag, 1999, Proc. 2nd Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI'99), pp. 664–671.
- [8] Y. Kita, D. L. Wilson, and J. A. Noble, "Real-time registration of 3D cerebral vessels to X-ray angiograms," in *Lecture Notes in Computer Science*, W. M. Wells, III, A. C. F. Colchester, and S. L. Delp, Eds. Berlin, Germany: Springer-Verlag, 1998, vol. 1496, Proc. 1st Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI'98), pp. 1125–1133.
- [9] S. Lavallee, J. Troccaz, P. Sautot, B. Mazier, P. Cinquin, P. Merloz, and J.-P. Chirossel, "Computer-assisted spinal surgery using anatomy-based registration," in *Computer-Integrated Surgery: Technology and Clinical Applications*, R. H. Taylor, S. Lavallee, G. Burdea, and R. Mösges, Eds. Cambridge, MA: MIT Press, 1996, pp. 425–449.
- [10] J. Weese, G. P. Penney, P. Desmedt, T. M. Buzug, D. L. G. Hill, and D. J. Hawkes, "Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery," *IEEE Trans. Inf. Technol. Biomed.*, vol. 1, no. 4, pp. 284–293, Dec. 1997.
- [11] A. Guezic, P. Kazanzides, B. Williamson, and R. H. Taylor, "Anatomy-based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot," *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 715–728, Oct. 1998.
- [12] M. Breeuwer, J. P. Wadley, H. L. T. de Bliet, P. A. C. Desmedt, F. A. Gerritsen, P. Gieles, J. Buurman, N. L. Dorward, B. Velani, N. D. Kitchen, D. G. T. Thomas, O. Wink, J. D. Blankensteijn, B. C. Eikelboom, W. P. T. M. Mali, M. A. Viergever, G. P. Penney, R. Gaston, C. R. Maurer Jr., D. L. G. Hill, D. J. Hawkes, F. Maes, D. Vandermeulen, R. Verbeek, P. Schmitz, J. Weese, J. Sabczynski, W. Zylka, C. Lorenz, T. M. Buzug, G. Schmitz, and M. H. Kuhn, "The EASI Project—Improving the effectiveness and quality of image-guided surgery," *IEEE Trans. Inf. Technol. Biomed.*, vol. 2, no. 3, pp. 156–168, Sep. 1998.
- [13] G. P. Penney, P. G. Batchelor, D. L. G. Hill, and D. J. Hawkes, "Validation of a two- to three-dimensional registration algorithm for aligning preoperative CT images and intraoperative fluoroscopy images," *Med. Phys.*, vol. 28, pp. 1024–1032, 2001.
- [14] R. Bansal, L. H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath, and J. S. Duncan, "Entropy-based dual-portal-to-3D-CT registration incorporating pixel correlation," *IEEE Trans. Med. Imag.*, vol. 22, no. 1, pp. 29–49, Jan. 2003.
- [15] W. Birkfellner, J. Wirth, W. Burgstaller, B. Baumann, H. Staedele, B. Hammer, N. C. Gellrich, A. L. Jacob, P. Regazzoni, and P. Messmer, "A faster method for 3D/2D medical image registration—A simulation study," *Phys. Med. Biol.*, vol. 48, pp. 2665–2679, 2003.
- [16] L. M. G. Brown and T. E. Boulton, "Registration of planar film radiographs with computed tomography," in *Proc. IEEE Workshop Mathematical Methods in Biomedical Image Analysis (MMBIA 1996)*, 1996, pp. 42–51.
- [17] D. Knaan and L. Joskowicz, "Effective intensity-based 2D/3D rigid registration between fluoroscopic x-ray and CT," in *Lecture Notes in Computer Science*, R. E. Ellis and T. M. Peters, Eds. Berlin, Germany: Springer-Verlag, 2003, vol. 2878, Proc. 6th Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2003), Pt. I, pp. 351–358.
- [18] H. Livyatan, Z. Yaniv, and L. Joskowicz, "Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1395–1406, Nov. 2003.
- [19] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2D–3D medical image registration," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 586–595, Aug. 1998.
- [20] D. Tomazevic, B. Likar, T. Slivnik, and F. Pernus, "3-D/2-D registration of CT and MR to X-ray images," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1407–1416, Nov. 2003.
- [21] L. Zöllei, W. E. L. Grimson, A. Norbash, and W. M. Wells III, "2D–3D rigid registration of X-ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, 2001, pp. 696–703.
- [22] D. LaRose, J. Bayouth, and T. Kanade, "Transgraph: Interactive intensity-based 2D/3D registration of X-ray and CT data," *Proc. SPIE—Medical Imaging 2000: Image Processing*, vol. 3979, pp. 385–396, 2000.
- [23] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH'96)*, 1996, pp. 31–42.
- [24] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The Lumigraph," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH'96)*, 1996, pp. 43–54.
- [25] G. W. Sherouse, K. L. Kovins, and E. L. Chaney, "Computation of digitally reconstructed radiographs for use in radiotherapy treatment design," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 18, pp. 651–658, 1990.

- [26] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Amsterdam, The Netherlands: Kluwer Academic, 1992.
- [27] S. Huang, L. Chen, and H. Chang, "A novel image compression algorithm by using Log-Exp transform," in *Proc. 1999 IEEE Int. Symp. Circuits Systems (ISCAS'99)*, vol. 4, 1999, pp. 17–20.
- [28] J. R. Adler Jr., M. J. Murphy, S. D. Chang, and S. L. Hancock, "Image-guided robotic radiosurgery," *Neurosurgery*, vol. 44, pp. 1299–1307, 1999.
- [29] M. J. Murphy, J. R. Adler Jr., M. Bodduluri, J. Dooley, K. Forster, J. Hai, Q. Le, G. Luxton, D. Martin, and J. Poen, "Image-guided radiosurgery for the spine and pancreas," *Comput. Aided Surg.*, vol. 5, pp. 278–288, 2000.
- [30] S. I. Ryu, S. D. Chang, D. H. Kim, M. J. Murphy, Q.-T. Le, D. P. Martin, and J. R. Adler Jr., "Image-guided hypo-fractionated stereotactic radiosurgery to spinal lesions," *Neurosurgery*, vol. 49, pp. 838–846, 2001.
- [31] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
- [32] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, pp. 137–154, 1997.
- [33] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [34] C. Studholme, D. L. G. 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.*, vol. 24, pp. 25–35, 1997.
- [35] T. Rohlfing, D. B. Russakoff, M. J. Murphy, and C. R. Maurer Jr., "An intensity-based registration algorithm for probabilistic images and its application for 2-D to 3-D image registration," *Proc. SPIE—Medical Imaging 2002: Image Processing*, vol. 4684, pp. 581–591, 2002.
- [36] *Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis*, M. Sonka and J. M. Fitzpatrick, Eds., SPIE Press, Bellingham, WA, 2000, pp. 447–513.
- [37] 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.*, vol. PAMI-9, pp. 698–700, 1987.
- [38] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, pp. 1–10, 1966.
- [39] J. B. 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 image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, pp. 554–566, 1997.
- [40] C. R. Maurer Jr., R. J. Maciunas, and J. M. Fitzpatrick, "Registration of head CT images to physical space using a weighted combination of points and surfaces," *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 753–761, Oct. 1998.
- [41] T. Rohlfing, C. R. Maurer Jr., D. Dean, and R. J. Maciunas, "Effect of changing patient position from supine to prone on the accuracy of a Brown-Roberts-wells stereotactic head frame system," *Neurosurgery*, vol. 52, pp. 610–618, 2003.
- [42] P. Lacroute and M. Levoy, "Fast volume rendering using a shear-warp factorization of the viewing transformation," in *Proc. 21st Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH'94)*, 1994, pp. 451–458.
- [43] T. Totsuka and M. Levoy, "Frequency domain volume rendering," in *Proc. 20th Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH'93)*, 1993, pp. 271–278.
- [44] M. Sramek and A. Kaufman, "Fast ray-tracing of rectilinear volume data using distance transforms," *IEEE Trans. Visualization Comput. Graph.*, vol. 6, no. 3, pp. 236–252, Jul. 2000.
- [45] K. J. Zuiderveld, A. H. J. Koning, and M. A. Viergever, "Acceleration of ray-casting using 3D distance transforms," *Proc. SPIE—Visualization in Biomedical Computing (VBC) 1992*, vol. 1808, pp. 324–335, 1992.
- [46] D. Laur and P. Hanrahan, "Hierarchical splatting: A progressive refinement algorithm for volume rendering," in *Proc. 18th Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH'91)*, 1991, pp. 285–288.
- [47] L. Westover, "Footprint evaluation for volume rendering," in *Proc. 17th Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH'90)*, 1990, pp. 367–376.
- [48] K. Mori, Y. Suenaga, and J. Toriwaki, "Fast software-based volume rendering using multimedia instructions on PC platforms and its application to virtual endoscopy," *Proc. SPIE—Medical Imaging 2003: Physiology and Function: Methods, Systems, and Applications*, vol. 5031, pp. 111–122, 2003.
- [49] W. Birkfellner, R. Seemann, M. Figl, J. Hummel, C. Ede, P. Homolka, X. Yang, P. Niederer, and H. Bergmann, "Wobbled splatting—A fast perspective volume rendering method for simulation of X-ray images from CT," *Phys. Med. Biol.*, 2005, to be published.
- [50] B. Cabral, N. Cam, and J. Foran, "Accelerated volume rendering and tomographic reconstruction using texture mapping hardware," in *Proc. 1994 Symp. Volume Visualization*, 1994, pp. 91–98.
- [51] W. Schroeder, K. Martin, and W. Lorensen, *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1998.