Accuracy Validation in Image-Guided Orthopaedic Surgery

D.A. Simon[1], R.V. O'Toole[2,1], M. Blackwell[1], F. Morgan[1], A.M. DiGioia[2,1], T. Kanade[1]
das@ri.cmu.edu, rvo@ri.cmu.edu, mkb@ri.cmu.edu, fxm@ri.cmu.edu, digioia@ri.cmu.edu, tk@ri.cmu.edu

[1]Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

[2]Center for Orthopaedic Research
Shadyside Hospital
Pittsburgh, PA 15232

Abstract

Validating the accuracy of image-guided surgical systems is a challenging and important problem which has received little attention in the literature. Potential sources of inaccuracy include: CT imaging noise, model generation errors, errors introduced by fixturing, intra-operative data noise, registration errors, and inaccuracies in surgical tools and actions. In this paper, we discuss our experience in validating an example image-guided application in orthopaedic surgery. Various sources of inaccuracy are identified and approaches for mitigating their effects are suggested. The difficult problem of generating a reliable ground-truth for evaluating the accuracy of surgical registration is discussed, and surface-based registration accuracy results are presented. A fiducial marker design which can be used to establish highly accurate ground-truth correspondence between pre- and intra-operative data is offered. Finally, the need for better accuracy metrics in image-guided surgery is noted, and shortcomings of metrics which are commonly used in the literature are illustrated.

Keywords: *image-guided surgery, accuracy validation, surface-based registration, accuracy metrics, fiducial marker design.*

1 Introduction

In image-guided surgery, pre-operative medical data are used to plan, simulate, guide or otherwise assist a surgeon in performing a medical procedure. In orthopaedic surgery, possible sources of pre-operative data include CT, MRI or X-Ray images. Using this data, a surgeon may develop a pre-operative plan which specifies how one or more tasks are to be performed during surgery. The plan is constructed in a coordinate system relative to the pre-operative data. The surgical procedure is performed in a coordinate system relative to the patient. Surgical registration is the process of establishing a transformation between the pre-operative data and plan, and the patient. Registration allows any 3-D point specified in the plan's coordinate system to be precisely located on the patient. Surgical execution is performed using either passive methods in which the surgeon is guided by information from the pre-operative plan, or active methods in which a semi-autonomous device, such as a robot, performs surgical tasks under the supervision of a surgeon.

This work was supported in part by a National Challenge grant from the National Science Foundation (award ECS-9422734).

A critical step required before the widespread clinical use of any novel image-guided surgical technique is the evaluation and validation of the method. While several researchers have addressed the validation problem in the context of particular systems [6][17], very little formal research has been done in this area. In this paper, we focus on the accuracy validation of pre-operative data and plans, registration methods, and surgical execution.

One difficulty in evaluating image-guided techniques is the need for highly accurate ground truth. For example, to validate a registration technique it is necessary to have a high quality estimate of the true registration transformation. Depending on the requirements, this can be a non-trivial problem requiring localization of complex patient anatomy with sub-millimeter accuracy.

In the literature, there has been little discussion of metrics for quantifying surgical task accuracy. Accuracy requirements and results are often reported as translational errors in distance units (e.g. mm) and rotational errors in degrees. In this paper we argue that there are potential ambiguities associated with these metrics due to a dependence upon the selected coordinate system. We propose that any meaningful measure of accuracy should be designed relative to the underlying task. For example, if the clinical goal is to cut a precise cavity in a femur, ultimately we are interested in whether the actual position of the cavity is within certain tolerances of the desired position.

This paper presents an overview of methods used in our lab for validating an image-guided application in orthopaedic surgery. We discuss sources of noise and error which arise in this system, and suggest approaches for dealing with these errors. Finally, we present accuracy validation results for the registration component of our system.

2 Prototypical System Overview

Figure 1 outlines the steps in a typical image-guided surgical system, using total hip replacement as the example application [24]. Initially, pre-operative data are acquired from the relevant patient anatomy. In our application, this data is acquired from a CT imager. After acquisition, the data is fed to a computer workstation where a surgeon can interactively generate a patient-specific pre-operative plan, potentially with the assistance of analysis and simulation tools. Since we

use a *surface-based* registration technique [22], the pre-operative data are also used to generate bounding surfaces of specified anatomical structures. Each of these processes are performed off-line, before surgery.

During surgery, the operations in the lower half of Figure 1 are performed. First, the relevant patient anatomy is either: 1) rigidly fixtured to prevent motion, as in [24], or 2) affixed with a tracking device that will allow the system to compensate for motion, as in [4][15]. Second, registration data are acquired from the relevant anatomy using a suitable sensor (e.g. digitizing probe, ultrasound). Third, the registration process determines the transformation between the surface model (in the coordinate system of the pre-operative plan), and the intra-operative surface data (in the coordinate system of the intra-operative sensor). Once this key transformation has been determined, the pre-operative plan may be executed within the reference frame that describes the patient's anatomy at the time of the surgery. Intra-operative actions may be performed either manually or autonomously. For manual actions, navigational guidance can be provided to the surgeon based on measurements of current tool locations, and knowledge of desired tool locations from the pre-operative plan. For autonomous actions, execution is performed by a tool such as a robot under the surgeon's supervision.

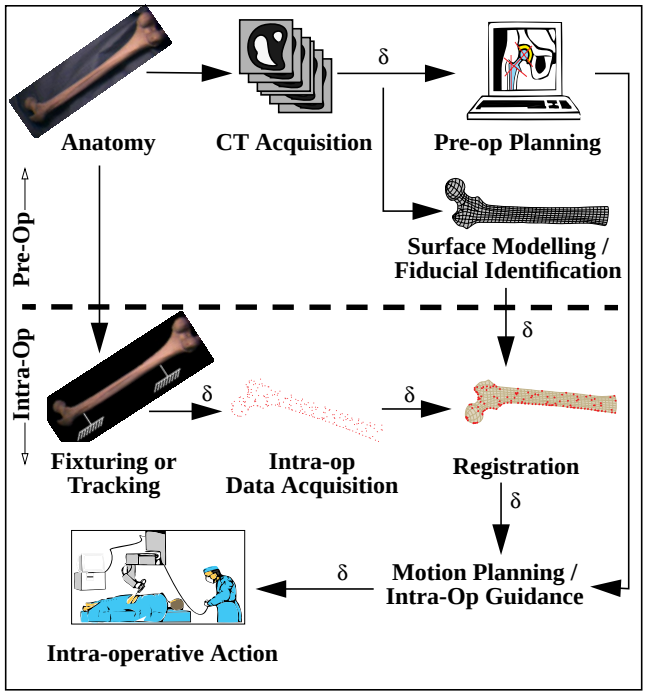

Figure 1: Image-Guided Surgical System Components

3 System Error Analysis

As with any complex system, there are many potential sources of error and noise in an image-guided surgical system. In Figure 1, a "δ" at the output of a component process indicates that the preceding operation may be a source of error or noise. These errors include:

- CT imaging errors
- Surface model generation errors
- Errors introduced by fixturing
- Intra-operative data noise
- Registration errors
- Inaccuracies in surgical tools and actions

Each of these error sources has an impact upon the accuracy of the executed task. The first goal of system validation is to measure overall task accuracy (e.g. how well does a cavity cut in a femur match the planned cavity location). The second goal is to understand how each of the component processes contribute to the final error. The remainder of this section discusses each of the aforementioned error sources.

3.1 CT Imaging Errors

During CT image creation, there are many opportunities for added error or noise [13]. Limitations in spatial and contrast resolution place bounds on the veracity of the resulting images. Artifacts caused by phenomena such as beam hardening, partial voluming and patient motion can make it difficult to interpret the underlying data. For our purposes, these sources of error are relevant because of their effect on the accuracy of the resulting surface models.

While we are not actively pursuing research in CT image formation, we are acutely aware of the potential for noise in this process. For the results in this paper we acquired high quality data from a General Electric "High-Speed Advantage" clinical CT imager (0.29mm x 0.29mm in-slice resolution, 1.0mm slice thickness, acquired at various inter-slice separations). To estimate the quality of the resulting data, we rely on evaluation of downstream operations. In Section 3.2, we describe an approach for validating surface model accuracy. In Section 4.2, we describe a method for analyzing the locations of fiducial markers extracted from images of a precisely known phantom. From these analyses we can indirectly estimate the accuracy of the underlying CT data.

Many parameters can be specified in clinical CT imagers. In-slice and inter-slice resolutions, beam power and choice of reconstruction method all influence the resulting CT images. We are currently studying the effect of CT parameters on the resulting surface model accuracy. We are currently designing CT acquisition protocols tailored to the problem of constructing surface models of bones. The challenge is to minimize the amount of CT data acquired (and thus radiation exposure to the patient) while ensuring sufficient accuracy of the resulting surface models.

3.2 Surface Model Generation Errors

Although there exist many techniques for creating surface models from medical images [12], little has been published about the validation of surface model accuracy. In medicine, surface models were originally used for diagnostic visualization tasks in which visual appearance, not geometric accura-

cy, was important. With the recent expansion of image-guided surgical techniques, surface model accuracy has taken on new importance [5][19].

In order to rigorously investigate error in surface models, a ground truth is needed. There are two methods for determining accurate ground truth: precisely manufacturing shapes for which a model exists, or accurately generating a computer surface model using an optical range sensor or other high accuracy measurement tool. Figure 2 shows a proposed extension to a method for analyzing the accuracy of surface models which was originally proposed by Geiger in [5]. Geiger argues that metrics such as difference volumes, Hausdorff distances, and measures of surface normal variation are useful for describing surface model accuracy. The primary difference between Geiger's original work and our proposed extension [19] is the use of real (as opposed to synthetic) CT data.

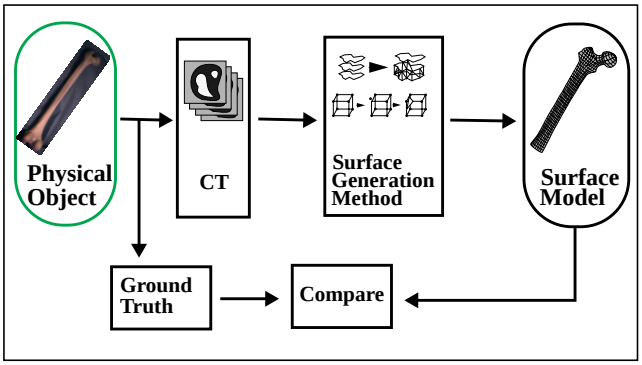

Figure 2: Schematic of surface model validation

3.3 Errors Introduced by Fixturing

Fixturing to patient anatomy establishes a fixed transformation between a patient-attached coordinate system and that of computer-assisted tools and sensors. Rigid fixturing ensures that the transformation between the pre-operative plan and the patient remains fixed during the procedure. An alternate approach is to rigidly fix a tracking device to the anatomy and use it to accommodate movements of the anatomy which occur during surgery.

The rigid fixation and tracking methods both rely on a firm connection to the patient anatomy. If there is movement between the anatomy and the fixation device or target, the registration result is no longer valid. Sensitive motion detectors can be used to insure that there is no motion between the anatomy and a fixed coordinate system [24]. This technique is less practical when the tracking method is used. Guaranteeing that a tracking target has not moved relative to the anatomy during surgery (e.g. from contact with the surgeon) is difficult, since it is impossible to differentiate motions of the target from motions of the anatomy.

An additional error source arises from the potential to deform the anatomy with the fixation device. This is of concern with the relatively massive external fixator devices used in orthopaedics. While tightening such devices onto the bone, it is possible that the bone will undergo large stresses and therefore deform. For surgeries where submillimeter accuracy is required, such deformations can impact the overall accuracy of the procedure since pre-operative surface models will no longer accurately represent the anatomy.

3.4 Intra-operative Data Noise

Data collected from patient anatomy during surgery are used by the registration process to establish the transformation between the patient and the pre-operative data. Noise present in this data will have a potentially harmful effect upon registration accuracy.

Several types of sensors have been used for acquiring intra-operative registration data. Research groups have reported on the use of: X-Ray imagers [14], ultrasonic sensors [1], optical digitizers [4][15][21][22], mechanical digitizers [10], optical range imagers [6][20][22], video cameras [3], and robots [24]. For our current applications, we are using an optical digitizer to provide intra-operative data (Optotrak - Northern Digital Inc.). While other groups have used this sensor in medical applications, little has been published regarding details of its use. Rohling et al. conducted a study comparing the accuracy of the Optotrak sensor to a mechanical digitizer [21]. We have independently verified many of the results presented in that paper [16]. Northern Digital has also published a technical report which discusses accuracy issues related to the use of several types of digitizing probes [18].

A potential source of error from Optotrak involves probe tip geometry. As explained in [18], Northern Digital uses a pivoting calibration to determine a probe's end-tip location. In order to minimize calibration errors due to motion during calibration, it is suggested to use ball-point rather than sharp-point probe tips. Ball-point tips also have the advantage that they are less likely to penetrate the surface of an object being digitized. However, due to the finite radius of a ball-point, data acquired using these probes are displaced from the true surface by a distance proportional to the ball radius. Depending upon the accuracy requirements of an application, these errors may or may not be significant. A possible fix for this problem is to acquire data such that the probe tip is well aligned with the underlying surface normal. In this case, measurements can be corrected to compensate for the ball radius, however, this approach may place an undue burden upon the data collector.

In our applications, intra-operative data are used solely for registration. Ideally, intra- and pre-operative data collected from the same underlying anatomy should be geometrically similar (i.e. if the two data sets are aligned via registration, overlapping data regions data should be coincident). For several reasons, this condition may not be met in practice. Spatial resolution differences between the underlying sensors (i.e. CT and Optotrak) may cause certain features to be visible in one modality, but not in the other. For example, small indentations present on bony surfaces may not be present in surface models generated from CT, but may be "visible" to an Op-

totrak point-tip probe. To compensate for this, a ball-probe tip can be used to spatially filter the intra-operative data (thus eliminating the surface indentations). It should be noted that it is more important that the two registration data sets be geometrically similar to each other, than it is for either data set to faithfully represent the underlying anatomy.

Other potential sources of Optotrak error which we have observed include:

- Displacement of apparent target (LED) centroids as a function of target rotation.
- Data acquisition near the sensor's field-of-view boundary.

While some effects of intra-surgical data noise can be removed via outlier elimination during registration, other noise (e.g. biases from finite diameter ball-probes) are difficult to remove. Therefore, it is important to carefully design the data acquisition process to minimize noise in the collected data.

3.5 Registration Errors

A key requirement in any registration process is the ability to extract and identify corresponding landmarks or features from the data sets being registered. This process may be facilitated using synthetic fiducial markers attached to the patient which are easily located in both pre- and intra-operative data. Fiducial-based registration has a number of clinical disadvantages including: attachment of markers may require additional surgery, and additional patient trauma at or near the marker attachment site. Therefore, many researchers have attempted to eliminate the need for fiducials by using landmarks which are intrinsic to the data itself [3][6][7][8][11]. Intrinsic landmarks used in registration include: bounding contours, ridge lines, discrete points, and surfaces. In the current work, we use a surface-based registration method described in [22].

The registration process is capable of introducing large errors into an image-guided surgical system. These errors may be caused by factors including:

- Convergence of the registration algorithm to local minima in the space of possible transformations.
- Poor geometric constraint between registration data sets [22]. This results in reduced sensitivity of the registration cost metric, and thus reduced accuracy.

In addition, registration accuracy is highly dependent upon the quality of the underlying data. Noisy data will often result in poor registration accuracy. In order to account for these factors, we take the following precautions during registration:

- Initial pose estimates should be as accurate as practical. Manually specified anatomical landmark correspondence provides one way to establish initial pose estimates. When good initial pose estimates are not available, precautions should be taken to ensure convergence to the global minima [2].
- Careful selection of intra-operative data can increase the sensitivity of the registration cost metric, and thus the resulting accuracy [22]. This is especially important when limited amounts of data are available, or the costs of data acquisition are high.
- Outlier elimination should be used to reduce the effect of noisy data upon registration. We are currently studying a "constraint-based" outlier elimination method which builds upon the work in [22].
- On-line, per-use verification checks of registration are highly desirable. For example, manual selection of landmarks on the patient, followed by automatic identification of those landmarks in CT data provides a means for verification.

3.6 Inaccuracies in Surgical Tools and Actions

The pre-operative plan manifests itself in the operating room as either semi-autonomous robotic motion, or as manual tool movements by a surgeon with the assistance of navigational guidance. In both cases, the resolution and accuracy with which a tool can be tracked impacts the precision of the executed procedure. Although manipulators have impressive repeatability specifications, calibration is usually needed to demonstrate high global accuracies. When the tool is manually controlled by a surgeon based upon visual guidance, execution inaccuracies may be large.

Tool errors become more significant when disturbance forces are encountered. If a robot is used for milling, only small cutting forces are tolerable as these forces produce deflections in the robot which are challenging to counteract. Since cutting or drilling forces increase with the speed of motion, a tradeoff exists in some applications between speed and accuracy.

4 Estimating Ground Truth

Figure 3 illustrates how we use ground truth to validate registration accuracy. Generating accurate ground truth is critical for reliable validation; however, it is also a difficult problem due to the submillimeter accuracies required.

Our approach is to compare surface-based to fiducial-based registration results. As with any fiducial-based registration approach, it is necessary to accurately extract the locations of fiducial markers from both pre-operative (CT) and intra-operative (Optotrak) data. Therefore, fiducial markers should be designed to facilitate accurate acquisition of this information. In our case, there were three primary fiducial marker design parameters: material, shape and size.

4.1 Fiducial Design

Fiducial Material: Choice of material was based on the goal of minimizing artifacts induced in CT, while still being sufficiently hard and rigid so as not to deform (i.e. during data acquisition or from humidity/temperature variations). Since validation was not performed on humans, biocompatability was not an issue. We chose an aluminum alloy (2017-T4) as

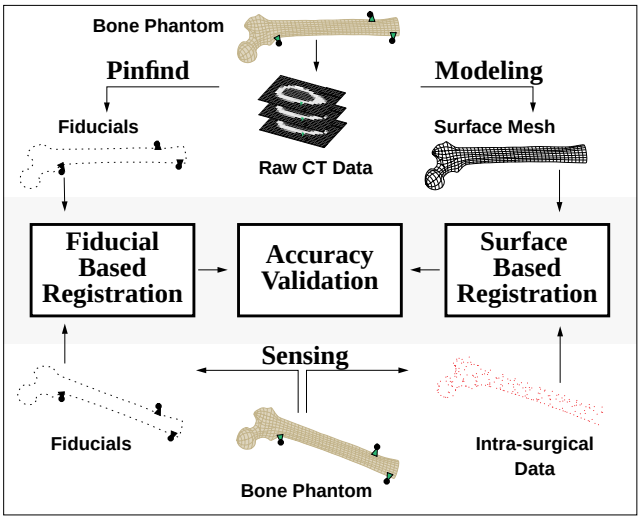

Figure 3: Validating Registration Accuracy

a material which avoids the imaging artifacts of harder materials (e.g. steel), while being more rigid and stable than softer materials (e.g. plastics). Aluminum's relatively high density also facilitates segmentation of the fiducials from CT data.

Fiducial Shape: The shape of a fiducial was selected to facilitate the accurate estimation of a *fixed point* on the fiducial from both pre- and intra-operative data. Due to data noise it is difficult to accurately extract a single point directly from the data. Rather, large quantities of data are used to infer a fixed point (i.e. centroid of a sphere, apex of a cone).

In CT, we observe that most of the error in locating a fiducial is due to beam-hardening and partial volume artifacts [13]. To limit the effects of partial voluming (which occurs when an image voxel only partially contains the fiducial), it is desirable to minimize the relative number of voxels on the surface of the fiducial. Thus, we seek a shape which minimizes the ratio of surface area to total volume (for a given size).

Beam-hardening and pluming artifacts in CT are influenced by shape and material of the imaged object. Thicker and denser materials result in larger potential for artifacts. However, small fiducials tend to have a high ratio of partial volume to non-partial volume voxels (approaching infinity as fiducial size approaches zero!). Therefore, a trade-off exists between beam-hardening and partial volume effects as a function of fiducial size.

We considered three possible fiducial shapes: cones, spheres, and cylinders. Spheres have many advantages, and for our application they clearly outperform cones and cylinders. Estimating sphere centroids from both CT and intra-operative data is straight forward. While artifacts in CT may change the apparent radius of a sphere, the sphere centroid tends to remain unchanged due to symmetry. The same is not true of cones or cylinders whose centroids (or other reference points) shift significantly as a result of artifacts. Spheres also have the smallest surface area for a given volume and are therefore less sensitive to partial voluming effects. Furthermore, spheres are much easier than cones or cylinders to localize intra-operatively, as described in Section 4.3.

Fiducial Description: A CT image of our final fiducial design mounted on a cadaver femur is shown in Figure 4. It is an aluminum sphere, 12.7mm in diameter (+/-25.4μm), mounted on a hollow Delrin cylinder which can be attached to bone using a screw or epoxy. The relatively large sphere diameter together with the moderate density of aluminum provides a trade-off between partial voluming and beam hardening / pluming. The Delrin mount provides a standoff between bone and aluminum, thus simplifying the CT segmentation task. The mounts have small holes which allow water to enter the hollow cylindrical mounts. Thus, the spheres can be almost completely surrounded by water during CT acquisition.

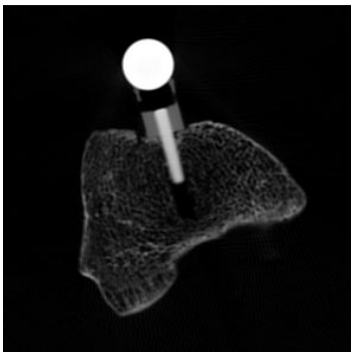

Figure 4: CT Image of a Fiducial Marker on Bone

4.2 Fiducial Locations from CT

We extract fiducial centroids from CT using a manual cropping procedure followed by an automatic threshold-based segmentation. The centroid can then be calculated directly from the thresholded data. We have experimented with several centroid location schemes which better utilize partial volume information, however due to the spherical symmetry of our fiducials, these techniques did not significantly improve the result.

We tested the accuracy of CT fiducial localization by creating a phantom with three fiducials, and measuring the centroid locations with a Coordinate Measuring Machine (CMM) to within 10μm. Inter-sphere distances averaged 49.6mm, with a maximum of 63.6mm. After CT scanning the phantom at 1mm slice intervals, we compared the extracted inter-fiducial distances to the corresponding known distances from the CMM. The average distance error was 0.1mm while the maximum distance error was 0.15mm. We also performed fiducial registration between the two data sets using the technique described in [9]. The resulting maximum residual error was 0.11mm, while the RMS residual error was 0.08mm. This sub-voxel accuracy was sufficient for our application and verifies the strength of the fiducial design and localization method.

Nolte et al. [17] have proposed an optimization technique which perturbs extracted fiducial locations from CT data so that the inter-fiducial distances agree with the corresponding

distances from a known ground truth. We are currently investigating this technique to determine whether it will improve our ground truth results.

4.3 Fiducial Locations from Intra-Operative Sensor

Using Optotrak, it is trivial to directly measure sphere centroids. Using a hollow cylindrical probe tip such as the one shown in Figure 5, a pivoting calibration [18] can be performed using a fiducial to provide the center of rotation. After calibration, the sensor will directly measure sphere centroids when the probe tip is mated to the sphere surface. By acquiring a large number of data points (e.g 15), very good centroid estimates are possible.

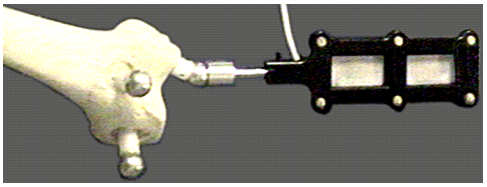

Figure 5: Centroid Locating Probe Tip

Using the phantom and techniques described in Section 4.2, we have compared fiducial centroids measured by Optotrak with CMM-based ground truth. The average distance error was 0.03mm while the maximum distance error was 0.05mm. Using the fiducial-based registration method, the maximum residual error was 0.03mm, while the RMS residual error was 0.03mm.

A second test of intra-operative fiducial localization was performed by estimating the centroids of 6 fiducials mounted on a bone phantom, twice without moving the bone. Since the transformation between the first and second data sets should be null, we can use the results to estimate the repeatability of our measurements. Using fiducial-based registration, the maximum residual error was 0.09mm, while the RMS residual error was 0.05mm. This resulted in an apparent motion of the bone of 0.09mm in translation, and 0.03 degrees of rotation about a coordinate system located at the centroid of all six fiducials.

4.4 Overall Accuracy of the Ground Truth

As a final measure of fiducial-based ground truth accuracy, we examined the residuals which result by registering CT and Optotrak fiducial centroids extracted from the bone phantom. The maximum resulting residual was 0.12mm, with an RMS error of 0.09mm.

5 Registration Validation Results

As outlined in Figure 3, we can use fiducial-based registration to estimate the accuracy of surface-based registration. For the purposes of discussing registration accuracy, we created a pre-operative plan for the placement of a femoral implant in total hip replacement surgery. Figure 6 is a schematic of a pre-operative plan which we have constructed using CT data from a cadaver bone. The goal of the planning process is to determine the proper position of the implant with respect to the femur in the CT coordinate system. The plan is constructed manually via a graphical user interface which overlays a model of the implant upon the CT data.

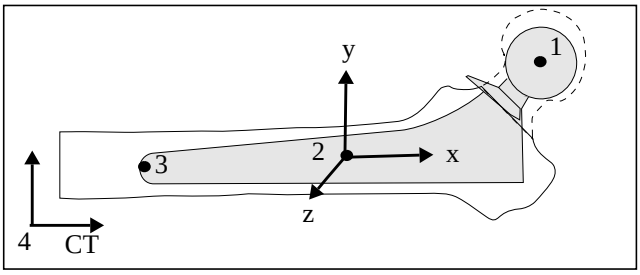

Figure 6: Pre-Operative Plan

As mentioned in the introduction, we believe that all accuracy results and requirements should be specified using metrics which have direct physical meaning to the task being performed. To demonstrate this concept, registration accuracy results are presented below using two different metrics.

In the first experiment, we constructed a surface model from CT images of a cadaveric femur scanned in water. 100 “intra-surgical” data points were collected within a roughly 65mm region of the proximal femur using Optotrak. Data collection was limited to areas of the bone which are clinically accessible during total hip replacement surgery. Surface-based registration was applied to the data in order to estimate the registration transformation. Fiducial-based registration was performed using 6 markers to calculate ground-truth.

Table 1 summarizes the results of this experiment using the conventional method of representing error. Each row of the table contains the registration error (i.e. difference between the fiducial-based and surface-based result) expressed as a transformation about a given coordinate system. The first column indicates the coordinate system, while the second and third columns indicate the magnitudes of the translation and rotation errors respectively. Coordinate systems 1-3 are all parallel, with the X-axis in the direction of the implant shaft’s central axis, the Y-axis defined by the projection of the femoral head centroid onto the X-axis, and the Z-axis defined as the cross product of the first two. The origin of each coordinate system is a point selected for its relevance to the implant placement task (centroid of the prosthetic femoral head, centroid of the implant, and distal tip of the implant, numbered 1-3 respectively in Figure 6). The fourth coordinate system in Table 1 is the one used by the CT imager.

In the table, the magnitude of the translation error component is a function of coordinate system, while the magnitude of the rotation component is constant. An explanation of this result is presented in the next paragraph. The important point is that each of the results in Table 1 refer to the *same transformation error*. This example demonstrates the potential ambiguity of this error metric. When using this metric, it is crucial that the selected coordinate system have a physically meaningful relation to the task being performed. Depending upon the selected coordinate system, the accuracy of our surface-based

Table 1: Registration Errors - Translation / Rotation

Coordinate System	‖T‖ (mm)	‖R‖ (deg)
1 - Head Centroid	0.25	0.97
2 - Implant Centroid	0.31	0.97
3 - Implant Distal Tip	0.17	0.97
4- CT	1.77	0.97

registration could either be reported as 0.17mm in translation or 1.77mm in translation, a factor of 10 difference! (Fortunately, meaningful error values for this task are closer to the smaller value).

The schematic in Figure 7 provides intuition into the results of Table 1. In the left half of the figure, the two lines labeled 1 and 2 represent a misalignment error such as that resulting from registration. With respect to a coordinate system positioned at point **a**, this error can be described as a pure rotation. In order to describe the same error with respect to a coordinate system positioned at point **b**, line 2 must first be rotated about point **b** to line 3, and then shifted back to point **a**. The angle of rotation about either point **a** or **b** are the same, however the translation differs. This same phenomenon applies to the results of Table 1, and explains the difference in translation magnitudes as a function of coordinate system.

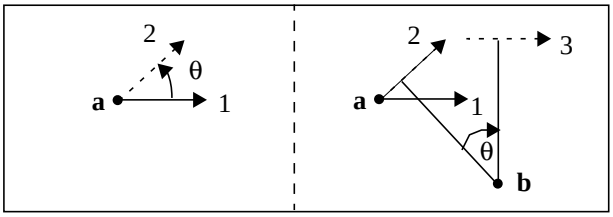

Figure 7: Error Dependence on Coordinate Frame

The second metric used to describe the results of the above experiment expresses registration error as a displacement induced at selected points on the implant. For each of the points numbered 1-3 in Figure 6, we calculated the displacement (i.e. change in position measured as a Euclidean distance) of the point which results by applying the registration error transformation of Table 1. These displacements are independent of the coordinate system about which the error is represented. Since this metric requires explicit specification of physically meaningful landmarks, it is potentially less ambiguous than the first metric. The resulting displacements are shown in Table 2 and can be interpreted as the implant misalignment at each point which would result if the implant were actually positioned using the registration transformation found above. It is no coincidence that the point displacement magnitudes are the same as the corresponding translation magnitudes from Table 1. (On the right side of Figure 7, the translation required to re-align line 3 with line 1 is the same as the displacement which would be induced at point **b** by rotating line 2 together with point **b**, about point **a** into line 1).

Table 2: Registration Errors - Point Displacement

Point Number	‖D‖ (mm)
1 - Head Centroid	0.25
2 - Implant Centroid	0.31
3 - Implant Distal Tip	0.17
4 - Extremum Point	0.54

The fourth point in Table 2 was selected to provide an upper bound on displacement among all points on the implant surface. This point was found by determining the screw axis representation of the error [23] (the axis about which the transformation can be expressed as a rotation about the axis coupled with a translation along it). The extremum point was then selected as the implant surface point farthest from the axis. The projection of this screw axis onto the pre-operative plan is shown in Figure 8. Note that the point displacements of Table 2 are a function of the distance between the axis and each of the points. We have found 3-D graphical renderings of the screw axis superposed on the relevant anatomy to be quite useful in visually interpreting registration error.

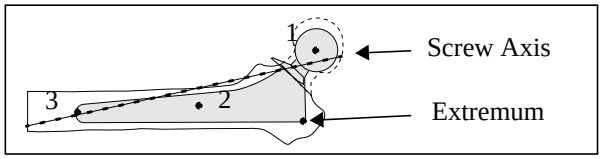

Figure 8: Screw Axis Representation of Error

6 Conclusions

We have presented several issues related to the accuracy validation of an image-guided surgical system. We believe that many of the issues discussed with respect to our example task of total hip replacement have application to a broader set of image-guided surgical problems.

A major issue which remains to be addressed is the specification of accuracy requirements for image-guided surgical applications. Fundamental questions remain to be answered.

- What are the best clinically meaningful measures of accuracy for a given task?
- How should accuracy specifications / tolerances for a given task and a given metric be determined?
- Are additional analysis tools required to allow clinicians to analyze task accuracy requirements? If so, what form should these tools take?

The consequences of under estimating accuracy requirements in an image guided surgical application could be potentially catastrophic. Accuracy requirements which are too strict may have negative consequences as well. In general, achieving higher accuracy will have higher associated monetary costs (i.e. additional pre-operative data, more accurate intra-opera-

tive sensors and tools, longer times spent in surgery acquiring data, etc.). Given the current economic climate in medicine, we can not afford to waste limited resources on unnecessary accuracy.

7 Acknowledgments

The authors wish to thank Rich Wallace of Shadyside Hospital for his invaluable help acquiring CT data. Jesse Easudes generated CAD models of the surgical implants used to construct pre-operative plans. Drew Shefman created much of the artwork in this paper. Lori Gregor and Nate Walk helped with data collection and experimental tasks.

8 References

[1] Barbe, C, Troccaz, J, Mazier, B, and Lavallee, S. Using 2.5D Echography in Computer Assisted Spine Surgery. *IEEE Engineering in Medicine and Biology Society Proceedings*, pp. 160-161. IEEE, San Diego, 1993.

[2] Besl, PJ and McKay, ND. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 14(2):239-256, February, 1992.

[3] Betting, F, Feldman, J, Ayache, N and Devernay, F. A New Framework for Fusing Stereo Images with Volumetric Medical Images. Ayache, N. (ed.), *Proc. CVRMed '95*, pp. 30-39. Springer-Verlag, Nice, France, April, 1995.

[4] Dessenne, V, Lavallee, S, Julliard, R, Orti, R, Martelli, S and Cinquin, P. Computer-Assisted Knee Anterior Cruciate Ligament Reconstruction: First Clinical Tests. *J. of Image Guided Surgery.* 1(1):59-64, April, 1995.

[5] Geiger, B. *Three-dimensional modeling of human organs and its application to diagnosis and surgical planning.* PhD thesis, Ecole des Mines de Paris, April, 1993.

[6] Grimson, WEL, Ettinger, GJ, White, SJ, Gleason, PL, Lozano-Perez, T., Wells WM, and Kikinis, R. Evaluating and Validating an Automated Registration System for Enhanced Reality Visualization in Surgery. Ayache, N. (ed.), *Proc. CVRMed '95*, pp. 3-12. Springer-Verlag, Nice, France, April, 1995.

[7] Hamadeh, A, Lavallee, S, Szeliski, R, Cinquin, P, and Peria, 0. Anatomy-based Registration for Computer Integrated Surgery. Ayache, N. (ed.), *Proc. CVRMed '95*, pp. 212-218. Springer-Verlag, Nice, France, April, 1995.

[8] Henri, CJ, Colchester, ACF, Zhao, J., Hawkes, DJ, Hill, DLG, and Evans, RL. Registration of 3-D Surface Data for Intra-Operative Guidance and Visualization in Frameless Stereotactic Neurosurgery. Ayache, N. (ed.), *Proc. CVRMed '95*, pp. 3-12. Springer-Verlag, Nice, France, April, 1995.

[9] Horn, BKP. Closed-Form Solution of Absolute Orientation Using Unit Quaternions. *J. of the Optical Society of America A.* 4(4):629-642, April, 1987.

[10] ISG Technologies Inc. *Viewing Wand Operator's Guide* ISG Technologies Inc., Mississauga, Canada, 1993.

[11] Jiang, H, Robb, RA and Holton, KS. A New Approach to 3-D Registration of Multimodality Medical Images by Surface Matching. *Visualization in Biomedical Computing - SPIE Vol 1808*, pp. 196-213. SPIE, 1992.

[12] Kalvin, AD. *A survey of algorithms for constructing surfaces from 3D volume data.* IBM Reseach Report RC 17600, IBM, January, 1992.

[13] Krestel, E. *Imaging Systems for Medical Diagnostics*. Siemens Aktiengesellschaft, Berlin, 1990.

[14] Lavallee, S and Szeliski, R. Recovering the Position and Orientation of Free-Form Objects from Image Contours Using 3D Distance Maps. *IEEE Trans. PAMI.* 17(4):378-390, April, 1995.

[15] Lavallee, S, Sautot, P, Troccaz, J, Cinquin, P, and Merloz, P. Computer-Assisted Spine Surgery: A Technique for Accurate Transpedicular Screw Fixation Using CT Data and a 3-D Optical Localizer. *J. Image Guided Surgery.* 1(1):65-73, April, 1995.

[16] Morgan, F., O'Toole, RV, Simon, DA, and Blackwell, MK. *Optotrak Validation Experiments.* Technical Report CMU-RI-TR-95-26, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, July, 1995.

[17] Nolte, LP, Zamorano, LJ, Zhaowei, J, Wang, Q, Langlotz, F, Arm, E, and Visarius, H. A Novel Approach to Computer Assisted Spine Surgery. *Proc. MRCAS '94*, pp. 323-328. Pittsburgh, PA, September, 1994.

[18] Northern Digital Inc. *Accuracy of Digitizing Probes.* Technical Report 3, Northern Digital Inc., Ontario, CAN, May, 1992.

[19] O'Toole, RV, Simon, DA, Jaramaz, B, Ghattas, O, Blackwell, MK, Kallivokas, L, Morgan, F, Visnic, C, DiGioia, AM, and Kanade, T. Towards More Capable and Less Invasive Robotic Surgery in Orthopaedics. *Proc. CVRMed '95*, pp. 123-130. Springer-Verlag, Nice, France, April, 1995.

[20] Peria, O, Francois-Joubert, A, Lavallee, S, Champleboux, G, Cinquin, P and Grand, S. Accurate registration of SPECT and MR brain images of patients suffering from epilepsy or tumor. *Proc. MRCAS '94*, pp. 58-62. Pittsburgh, PA, September, 1994.

[21] Rohling, R, Munger, P, Hollerbach, J and Peters, T. Comparison of Relative Accuracy Between a Mechanical and an Optical Position Tracker for Image-Guided Neurosurgery. *Proc. MRCAS '94*, pp. 277-282. Pittsburgh, PA, September, 1994.

[22] Simon, DA, Hebert, M and Kanade, T. Techniques for Fast and Accurate Intra-Surgical Registration. *J. Image Guided Surgery.* 1(1):17-29, April, 1995.

[23] Spoor, CW and Veldpaus, FE. Rigid body motion calculated from spatial co-ordinates of markers. *J. Biomechanics.* 13:391-393, 1980.

[24] Taylor, RH , Mittelstadt, BD, Paul, HA, Hanson, W, Kazanzides, P, Zuhars, JF, Williamson, B, Musits, BL, Glassman, E and Bargar, WL. An Image-Directed Robotic System for Precise Orthopaedic Surgery. *IEEE Trans. Robotics and Automation.* 10(3):261-275, June, 1994.