

Anatomy-Based Registration of CT-Scan and Intraoperative X-Ray Images for Guiding a Surgical Robot

A. Guézic,* P. Kazanzides, *Member, IEEE*, B. Williamson, and R. H. Taylor

Abstract— We describe new methods for rigid registration of a preoperative computed tomography (CT)-scan image to a set of intraoperative X-ray fluoroscopic images, for guiding a surgical robot to its trajectory planned from CT. Our goal is to perform the registration, i.e., compute a rotation and translation of one data set with respect to the other to within a prescribed accuracy, based upon bony anatomy only, without external fiducial markers.

With respect to previous approaches, the following aspects are new: 1) we correct the geometric distortion in fluoroscopic images and calibrate them directly with respect to the robot by affixing to it a new calibration device designed as a radiolucent rod with embedded metallic markers, and by moving the device along two planes, while radiographs are being acquired at regular intervals; 2) the registration uses an algorithm for computing the best transformation between a set of lines in three space, the (intraoperative) X-ray paths, and a set of points on the surface of the bone (imaged preoperatively), in a statistically robust fashion, using the Cayley parameterization of a rotation; and 3) to find corresponding sets of points to the X-ray paths on the surfaces, our new approach consists of extracting the surface *apparent contours* for a given viewpoint, as a set of closed three-dimensional nonplanar curves, before registering the apparent contours to X-ray paths.

Aside from algorithms, there are a number of major technical difficulties associated with engineering a clinically viable system using anatomy and image-based registration. To detect and solve them, we have so far conducted two experiments with the surgical robot in an operating room (OR), using CT and fluoroscopic image data of a cadaver bone, and attempting to faithfully simulate clinical conditions. Such experiments indicate that intraoperative X-ray-based registration is a promising alternative to marker-based registration for clinical use with our proposed method.

Index Terms— Anatomy- and image-based registration, CT, revision total hip replacement surgery, ROBODOC®, X-ray fluoroscopy.

I. INTRODUCTION

WE STUDY the problem of registering computed tomography (CT) preoperative data to intraoperative X-ray data. Such X-ray data is currently acquired via X-ray

Manuscript received March 23, 1998; revised July 2, 1998. This work was supported in part by NIST Advanced Technology Program (ATP) under Cooperative Agreement 70NANB5H1088. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Viergever. *Asterisk indicates corresponding author.*

*A. Guézic is with the IBM T.J. Watson Research Center, 30 Sawmill River Road, Hawthorne, NY 10532 USA (e-mail: gueziec@watson.ibm.com).

P. Kazanzides and B. Williamson are with Integrated Surgical Systems, Davis, CA 95616 USA.

R. H. Taylor is with the Johns Hopkins University, Computer Science Department, Baltimore, MD 21218 USA.

Publisher Item Identifier S 0278-0062(98)09094-6.

fluoroscopy and image intensification. Our work is part of a joint study with Integrated Surgical System (ISS), of Davis, CA, to extend the ROBODOC® system from primary total hip replacement (PTHR) surgery to revision total hip replacement (RTHR) surgery. RTHR is performed after a patient had PTHR and the implant failed for some reason; RTHR is a much more difficult operation, because less bone tissue remains, and because leftover bone cement must be detected and removed [46].

For PTHR and RTHR, the surgical robot of the ROBODOC® system accurately mills the cavity for the femoral implant. The robot trajectory is planned preoperatively using the ORTHODOC® software, based upon a CT-scan of the hip and a computer-aided design (CAD) model of an appropriate femoral implant. To register the surgical robot to its planned trajectory, the current clinical protocol uses two or three metallic markers surgically inserted in the femur [1], [2]. These are preoperatively imaged using CT and intraoperatively exposed and located by the robot. We study the feasibility of substituting external marker-based registration with anatomy-based and X-ray-based registration. Our goal is sometimes referred to as three-dimensional (3-D) to two-dimensional (2-D) “3-D to 2-D” registration [3]–[7]. Although related registration methods have been developed in the past, there are remaining significant challenges as follows.

- 1) Developing algorithms that work accurately and repeatably with the data that can be acquired clinically.
- 2) Solving a number of difficult engineering problems before building a clinically viable system. So far, we have identified: 1) a timing problem: can the image acquisition, calibration, segmentation and registration occur in the operating room (OR) within the allotted time (a few minutes at most)?; 2) a workspace problem: can the regions of interest be actually imaged during surgery (Fig 1: the surgical robot and various equipment take significant room); 3) a system integration problem: it is preferable to have a simple registration chain (possibly avoiding optical tracking systems requiring additional equipment and skilled personnel); and 4) a radiation dose problem: can the patient and personnel dose be determined? Do the benefits of the method justify this dose?

A. Method

Our method works by extracting surface models of the bone from CT data and contour models from fluoroscopic

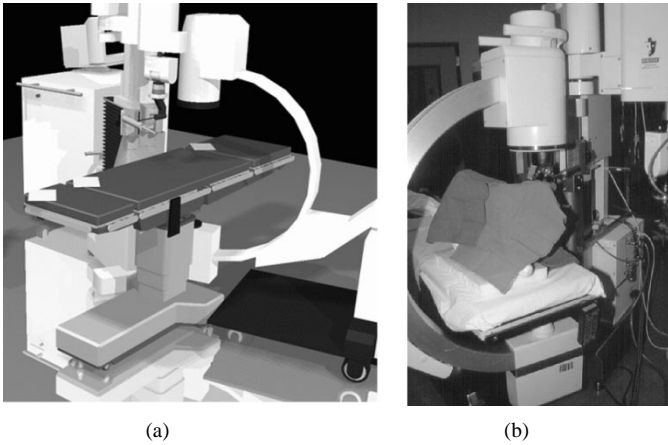


Fig. 1. Workspace limitations: (a) simulation and (b) actual experiment.

images (in Section III), converting such contours to bundles of X-ray paths using image calibration and distortion correction (as explained in Section II), and registering the surface to the X-ray paths (in Section IV). The fluoroscopic image geometric distortion correction and calibration of Section II are performed using a variation of the two-plane calibration method: the robot and a PC were used for collecting multiple images of a radiolucent rod. Inspired by the distal markers used clinically, we exploit the rigid structure of the femur by collecting images of the knee to perform the registration on the proximal femur. One advantage of this method is that the anatomy of the knee is very often intact (unlike the proximal femur) for patients undergoing THR. Our pose estimation algorithm of Section IV is new, using surface apparent contours for determining correspondences between lines and surface points. In Section IV-A, to search for correspondences between X-ray lines and surface apparent contours, we use an algorithm that experimentally performs in constant to logarithmic time. In Section IV-B, we use a robust method for computing the registration of points to lines.

In principle our method applies to both PTHR and RTHR; It is maybe better justified for RTHR, because there is less bone for attaching fiducials or probing for use in a point-to-surface registration. Also, intraoperative X-rays are routinely used in RTHR, so our proposed protocol would not require additional X-ray equipment in the OR. However, successful application to RTHR requires preoperative CT scan images that are not too corrupted by metal-induced artifacts in the areas used for registration [9].

We prototyped and experimented our method on THR surgery, but there are no major obstacles preventing its application to rigid anatomy-based registration for orthopaedics, neurosurgery or radiotherapy between 3-D preoperative imagery and 2-D intraoperative imagery. However, some of our choices are particular to our application, such as performing the image calibration and registration directly in the coordinate system of the robot.

B. Experiments

Based upon two experimental studies using a cadaver bone, we compare in Section V the accuracy of our method with

the current clinical registration method (using markers), that represents the ideal, or “ground truth,” registration for this study. In both experiments, the markers were located by the robot, providing the ideal position of the CT bone image in robot coordinates. The CT bone image was also positioned by applying our registration method to four fluoroscopic views. (In the first study, two fluoroscopic views were used in combination with one proximal marker.) With respect to the ideal position we observed a maximum positional error of approximately 1–3 mm, measured at the marker locations, depending on the initial positioning prior to registration. We tested various initial positions, corresponding to a 5° – 7° rotation of the ideal position.

II. ACQUISITION AND CALIBRATION OF FLUOROSCOPIC IMAGES

For each pixel of a digital X-ray image collected with our method, we wish to determine an equation of the straight line in 3-D space corresponding to the path of the X-rays that projected to that pixel. We call this line an *X-ray path*. This process is generally called *calibrating* the X-ray image, and is very related to calibrating a camera as is done in computer vision and related fields (consult [10] and our bibliography for additional references on this general problem). If the X-rays are captured on film, the X-ray source being very small relative to the film size, a good model for the projection is a perspective, or “pin-hole” projection.

In fluoroscopy, X-ray photons are converted to visible light photons using a phosphor screen. Then, inside an image intensifier (II) electrons are produced when visible light photons reach a photocathode, accelerated in a magnetic field and converted to visible light with an output phosphor. In conventional fluoroscopes marketed at the time this article is written, the intensified visible light is generally captured using a standard charge-coupled device (CCD) camera. Two main types of geometric distortion affect such X-ray images: “pincushion” distortion caused by the curvature of the photocathode, and “S-shaped” distortion caused by interactions of stray magnetic fields with electrons in the II [11].

We are aware of two main approaches for calibrating such images: 1) correction of distortion [12]–[17] followed by determination of the parameters of the perspective projection [4], [18]–[20]; and 2) simultaneous distortion correction and determination of the projection parameters using multiplane methods [21]. In the first approach 1), all methods for correcting the distortion that we are aware of compare a known grid pattern with its X-ray image, and record the displacements of each relevant grid node caused by distortion. In few methods, global parametric functions are used to model the distortion for the entire image (see [12] for the pincushion component and for the “S” component). In most methods, however, grid node displacements are interpolated with a suitable function (linear [11], [13], bilinear [14], spline [21], or cubic polynomial [23]), and generally stored as a look-up table. Studies (and our observations) indicate that the geometric distortion characteristics change significantly when the fluoroscope changes position or orientation [14], [16].

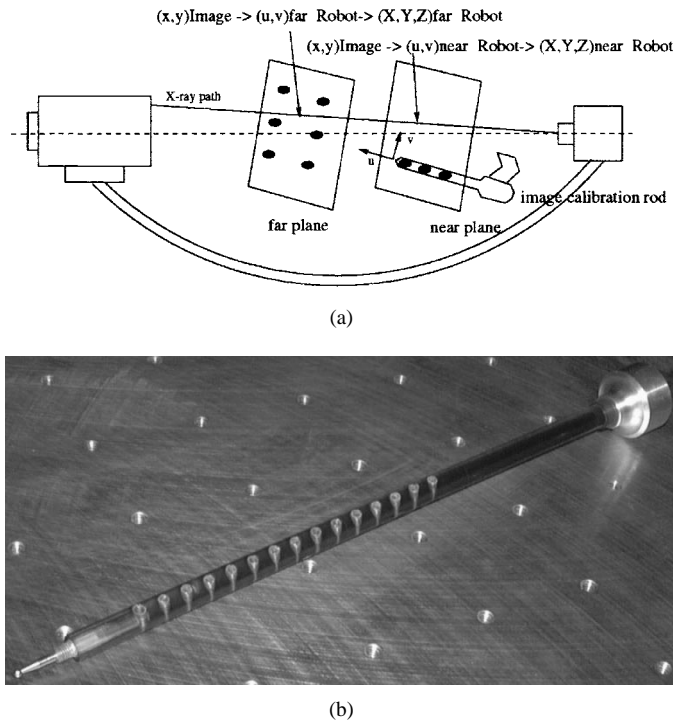


Fig. 2. (a) Two-plane calibration method and (b) radiolucent image calibration rod.

Hence, a new distortion correction map must be determined for each pose of the fluoroscope. While one solution would require that a grid of markers be permanently attached to the fluoroscope, our method does not require this.

We have chosen the second approach using a two plane calibration method, which is conceptually very simple (see Fig. 2). We believe that the parameters of the projection can be obtained at little additional cost during the distortion calibration, by essentially imaging two grids instead of one [24]: first, a “near” calibration grid is imaged: for each pixel representing a grid point the 3-D position of the grid point is known; in between such pixels a suitable interpolation method is used to associate a 3-D point $(x, y, z)_{\text{near}}$ to intermediate pixels. Second, a “far” grid is imaged, and a similar interpolation is used. Then each image pixel can be associated to two 3-D points $(x, y, z)_{\text{near}}$ and $(x, y, z)_{\text{far}}$, i.e., a straight line, which is our goal in this section.

Our practical approach is related to the NPBS method [21] with two important differences. In [21], distortion-induced marker displacements are interpolated with particular spline functions termed bicubic pseudo thin plate splines (TPS’s). Calibration grids are located with an optical tracking system (OPTOTRAKTM). A first difference is that we interpolate marker displacements using TPS functions. Invented by J. Duchon in 1976 [25], TPS functions are the mathematical model for a infinite thin metallic plate. TPS functions can interpolate scattered points particularly smoothly. A discussion on the physical meaning of TPS functions, and a description of how to implement TPS interpolations of data can be found in [26]. It is not clear that the bicubic pseudo TPS’s have any advantage over TPS functions except that they are less expensive to compute [21]: typical TPS interpolation implementations

involve a singular value decomposition (SVD). We did not find that this was a limitation, due to considerable progress in computing power of personal computers in the recent years and to the availability of high quality software packages for computing SVD, such as LAPACK [27].

Mathematically, for each plane, we wish to determine two functions of the pixel coordinates x, y : the robot coordinates $u(x, y)$ and $v(x, y)$ (see Fig 2: this is more efficient than determining three separate functions for x, y , and z). It is sufficient to discuss how $u(x, y)$ is determined: we write

$$u(x, y) = \sum_{i=0}^{n-1} a_i k(x, y, x_i, y_i) + a_n + a_{n+1} x + a_{n+2} y \quad (1)$$

where n is the total number of markers in the calibration plane, $k(x, y, x_i, y_i)$ is the TPS function $k(x, y, x_i, y_i) = (1/16)\pi r_i^2 \log r_i^2$, $r_i^2 = (x - x_i)^2 + (y - y_i)^2$ and the a_i are unknown coefficients. Since u is known at the marker locations (x_i, y_i) , we can obtain the a_i by solving the following linear system of equations (preferably using SVD, as advocated in Golub and Van Loan, [28, p. 257]):

$$\begin{bmatrix} K + n\lambda I & P \\ P^T & 0 \end{bmatrix} A = U$$

where $K = [(k(x_i, y_i, x_j, y_j))]$ is an $n \times n$ “TPS” matrix, I is an $n \times n$ identity matrix, $P = [(1 x_i y_i)]$ is an $n \times 3$ matrix,¹ $U = [u(x_i, y_i) | 0, 0, 0]^T$ is an $n+3$ vector, where the $u(x_i, y_i)$ are n known robot coordinates for each marker, A is the $n+3$ vector of unknowns, and λ is a smoothing parameter used to alleviate numerical problems: The TPS matrix is nearly singular, and we have found that using a positive value for λ , even very small with respect to one and x, y coordinates, is helpful (in practice, we have used 0.01). Using SVD, we can then easily solve the above linear system.

A second difference is that instead of using a physical grid of markers, we construct a “virtual” grid by moving a radiolucent (ULTEMTM) rod with embedded radio opaque (stainless steel) beads along a plane [see Fig 2(b)]. The radiolucent rod is attached directly to the surgical robot, allowing the calibration parameters to be directly acquired in the coordinate system of the robot, and eliminating the need for an OPTOTRAKTM or similar device in the OR. For each rod position, an image as in Fig. 3(a) is acquired by digitizing the video output of the fluoroscope.² In our current setup, X-rays are continuously activated. It is also possible to switch X-rays on just before each individual image is acquired, by altering the fluoroscopy protocol.

Using a grid of markers instead of a rod would allow to reduce X-ray exposure. We have chosen this method of sweeping the rod for the following reasons. Sweeping the rod simplifies the calibration of the grid with respect to the robot

¹The P^T component of the system specifies the behavior of the TPS and its x and y derivatives at infinity.

²Direct access to the internal digital format used by fluoroscope manufacturers will not necessarily be advantageous in the short term because such formats are likely to be different for each manufacturer, resampling may be necessary, and individual CCD pixels may be incorrect; since images are black and white, NTSC and equivalent video formats may be of sufficient quality. The situation will be different when direct digital acquisition of X-rays (without image intensifiers) becomes available in future years

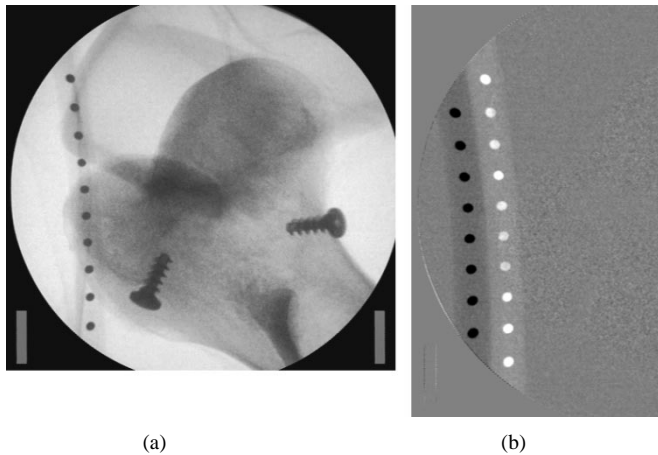


Fig. 3. (a) Geometrically distorted X-ray fluoroscopic image showing markers of our calibration rod and (b) taking image differences for extracting markers and detecting motion.

because it is only necessary to determine the rod position and orientation (i.e., a point and line) in the robot coordinate system [1]. In our current system, this simplification was required because the robot has five degrees of freedom and, thus, cannot arbitrarily position a plane in space.

Another advantage of our method is that between two images, only the rod has moved (this takes typically 1 s) and the detection of markers is greatly facilitated by taking image differences, as shown in Fig. 3(b). Combined with the fact that a single row of markers is detected each time (or two rows at a time) marker detection and identification can be fully automated. Although this problem is seldom documented, our experience is that with a conventional grid, identification of individual markers is error prone, and is sometimes done by an operator. After suitable system integration, we believe that our calibration method can be completely automated.

To validate our method, after collecting a series of images corresponding to two calibration planes, a row of markers was imaged in a different position and orientation. (Such markers were not used for calibration.) We computed the distances between the known 3-D positions of the markers and X-ray paths obtained by image calibration. Such distances were found to be between 0.1 and 0.4 mm. Fig. 4 shows the data used for the validation experiment. A row of markers can be seen on Fig. 4(a) together with some of the distal femur anatomy (and pins used by ROBODOC®, which are irrelevant here) after contour extraction from one X-ray image. Further processing [Fig. 4(b)] allows to extract only the contours corresponding to the markers (which are approximately circular), and to define for each contour x and y coordinates for the marker position in the image. Using the two-plane calibration information, for each marker position we can compute the equation of an X-ray path in three dimensions; such X-ray paths corresponding to the markers are shown in [Fig. 4(c)]. The distances reported were computed from the X-ray paths in 3-D to the known 3-D marker positions (ideally, they should be zero, as the X-ray path corresponding to a marker must intercept it).

It is not clear what an optimum number and distribution of markers across the image is. In the literature cited, anywhere

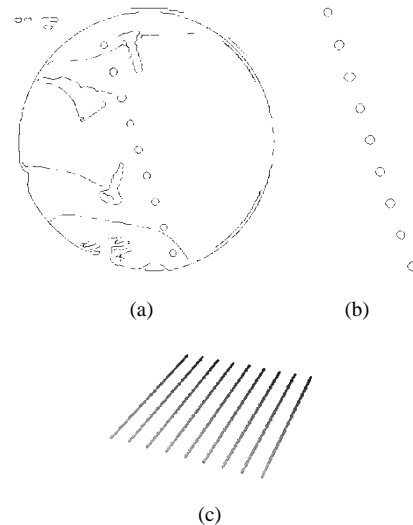


Fig. 4. Data used for calibration validation experiment: (a) row of markers, anatomical contours, and ROBODOC® pins after contour extraction from the image, (b) further processing allows to extract markers from A, and (c) X-ray paths computed from the x , y coordinates of data in (b) using X-ray image calibration information.

from four to several hundred markers were used to correct similar distortion patterns. In our case, we wish to limit the number of rod positions because of time and X-ray exposure constraints. We have found that anywhere between 6×6 and 9×9 markers was a good compromise. Assuming 1 s (or slightly more) for acquiring an image and moving the rod under continuous X-ray activation, we plan to collect data corresponding to one view in between 12 and 18 s. This yields reasonable X-ray dosages (an estimate of the effective X-ray dose would be on the order of 40 to 50 mrem/min of fluoroscopy assuming a 9-in field of view). Additional experiments may be required to determine the optimal number and distribution of markers across images for a prescribed accuracy.

Image intensifiers (II) may be gradually abandoned in the future in favor of direct digital X-ray image capture devices (using amorphous silicon or selenium for instance), provided that these technologies become mature for real-time imagery in the OR. These devices will probably introduce some amount of geometric distortion, albeit less than II's. Although our proposed method would be easily adapted to interface with these devices, it is unclear at the time this article is written whether other calibration methods could be more effective.

Aside from our use of the surgical robot to operate the image calibration device, which allows a very effective simplification of the registration chain, various aspects of our X-ray calibration process are general and can be used for other applications, such as the two-plane calibration, our use of TPS functions, and the process of collecting multiple images per plane that simplifies image processing.

III. EXTRACTION OF ANATOMICAL STRUCTURES FROM CT AND X-RAY DATA

Our goal is to extract a set of "surfaces" from CT representing anatomical structures of interest as well as "contours" from X-rays. We believe that the type of computer representation

used for anatomical structures to be registered influences the accuracy of the registration. Representations that were proposed include a collection of scattered points [29] or a connected set of voxels forming a solid inside a 3-D volume [3], a middle ground being a piecewise planar surface model (mesh of triangles or polygons) [30] or a differentiable surface using splines or other suitable smooth functions.

In this article, we mean by surface a set of vertices together with a connected set of triangles, each being defined as an ordered triple of references to vertices. In addition, surfaces will be consistently oriented and will satisfy the “manifold property,” i.e., for each vertex, the configuration of incident triangles can be continuously deformed to a disk or to a half disk for a boundary vertex (see, for instance, [31]). This property enforces an intuitive surface representation, very suitable for anatomical structures (at the scale we use for observation) that rules out edges shared by more than two triangles and vertices shared by more than one connected fan of triangles. Although piecewise planar, if enough vertices and triangles are used, this model can faithfully represent smooth anatomical surfaces.³ The benefits of this representation will become apparent in Section IV on 2-D to 3-D registration. When using points only, the connectivity information is absent. It can potentially be retrieved using triangulation methods [33], but then one realizes that much fewer triangles are typically required to approximate the same geometry sufficiently closely. When using volumetric models, the assumption is that accurate solid models of the anatomical structures are available, which is not necessarily true, because of missing data.

In the CT scan data that is clinically acquired for THR, the slice spacing typically varies significantly, from 1 mm in the vicinity of the fiducial markers to 6 mm or more in the femur shaft, to maximize the detail in the critical areas while limiting the X-ray dosage. Direct extraction of bone surfaces from such CT data using iso-surface building software yields poor contour definition at the slice level and poor surface tiling in between contours. Instead, our current approach for extracting a surface model from the CT data is to use a classical semiautomated method wherein contours are drawn on each relevant slice by an operator and are then tiled to form a surface [30], [33], [34].

A. 2-D Contour Extraction (CT Slice or Fluoroscope Image)

For each slice, we use a deformable model technique to detect the 2-D contour of the bone. We have implemented the technique of Kass *et al.* [35] with minor modifications. In [35] an energy is defined that incorporates the stretching and bending energies of the contour as well as a potential field measuring closeness to the data. We have chosen a smoothed image gradient norm for the potential field. The contour is then modeled with a series of points, and the energy is minimized by solving a partial differential equation using a finite differences method. In our implementation, the user is asked to select a few points in the vicinity of the structures of interest. The system then connects such points by a polygon, samples new points on that polygon, and uses

³The notion of smoothness applies to piecewise planar surfaces [32].

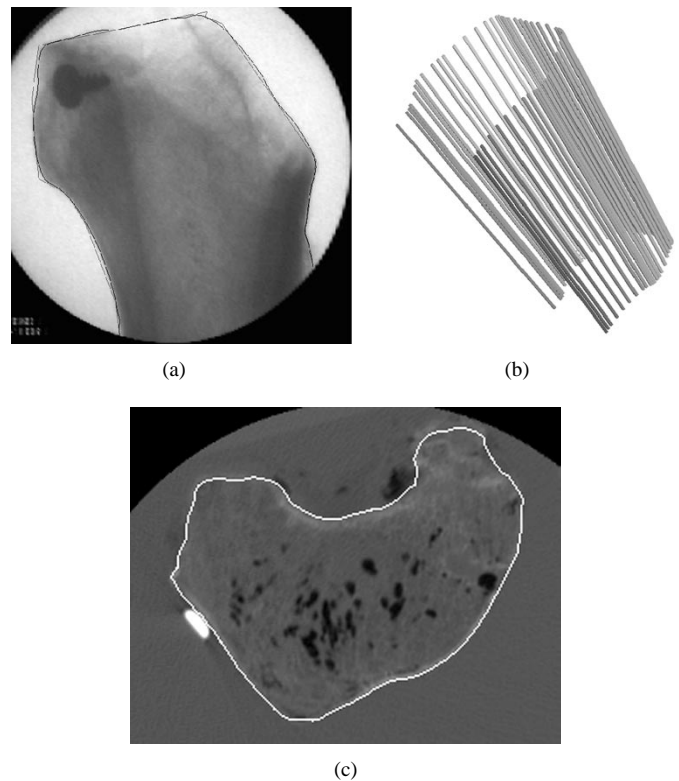


Fig. 5. (a) extracting contours from X-ray data, (b) X-ray paths obtained from (a) after calibration, and (c) extracting distal femur contours from CT.

such sampled points as an initial estimate for the deformable model [Fig. 5(a)–(c)].

In recent years, a considerable research effort was devoted to extend [35] for attracting the contours from a long range, and for automatically changing the topology of the deformable model. For our particular application to the femur bone, the contour topology is relatively simple, formed of either one simple polygon (most of the time) or two. To define a longer range potential, we have implemented a two step process, whereby a first deformable model is attracted by a smoothed low resolution version of the potential; the result is then used as an initialization for a second deformable model that is attracted by the full resolution potential. This process was particularly useful to reuse the same initial points from slice to slice and limit the user input. Such reuse of the input points is shown in Fig. 6. The same method is used to extract contours from 2-D X-ray image data, as shown in Fig. 5(a).

B. Surface Construction from Contours

Several approaches address the general problem of building a suitable surface model from any collection of contours with complex branching situations [33], [34], [36]. The full generality of such methods is not necessary for most of the femur length, as contours are in one to one correspondence. Branching contours exist at the head of the femur and condyles (in fact, the femur head should not be incorporated in the model since it is surgically removed during the operation).

When the correspondence between two contours is determined, we use a variation of a method described in [37] to build a piece of surface connecting the two contours. We

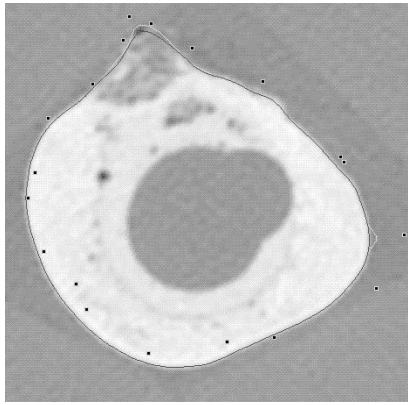


Fig. 6. Hierarchical deformable contour that can cope with incorrect user input, or reuse such input for multiple slices.

define a measure of surface quality as the sum of the measures of each individual triangle quality: typical measures we used were triangle area and “compactness,” [31], which is the ratio of the positive area by the sum of squared side lengths, scaled such that an equilateral triangle has a compactness of one. We then search for a true optimum of this surface quality measure: we consider a rectangular graph (grid) whose $n \times m$ nodes represent edges between, respectively, n and m vertices in both adjacent contours. Edges $(1, 1)$ and (m, n) are chosen as a closest point pair between the contours. A path in this graph from node $(1, 1)$ to node (m, n) represents a surface connecting the two contours. All possible paths are constructed using dynamic programming and the best one according to our criterion is retained.

The resulting surfaces are simplified with a tolerance between 0.0 and 0.3 mm using the variable tolerance method of Guéziec [31]. Two distinct surfaces are typically built, shown in Fig. 11: one for the proximal femur, and one for the distal femur (the shaft of the femur is not used for registration and typically not CT-scanned). Both surfaces have holes at either or both ends (are not solids). For our second experiment, the proximal surface of Fig. 11 was formed of 1664 triangles and the distal surface of 3913 triangles.

IV. 2-D TO 3-D REGISTRATION

The following methods have been investigated by other researchers: Lavallee *et al.* [3] register 2-D images (initially video and subsequently, X-ray) to 3-D solid anatomical models. An elegant and powerful aspect of their work is that they use a hierarchical data structure to quickly query the closest point from anatomical surfaces to X-ray paths. The Euclidean distance from an X-ray path to a surface that it intersects is by definition zero. To constrain such X-ray paths to be tangent after registration instead of intersecting, they define a negative distance inside the 3-D model; this implies that there is an inside and outside, i.e., that the 3-D model is a solid. We believe that this is a significant limitation as in practice only selected slices of data through the anatomy are available. To complete such data so as to define a solid volume, it would be necessary to add information that is not in the data. To resolve this issue, in our method, we first compute apparent contours

of our surface models and only then compute closest points. Similarly to [3] our search for closest points on the apparent contours is hierarchical.⁴

Lee [4] uses stereo pairs of radiographs to track in real time the position of a patient’s skull during radiotherapy delivery. He uses localized bony features segmented from a CT-scan. Feldmar *et al.* [5] register surfaces to projected contours, by defining image to surface correspondences and by minimizing a least squares criterion using iterative methods. The criterion incorporates contour and surface normals. In their method, the calibration parameters can be optimized as well during the registration process. Hamadeh [6] extends [3] by combining the X-ray segmentation and registration steps. Liu *et al.* [7] synthesize a 3-D model from X-ray views that they then register to another 3-D model.

A. Using Apparent Contours of the Surfaces

For each fluoroscopic view, our calibration method explained above (Section II) provides a bundle of lines from any collection of pixels. Assuming that the distortion was compensated for, the geometry of this line bundle is that of a perspective projection, i.e., all lines intersect at a *center of perspective*, (possibly at infinity for a parallel projection). Using the method of Section IV-B, we compute the position of this center of perspective for each view.

Given the center of perspective (possibly very far from the surface for a parallel projection) we can now define surface apparent contours: for each surface triangle, the “viewing direction” is defined as the vector originating from the center of perspective to the triangle centroid. If the triangle normal (defined by the cross product of ordered oriented triangle edges) makes an obtuse angle with the viewing direction, the triangle is said to be *visible* and invisible otherwise. Surface apparent contours are a subset of surface edges, such that the triangle on one side of the edge is visible and the triangle on the other side of the edge is invisible. The apparent contours are such that the edges are linked to form (nonplanar) closed polygonal curves in three dimensions.

To build the apparent contours, we first identify all edges belonging to any apparent contour using the criterion defined above, and add such edges to a list. We orient edges such that the visible triangle is on the left side of the edge as shown in Fig. 7(a), thus defining an edge origin and an edge destination. We then iterate on the following process: 1) we take the first edge in the list; we create a new apparent contour starting with that edge; 2) we complete the apparent contour containing that edge as follows: starting from the destination of the current edge we determine a next edge as shown in Fig. 7(a): we visit the triangles incident to the destination vertex in a counter-clockwise fashion and determine the first edge that belongs to the list of apparent contour edges [Fig. 7(a): this is necessary because there may be several such edges]. We reapply Step 2) until the next edge is the same as the first edge that was processed in 1); and 3) we remove all the edges forming that

⁴Such methods share the spirit of the iterative closest point methods [38], [39], with the important difference that in our case, a shape is matched to its projection, and the notion of a “closest,” or corresponding point is not straightforward.

contour from the list of apparent contour edges. We reapply Steps 1) to 3) until the list of apparent contour edges is empty. A typical result is illustrated in Fig. 7(b).

Then, we find the closest point from each X-ray path (line l) on the apparent contour(s): we first compute the closest point from the line in all apparent contours in Step A; once all the closest points have been determined, in Step B we select a particular apparent contour and the corresponding closest point. In Step A, we use a hierarchical decomposition of the polygonal (apparent) contours in segments with an associated “bounding region” R that completely encloses the contour portion that the segment defines [R consists of two hemispheres linked with a cylinder centered on the segment (a, b) shown in Fig. 7(c)]. We start with the first segment and compute a distance d to the line l . We consider the interval $[d-e, d+e]$, e being the maximum deviation attached to the segment, and add this interval to a priority queue keyed with $d - e$. We loop on the operation of taking the interval with the lowest key from the priority queue, of splitting in two the segment corresponding to the interval according to the information in the hierarchical decomposition of the polygonal contour, and of computing distances to both segments obtained. We stop the looping process when it is determined that the interval with the lowest key in the priority queue $[d_1 - e_1, d_1 + e_1]$ (Segment 1) has an empty intersection with the next interval in the priority queue $[d_2 - e_2, d_2 + e_2]$, meaning that $d_1 + e_1 < d_2 - e_2$. We record the point \mathbf{p} on Segment 1 that is closest to the line. We next define $r = (d_2 - e_2 - d_1 - e_1)/2$. We also record the point \mathbf{q} on the line that is closest to \mathbf{p} . When computing the closest point to the same polygonal contour from a second line, we first compute the distance from the second line to the point \mathbf{q} . If that distance is less than r then the closest point is directly searched for inside Segment 1. We observed experimentally that the average complexity of determining the closest point using that method is proportional to $\log(n)$ where n is the number of vertices in the polygonal contour.

In Step B, once all the closest points from a line to all the apparent contours have been determined according to Step A, we first determine if the line “stabs” the apparent contour or does not stab it. As shown in Fig. 7(d), the line stabs the apparent contour if the vector $\mathbf{q} - \mathbf{p}$ pointing from the closest point \mathbf{p} on the apparent contour to the point \mathbf{q} on the line (with a right angle at \mathbf{q}), the oriented edge \mathbf{e} starting from the closest point and the line direction \mathbf{v} , form a right handed frame. For a given line, if it corresponds to an external bone pixel in the X-ray image (“internal contours” can be used as well, because X-rays allow to visualize opaque objects as transparent), we determine whether there is any apparent contour that it stabs; if so, we retain the stabbed apparent contour for which the distance to the closest point is the largest. Otherwise, we retain the apparent contour for which the distance to the closest point is the smallest.

B. Robust Matching of 3-D Points to 3-D lines

At this point, starting with a set of 3-D lines (bundles of X-ray paths from different X-ray views, whose equations are

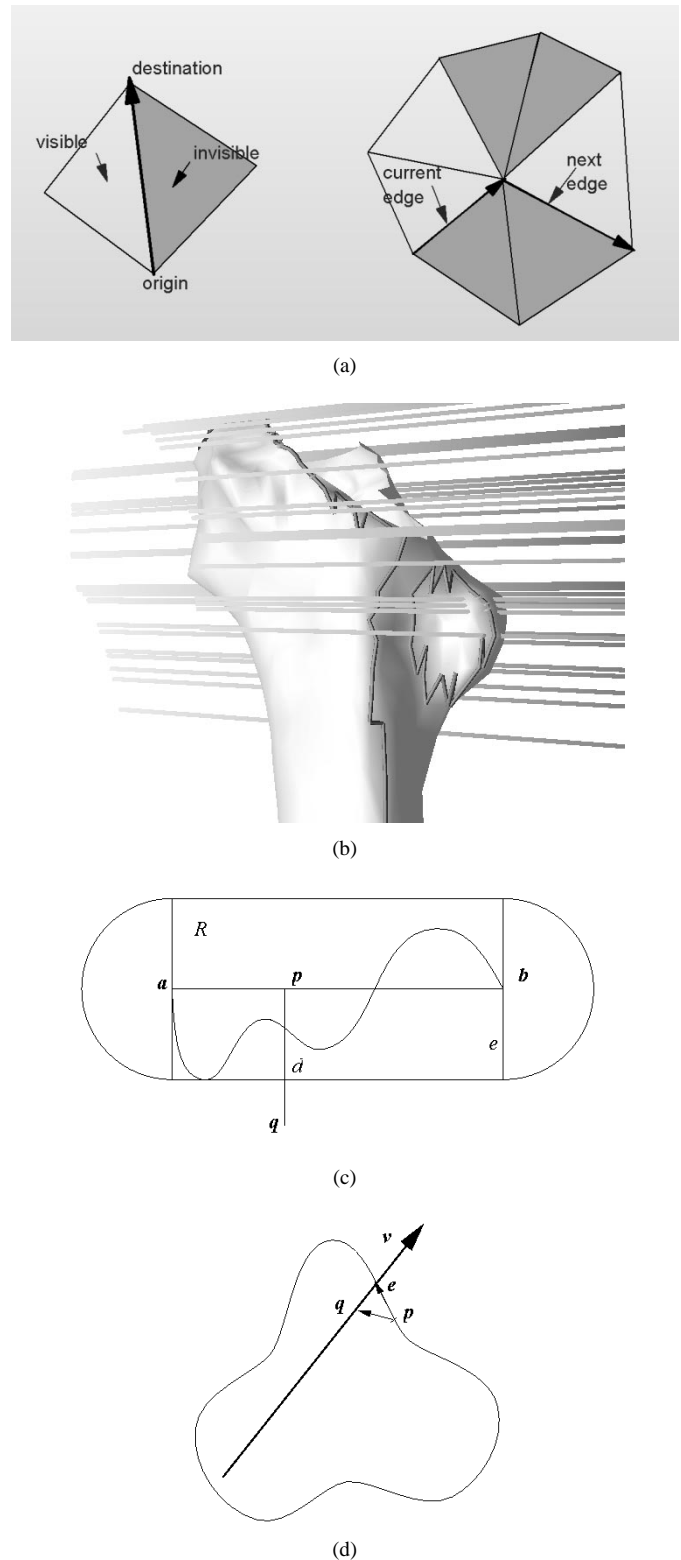


Fig. 7. (a) Notations and basic step for extracting surface apparent contours, (b) surface apparent contours shown together with X-ray paths, (c) bounding region R completely enclosing the portion of an apparent contour between the two points \mathbf{a} and \mathbf{b} ; a hierarchy of such bounding regions is build, and (d) determining if the X-ray path “stabs” the apparent contour.

known in the coordinate system of the robot, such as those of Fig. 5 and a set of corresponding 3-D points (closest points on CT surface apparent contours) we must determine a rotation

and translation that brings each point as close as possible to its corresponding line. This is a *pose estimation* problem.

Kumar and Hanson [40] analyze different formulations for the pose estimation problem and provide several robust methods to weigh out outliers. In addition, there is a rich literature on the problem of pose estimation using points or line tokens [41]. We compare several alternatives in the sequel.

1) *Formulation*: Assuming that the imagery is calibrated (which was done in Section II), using the following formulation we can treat parallel projections (not only perspective) and process simultaneously several views, when several centers of perspective are present.

We minimize the sum of squared distances between a set of m 3-D vertices $\{\mathbf{p}_i\}$ and corresponding m 3-D lines $\{(\mathbf{c}_i, \mathbf{v}_i)\}$, passing through a given center of perspective \mathbf{c}_i and having the direction of the unit vector \mathbf{v}_i , $\|\mathbf{v}_i\| = 1$ (see Fig. 8). In the following, \mathbf{Q} represents a 3-D rotation matrix and \mathbf{t} represents a 3-D translation vector. \mathbf{Q} and \mathbf{t} are unknown and all the other symbols have known values or expressions. The distance d_i between the transformed 3-D point $\mathbf{Q}\mathbf{p}_i + \mathbf{t}$ and the i th. 3-D line $(\mathbf{c}_i, \mathbf{v}_i)$ is obtained by taking the norm of the cross product between the vector joining \mathbf{c}_i and $\mathbf{Q}\mathbf{p}_i + \mathbf{t}$ and the unit vector \mathbf{v}_i that specifies the direction of the line. The problem can be written as follows:

$$\min_{\mathbf{Q}, \mathbf{t}} \sum_{i=1}^m d_i^2 = \min_{\mathbf{Q}, \mathbf{t}} \sum_{i=1}^m \|\mathbf{V}_i(\mathbf{c}_i - (\mathbf{Q}\mathbf{p}_i + \mathbf{t}))\|^2$$

$$\mathbf{V}_i = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \quad (2)$$

where \mathbf{v}_i denotes the skew symmetric matrix implementing the vector cross product with $(v_1, v_2, v_3)^t$. Solutions exist for any number of data points and lines provided that there is at least one point. However, it is known that a minimum of three points are required to provide a unique solution [42].

One parameterization of the rotation that is convenient for our formulation is the one by Cayley that we explain next. If \mathbf{U} denotes a skew symmetric matrix obtained from the vector \mathbf{u} , i.e.,

$$\mathbf{U} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}$$

then the 3×3 matrix $\mathbf{Q} = (\mathbf{I} - \mathbf{U})(\mathbf{I} + \mathbf{U})^{-1}$ is a rotational matrix. It is a rational parameterization. Although it is rather not standard in the literature, we have chosen it for various reasons: because of its rational nature, the computations are fast and more accurate than with parametrizations involving sines and cosines; going back and forth between a linearized (simply derived as $\mathbf{Q} \sim \mathbf{I} - 2\mathbf{U}$) and exact form is easy; when the rotation is small, the parameter values correspond to the magnitude of rotation along the Cartesian axes; computing the derivative of the rotation with respect to the parameters which is necessary for the Levenberg-Marquardt (LM) method discussed below is very manageable.

We now present three different methods, called LM, linear, and robust, to minimize (2) and obtain a rigid transformation.

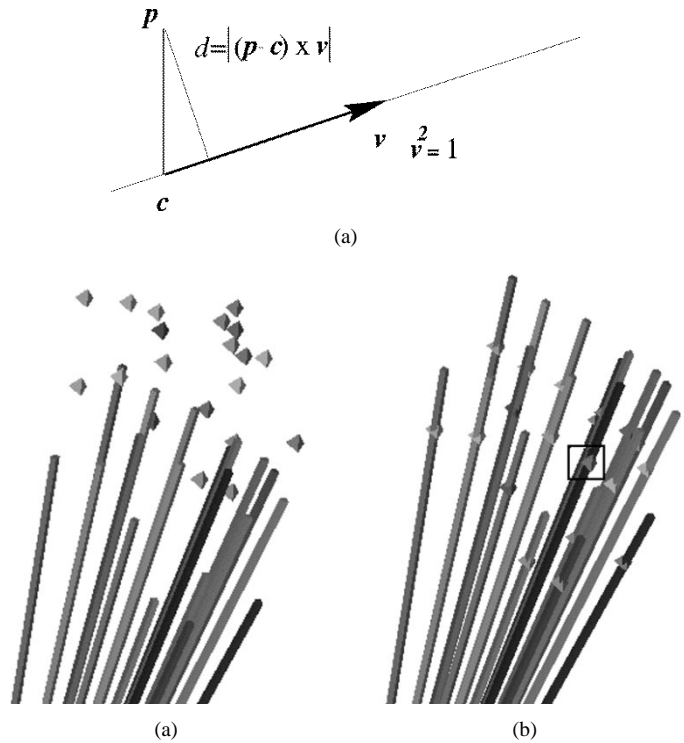


Fig. 8. (a) Euclidian distance d between a 3-D line defined with a vertex \mathbf{c} , a unit direction \mathbf{v} and a 3-D point \mathbf{p} , (b) sets of points and corresponding lines before registration, and (c) set of (b) after registration, regardless of the outlier marked with a rectangle. (The shades of grey associated to the different lines and points have no particular meaning.)

We then compare experimentally the methods in order to choose the most appropriate for our application.

2) *Nonlinear Minimization (LM)*: Suppose that we consider (2) as the sum of squares of m real-valued functions $f_i(\mathbf{x})$ of $n = 6$ variables, $\mathbf{x} = (u_1, u_2, u_3, t_x, t_y, t_z)^t$, with

$$f_i(\mathbf{x}) = d_i = \|\mathbf{V}_i(\mathbf{c}_i - (\mathbf{Q}\mathbf{p}_i + \mathbf{t}))\|. \quad (3)$$

We can rewrite (2) as

$$\min_{\mathbf{x}} \sum_{i=1}^m \|f_i(\mathbf{x})\|^2 = \min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x})\|^2 \quad (4)$$

where \mathbf{f} is the vector $(f_1, \dots, f_m)^t$. To minimize (4), we can use the LM method, which consists of switching gradually from Newton's minimization method, to the steepest descent method. Since details on these methods are available elsewhere [43], we briefly recall the principles and concentrate on providing suitable input to library routines that implement these methods.

Newton's method supposes that a quadratic approximation of $(\|\mathbf{f}(\mathbf{x} + \mathbf{y})\|^2)/2$ is available for a small $\|\mathbf{y}\|$

$$\frac{\|\mathbf{f}(\mathbf{x} + \mathbf{y})\|^2}{2} \simeq \frac{\|\mathbf{f}(\mathbf{x})\|^2}{2} + \mathbf{g}^t \mathbf{y} + \mathbf{y}^t \mathbf{H} \mathbf{y}$$

where \mathbf{g} denotes the gradient of $(\|\mathbf{f}\|^2)/2$ and \mathbf{H} the Hessian, or symmetric matrix of second-order partial derivatives. By application of the chain rule, we get $\mathbf{g} = \mathbf{J}^t \mathbf{f}$, where \mathbf{J} is the gradient of \mathbf{f} , which is a $m \times 6$ matrix. Further differentiation leads to $\mathbf{H} \simeq \mathbf{J}^t \mathbf{J}$, provided that the error residual $\|\mathbf{f}\|$ is small with respect to $\|\mathbf{J}\|$. Newton's method consists of

computing the value of \mathbf{y} that will minimize the quadratic term, which corresponds to solving $\mathbf{J}^t \mathbf{J} \mathbf{y} + \mathbf{J}^t \mathbf{f} = \mathbf{0}$. For numerical reasons, it is preferable to minimize the following, which is less sensitive to roundoff errors:

$$\min_{\mathbf{y}} \|\mathbf{J} \mathbf{y} + \mathbf{f}\|.$$

The LM method consists of solving

$$(\mathbf{J}^t \mathbf{J} + \lambda \mathbf{I}) \mathbf{y} + \mathbf{J}^t \mathbf{f} = \mathbf{0}.$$

For a small λ it corresponds to Newton's method, and for a large λ , \mathbf{y} is determined by the formula $\mathbf{y} = -(\mathbf{J}^t \mathbf{f} / \lambda)$, which corresponds to the steepest descent method, along the direction of the gradient. In practice, λ is set to a very small value at the beginning and gradually increased as the solution vector is approached.

It is necessary to provide the LM implementation with procedures for computing both \mathbf{f} and \mathbf{J} for any input parameters \mathbf{x} . \mathbf{f} is easily computed by first deriving the rotation \mathbf{q} from (u_1, u_2, u_3) and by using (3). The gradient requires more work: the derivatives with respect to the translation parameters are the simplest

$$\frac{\partial f_i}{\partial \mathbf{t}} = \frac{\mathbf{V}_i^t \mathbf{V}_i (\mathbf{Q} \mathbf{p}_i - \mathbf{c}_i + \mathbf{t})}{f_i}.$$

In order to compute the derivative of f_i with respect to \mathbf{u} , considerable simplification is obtained by computing the derivatives of the vector $\mathbf{Q} \mathbf{p}_i$ with respect to \mathbf{u} , which is a 3×3 matrix: $\partial \mathbf{Q} \mathbf{p}_i / \partial \mathbf{u}$. We have used Mathematica to do that (the formula can be found in [45]). We can write

$$\frac{\partial f_i}{\partial \mathbf{u}} = \left(\frac{\partial \mathbf{Q} \mathbf{p}_i}{\partial \mathbf{u}} \right)^t \frac{\mathbf{V}_i^t \mathbf{V}_i (\mathbf{Q} \mathbf{p}_i - \mathbf{c}_i + \mathbf{t})}{f_i}.$$

The method requires that the 6×6 matrix $\mathbf{J}^t \mathbf{J}$ be of full rank. It is thus necessary to provide at least six corresponding points and lines. We will see that the linear method exposed next does not have this requirement.

3) *Linearization and Constrained Minimization*: A second method, which we refer to as the linear method, consists of linearizing the rotation. We subject the minimization to the constraint that the rotation will be small, for the linearized equation to be valid. After linearizing the rotation \mathbf{Q} with $\mathbf{I} - 2\mathbf{U}$, and constraining $\|\mathbf{u}\|$ to be smaller than r , real value, (such as 0.1, in order to guarantee that the approximation holds true up to at least two decimals in each rotation entry) (2) becomes

$$\min_{\mathbf{U}, \mathbf{t}} \sum_i \|\mathbf{V}_i (\mathbf{c}_i - (\mathbf{I} - 2\mathbf{U}) \mathbf{p}_i - \mathbf{t})\|^2 + \lambda (r^2 - \|\mathbf{u}\|^2). \quad (5)$$

We transform the expression (5) and introduce the skew symmetric matrices \mathbf{P}_i (the relationship between \mathbf{P}_i and \mathbf{p}_i is the same as the one between \mathbf{U} and \mathbf{u})

$$\min_{\mathbf{u}, \mathbf{t}} \sum_i \|\mathbf{V}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})\|^2 + \lambda (r^2 - \|\mathbf{u}\|^2) \quad (6)$$

which we can rewrite in matrix form

$$\min \|\mathbf{A} \mathbf{x} - \mathbf{b}\| \quad \text{subject to } \|\mathbf{B} \mathbf{x}\| < r \quad (7)$$

with

$$\mathbf{A} = \begin{bmatrix} 2\mathbf{V}_1 \mathbf{P}_1 & \mathbf{V}_1 \\ \cdots & \cdots \\ 2\mathbf{V}_n \mathbf{P}_n & \mathbf{V}_n \end{bmatrix}$$

$3n \times 6$ matrix, $\mathbf{x}^t = (\mathbf{u}^t, \mathbf{t}^t)$, $\mathbf{b} = [\mathbf{V}_i (\mathbf{c}_i - \mathbf{p}_i)]$, $3n \times 1$ matrix, and finally $\mathbf{B} = [\mathbf{I} \quad \mathbf{0}]$ 3×6 matrix.

This method is iterative: once \mathbf{U} , \mathbf{t} have been estimated, we compute \mathbf{Q} from \mathbf{u} (which has a rational expression), set the new 3-D points $\mathbf{p}_i := \mathbf{Q} \mathbf{p}_i + \mathbf{t}$ and solve (7) until the increments of \mathbf{u} , \mathbf{t} are below a certain small value or until a maximum number of iterations has been exceeded. Contrary to the LM method, any number of points greater than two can be fed to (7). An answer could normally be obtained with one point only, but we have chosen to only consider problems with more equations than unknowns. The method works very well for three points and thus offers a good alternative to closed form solutions [42].

4) *Use of a Robust M-Estimator*: In this section, we describe the third method which we refer to as the Robust method. Robust means that the method will be resistant to outliers [44]. We start with the expression minimized in (6); then each error term (3-D distance point-line) is weighted by (specifically, becomes the argument of) a function ρ , such that measurements that correspond to an error exceeding the median of the errors by some factor will have a constant contribution. This formulation was introduced by Tukey, and ρ is called the Tukey weighting function. We obtain

$$\min_{\mathbf{u}, \mathbf{t}} \Delta = \sum_i \rho \left(\frac{\|\mathbf{V}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})\|}{s} \right) \quad (8)$$

where s is a scale factor. Let us note $e_i = \|\mathbf{V}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})\|$ the error. The constant s can be defined as $s = (\text{median}(e_i)) / 0.6745$. It is not necessary to know the exact expression of ρ . We will rather use the derivative of ρ , Ψ [40], [41]: $\Psi(u) = u(1 - u^2/a^2)^2$ if $\|u\| \leq a$ and zero, otherwise. Following [40] we set $a = 2$. Note that the choice of a as well as the choice of s influence the proportion of data points that will be considered outliers. We differentiate (8) with respect to \mathbf{u} , \mathbf{t} . The first step is to differentiate e_i with respect to \mathbf{u} and \mathbf{t}

$$\begin{aligned} \frac{\partial e_i}{\partial \mathbf{u}} &= \frac{\partial \sqrt{(\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})^t \mathbf{V}_i^t \mathbf{V}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})}}{\partial \mathbf{u}} \\ &= \frac{2\mathbf{P}_i^t \mathbf{V}_i^t \mathbf{V}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})}{2e_i} \\ \frac{\partial e_i}{\partial \mathbf{t}} &= \frac{\mathbf{V}_i^t \mathbf{V}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})}{2e_i}. \end{aligned} \quad (8a)$$

At this point, we introduce the projection matrix $\mathbf{W}_i = \mathbf{V}_i^t \mathbf{V}_i = \mathbf{I} - \mathbf{v}_i \mathbf{v}_i^t$. The derivative of Δ with respect to the vector \mathbf{u} yields, once the constant terms have been removed from the summation

$$\begin{aligned} \frac{\partial \Delta}{\partial \mathbf{u}} = 0 &= \sum_i \Psi \left(\frac{e_i}{s} \right) \frac{\mathbf{P}_i^T \mathbf{W}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})}{e_i} \\ \frac{\partial \Delta}{\partial \mathbf{t}} = 0 &= \sum_i \Psi \left(\frac{e_i}{s} \right) \frac{\mathbf{W}_i (\mathbf{p}_i - \mathbf{c}_i + 2\mathbf{P}_i \mathbf{u} + \mathbf{t})}{e_i} \end{aligned}$$

which we can rewrite in matrix form $\mathbf{C}\mathbf{x} = \mathbf{d}$, with

$$\mathbf{C} = \sum_i \Psi\left(\frac{\mathbf{e}_i}{s}\right) \begin{bmatrix} 2\mathbf{P}_i\mathbf{W}_i\mathbf{P}_i & \mathbf{P}_i\mathbf{W}_i \\ 2\mathbf{W}_i\mathbf{P}_i & \mathbf{W}_i \end{bmatrix}$$

6×6 matrix

$$\mathbf{x}^T = (\mathbf{u}^T, \mathbf{t}^T), \quad \mathbf{d} = \sum_i \Psi\left(\frac{\mathbf{e}_i}{s}\right) \begin{bmatrix} \mathbf{P}_i\mathbf{W}_i(\mathbf{c}_i - \mathbf{p}_i) \\ \mathbf{W}_i(\mathbf{c}_i - \mathbf{p}_i) \end{bmatrix}$$

6×1 vector (noting that $\mathbf{P}_i^T = -\mathbf{P}_i$). The method can be applied to any number (≥ 1) of points. However, in the case of three points, an exact match for all points is possible. We have observed that the linear method is more accurate in that case, since the normal least squares equations are used rather than the direct minimization.

In practice, the linear method, (7), has been implemented with an explicit constraint of $\|\mathbf{u}\|$ being smaller than r , while the robust method has been implemented without constraint. The technique for implementing that constraint is generalized singular value decomposition (GSVD), which is explained in Golub and Van Loan [28].

5) *Experimental Comparison of the Methods:* To compare the LM, linear and robust methods, we generate a set of lines and points simulating a typical configuration that would occur when solving a pose estimation problem specialized to radiographic imagery.

The first step is to generate the base configuration of points and lines. We first choose m random points $\{p_i\}_{i=1, \dots, m}$ inside the unit cube: $0 \leq x, y, z \leq 1$. We assume a perspective projection. The following step is to choose a center of perspective \mathbf{c} , where all lines will intersect. For this experiment, the point $(-2, -2, -2)$ was chosen, which corresponds to a maximum angle for the bundle of lines of approximately 20° . In order for the experiment to be independent of our particular choice for the origin and axes, we rotate the entire configuration by a rotation of random axis and of arbitrary angle (30°). We then rotate the points $\{p_i\}$ only by a random rotation of 30° , which is within the range of expected rotations for the pose estimation problem. We translate the points by a random vector that has the same magnitude as the points (drawn inside the unit cube). This completes the construction of the base configuration. A configuration of 20 points and lines is shown in Fig. 8.

The next step is to take the base configuration, register the points to the lines and compare the transformation obtained with the one that was effectively applied. In order to make that comparison, we apply both transformations to landmark points and compute the distance between the two transformed points. We used the following vertices of the unit cube as landmark points $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. The maximum distance is chosen to measure the quality of registration (maximum registration error). The last step is to add noise to selected points. We add zero mean Gaussian noise to the points. The maximum standard deviation of the noise varies from 0–0.33, which is roughly one-third of the diameter of the cloud of points. Such a large amount of noise would rarely be observed in practice, if at all. We add noise to a selected number of points, up to 50%.

In Fig. 9(a)–(c) we show the maximum registration error with each method. The figure shows that the LM and linear methods typically tolerate noise that does not exceed a tenth of the diameter of the cloud of data points (assuming that a significant portion of the points are corrupted by the noise). Noise of higher magnitude can be coped with provided that less than 5% of the points are corrupted. While in theory, the robust method should still work with 50% of spurious points, in practice, such a large number of outliers is only tolerated with small noise levels, and we have observed a successful registration with up to approximately 40% of outliers corrupted with noise of magnitude a third of the diameter of the cloud of data points. Due to the iterative evaluation process for the rotation and translation, significant errors can remain at early stages of the registration, making it difficult to isolate outliers based solely upon the residual error. Valid data points are more likely to be misinterpreted as outliers, resulting in an incorrect registration.

Table I provides us with the average execution times (using an IBM RS6000 UNIX workstation) for the three different methods on data sets that comprise 100 points and 20 points, respectively. All three methods execute in less than one second for datasets of fewer than 100 points, which is much more points than expected in practice, since in a practical situation, it is difficult to extract from image and model data a large number of 3-D points and lines that unambiguously correspond to each other. Due to an advantageous complexity for the numerical computation of the six parameters, the robust method executes significantly faster than the other two and is only linearly dependent on the number of points. The linear (constrained) method uses a version of GSVD, that has a component cubic in the number of data points, which is why the computing time grows faster than with the other methods. The LM method uses a linear system that is also 6×6 , but has a significant overhead of computing the gradient of each of the functions. It thus starts with the highest cost, but the cost does not increase dramatically with the number of points m . The robust method has two parts: the first part, whose complexity is linear in the number of data points, consists in determining the median of error measurements and forming a linear system; the system, which is of size 6×6 , is solved in the second part, in constant time. The robust method will execute in less than 1 s for virtually any data set. The method, applied to 1000 points, computed registrations in an average of 0.25 s.

C. Registration of Surface Models to Bundles of X-Ray Paths

Although all three methods among LM, linear, and robust are reasonably fast (can complete in well below 1 s on typical data sets), the robust method was chosen because of its advantageous computational complexity, and of its capability to cope with about 40% of outliers (as it appears, independently of the magnitude of the noise).

To register the CT surface models to the X-rays (and to the robot), we iterate on 1) solving the equation $\mathbf{C}\mathbf{x} = \mathbf{d}$ of Section IV-B-4, 2) applying the resulting rotation and translation to the surfaces, 3) recomputing apparent contours, and 4) closest points on those, until the residual error falls

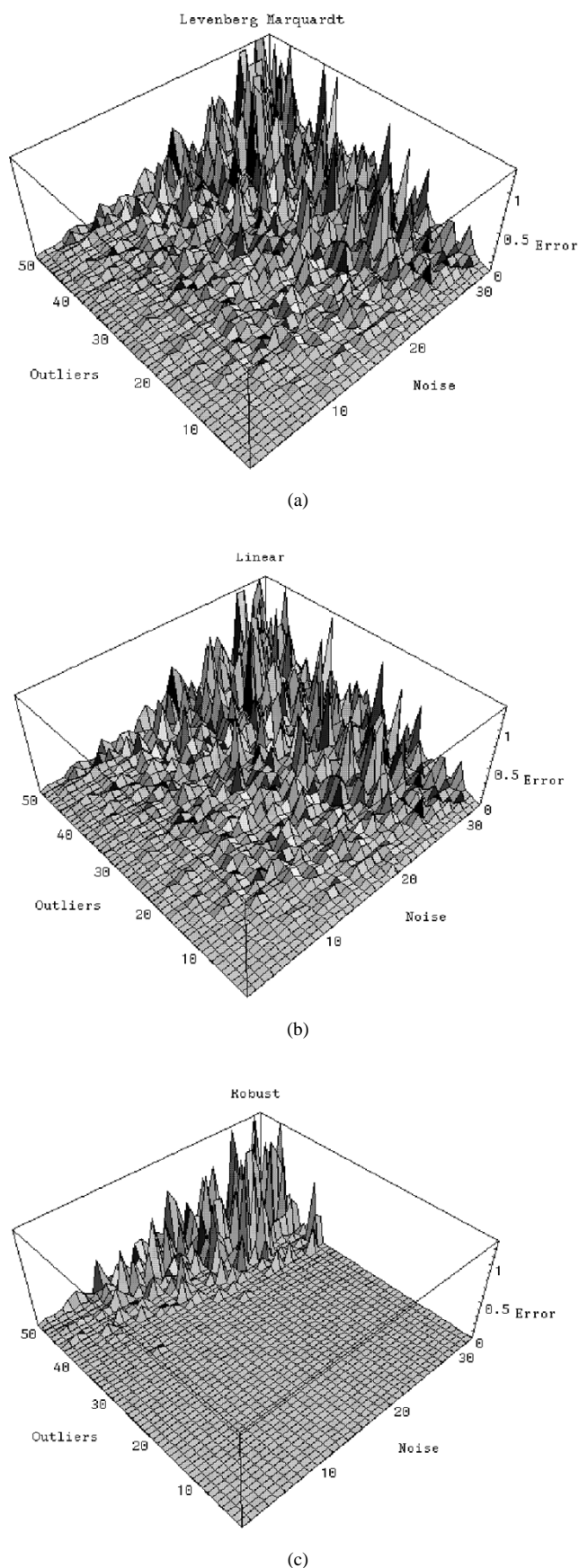


Fig. 9. (a) Maximum registration error plotted versus noise magnitude and percentage of outliers for the LM method, (b) maximum registration error for the linear method, and (c) idem for the robust method.

TABLE I
AVERAGE EXECUTION TIMES IN MS FOR THE THREE REGISTRATION METHODS APPLIED TO DATA SETS THAT COMPRISE 100 POINTS (TOP) AND 20 POINTS (BOTTOM)

Number Points/Method	LM	Linear	Robust
100 points (CPU time)	790	690	28
20 points (CPU time)	200	42	9.6

below a given value, or until we exceed a maximum number of iterations, or until the magnitude of the incremental rotation and translation falls below a minimum threshold. Results obtained with this method are presented in the next section.

V. EXPERIMENTS AND DISCUSSION

Much of the prior work, with the notable exception of Lee’s, uses high-quality 3-D and projection images for experiments, such as high resolution CT scans of dry bones, or simulations of radiographs using video images. Such data are only available in a controlled laboratory test. In contrast, we have used the same data that would be clinically available. The CT slices show soft tissue, present notable artifacts, and are of wide and unequal spacing, to minimize the X-ray dose delivered to the patient. Precise segmentation of such data is challenging. The fluoroscopic images were obtained with a Ziehm Exposcop Plus (R).

We have performed two experiments, with increasing realism. The first experiment was conducted in June 1996 using the ROBODOC® system at Sutter Hospital, Sacramento, CA. We CT-scanned a cadaver bone having three implanted fiducial markers. We then brought the cadaver bone to the OR and fixated it to the robot. We positioned the robot so that the calibration rod and cadaver femur were both in the field of view of the fluoroscope. We then captured sets of fluoroscopic images of the cadaver femur and calibration rod; the robot could move the rod to form two planar grids which were imaged. We repeated this operation for the proximal and distal femurs. The purpose of this experiment was to determine the geometric relationship of these images with respect to the robot (X-ray to robot registration) and with respect to the CT-scan image of the bone. The results were compared with the ROBODOC® registration technique using three markers implanted in the femur. Since the object of this experiment was to test the algorithms, we were not overly concerned with the clinical viability of the test setup. This experiment was performed without an OR table, and we positioned the robot and fluoroscope in places that would not be available if an OR table and patient were present.

The results of this experiment were as follows: starting with the registered position computed using the markers (current ROBODOC® protocol), we perturbed it with a rotation of 9° and started the registration process from that perturbed position. The registration was completed in about 6 s on an IBM RS6000 580. We then measured the distance between the marker locations specified by our algorithm and the locations specified by the marker-based ROBODOC® registration. In addition to the fluoroscope images, we used one of the markers for registration (the proximal marker): since the position of this

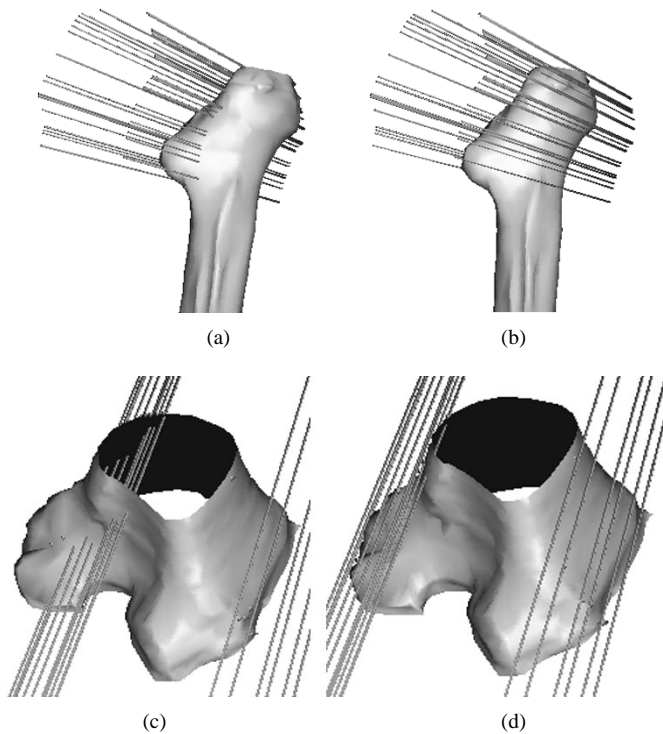


Fig. 10. (a) Proximal and (c) distal femur models and X-ray paths of our first experiment (b) before registration and (d) after registration (using the proximal marker in addition to X-ray images).

marker was prescribed, the registration process optimized the rotational component only. In this case we found a difference of 2 mm or less in the positions of the distal markers (i.e., the positions of the distal markers found by the robot versus those calculated using the proximal marker/fluoroscopy registration). The registration error was somewhat higher (about 5mm) if we used only the fluoroscope images. Configurations before and after registration are shown in Fig. 10.

We performed another experiment in November 1996. It was similar to the previous experiment in that the robot, calibration rod, fluoroscope, and cadaver femur were used. The main goals of this experiment were to test the feasibility of the registration method for THR. A new design of the calibration rod was used: it was elongated, and made of sterilizable ULTEMTM; a metallic tip allowed calibration of the rod coordinates by the robot. We draped a cadaver bone along with a foam bone and placed them on an OR table to simulate the Antero-Lateral approach to THR [Fig. 1(b)]. This was based on the judgement of an OR nurse and an ISS engineer. We avoided any position of the equipment that would not be possible if a patient's body was on the table. We collected images corresponding to four different positions and orientations of the fluoroscope, two views proximally and two views distally (Fig. 11); for each view the robot moved the calibration rod along two planes to obtain calibrated image information. On the X-ray images, the anatomy was sometimes occluded by various OR instruments [Fig. 11(a)].

The results of this second experiment are gathered in Table II. Starting with the registered position computed using the markers, we perturbed it with 11 various rotations and translations, and started the registration process. We limited

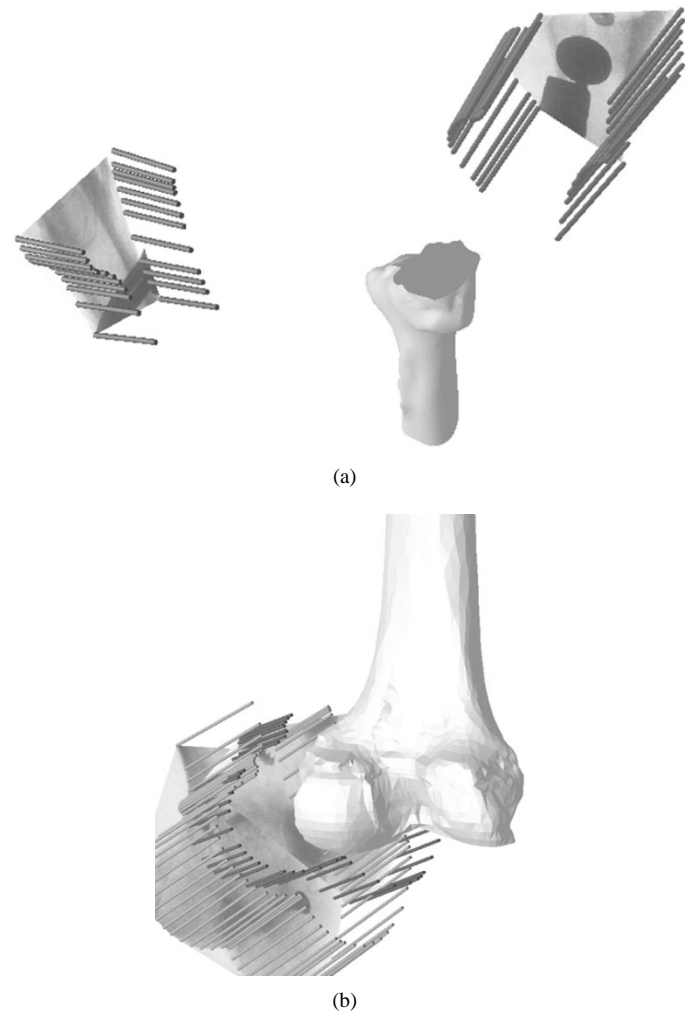


Fig. 11. (a) Proximal and (b) distal CT femur models from our second experiment, in registered position with X-ray fluoroscopy, showing X-ray paths (dark lines emanating from 2-D image bone contours) and distortion-corrected X-ray images projected on "far" calibration plane. In (b), calibration planes corresponding to different fluoroscope orientations happen to intersect, and so do images.

the rotation angle to 7° (which, when applied to a 450 mm femur, can perturb the position by as much as 55 mm; a 10° rotation could perturb the position in excess of 80 mm. Hence, we applied very significant perturbations, exceeding considerably the expected misregistration after suitable initial positioning in a practical situation, of 10 to 20 mm at the most). The translation was set so as to leave the centroid of the pin positions invariant. For this experiment, only the fluoroscopic views were used. The maximum registration error (measured at each marker location) was between 1.2 and 3.7 mm, with the exception of one initial position (7° , position four) for which the registration error reached 4.8 mm (these 4.8 mm, however, were measured at one of the distal pins; the error on implant placement, which is proximal, would be closer to 3.5 mm which is the error measured proximally).

The following two remarks help interpret these data: first, for a femur that is about 450-mm long, an angular misregistration of 1° can correspond to an error of 8 mm (4 mm if the center of rotation is situated in the medial part of the femur).

TABLE II
REGISTRATION RESULTS: DISTANCE MEASURED AT FIDUCIALS AFTER
REGISTRATION. WE START WITH THE REGISTERED POSITION OF THE ANATOMICAL
SURFACES PROVIDED BY THE PIN-BASED METHOD AND PERTURB THIS
POSITION WITH SEVERAL DIFFERENT ROTATIONS AND TRANSLATIONS

Transformation	proximal marker	distal marker 1	distal marker 2
identity	1.16	2.38	2.4
rotation 5 degrees 1	1.7	1.56	3.03
rotation 5 degrees 2	1.19	1.28	1.53
rotation 5 degrees 3	3.05	2.91	2.95
rotation 5 degrees 4	2.14	1.61	2.66
rotation 5 degrees 5	2.76	3.06	2.69
rotation 7 degrees 1	3.08	3.66	2.96
rotation 7 degrees 2	2.7	2.52	3.22
rotation 7 degrees 3	2.56	2.84	2.38
rotation 7 degrees 4	3.48	4.81	2.83
rotation 7 degrees 5	2.59	2.32	3.18

The above figures indicate that the registration is correct within 1° or less of rotation. The main reason the numbers would seem large is the size of the object that we register. Second, in practice it is very unlikely that we will start the registration process with an initial position representing an error in excess of 50 mm. More likely, we will have initial errors of 10 to 20 mm to correct for, and the accuracy numbers quoted here are thus conservative.

These data do not indicate how the registration accuracy depends upon the initial position (as opposed to the angular misregistration alone).

A. Potential Sources for Inaccuracies in X-Ray and Anatomy-Based Registration

There are potential sources for inaccuracies in each step of the process: X-ray image calibration errors, 2-D and 3-D image segmentation errors, and registration errors.

Potential image calibration errors could have two effects. First, if the relationship is incorrect between image and device to be registered (surgical robot), a correct registration would be corrupted by some bias. Second, a wrongful correction of the image distortion can result in a lack of correspondences between the 2-D and 3-D data to be registered, having similar effects as image segmentation errors. However, the tests described in Section II indicate that our 2-D image to robot calibration step is very accurate.

Potential image segmentation errors affect the accuracy of registration because they affect negatively the potential to discover correspondences between surface and contours to be registered (potentially corresponding points are eliminated or displaced, and spurious points may be added). Ideally, once a segmentation procedure is established, one way to quantify these errors would be to perform a large number of experiments on various datasets. This is, however, outside the scope of this article. Another way to quantify these errors is to test the algorithm on synthetic data. We performed such tests in [45], and obtained very accurate registrations (submillimeter). This would indicate that segmentation is the foremost source of errors.

The last source of errors is the registration algorithm itself. The registration algorithm establishes point correspondences (using apparent contours, and a novel search algorithm on the apparent contours). These correspondences could be incorrect. For instance, the closest point on the apparent contour is not necessarily the best match. This area is open for improvements in the future. Also, the registration would fail in situations where the initial positioning brings the surfaces in a location that is too distant from the correct position for the registration algorithm to select a majority of correct (line, apparent contour point) correspondences, which is required for a successful registration when using a robust method (which can correct some errors, but fails when the majority of data points are spurious).

VI. CONCLUSION

Based upon our preliminary findings, fluoroscopy based registration is a promising alternative to marker based registration for our application: 1) our proposed use of fluoroscopy would be compatible with time, workspace and x-ray exposure constraints in the operating room; 2) our proposed registration method would cope with clutter and distortion in clinical x-ray fluoroscopic images, and nonlinear correspondences between fluoroscopy and CT image brightnesses (because it is feature-based); 3) the image-based registration accuracy would satisfy prevalent guidelines (clinical studies involving patient data would be required to confirm this).

To develop and test this method further, a next step would be to conduct studies involving patient data. A number of issues may be investigated, notably evaluating the robustness of the method when applied to clinical image data, assessing the registration accuracy without markers (that were necessary in the present study for establishing a ‘‘ground truth’’), developing methods for monitoring and refining the registration during the operation, using additional intraoperative images. Another logical step would be to integrate the different software and hardware components involved into a complete system.

ACKNOWLEDGMENT

The authors would like to thank T. Beck from John Hopkins University for evaluation of the X-ray dosage estimates. They would also like to thank G. Taubin for suggesting the use of the Cayley parameterization of the rotation. Finally, they would like to thank J. Funda who participated in the evaluation of various points-to-lines registration methods. The 2-D contour detection software was written together with S. Gomory.

REFERENCES

- [1] R. H. Taylor, H. A. Paul, P. Kazanzides, B. D. Mittelstadt, W. Hanson, J. F. Zuhars, B. Williamson, B. L. Musits, E. Glassman, and W. I. Bargar, ‘‘An image-directed robotic system for precise orthopaedic surgery,’’ *IEEE Trans. Robot. Automat.*, vol. 10, pp. 261–275, 1994.
- [2] B. D. Mittelstadt, P. Kazanzides, J. Zuhars, B. Williamson, P. Cain, F. Smith, and W. L. Bargar, ‘‘The evolution of a surgical robot from prototype to human clinical use,’’ in *Computer-Integrated Surgery*, R. H. Taylor *et al.*, Eds. Cambridge, MA: MIT Press, 1996, pp. 397–407.
- [3] S. Lavalley, R. Szeliski, and L. Brunie, ‘‘Matching 3-D smooth surfaces with their 2-D projections using 3-D distance maps,’’ in *Proc. Geometric Methods in Computer Vision*, SPIE, San Diego, July 25–26, 1991, vol. 1570, pp. 322–336.

- [4] B. Lee, "Stereo matching of skull landmarks," Ph.D. dissertation, Stanford University, Stanford, CA, 1991.
- [5] F. Betting, J. Feldmar, N. Ayache, and F. Devernay, "A new framework for fusing stereo images with volumetric medical images," in *Computer Vision, Virtual Reality and Robotics in Medicine, Lecture Notes in Computer Science*, N. Ayache, Ed., vol. 905. Nice, France: Springer-Verlag, Apr. 1995, pp. 30–39.
- [6] A. Hamadeh, P. Sautot, S. Lavallée, and P. Cinquin, "Toward automatic registration between CT and X-ray images cooperation between 3-D/2-D registration and 2-D edge detection," in *Proc. 2nd Int. Symp. on Medical Robotics and Computer Assisted Surgery*, Baltimore, MD, Nov. 1995, pp. 39–46.
- [7] A. Liu, E. Bullitt, and S. Pizer, "Surgical instrument guidance using synthesized anatomical structures," in *Computer Vision and Virtual Reality in Medicine II—Medical Robotics and Computer Assisted Surgery III, Lecture Notes in Computer Sciences*, vol. 1205. Grenoble, France: Springer-Verlag, 1997, pp. 99–108.
- [8] R. V. III O'Toole, D. A. Simon, B. Jaramz, O. Ghattas, M. K. Blackwell, L. Kallivokas, F. Morgan, C. Visnic, A. M. DiGioia, III, and T. Kanade, "Toward more capable and less invasive robotic surgery in orthopedics, a new framework for fusing stereo images with volumetric medical images," in *Computer Vision, Virtual Reality and Robotics in Medicine, Lecture Notes in Computer Science*, N. Ayache, Ed., vol. 905. Nice, France: Springer-Verlag, 1995, pp. 123–130.
- [9] A. Kalvin and B. Williamson, "Using scout images to reduce metal artifacts in CT with applications to revision total hip surgery," in *SPIE: Medical Imaging: Image Processing*, K. M. Hanson, Ed., Feb. 1997, vol. 3034, pp. 1017–1028.
- [10] O. Faugeras, *Three-Dimensional Computer Vision*. Cambridge, MA: MIT Press, 1994.
- [11] B. Schueler and X. Hu, "Correction of image intensifier distortion for three-dimensional X-ray angiography," in *SPIE: Medical Imaging*, 1995, vol. 2432, pp. 272–279.
- [12] S. Rudin, D. R. Bednarek, and W. Wong, "Accurate characterization of image intensifier distortion," *Med. Phys.*, vol. 18, no. 6, pp. 1145–1151, 1991.
- [13] J. M. Boone, J. A. Seibert, W. A. Barrett, and E. A. Blood, "Analysis and correction of imperfections in the image intensifier-TV-digitizer imaging chain," *Med. Phys.*, vol. 18, no. 2, pp. 236–242, 1991.
- [14] L. Launay, C. Picard, E. Maurincommé, R. Anxionnat, P. Bouchet, and L. Picard, "Quantitative evaluation of an algorithm for correcting geometrical distortions in DSA images: Applications to stereotaxy," *SPIE: Medical Imaging: Image Processing*, Murray H. Loew, Ed., Mar. 1995, vol. 2434.
- [15] M. Roth, Ch. Brack, A. Schweikard, H. Götte, J. Moctezuma, and F. Gossé, "A new less invasive approach to knee surgery using a vision-guided manipulator," presented at *Virtual Reality*, Montpellier, France, May 1996.
- [16] S. Schreiner, J. Funda, A. C. Barnes, and J. H. Anderson, "Accuracy assessment of a clinical biplane fluoroscope for three-dimensional measurements and targeting," *SPIE: Med. Imag.: Image Display*, K. M. Hanson, Ed., Feb. 1997, vol. 3031, pp. 160–166.
- [17] K. R. Hoffman, B. B. Williams, J. Esthappan, S. Y. J. Chen, J. D. Carroll, H. Harachi, V. Doerr, G. N. Kay, A. Eberhardt, and M. Overland, "Determination of 3-D positions of pacemaker leads from biplane angiographic sequences," *Med. Phys.*, vol. 24, no. 12, pp. 1854–1862, 1997.
- [18] A. Rougee, C. Picard, Y. Troussset, and C. Ponchut, "Geometrical calibration for 3-D X-ray imaging," in *SPIE: Med. Imag.: Image Capture, Formatting, and Display*, 1993, vol. 1897, pp. 161–169.
- [19] L. M. Brown, "Registration of planar film radiographs with computed tomography," in *Mathematical Methods in Biomedical Image Analysis*. San Francisco, CA: IEEE Press, June 1996, pp. 42–51.
- [20] N. Navab, A. Bani-Hashemi, M. Mitschke, D. W. Holdsworth, R. Fahrig, A. J. Fox, and R. Graumann, "Dynamic geometrical calibration for 3-D cerebral angiography," in *SPIE—Med. Imag.: Phys. Med. Imag.*, R. Van Metter and J. Beutel, Eds., Feb. 1996, vol. 2708, pp. 361–370.
- [21] G. Champelboux, S. Lavallée, P. Sautot, and P. Cinquin, "Accurate calibration of cameras and range imaging sensors: the NPBS method," in *Proc. Int. Conf. on Robotics and Automation*, 1992, pp. 1552–1557.
- [22] R. Ning, J. K. Riek, and D. L. Conover, "An image intensifier-based volume tomographic angiography imaging system: Geometric distortion correction," in *SPIE: Medical Imaging: Physics of Medical Imaging*, 1994, vol. 2708, pp. 199–210.
- [23] J. B. Garrison, W. L. Ebert, R. E. Jenkins, S. M. Yionoulis, H. Malcom, and G. A. Heyler, "Measurement of three-dimensional positions and motions of large numbers of spherical radiopaque markers from biplane cinematograms," *Comput., Biomed. Res.*, vol. 15, pp. 76–96, 1982.
- [24] H. A. Martins, J. R. Birk, and R. B. Kelley, "Camera models based on data from two calibration planes," *Comput. Graphics, Image Processing*, vol. 17, pp. 173–180, 1981.
- [25] J. Duchon, "Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces," *Revue Francaise d'Automatique, Informatique, Recherche Operationnelle (RAIRO) Analyze Numerique*, vol. 10, no. 12, pp. 5–12, 1976.
- [26] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 6, pp. 567–585, June 1989.
- [27] E. Anderson, Z. Bai, C. Bischof, J. Demmel, and J. Dongarra, *Lapack User's Guide*, 2nd ed. Philadelphia, PA: SIAM, 1995.
- [28] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1993, pp. 562–564.
- [29] C. A. Pelizzari, G. T. Y. Chen, R. R. Spelbring, D. R. Weichselbaum, and C. T. Chen, "Accurate 3-D registration of CT, PET, and/or MR images of the brain," *J. Comput. Assist. Tomogr.*, vol. 13, no. 2, pp. 20–26, 1989.
- [30] C. R. Maurer, G. B. Aboutanos, B. M. Dawant, R. J. Maciunas, and J. M. Fitzpatrick, "Registration of 3-D images using weighted geometrical features," *IEEE Trans. Med. Imag.*, vol. 15, pp. 836–849, Dec. 1996.
- [31] A. Guéziec, "Surface simplification with variable tolerance," in *Proc. 2nd Int. Symp. on Medical Robotics and Computer Assisted Surgery*, Baltimore, MD, Nov. 1995, pp. 132–139.
- [32] G. Taubin, "A signal processing approach to fair surface design," in *Siggraph'95 Conf. Proc.*, Los Angeles, CA, Aug. 1995, pp. 351–358.
- [33] B. Geiger, "Three-dimensional modeling of human organs and its application to diagnosis and surgical planning," INRIA, Sophia-Antipolis, Valbonne, France, Tech. Rep. 2015, 1993.
- [34] D. Meyers, S. Skinner, and K. Sloan, "Surfaces from contours," *ACM Trans. Graphics*, vol. 11, no. 3, pp. 228–258, July 1992.
- [35] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proc. Int. Conf. on Computer Vision*, London, U.K., June 1987, pp. 259–268.
- [36] J. D. Boissonnat, "Shape reconstruction from planar cross sections," *Comput. Vision, Graphics, Image Processing*, vol. 44, no. 1, pp. 1–29, 1988.
- [37] J. Fuchs, Z. M. Kedem, and S. P. Uselton, "Optimal surface reconstruction from planar contours," *Commun. ACM*, vol. 20, no. 10, pp. 693–702, Oct. 1977.
- [38] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 239–256, 1992.
- [39] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vision*, vol. 13, no. 2, pp. 119–152, 1994.
- [40] R. Kumar and A. R. Hanson, "Robust methods for estimating pose and a sensitivity analysis," *Comput. Vision, Graphics, Image Processing—IU*, vol. 60, no. 3, pp. 313–342, 1994.
- [41] R. M. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, pp. 1426–1446, Nov./Dec. 1989.
- [42] M. A. Fishler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Graphics, Image Processing*, vol. 24, no. 6, pp. 381–395, June 1981, ACM.
- [43] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. New York: Academic, 1993.
- [44] P. J. Huber, *Robust Statistical Procedures*, Eidgenössische Technische Hochschule Zürich, SIAM, 1977.
- [45] A. Guéziec and J. Funda, "Evaluating the registration of 3-D points to 3-D lines with application to the pose estimation in a projective image," IBM, T. J. Watson Research Center, Tech. Rep. 20560, 1996; Available: <http://www.research.ibm.com/cyberdig>.
- [46] R. H. Taylor, L. Joskowicz, B. Williamson, A. Guéziec, A. Kalvin P. Kazanzides, R. Van Vorhis, R. Kumar, J. Yao, A. Bzostek, A. Sahay, M. Börner, and A. Lahmer, "Computer integrated revision total hip replacement surgery: Concept and preliminary results," in *Medical Image Analysis*. Oxford, U.K.: Oxford Univ. Press, to be published.